



Identification of Novel miRNAs in the F8 Gene Via Bioinformatics Tools

Halimeh Rezaei, Majid Motovali-Bashi*, Sheyda Khalilian

Department of Cell and Molecular Biology and Microbiology, Faculty of Biological Science and Technology, University of Isfahan, Iran.

*Corresponding author: Majid Motovali-Bashi, Department of Cell and Molecular Biology and Microbiology, Faculty of Biological Science and Technology, University of Isfahan, Iran. Tel: +98-3137932474, Fax: +98-3137932456, E-mail: mbashi@sci.ui.ac.ir

Background: Hemophilia A is an X-linked bleeding disorder resulting in a deficiency of plasma clotting factor VIII and caused by mutations in the *FVIII gene (F8 gene)*. MicroRNAs (miRNAs) in body fluids are promising biomarker candidates for Hemophilia A, due to their stability in body fluids and accessibility by non- or minimally-invasive procedures. Therefore; Advances in miRNA analysis methods resulted in a wide range of publications on miRNAs as putative biomarkers.

Objective: Here we tried to scan the *F8 gene* region to predict a novel miRNA and identify it as a regulator of the *F8 gene*.

Materials and Methods: To this aim, the ability to express novel miRNAs in *F8* locus was assessed via reliable bioinformatics databases such as SSCprofler, RNAfold, miREval, FOMmiR, MaturBayes, miRFIND, UCSC genome browser, Deep Sequencing, and miRBase.

Results: Data analysis from the relevant databases offers one stem-loop structure that is predicted to express a novel miRNA.

Conclusions: The diagnosis of Hemophilia A with the help of these types of biomarkers is a non-invasive procedure that has been demonstrated to have a significant role in the early diagnosis of the disease. Hopefully, the proposed candidate sequence will be confirmed in vitro and become a non-invasive biomarker in the near future.

Keywords: Bioinformatics; Database; Hemophilia; MicroRNA

1. Background

Hemophilia A is a common recessive X-linked bleeding disorder. It is caused by a quantitative or qualitative defect of the plasma protein involved in hemostasis called Factor VIII (FVIII), with a frequency of one in 5,000 to 10,000 live births in males. Based on the natural function of FVIII, hemophilia A can be classified into three levels: mild (5-25%), moderate (1.5%), and severe (1%). FVIII is a necessary plasma protein in the blood coagulation system, which is converted into an inactive form by connecting with Von Willebrand Factor while the clotting process. Coagulation Factor VIII activates and separates the von Willebrand factor in response to injury. It also interacts with another coagulation factor called factor IX. This interaction sets off a chain of additional chemical reactions that leads to blood coagulation (1). The Human *FVIII gene (F8 gene)* is located in Xq28 locus and encodes the coagulation Factor VIII. This gene with an approximate length of

186 kb, including 26 exons, produces two alternatively spliced transcripts (isoform a and isoform b). Up to now, more than 1,000 mutations have been registered in the Hemophilia World Databank called HAMSTeRs (2).

MicroRNAs (miRNAs) are short non-coding RNAs with 18-25 nucleotides in length. They are evolutionally conserved and act principally as post-transcriptional gene expression regulators. On the other hand, mature miRNA is incorporated into an RNA-induced silencing complex (RISC) and binds to the 3' untranslated region (UTR) of the target messenger RNAs (mRNAs) to mediate translational repression (3). Since a single miRNA can target up to hundreds of genes, it is believed that about 60% of all human genes are known as putative targets for individual miRNAs. Moreover, individual genes may contain multiple binding sites for different miRNAs, resulting in a complex regulatory network

(4). The discovery of a functional role of miRNAs in the pathogenesis of a wide range of human diseases, and the indication of their tissue-specific expression patterns inclined researchers to investigate the role of miRNAs as potential non-invasive biomarkers including diagnostic, prognostic, monitoring, risk, and safety biomarkers in diagnosis and treatment. It has been recognized that circulating miRNA levels are different between healthy and diseased individuals suggesting that miRNA alterations may serve as an important indicator of various pathologies (5). The serum miRNA studies started with the associated levels of miR-21 in patients with diffuse large B-cell lymphoma (6). The expression of the miRNA profiles for autoimmune disorders, some types of cancer, diabetes mellitus, and other diseases have also assumed relevance. Presently, the circulating miRNAs are easy to detect and quantitate; Hence, in addition to registered miRNAs, predicting and identifying parts of the genome that are susceptible to the expression of new miRNAs has created a fascinating field for molecular studies of miRNAs. These studies are feasible by bioinformatics and molecular laboratory techniques (7).

