IJMS

Original Article

Bioinformatics Prediction and Experimental Confirmation of a Novel miRNA in the Fibroblast Growth Factor Receptor 1 (*FGFR1*) Gene

Maryam Parvaresh[1], PhD candidate;
Majid Motovali-Bashi[1], PhD;
Sadat Dokaneheifard[2], PhD

[1]Division of Genetics, Department of Cell and Molecular Biology and Microbiology, Faculty of Biological Science and Technology, University of Isfahan, Isfahan, Iran;
[2]Biomedical Research Building, Department of Human Genetics, Sylvester Comprehensive Cancer Center, University of Miami Miller School of Medicine, Miami, FL, 33136, USA

Correspondence:
Majid Motovali-Bashi, PhD;
University of Isfahan, Azadi Sq.,
Postal code: 81746-73441, Isfahan, Iran
Tel: +98 31 37932474
Email: mbashi@sci.ui.ac.ir
Received: 08 November 2025
Revised: 30 May 2026
Accepted: 23 June 2026

What's Known

- MicroRNAs (MiRNAs) play an important role in gene expression, influencing diverse biological processes and diseases, including cancer. However, many of them remain undiscovered.
- Many fragile sites and cancer-associated genomic regions contain miRNA genes. Analyzing cancer-amplified regions such as the fibroblast growth factor receptor 1 (*FGFR1*) genomic region helps identify novel miRNAs.

What's New

- Novel miRNAs were predicted within the *FGFR1* genomic region using several bioinformatics tools, and their predicted targets were analyzed.
- To the best of our knowledge, we experimentally validated one predicted miRNA (*FGFR1-miR1*) for the first time. Its effect on the expression of the predicted target gene, breast cancer 1 (*BRCA1*), may indicate a potential association with breast cancer.

Abstract

Background: MicroRNAs (miRNAs) are small non-coding RNAs that regulate multiple key cellular processes, including proliferation and differentiation. Since many miRNAs arise from cancer-related genomic regions, we examined the fibroblast growth factor receptor 1 (*FGFR1*) genomic region, a locus frequently altered in several cancers, to identify potential novel miRNAs.

Methods: This study was conducted at the University of Isfahan, Iran, between 2019 and 2022. The *FGFR1* region was analyzed using sequence-based miRNA prediction tools, sequence read archive (SRA) expression screening, and target-gene prediction algorithms. Among predicted precursors, one high-confidence candidate (*FGFR1-premir1*) was selected for experimental validation. *FGFR1-premir1* was transfected into HEK293T cells to assess its mature form (*FGFR1-miR1*) processing and impact on predicted target genes using real-time quantitative PCR (qPCR). Statistical analyses were conducted using GraphPad Prism (version 8.0.2), employing one-way analysis of variance (ANOVA) with Tukey's *post hoc* test and Pearson's correlation coefficient test. A $P<0.05$ was considered statistically significant.

Results: Bioinformatics screening identified 21 putative pre-miRNAs, with 11 prioritized for further analysis. Approximately 43% of predicted target genes involved cancer-associated pathways. Several precursors (including put-premir 11, 12, 16, 17, and 20) emerged as promising candidates. Experimental assays confirmed *FGFR1-premir1* produced mature *FGFR1-miR1*, and its overexpression significantly reduced breast cancer 1 (*BRCA1*) mRNA levels. Functional analysis using the Database for Annotation, Visualization and Integrated Discovery (DAVID) indicated enrichment in invasion and metastasis pathways.

Conclusion: *FGFR1-miR1* was identified as a novel miRNA derived from the *FGFR1* genomic region. Its *BRCA1* modulation suggests possible functional relevance, although further studies are necessary to validate its role and to characterize additional predicted miRNAs.

Please cite this article as: Parvaresh M, Motovali-Bashi M, Dokaneheifard S. Bioinformatics Prediction and Experimental Confirmation of a Novel miRNA in the Fibroblast Growth Factor Receptor 1 (*FGFR1*) Gene. Iran J Med Sci. doi: 10.30476/ijms.2026.109524.4486.

Keywords ● MicroRNAs ● Computational biology ● *BRCA1* ● Neoplasms ● Gene expression

Copyright: ©Iranian Journal of Medical Sciences. This is an open-access article distributed under the terms of the Creative Commons Attribution-NoDerivatives 4.0 International License. This license allows reusers to copy and distribute the material in any medium or format in unadapted form only, and only so long as attribution is given to the creator. The license allows for commercial use.

Introduction

microRNAs (miRNAs) are small (18-27 nucleotides in length) non-coding RNA molecules that function as critical post-transcriptional regulators of gene expression, primarily through mRNA degradation or translational repression.[1] Mature miRNAs, generated through sequential nuclear and cytoplasmic processing (Drosha/Dicer), are loaded into the RNA-induced silencing complex (RISC) to target mRNAs.[2] Since miRNAs can interact with multiple target genes, they are involved in many fundamental cellular processes such as proliferation, differentiation, immune responses, and apoptosis.[3] Dysregulation of miRNA expression is detected in numerous human diseases, especially in cancer, making them promising targets for novel therapeutic intervention.[4] The distribution of miRNAs is nonrandom in the human genome. A significant proportion of miRNA genes originate from genomic regions associated with cancer, such as fragile sites and cancer-related genomic regions.[5] miRNAs located within or near these regions appear to be related to chromosomal events leading to carcinogenesis. Gene amplification, a frequent genomic alteration in cancers, is a key mechanism that drives oncogene overexpression within these regions.[5]

The fibroblast growth factor receptor 1 (*FGFR1*) gene, located in the chromosomal region 8p11-12, is one of the most commonly amplified oncogenes in human cancers. *FGFR1* amplification is reported in approximately 10% of breast cancers and 19% of lung squamous cell carcinomas (LSCC). This receptor tyrosine kinase plays an essential role in multiple biological processes such as proliferation, differentiation, wound healing, regulation of cell growth, and migration.[6, 7] Given that *FGFR1* resides in a cancer-related genomic region,[6] and considering the frequent clustering of miRNA genes in such regions,[5] a thorough investigation of this locus is a logical step for identifying novel miRNAs that may regulate this key oncogene or its associated pathways.

Although over 2,600 human miRNAs have been registered in miRBase (v22.1), many more, particularly those expressed in specific tissues or developmental stages, remain undiscovered.[8, 9] Experimental methods for miRNA discovery, such as cDNA cloning, Northern blotting, microarray, and *in situ* hybridization, can be laborious and time-consuming. High-throughput methods are costly and usually limited by tissue type and developmental stage of samples. Computational approaches based on the biological features of miRNAs, such as hairpin secondary structure, thermodynamic stability, and evolutionary conservation, serve as a fast and cost-effective complementary strategy for predicting candidate miRNAs before experimental validation.[3, 10]

Our search in major biomedical databases, namely PubMed (including MEDLINE and PMC) and Scopus, and miRBase revealed numerous studies on miRNAs targeting *FGFR1*;[11, 12] however, none focused on identifying miRNAs encoded within the *FGFR1* gene. This lack of prior investigation highlights a significant gap in our understanding of the full regulatory landscape of this important oncogene. Therefore, we used bioinformatics tools to identify hairpin structures located within the human *FGFR1* gene and applied several miRNA prediction software programs to detect and analyze novel candidate miRNAs and their target genes. Among the predicted candidates, the miRNA with the highest combined score across multiple prediction tools was selected for experimental validation, including detection of the mature miRNA and assessment of its impact on predicted target genes upon overexpression.