2. Objectives

Considering the fact that hemophilia A is a monogenic disorder, most of the known and registered miRNAs in human genes affect the expression levels of its target gene, the present study aimed to search for the miRNAs embedded within the sequence of the F8 gene to identify and control the progression of hemophilia A.

3. Materials and Methods

Algorithms applied in databases and bioinformatics tools that are used to identify and predict new miRNAs are first identified based on the sequence of known registered miRNAs. Then these programs are used to scan the genome and detect the sequence of putative novel miRNAs. In other words, collecting, studying, and merging the high load of information about known and registered miRNAs, reveals similar characteristics. Such characteristics are the stem-loop structure length, thermodynamic stability, the bulge size and position, nucleotide content, sequence complexity, and repetitive elements. They exist in the genes encoding miRNAs and are used in their prediction.

3.1. SSCprofiler Database

The present study investigated a novel stem-loop structure in the F8 gene by using the SSCprofiler database (<http://mirna.imbb.forth.gr/SSCprofiler.html>) (8).

SSCprofiler database provides biological information based on sequence, structure, and protection of human miRNAs. The sensitivity and specificity accuracy of the output data from the SSCprofiler database is 95.88% and 84.16%, respectively (9).

3.2. RNAfold Web Server

The RNAfold server (<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>) (10) was used to study the stem-loop structure and the stability provided by the SSCprofiler database. It predicts the secondary structure of single-stranded RNA, calculates the partition function and base pairing probability matrix as well as the minimum free energy (MFE) structure.

3.3. MiREval, FOMmiR, and MatureBayes Web Tools

Three webservers, miREval (<http://mimirna.centenary.org.au/mireval>) (11), FOMmiR (<http://app.shenwei.me/cgi-bin/FOMmiR.cgi>) (12) and MatureBayes (<http://mirna.imbb.forth.gr/MatureBayes.html>) (13) were used to evaluate the accuracy of the predicted stem-loop structure. The miREval web server uses a Support Vector Machine (SVM) that trains 57 features including secondary structure, free energy, and sequence composition. There are two positive and negative information categories for the SVM enabling the miREval database to distinguish miRNA stem-loops from stem-loops in other non-coding RNAs. The FOMmiR database is not only able to distinguish the miRNA precursors from the stem-loops, but it also locates the position and strand of the mature miRNA. This database, therefore, presents a new understanding of the biological recognition which might be closely associated with the enzyme cleavage mechanism during the miRNA maturation.

On the other hand, the MatureBayes database is a tool for detecting mature miRNA within stem-loop structures using a Naive Bayes classifier.

3.4. MiRFIND Database

Using miRFIND database (<http://140.120.14.132:8080/MicroRNAProject-Web>) (14), the function of the Drosha and Dicer enzymes were further analyzed according to the stem-loop structure sequence.

3.5. UCSC Genome Browser

The conservation of the candidate stem-loop structure in the vertebral genome was also studied by using the UCSC Genome Browser database (<https://genome.ucsc.edu>) (15). RNA expression profiling in 14 different human cell lines was investigated (by Deep Sequencing), and the probability of the novel miRNA

expression in the candidate sequence was evaluated as well.

3.6. *MiRBase Database*

Finally, miRBase database (<http://www.mirbase.org/>) (16) was used to confirm the novelty of the candidate sequence as mature miRNA. The miRNA gene candidate did not demonstrate an apparent sequence similarity to the known miRNA genes. More than 12,000 mature

miRNAs from 600 species were previously identified by other researchers and registered on this database.

4. **Results**

The *F8 gene* associated with hemophilia A was scanned to identify and predict the stem-loop structure. In this regard, databases and high reliable bioinformatics servers were used. According to the results, the stem-loop structure with the sequence of “TGTAAGGCT-

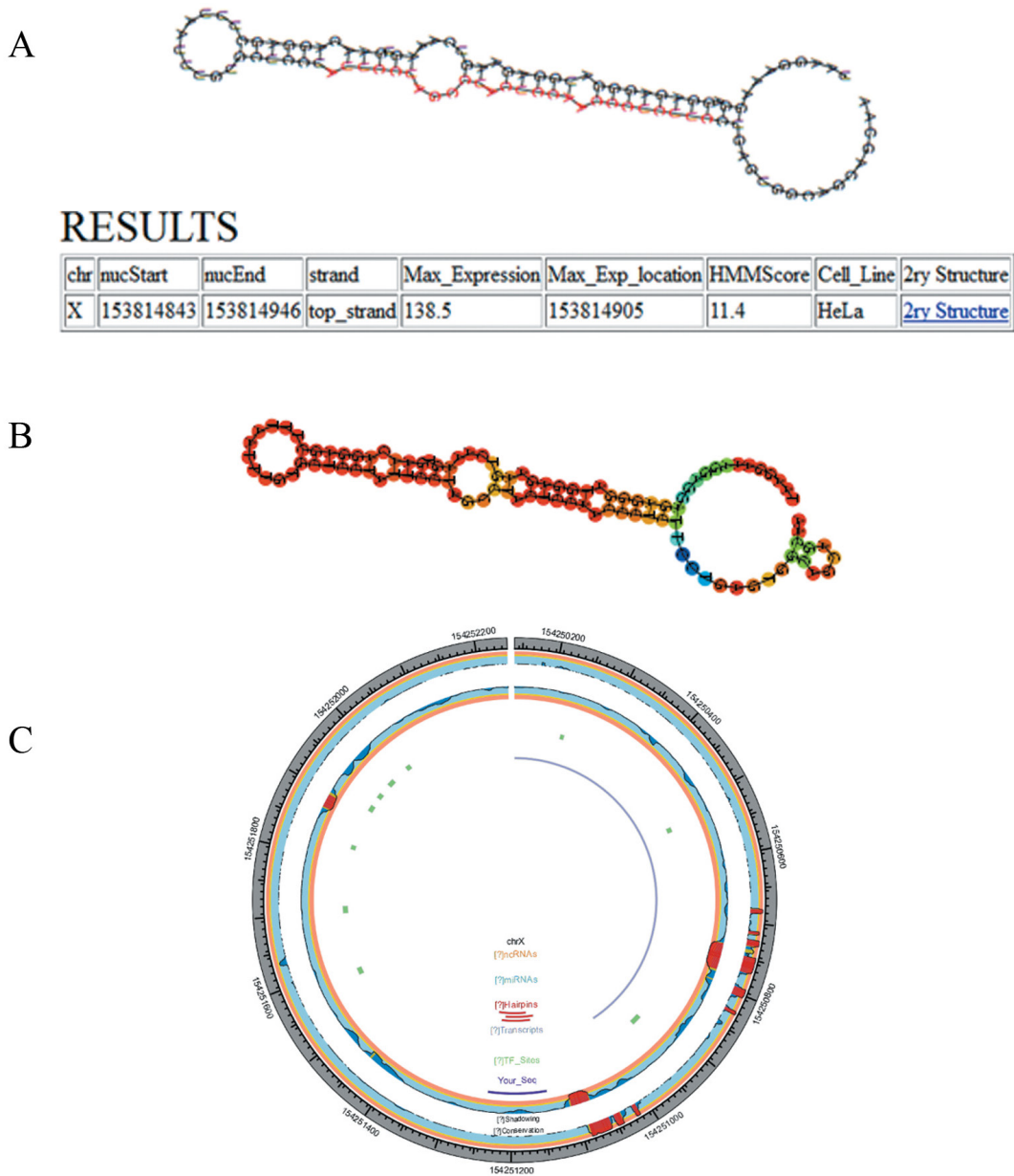


Figure 1: Prediction of candidate-miR within the first intron of the human F8 gene. (A) Results of SSC profiler for candidate-miR. Hairpin structure containing a probable sequence of mature miR (Red) is shown, and HMM score related to this structure is shown in the table. Furthermore, maximum expression (Max-Expression) according to a full genome tiling array in HeLa cell line is presented for this sequence. (B) Graphical output of hairpin structure in RNAfold web server. The secondary structure result of candidate-miR is depicted. (C) miREval output data. 1000 base pairs around our inquiry sequence are displayed as a circle graph by miREval

CATAAAAGTTGAGGAAGCCATTTGGGCTCgtc-tactccagcatggtccacagaccaggagtagcagcatcacctgagg-gcaattcaaatgca “, located in the first intron of the *F8 gene*, was predicted and introduced for experimental verification.

4.1. SSCprofler Database

This database predicts and identifies the stem-loop

structure in the *F8 gene*. It also uses a hidden Markov model (HMM) to model secondary structural features in each position of miRNA stem-loop. This score considers the sequence, structure, and conservation of the miRNA coding genes simultaneously in the statistical models. Thus, the higher the score, the greater the chance that the potential candidate structure belongs to the real miRNA (**Fig. 1A**).