Materials and Methods

Bioinformatics Prediction of miRNA

The complete genomic sequence of the *FGFR1* gene plus 600 bp upstream and downstream was scanned by SSCprofiler software (http://mirna.imbb.forth.gr/SSCprofiler.html)[13] in both DNA strands to identify potential miRNA precursors. Candidates exhibiting high hidden Markov model (HMM) scores, high expression levels (max_expression derived from tiling array data), and a hairpin pre-miRNA-like structure are more prominent and suitable options for pre-miRNA candidate selection.[13]

SSCprofiler-predicted precursors were subsequently analyzed in the RNAfold web server (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi) for secondary structure and minimum free energy (MFE). Candidates with a stable stem-loop structure and an MFE value below -20 kcal/mol were retained. The stem-loop structure of these candidates was further evaluated using MiREval (http://mimirna.centenary.org.au/mireval/)[14] and miRNA Spotter (http://mcube.nju.edu.cn/jwang/lab/soft/miRNA/home_miRNA.html). PHDcleav (https://webs.iiitd.edu.in/raghava/phdcleav/)[15] was used to predict the potential Dicer processing sites.

To distinguish real pre-miRNAs from pseudo hairpins, a consensus approach was employed using a suite of computational tools with different algorithms, including MiPred online software (http://server.malab.cn/MiPred/predict.

jsp) (which uses a random forest (RF) algorithm with combined features and has demonstrated 98.21% specificity and 95.09% sensitivity),[16] iMcRNA (including iMcRNA-PseSSC and iMcRNA-EXPseSSC methods) (http://bliulab.net/iMcRNA/server),[17] iMiRNA-PseDPC (http://bliulab.net/iMiRNA-PseDPC/),[18] miRNA-dis (http://bliulab.net/miRNA-dis/),[19] iMiRNA-SSF (http://bioinformatics.hitsz.edu.cn/iMiRNA-SSF/),[20] miRNA-deKmer (http://bioinformatics.hitsz.edu.cn/miRNA-deKmer/) web server.[21] A positive prediction was defined as a "real pre-miRNA" output, while for the highly stringent MiPred tool, a "pseudo pre-miRNA" classification was also considered a supporting positive result.

Mature miRNA sequences were predicted by MatureBayes (http://mirna.imbb.forth.gr/MatureBayes.html) using a Naïve Bayes classifier (NBC),[22] and FOMmiR (http://app.shenwei.me/cgi-bin/FOMmiR.cgi), which uses a Fixed-Order Markov Model based on the secondary structure pattern.[23]

Sequence similarity of predicted miRNAs against registered human and other species' miRNAs in the miRBase database (http://www.mirbase.org/)[8] was assessed via the Basic Local Alignment Search Tool for nucleotide sequences (BLASTN). The evolutionary conservation status of precursor and mature sequences was examined using the UCSC genome browser database (https://genome.ucsc.edu/). The presence of small RNA reads corresponding to the candidate was investigated in the NCBI sequence read archive (SRA) database (https://www.ncbi.nlm.nih.gov/sra). The final selection criterion required a candidate to achieve a favorable score in the majority of these classification tools.

Bioinformatics Target Gene Prediction

Target gene prediction for the candidate mature miRNAs was performed by miRDB (http://www.mirdb.org/)[24] and DIANA MR-MicroT (https://mrmicrot.imsi.athenarc.gr/)[25] tools. The involvement of *FGFR1* in various pathways was investigated using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (https://www.genome.jp/kegg/). Gene ontology (GO) annotation and KEGG pathway enrichment analysis for the predicted target genes were performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID) web server (https://david.ncifcrf.gov/).[26] The plots of the analysis results were obtained through the SRplot online server (https://www.bioinformatics.com.cn/en).

The predicted target genes were manually screened using databases, including GeneCards (https://www.genecards.org), online mendelian inheritance in man (OMIM) (https://omim.org/), network of cancer genes home (NCG7.1) (http://ncg.kcl.ac.uk/), tumor suppressor gene database (TSGene) (https://bioinfo.uth.edu/TSGene/), oncogene database (ONGene) (http://ongene.bioinfo-minzhao.org/), Cancer Gene Census (CGC) provided by the catalogue of somatic mutations in cancer (COSMIC) (https://cancer.sanger.ac.uk/census), database of disease-gene associations with annotated relationships among genes (eDGAR) (http://edgar.biocomp.unibo.it/gene_disease_db/index.html) and candidate cancer gene database (CCGD) (http://ccgd-starrlab.oit.umn.edu/index.html) to identify cancer-associated genes, including oncogenes, tumor suppressor genes, and candidate cancer genes.

Cell Culture

HEK293T cell line was cultured in DMEM-HG (Biowest, France), supplemented with 10% fetal bovine serum (FBS) (Biowest, France), 100 *μ*g/mL streptomycin (Biosera, France), and 100 U/mL penicillin (Biosera, France). The cells were incubated at 37 °C with 5% CO_2.

DNA Constructs and Sequencing

This study was approved by the Research Ethics Committee of the University of Isfahan, Iran, under code IR.UI.REC.1399.085. Written informed consent was obtained from the participants. Genomic DNA was extracted from healthy human whole blood based on a salting-out method.[27] The genomic region containing the stem-loop structure corresponding to the *FGFR1-premir1* candidate was PCR-amplified using Taq DNA polymerase (Ampliqon, Denmark) with specific primers (table 1). The PCR product was purified using a Gel Purification Kit (GeNetBio, Korea), cloned into the pUCm-T vector (Bio Basic, Canada), and subsequently subcloned into the pEGFP-C1 expression vector (Clontech, CA) downstream of the green fluorescent protein (GFP) gene after digestion with *EcoRI* and *BamHI* restriction enzymes (Thermo Scientific, USA). Recombinant vectors were confirmed by colony PCR and direct sequencing (Pishgam Biotech Co., Iran). A previously described scrambled sequence was cloned into pEGFP-C1,[28] and empty pEGFP-C1 vectors (mock) were used as negative controls.

Additionally, cDNAs from HEK293T cells were amplified with specific primers to potentially detect the mature miRNA. The resulting ~80 bp product was purified using a Gel Purification Kit (GeNetBio, Korea), cloned into the pDrive vector (Qiagen, USA), and sequenced to confirm the correct insert (Pishgam Biotech Co, Iran).