A

B

MatureBayes

The results of the query

>Precursor hairpin:46:54	
Sequence	UAAGGAAGGAGGAGGGAUGGAGAAGUCAAGUGAACAGGAGCUAAUUAUUGUCUCCUAUUCUAGCCUACUCCAACCCUUCUUCGAGUGGCAGCAGCAA
Duplex	Position 17 (5' stem) Sequence: GGAUGGAGAAGUCAAGUGAAC Position 60 (3' stem) Sequence: UAUUCUAGCCUACUCCAACCC
Mature 5' stem	Position 9 Sequence: GAGGAGGGAUGGAGAAGUCA
Mature 3' stem	Position 60 Sequence: UAUUCUAGCCUACUCCAACCC

C

Figure 2: The results of other databases used to confirm the presence of novel miRNA. **(A)** FOMmiR database information. The predicted mature miRNA sequence is observed in red in the candidate stem-loop structure. **(B)** MatureBayes database output. The 3p and 5p sequence of mature miRNA has been identified by a nucleotide position in the candidate sequence. **(C)** Results of UCSC genome browser on Human Feb.2009 (GRCH37/hg19) Assembly. Conservation levels are shown with blue columns.

4.2. RNAfold Web Server

In this server, the stem-loop structure proposed from the SSCProfiler database was introduced to make more precise studies on the stability of the secondary structure. The stability of the proposed secondary structure has been investigated according to the minimum free energy (MFE) allocated to any structure (**Fig. 1B**). The MFE for candidate-miR is $-35.10 \text{ kcal.mol}^{-1}$.

4.3. MiREval, FOMmiR, and MatureBayes Web Tools

The accuracy of the stem-loop structure was evaluated through the miREval web tool (**Fig. 1C**) and the prediction of mature miRNA in the candidate sequence was investigated through the FOMmiR and MatureBayes web tools (**Fig. 2A /2B**).

4.4. MiRFIND Database

Drosha and Dicer cleavage sites were identified in the candidate sequence. The mature-miRNA Drosha/Dicer processing sites are 17/38 and 84/6. The mature-miRNA sequences are 5-AAGCCAUUUGGGCUCUGCUACU-3' and 5-AGUAGCAGCAUCACCUGAGGGCAA-3'. Predicted seed sites corresponding to mature miRNA-5p and-3p, respectively.

4.5. UCSC Genome Browser

The percentage of candidate sequence conservation among 100 vertebral genomes (**Fig. 2C**) and also Deep sequencing data were analyzed and it was found that the candidate-miR was expressed in SKMC TAP-only, IMR90 CIP-TAP, and IMR90 whole cells.

4.6. MiRBase Database

The miRBase database was used to ensure that the candidate sequence was not reported as mature miRNA in other studies.

5. Discussion

Today, the most common use of predictive and diagnostic biomarkers is for the treatment decisions and detection of recurrent disease and monitoring therapy. Appropriate biomarkers should be stable and non-invasive. They also must be disease-specific for reliable and accurate measurement across a range of diseased populations (17). Given the prevalence of miRNA regulation, it is suggested that miRNAs have been involved in a wide range of human diseases. For instance, miR-139-5P (18) and miR-182, and miR-187 (19) are among the detected miRNAs that are used as biomarkers in cancer diagnosis. It should be noted that down-regulation of the biogenesis factors, the mutation

in the miRNA locus, or epigenetic changes such as hyper methylation, can perturb the miRNA function (20).

Before the investigation of the miRNA role in a disease, it must be predicted and annotated following its specific expression pattern. After that, by artificially changing the expression level of miRNAs, the initiation and progression of diseases can be controlled. Therefore, miRNA prediction is an essential and early step of analysis in the clinical context. The discovery of novel miRNAs can also increase the use of miRNA-based therapy and ultimately lead to a change in treatment attitudes, improved clinical outcomes, and better allocation of health care resources (21). On the other hand, the cloning of some miRNAs is not easily achieved which is due to their physical properties, including nucleotide sequence or post-translational modifications (editing, methylation, etc.). Expensive and time-consuming cloning techniques added more limitations as well. The computational algorithms provide quick, efficient, and inexpensive methods for detecting and predicting miRNAs coding sequences in the genome. This should be confirmed in vitro by examining the expression of the endogenous miRNA mature form (22, 23).