Table 1: List of primers and oligo sequences used in the study

Primer name	Primer sequence (5′ to 3′)	Amplicon size (base pairs)	Usage
FGFR-12-F5	CCTGGAGTTTCTGGCTGTCCTT		Amplification of predicted mature miRNA
FGFR-12-F5-2	AGTTTCTGGCTGTCCTTGGAA		
FGFR-12-F8	GGTTCCAGGAAAACCAGCCTC		
Universal-outer	GCGTCGACTAGTACAACTCAAG		
Anchored OligodT	GCGTCGACTAGTACAACTCAAGGTTCTTCCAGTCACGA CG(T)18N		cDNA synthesis
FGFR1-12-Intron2	Int2-F: GAGAGCAAACCACCTAAAACCC Int2-R: CCACCCCAACTACCAGCAG	524	Amplification of predicted miRNA precursor
SNORD47-F (*U47*)	Forward: ATCACTGTAAAACCGTT		Amplification of reference gene
SNORD48-F (*U48*)	Forward: TGACCCCAGGTAACTCTGAGTGTGT		
GAPDH	Forward: GCCACATCGCTCAGACAC Reverse: GGCAACAATATCCACTTTACCAG	115	
β-actin	Forward: GTCATTCCAAATATGAGATGCGT Reverse: CAATGCTATCACCTCCCCTG	125	
BRCA1	Forward: TTTGCCAGAAAACACCACATCAC Reverse: CCGTTCACACACAAACTCAGC	96	Amplification of target gene
SMAD2	Forward: GGAGCAGAATACCGAAGGC Reverse: CTTTTCATGGGACTTGATTGGTG	166	
TA vector primer	M13 Forward: GTAAAACGACGGCCAGT M13 Reverse: CAGGAAACAGCTATGAC		Colony PCR
pEGFP-C1 primer	Forward: TACAAATGTGGTATGGCTGATTATG Reverse: AGTCCGCCCTGAGCAAAG		

U47: Small nucleolar RNA, C/D box 47; *U48*: Small nucleolar RNA, C/D box 48; *GAPDH*: Glyceraldehyde-3-phosphate dehydrogenase; *β-actin*: Actin beta; *BRCA1*: Breast cancer 1; *SMAD2*: SMAD family member 2; miRNA: MicroRNA

Plasmids for sequencing were isolated using a Plasmid DNA Extraction Kit (GeNetBio, Korea).

Transfection of Candidate Pre-miRNA

HEK293T cells were seeded in 6-well plates and transfected at 70-80% confluency with the pEGFP-C1 expression vector carrying the *FGFR1-premir1* sequence using TurboFect™ Transfection Reagent (Thermo Scientific, USA) following the manufacturer's instructions. Mock and scrambled vectors were transfected as negative controls. The experiments were performed in three independent biological replicates. Successful transfection and GFP expression were confirmed 48 hours post-transfection by inverted fluorescence microscopy (ZEISS Axiovert 25, Germany), indicating a high transfection efficiency, with over 70% of HEK293T cells showing robust GFP expression (figure 1A).

RNA Extraction, cDNA Synthesis, and Real-Time Quantitative PCR (qPCR)

Total RNA was extracted using Trizol reagent according to the manufacturer's instructions (Life Biolab, Germany), and treated with RNase-free DNase I (Thermo Scientific, USA). The purity and concentration of extracted RNAs were estimated by NanoDrop (Thermo Scientific, USA), with quality confirmed by 2% agarose gel electrophoresis. The 3′-Poly-A tail was added to RNAs using Poly-A polymerase (Bonyakhteh, Iran). Complementary DNA (cDNA) was synthesized from polyadenylated RNAs using reverse transcriptase (Bonyakhteh, Iran) and the specific anchored-oligo-dT primer (table 1).

The obtained cDNAs were amplified to detect candidate mature miRNA (*FGFR1-miR1*) using specific primers in a StepOne™ Real-Time PCR system (Applied Biosystems, USA). Amplification was done under the following conditions: initial denaturation at 95 °C for 15 min, followed by 45 cycles of 95 °C for 5 sec, 61.5 °C for 10 sec, and 72 °C for 15 sec, with RealQ Plus 2x Master Mix Green (Ampliqon, Denmark). Each experiment was replicated at least three times.

Primer efficiency was calculated through LinRegPCR software (version 2017.1). For data normalization, small nucleolar RNA, C/D box 47 (*SNORD47* [*U47*]), small nucleolar RNA, C/D box 48 (*SNORD48* [*U48*]), glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), and actin beta (*β-actin*) were used as the reference genes. These genes were selected based on previous studies revealing their stable expression in human cell lines and tissues,[29, 30] and their successful use in HEK293T cells was demonstrated.[31]

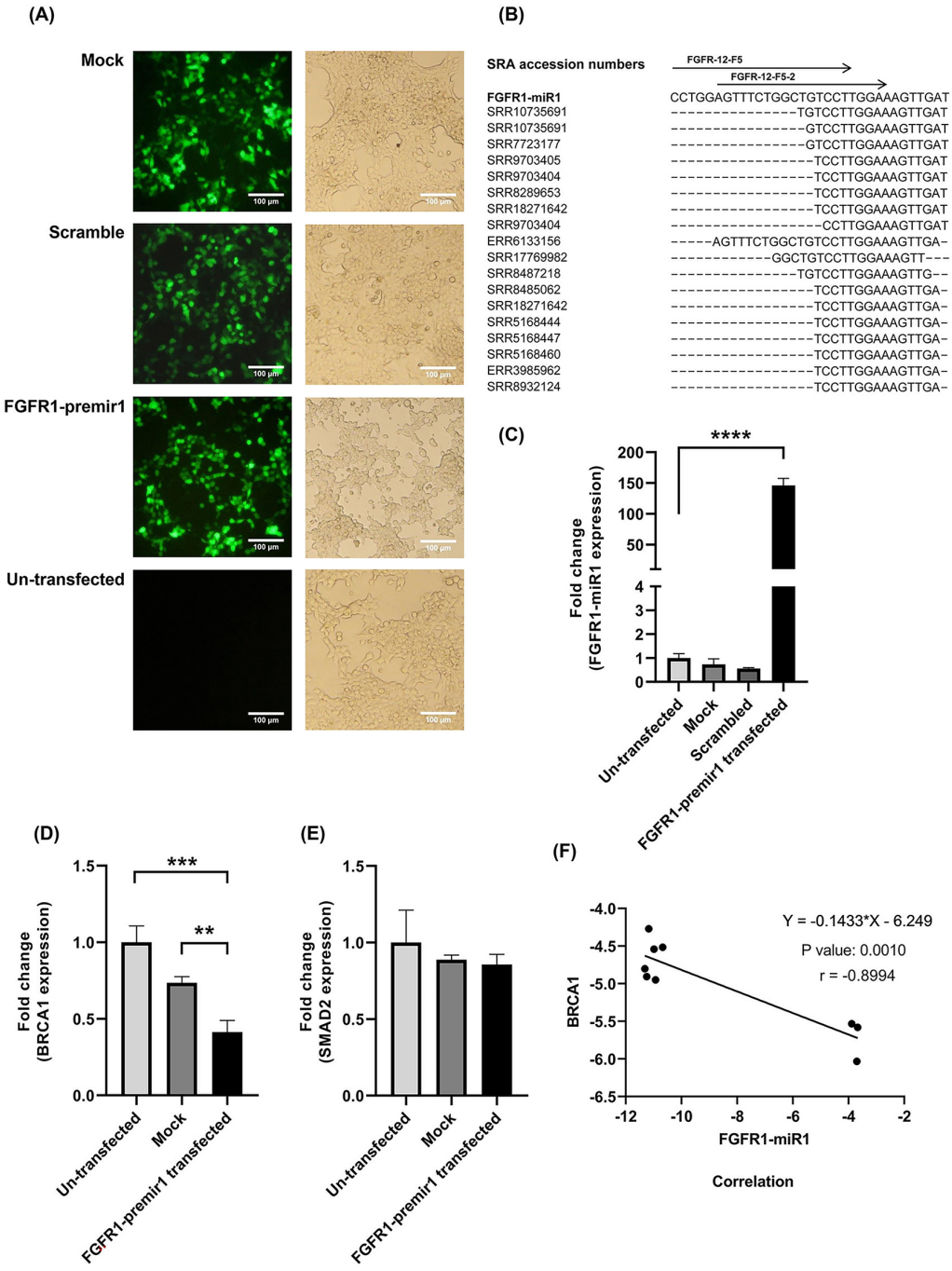


Figure 1: Estimation of transfection efficiency through GFP emission, SRA analysis of the *FGFR1-miR1* sequence, and evaluation of exogenous *FGFR1-miR1* expression and its predicted target genes in the HEK293T cell line are presented. (A) Transfection efficiency is demonstrated by observing the expression of GFP protein in HEK293T cells transfected with the expression vector containing *FGFR1-premir1*, mock, and scramble vectors, using fluorescence microscopy (at 10x magnification). (B) This panel shows the identified reads from SRA analysis covering the 3′ side of the *FGFR1-miR1* sequence. The black arrows indicate the sequence region of the two FGFR-12-F5 and FGFR-12-F5-2 primers. (C) The result of RT-qPCR indicated an approximately 146-fold overexpression of *FGFR1-miR1* in the transfected HEK293T cell line containing the put-premir12 precursor with a significant P value of <0.0001 compared to the control. (D) The *BRCA1* target gene is shown to be downregulated following the *FGFR1-miR1* overexpression in comparison with the related controls. (E) The *SMAD2* target gene expression level was not significantly altered following the overexpression of *FGFR1-miR1* in the HEK293T cells. *U47* RNA, *U48* RNA, *GAPDH*, and *β-actin* were used as internal controls. Error bars indicate the standard deviation (SD) of triplicate experiments. Statistical significance for expression analyses was determined using one-way ANOVA. (F) The graph demonstrates that a significant negative correlation was obtained between *FGFR1-miR1* and *BRCA1* expression levels, with a correlation coefficient of r=-0.8994 (P=0.0010), analyzed by Pearson's correlation test. (**P<0.01; ***P<0.001; ****P<0.0001). GFP: Green fluorescent protein; SRA: Sequence read archive; RT-qPCR: Reverse transcription-quantitative polymerase chain reaction; *BRCA1*: Breast cancer 1; *SMAD2*: SMAD family member 2; *GAPDH*: Glyceraldehyde-3-phosphate dehydrogenase; *β-actin*: Actin beta; *U47*: Small nucleolar RNA, C/D box 47; *U48*: Small nucleolar RNA, C/D box 48; ANOVA: Analysis of variance; SD: Standard deviation

Additionally, *U47* has been confirmed as a suitable internal control by Bonyakhteh Company (https://bonbiotech.ir). Expression data were analyzed using the $2^{-\Delta CT}$ and $2^{-\Delta\Delta CT}$ methods.[32]

Statistical Analysis

Statistical analysis was performed by GraphPad Prism software, version 8.0.2 (GraphPad, San Diego, CA, USA). The one-way analysis of variance (ANOVA) followed by Tukey's *post hoc* test was utilized for multiple group comparisons. Prior to ANOVA analysis, assumptions of normality and homogeneity of variances were verified using the Shapiro-Wilk test and Brown-Forsythe test, respectively. A $P<0.05$ was considered a significant difference. Pearson's correlation coefficient test was used to evaluate the existence of a correlation between candidate miRNA and the target gene expression levels. This analysis was performed on the mean expression values derived from three independent experiments for each of the three cell groups (Un-transfected, Mock, and Transfected), resulting in a total of nine data pairs for the analysis.

Results

Bioinformatics Prediction of Novel miRNAs

The *FGFR1* genomic sequence was scanned by the SSCprofiler tool, which identified 24 and 29 precursor miRNAs on the direct and reverse strands, respectively, based on their characteristic features.[13] Considering that nearly half of the known human miRNAs are located in intronic regions[33] and are often evolutionarily conserved, we selected 21 conserved pre-miRNA candidates located in the intronic and untranslated regions (UTRs) of the *FGFR1* gene for further evaluation. These candidates, designated put-premir1 to put-premir21 (table 2), were subjected to comprehensive bioinformatics analysis to predict hairpin structures and distinguish real pre-miRNAs from pseudo-miRNAs (supplementary tables S1 and S2). The stability of their hairpin secondary structures was confirmed using the RNAfold program, showing high similarity with the SSCprofiler structure predictions for a majority of candidates (figure 2 and supplementary figure S1). Subsequently, the candidates were evaluated for the prediction of their mature miRNAs using tools such as FOMmiR and MatureBayes (supplementary figure S2 and supplementary table S3). FOMmiR confirmed stem-loop structures and mature miRNA sequences within the precursor for only two candidates (put-premir12 and 19) (supplementary figure S2). In addition, Dicer processing sites were predicted by PHDcleav for 13 of the pre-miRNA sequences (supplementary table S2). Using the UCSC genome browser, the presence of small RNA sequences across various