V. Kim *et al.* Showed that a small RNA is considered to be the true miRNA according to the following points:

- 1) The miRNA expression should be confirmed by Real-time PCR.
- 2) The small RNA sequence should be located in a 60-80 nucleotide stem-loop proximal arm and without a large internal bulge or loop.
- 3) The small RNA sequence should be phylogenetically conserved.
- 4) By decreasing the function of the Dicer enzyme, the number of miRNA precursors should increase.

Finally, with further studies, they predicted and introduced about 38 novel miRNAs.

Furthermore, Berezikov *et al.*, in a later study, presented 69 candidate sequences regarding the conservation profiles and RNA folding criteria. They confirmed the expression of 16 mature human miRNAs by Northern Blot analysis (24).

Hoballa *et al.*, consistent with the bioinformatics prediction, introduced two novel miRZa-3p and miRZa-5p, which target SMAD3 and IGF1R genes and increase the cell population in the sub-G1 stage (25).

In addition, Dokanehifard *et al.* used the SSCProfiler, UCSC genome browser, and several other databases. They identified, predicted, and validated two novel miRNAs in TrkC gene as well as has-miR-6165 in

the NGFR gene associated with colorectal cancer (26).

The purpose of this study was to scan the *F8 gene* to identify and predict a candidate sequence for the expression and production of a mature miRNA, considering that most of the known and recorded human miRNAs in human genes affect the expression levels of their coding genes, and bearing in mind that hemophilia A is a monogenic disorder. The present study used a highly accurate and reliable database. Hopefully, the proposed candidate sequence will be experimentally approved in future studies and will play a key role in the development of miRNA-based drugs for the treatment of hemophilia a patients.

Acknowledgments

This study was done at the University of Isfahan and supported by Molecular Genetics Department. The authors are sincerely grateful to the department for their support.

Authors Contributions

All authors have contributed equally to the work.

Financial Disclosure

The authors declare that they have no financial conflicts of interest in relation to the study in this paper.