Table 2: List of the pre-miRNAs predicted by SSCprofiler

Number	Predicted pre-miRNA	Position on genome	Strand	Max_ Expression[a]	Max_Exp_ location[b]	HMMScore[c]	MFE[d]
1	Put-premir1	chr8, 38698654 38698757	top_strand	17.0	38698717	4.7	-36.0
2	Put-premir2	chr8, 38706497 38706600	top_strand	75.0	38706533	3.4	-26.9
3	Put-premir3	chr8, 38707482 38707585	top_strand	32.0	38707497	7.0	-32.7
4	Put-premir4	chr8, 38711691 38711794	top_strand	23.0	38711754	9.6	-31.1
5	Put-premir5	chr8, 38719395 38719498	top_strand	11.0	38719488	11.4	-32.3
6	Put-premir6	chr8, 38742910 38743013	top_strand	25.0	38742949	6.7	-26.8
7	Put-premir7	chr8, 38716112 38716215	bottom_strand	27.0	38716172	10.5	-29.8
8	Put-premir8	chr8, 38717252 38717355	bottom_strand	12.5	38717311	6.9	-33.6
9	Put-premir9	chr8, 38727409 38727512	bottom_strand	109.0	38727414	5.5	-34.1
10	Put-premir10	chr8, 38727597 38727700	bottom_strand	7.0	38727625	12.3	-29.1
11	Put-premir11	chr8, 38728456 38728559	bottom_strand	21.5	38728467	9.4	-41.4
12	Put-premir12	chr8, 38728989 38729092	bottom_strand	15.5	38729051	14.7	-49.2
13	Put-premir13	chr8, 38737818 38737921	bottom_strand	449.5	38737917	8.1	-28.52
14	Put-premir14	chr8, 38741486 38741589	bottom_strand	31.0	38741583	3.1	-25.8
15	Put-premir15	chr8, 38741492 38741595	bottom_strand	31.0	38741583	3.1	-27.4
16	Put-premir16	chr8, 38742644 38742747	bottom_strand	12.5	38742666	7.6	-35.4
17	Put-premir17	chr8, 38744473 38744576	bottom_strand	597.5	38744528	9.3	-35.5
18	Put-premir18	chr8, 38690440 38690543	bottom_strand	33.5	38690453	4.9	-38.6
19	Put-premir19	chr8, 38745116 38745219	top_strand	70.0	38745216	7.6	-56.6
20	Put-premir20	chr8, 38745349 38745452	top_strand	158.0	38745394	11.6	-43.2
21	Put-premir21	chr8, 38745629 38745732	bottom_strand	738.0	38745640	10.0	-54.3

Positions on the genome are based on the human genome coordinates version – hs1; [a]Max_Expression: Maximum expression; [b]Max_Exp_location: Maximum expression location; [c]HMM: Hidden Markov model; [d]MFE: Minimum free energy; hs1: *Homo sapiens* 1 (T2T-CHM13 genome assembly)

cell lines and tissues was detected within 17 candidates (except the put-premir4, 5, 8, and 16) providing valuable clues regarding the potential locations of the mature miRNA sequences within the precursors. A BLAST search against the miRBase database revealed that none of the predicted miRNA sequences were similar or identical to previously known miRNAs.

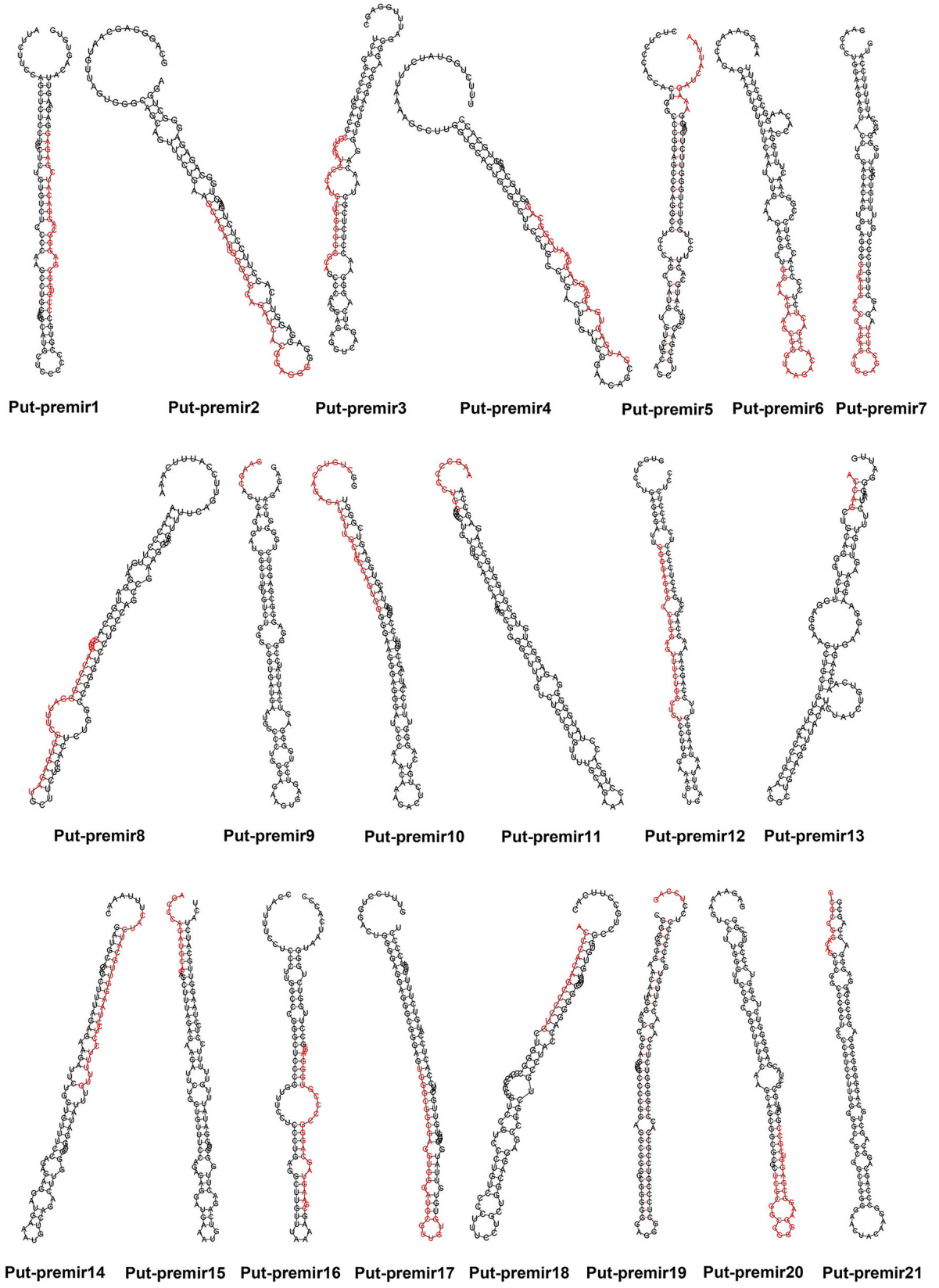


Figure 2: This figure illustrates the secondary structures of pre-miRNAs predicted by the SSCprofiler tool. Sequences of the putative mature miRNAs are shown in red. miRNA: MicroRNA

Based on consistent validation across multiple prediction tools, 11 pre-miRNA candidates (put-premir1, 4, 7, 9, 11, 12, 16, 17, 19, 20, and 21) were chosen for further in-depth analysis. To find evidence of their expression, small RNA sequencing data from the NCBI-SRA database related to different tumors and breast cancer cell lines were analyzed (Online table S4). This analysis revealed that six of the eleven candidate pre-miRNAs (put-premir20, 7, 11, 12, 16, and 17) had a higher number of aligned reads compared to the others, particularly put-premir11 and 16, which revealed 140 and 169 aligned reads (corresponding to the MatureBayes-detected miRNAs put-miR11-2 and put-miR16-2), respectively (table 3, supplementary figures S3 and S4). While the number of aligned reads was generally low, as is common for unknown predicted or condition-specific miRNAs, their detection can provide initial support for the expression of these predicted sequences.

Selection of Candidate Precursor for Experimental Assays

Among the 11 predicted pre-miRNA candidates, put-premir12 was selected for experimental validation due to having the highest level of confirmation across the utilized bioinformatics tools. Its pre-miRNA structure was validated by the majority of prediction tools using different algorithms, exhibiting a highly stable secondary structure with the highest HMMScore among the evaluated candidates (table 2), which indicates a high likelihood and specificity of being a true miRNA precursor.[13] Crucially, prediction tools such as FOMmiR not only validated its hairpin structure but also successfully identified a potential mature miRNA sequence within it. Moreover, the alignment of several reads from the SRA datasets to this sequence provided initial evidence of its expression.

Table 3: BLAST results of candidate sequences obtained from the six pre-miRNAs with selected SRXs in the SRA database

Pre-miRNA	Candidate region		Candidate sequence used in SRA with aligned read sequences	SRA accession number	Spots number	Aligned read length
Put-premir20	CSHL small RNA-seq track in UCSC	Candidate sequence:	AAAAGTCCTTGGGTTCCGCGGCT			
		Aligned reads:	--AAGTCCTTGGGTTCCG----- -AAAGTCCTTGGGTTCCG----- -AAAGTCCTTGGGTTCCG----- -AAAGTCCTTGGGTTCCG-----	SRR5230626 SRR5230626 SRR1973961 SRR1973966	11 8 4 2	16 nt 17 nt 17 nt 17 nt
	Mature Bayes (3′ stem)	Candidate sequence:	CGAGGGGTCTCGGTCCCGTCCG			
		Aligned reads:	CGAGGGGTCTCGGCCCCGT--- CGAGGGGTCTCGGCCCCGTC-- CGAGGGGTCTCGGTCC------	SRR1752097 SRR1752097 SRR4103657	7 1 4	19 nt 20 nt 16 nt
	Mature Bayes (5′ stem)	Candidate sequence:	TTTCAAGCAGCGGCGCGCTCGC			
		Aligned read:	-----AGCAGCGGCGCGCTCG-	SRR4103583	3	16 nt
Put-premir7	SSCprofiler putative mature miRNA	Candidate sequence:	GCCAGGACCCAGCAGATGCAGCCTCT			
		Aligned reads:	---------CAGCAGATGCAGCCTC- -CCAGGACCCAGCAGATGCAG-----	SRR4103657 SRR5329396	9 1	16 nt 20 nt
	Mature Bayes (5′ stem)	Candidate sequence:	AGTACCTGGACACAGTGGAGGG			
		Aligned reads:	---ACCTGGACACAGTGGA--- ---ACCTGGACACAGTGGA--- ---ACCTGGACACAGTGGA---	DRR036718 DRR036709 SRR2915928	6 4 3	16 nt 16 nt 16 nt
Put-premir11	Mature Bayes (3′ stem)	Candidate sequence:	CCTATGGGGGAGAGGCTGTGCG			
		Aligned reads:	----TGGGGGAGAGGCTGTGCG ----TGGGGGAGAGGCTGTGCG ----TGGGGGAGAGGCTGTG-- ----TGGGGGAGAGGCTGTG-- ----TGGGGGAGAGGCTGTG--	SRR6512913 SRR4103657 SRR5329396 SRR5329398 SRR5689213	140 2 20 8 3	18 nt 18 nt 16 nt 16 nt 16 nt
	26nt selected sequence	Candidate sequence:	ATGGGGGAGAGGCTGTGCGTGGTGCC			
		Aligned reads:	-TGGGGGAGAGGCTGTGCGTGG---- -TGGGGGAGAGGCTGTGCGT------	SRR6512913 SRR4103657	13 1	21 nt 19 nt
	30nt selected sequence (including Mature Bayes (5′ stem))	Candidate sequence:	GTCTCTGTGTTTTGCAGAAACCTGCACCTA			
		Aligned reads:	GTCTCTGTGTTTTGCAGA---------- GTCTCTGTGTTTTGCAGA---------- ----CTCTGTGTTTTGCAGAA--------- ----CTCTGTGTTTTGCAGA---------- ----CTCTGTGTTTTGCAGA----------	SRR5230577 SRR5230576 SRR5230576 SRR5230576 SRR5230626	1 1 1 3 3	18 nt 18 nt 17 nt 16 nt 16 nt

Pre-miRNA	Candidate region		Candidate sequence used in SRA with aligned read sequences	SRA accession number	Spots number	Aligned read length
Put-premir12	Mature Bayes (5′ stem)	Candidate sequence:	TGAGGGATTGGGGAGGGCCTGG			
		Aligned reads:	-GAGGGATTGGGGAGGG-----	SRR5329396	2	16 nt
			--AGGGATTGGGGAGGG-----	SRR5329396	4	15 nt
			--------TGGGGAGGGCCTGG	SRR6512913	19	14 nt
	CSHL small RNA-seq track in UCSC	Candidate sequence:	CCTGGAGTTTCTGGCTGTCCTTG			
		Aligned reads:	CCTGGAGTTTCTGGCTGTCCTTG	SRR4103583	1	23 nt
			------GTTTCTGGCTGTCCT--	DRR036709	3	15 nt
	30nt selected sequence	Candidate sequence:	GAGTTTCTGGCTGTCCTTGGAAAGTTGATT			
		Aligned reads:	GAGTTTCTGGCTGTCCTTG----------	SRR4103583	1	19 nt
			-----TCTGGCTGTCCTTGGAA--------	DRR036718	5	17 nt
			-----TCTGGCTGTCCTTGGAA--------	DRR036709	2	17 nt
			------CTGGCTGTCCTTGGAA--------	DRR036709	15	16 nt
			------CTGGCTGTCCTTGGAA--------	DRR036718	12	16 nt
			------CTGGCTGTCCTTGGAA--------	SRR2915928	11	16 nt
			----------TCCTTGGAAAGTTGA--	SRR5168444	10	15 nt
			-------TGGCTCTCCTTGCAAAGTTGAT-	SRR5329398	19	22 nt
	FOMmiR (seq_3p)	Candidate sequence:	TGATTTAATAAGGGTTCCAGGAAA			
		Aligned reads:	------AATAAGGGTTCCAGGA--	DRR036709	1	16 nt
			---------AAGGGTTCCAGGAAA	SRR2061211	23	15 nt
Put-premir16	Mature Bayes (3′ stem)	Candidate sequence:	CCGTGGGACGCGCCTTGGTCTG			
		Aligned reads:	--GTGGGACGCGCCTTGG----	SRR1646493	169	16 nt
			--GTGGGACGCGCCTTGG----	SRR4103657	7	16 nt
			--GTGGGACGCGCCTTGG----	SRR4103533	5	16 nt
			----GGGACGCGCCTTGGTC--	SRR1646493	5	16 nt
			--GTGGGACGCGCCTTGGTC--	SRR4103533	1	18 nt
Put-premir17	Mature Bayes (3′ stem)	Candidate sequence:	TGTGTTGCATTCCACTCCATGT			
		Aligned reads:	---GTTGCATTCCACTCCAT--	SRR4103657	3	17 nt
			----TTGCATTCCACTCCAT--	SRR4103657	11	16 nt
			---GTTGCATTCCACTCCA---	SRR4103657	11	16 nt
	CSHL small RNA-seq track in UCSC	Candidate sequence	GTTGTGTGTGTTATGTGTGTGTGTTG			
		Aligned reads:	---GTGTGTGTTGTGTGTGTGTGTT-	SRR6512913	15	22 nt
			--TGTGTGTGTTGTGTGTGTGTGT--	SRR4103657	4	22 nt
			--TGTGTGTGTTATGTGTGTGT----	SRR4103533	4	20 nt
			--------TGTTATGTGTGTGTGTT-	SRR5168449	1	17 nt