References

- Khalilian S, Motovali-Bashi M, Rezaie H. Factor VIII: Perspectives on Immunogenicity and Tolerogenic Strategies for Hemophilia a Patients. *IJMCM*. 2020;**9**(1):33-50. doi.org/10.22088/IJMCM.BUMS.9.1.33
- Halldén C, Nilsson D, Säll T, LIND-HALLDÉN C, Lidén AC, Ljung R. Origin of Swedish Hemophilia A mutations. *JTH*. 2012;**10**(12):2503-2511. doi.org/10.1111/jth.12010
- Rezaei H, Motovali-Bashi M, Khalilian S. MicroRNA Prediction in the FVIII Gene Locus: A Step Towards Hemophilia a Control. *GCT*. 2020;**7**(3):e103096. doi.org/10.5812/gct.103096
- Chen L, Hu N, Wang C, Zhao H, Gu Y. Long non-coding RNA CCAT1 promotes multiple myeloma progression by acting as a molecular sponge of miR-181a-5p to modulate HOXA1 expression. *Cell Cycle*. 2018;**17**(3):319-329. doi.org/10.1080/15384101.2017.1407893.
- Pogribny IP. MicroRNAs as biomarkers for clinical studies. *EBM*. 2018;**243**(3):283-290. doi.org/10.1177/1535370217731291.
- Velu VK, Ramesh R, Srinivasan AR. Circulating MicroRNAs as Biomarkers in Health and Disease. *J clin diagn*. 2012;**6**(10):1791-1795.
- Fang W, Bartel DP. The menu of features that define primary microRNAs and enable de novo design of microRNA genes. *MolCell*. 2015;**60**(1):131-145. doi.org/10.1016/j.molcel.2015.08.015.
- Gomes CP, Cho JH, Hood L, Franco OL, Pereira RW, Wang K. A Review of Computational Tools in microRNA Discovery. *Front Genet*. 2013;**4**:81. doi.org/10.3389/fgene.2013.00081.
- Oulas A, Boutla A, Gkirtzou K, Reczko M, Kalantidis K, Poirazi P. Prediction of novel microRNA genes in cancer-associated genomic regions—a combined computational and experimental approach. *Nucleic Acids Res*. 2009;**37**(10):3276-3287. doi.org/10.1093/nar/gkp120.
- Lorenz R, Bernhart SH, Zu Siederdisen CH, Tafer H, Flamm C, Stadler PF, *et al.* ViennaRNA Package 2.0. *Algorithms Mol Biol*. 2011;**6**(1):26. doi.org/10.1186/1748-7188-6-26.
- Rajendiran A, Chatterjee A, Pan A. Computational approaches and related tools to identify MicroRNAs in a species: A Bird's Eye View. *Interdiscip Sci*. 2018;**10**(3):616-635. doi.org/10.1007/s12539-017-0223-x.
- Shen W, Chen M, Wei G, Li Y. MicroRNA prediction using a fixed-order Markov model based on the secondary structure pattern. *PLoS One*. 2012;**7**(10):e48236. doi.org/10.1371/journal.pone.0048236.
- Saini S, Thakur CJ, Kumar V, Tandon S, Bhardwaj V, Maggar S, *et al.* Computational prediction of miRNAs in Nipah virus genome reveals possible interaction with human genes involved in encephalitis. *Mol Biol Res Commun*. 2018;**7**(3):107.
- Pan R-H, Tseng L-Y, Liao I-E, Chan C-L, Lai KR, Lin K-B, editors. Design of an NGS MicroRNA predictor using multilayer hierarchical MapReduce framework. *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. 2015;IEEE. doi.org/10.1109/dsaa.2015.7344862.
- Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, *et al.* The human genome browser at UCSC. *Genome Res*. 2002;**12**(6):996-1006. doi.org/10.1101/gr.229102.
- Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res*. 2006;**34**(suppl_1):D140-144. doi.org/10.1093/nar/gkj112.
- Byrnes SA, Weigl BH. Selecting analytical biomarkers for diagnostic applications: a first principles approach. *Expert Rev Mol Diagn*. 2018;**18**(1):19-26. doi.org/10.1080/14737159.2018.1412258.
- Miyoshi J, Toden S, Yoshida K, Toiyama Y, Alberts SR, Kusunoki M, *et al.* MiR-139-5p as a novel serum biomarker for recurrence and metastasis in colorectal cancer. *Sci Rep*. 2017;**7**(1):1-13. doi.org/10.1038/srep43393.
- Casanova-Salas I, Rubio-Briones J, Calatrava A, Mancarella C, Masiá E, Casanova J, *et al.* Identification of miR-187 and miR-182 as biomarkers of early diagnosis and prognosis in patients with prostate cancer treated with radical prostatectomy. *J Urol*. 2014;**192**(1):252-259. doi.org/10.1016/j.juro.2014.01.107.
- Davalos V, Moutinho C, Villanueva A, Boque R, Silva P, Carneiro F, *et al.* Dynamic epigenetic regulation of the microRNA-200 family mediates epithelial and mesenchymal transitions in human tumorigenesis. *Oncogene*.

- 2012;**31**(16):2062-2074. doi.org/10.1038/onc.2011.383.
21. Monroig PdC, Calin GA. MicroRNA and epigenetics: diagnostic and therapeutic opportunities. *Curr. Pathobiol Rep.* 2013;**1**(1):43-52. doi.org/10.1007/s40139-013-0008-9.
22. Vanas V, Haigl B, Stockhammer V, Sutterlüty-Fall H. MicroRNA-21 increases proliferation and cisplatin sensitivity of osteosarcoma-derived cells. *PloS one.* 2016;**11**(8):e0161023. doi.org/10.1371/journal.pone.0161023.
23. Jalali-Qomi s, motovali-bashi m, Rezaei H, khalilian s. Experimental validation of a predicted microRNA within human FVIII gene. *Mol Biol Res Commun.* 2021:45-53. doi.org/10.22099/MBRC.2021.39067.1573
24. Berezikov E, Guryev V, van de Belt J, Wienholds E, Plasterk RH, Cuppen E. Phylogenetic shadowing and computational identification of human microRNA genes. *Cell.* 2005;**120**(1):21-24. doi.org/10.1016/j.cell.2004.12.031
25. Hoballa MH, Soltani BM, Mowla SJ, Sheikhpour M, Kay M. Identification of a novel intergenic miRNA located between the human DDC and COBL genes with a potential function in cell cycle arrest. *Mol Cell Biochem.* 2018;**444**(1-2):179-186. doi.org/10.1007/s11010-017-3242-3.
26. Dokanehiifard S, Yasari A, Najafi H, Jafarzadeh M, Nikkhah M, Mowla SJ, *et al.* A novel microRNA located in the TrkC gene regulates the Wnt signaling pathway and is differentially expressed in colorectal cancer specimens. *J Biol Chem.* 2017;**292**(18):7566-7577. doi.org/10.1074/jbc.m116.760710.