Different nucleotides compared to the candidate sequences are shown in red and underlined; SRA: Sequence Read Archive; SRX: SRA experiment set; CSHL: Cold Spring Harbor Laboratory

Furthermore, its location within a highly conserved region of the second intron of the *FGFR1* gene suggested its potential functional importance. These multi-feature criteria made it a highly promising candidate for experimental investigation.

Experimental Detection of Mature miRNA from the Put-Premir12

The pEGFP-C1 expression vector containing the sequence of predicted put-premir12 was transfected into the HEK293T cell line. Transfection efficiency was confirmed through GFP fluorescence (figure 1A). Following RNA extraction and cDNA synthesis, PCR amplification using specific primers designed against the predicted mature regions (FGFR-12-F5 and FGFR-12-F8) produced a distinct ~80 bp product on a 3% agarose gel. Amplified PCR products were purified, cloned, and sequenced to confirm the identity of the predicted miRNA.

Sequencing of three out of four independent clones from FGFR-12-F5 amplification revealed an identical sequence to the precursor, confirming the presence of the predicted miRNA. The sequence was submitted to the NCBI database under the accession number OP795928. The predicted mature miRNA and its precursor were named FGFR1-miR1 and FGFR1-premir1, respectively. The physical location, hairpin structure, and high evolutionary conservation of *FGFR1-premir1* and its precursor are shown in figures 3A, 3B, and 3C. It was noted that the initial FGFR1-miR1 sequence obtained was longer than a natural mature miRNA, likely due to the primer binding site to the miRNA sequence. Further confirmation with a new primer (FGFR-12-F5-2), designed closer to the 3′ end of the miRNA sequence, yielded the expected 80 bp PCR product.

Furthermore, multiple attempts to analyze SRA from different studies revealed several reads aligned with the FGFR1-miR1 sequence, which may support its expression in other studies.

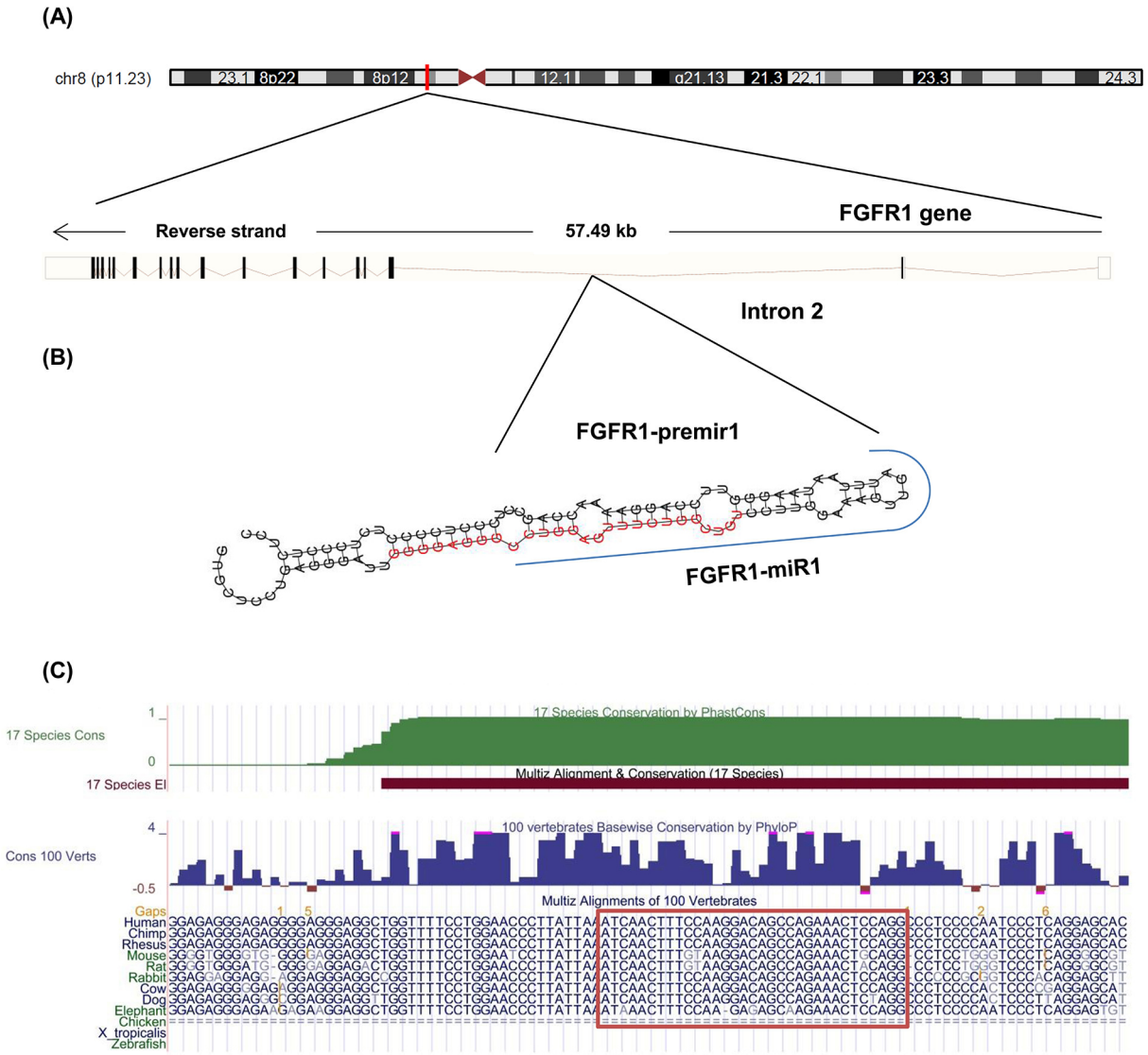


Figure 3: The figure depicts the prediction of *FGFR1-miR1* and its precursor within the second intron of the human *FGFR1* gene. (A) A schematic presentation shows the *FGFR1* gene exons (rectangular shape) and introns (broken lines) adapted from Ensembl. The approximate position of the predicted precursor (*FGFR1-premir1*), located on the short arm of chromosome 8, is displayed within the second intron of the *FGFR1* gene. (B) This panel illustrates the predicted stem-loop structure of the *FGFR1-premir1* encoding the *FGFR1-miR1*. The red sequence represents the predicted mature form of miRNA identified by the SSCprofiler. (C) UCSC Genome Browser BLAT search demonstrates strong conservation of the *FGFR1-premir1* and *FGFR1-miR1* among several organisms. The bluish, brown, and green area diagrams display the conservation status. The red box indicates the *FGFR1-miR1* sequence area. *FGFR1*: Fibroblast growth factor receptor 1; miRNA: MicroRNA; BLAT: BLAST-Like Alignment Tool; Chr8: Chromosome 8

Specifically, we detected several reads that cover the 3′-side of the sequence. Among these, some aligned reads were one to three nucleotides shorter at the 3′ end compared to the candidate miRNA (figure 1B).

To evaluate *FGFR1-miR1* expression in HEK293T transfected cells compared to the mock, scramble, and non-transfected cells, we used real-time qPCR. *U47* and *U48* were used as the reference genes. The transfected cells containing *FGFR1-premir1* revealed a 146-fold (P<0.0001) overexpression of *FGFR1-miR1* versus the non-transfected cells (figure 1C).

Target Gene Prediction of Candidate miRNAs

Using the miRDB database, target genes for the candidate mature miRNAs (table 4) were predicted based on 3′ UTR seed sequence complementarity through the MirTarget algorithm.[24] Initial screening identified 1014 target genes with scores above 90 for all predicted mature miRNAs (supplementary table S5). Manual screening of target genes with the mentioned databases revealed 435 cancer-related genes (~42.9%). This included genes directly implicated in breast and lung cancers, as well as known tumor suppressors

and oncogenes related to breast cancer (supplementary tables S6 and S7). Among the candidates, miRNAs derived from put-premir12, 11, and 20 (specifically put-miR12-4, 11-1, 20-1, and 11-3) exhibited the highest number of target genes with a score >90. Notably, put-miR20-1 targeted the highest percentage of genes related to breast and lung cancers, with 10.3% and 12.3%, respectively.

Identification of FGFR1-miR1 Target Genes

The experimentally detected sequence of *FGFR1-miR1* was longer than the typical miRNA length due to the incorporation of the primer binding site, which obscured the precise start of its 5′ end. This 5′ region, containing the miRNA seed sequence (first 2-7 nucleotides), is essential for target gene identification and is the basis of the main functionality of target gene prediction tools.[3, 34] Therefore, even a single nucleotide alteration at the 5′ end can significantly change predicted targets.

To mitigate this limitation and identify potential *FGFR1-miR1* target genes, we evaluated ten proposed miRNA sequences ranging from 18 to 27 nucleotides in length from the 3′ side of the detected sequence using the DIANA MR-microT tool. Among the detected genes, 2834 common target genes were extracted across all ten proposed sequence variations. From these, breast cancer 1 (*BRCA1*) and SMAD family member 2 (*SMAD2*) were selected as suitable candidates for experimental validation because both (a) appeared among the common targets of all ten variants, (b) received high prediction scores (mean scores of 0.84 and 0.91, respectively), and (c) are well-known tumor suppressors/cancer-associated genes, implicated in the enriched cancer-related pathways. Notably, the prediction scores varied with the length of the miRNA sequence. *SMAD2* exhibited higher scores at 18 nucleotides and for lengths between 20-23 nucleotides, whereas *BRCA1* showed higher scores at 19 nucleotides and for lengths between 24-27 nucleotides.

Table 4: List of putative mature miRNA sequences used for target prediction

Predicted pre-miRNA	Putative mature miRNA (Candidate region)	Mature miRNAs sequence
Put-premir1	Put-miR1-1 (MatureBayes 5'stem)	CATGCTCCCCCCGTGCCCGTGG
	Put-miR1-2 (MatureBayes 3'stem)	CGTGGCGAGGGCAGGACATCGA
	Put-miR1-3 (UCSC putative mature miR)	TCTCCCCAAGCCTGGAAATG
Put-premir4	Put-miR4-1 (MatureBayes 3'stem)	GAGGAGCAGGAATGGGCAGACT
	Put-miR4-2 (SSCpofiler putative mature miR)	GATGAGTGAGGAGCAGGAATGGGCAG
Put-premir19	Put-miR19-1 (MatureBayes 5'stem)	TGGTGCCCCGGAGGCGGGGCGG
	Put-miR19-2 (FOMmiR putative mature miR 3p)	TCCCGTCCGCCACCCGGGGT
Put-premir20	Put-miR20-1 (MatureBayes 5'stem)	TTTCAAGCAGCGGCGCGCTCGC
	Put-miR20-2 (MatureBayes 3'stem)	CGAGGGGTCTCGGTCCCGTCCG
	Put-miR20-3 (UCSC putative mature miR)	AAAGTCCTTGGGTTCCGCGGC
Put-premir7	Put-miR7-1 (MatureBayes 5'stem)	AGTACCTGGACACAGTGGAGGG
	Put-miR7-2 (MatureBayes 3'stem)	AGAGCTTGTCCCTGTTTTGTGA
	Put-miR7-3 (SSCpofiler putative mature miR)	GCCAGGACCCAGCAGATGCAGCCTCT
Put-premir9	Put-miR9-1 (MatureBayes 5'stem)	ATGGCTTGTGTCTGGGCGGTGA
	Put-miR9-2 (MatureBayes 3'stem)	TGGGGAGTCATTATCCGGGAGG
Put-premir11	Put-miR11-1 (MatureBayes 5'stem)	TTTTGCAGAAACCTGCACCTAT
	Put-miR11-2 (MatureBayes 3'stem)	CCTATGGGGGAGAGGCTGTGCG
	Put-miR11-3 (UCSC putative mature miR)	TGGGGGAGAGGCTGTGCGTGGT
Put-premir12	Put-miR12-1 (MatureBayes 5'stem)	TGAGGGATTGGGGAGGGCCTGG
	Put-miR12-2 (UCSC putative mature miR)	CCTGGAGTTTCTGGCTGTCCTTG
	Put-miR12-3 (Selected sequence)	TCTGGCTGTCCTTGGAAAGTTGAT
	Put-miR12-4 (FOMmiR (seq_3p))	TGATTTAATAAGGGTTCCAGGAAA
Put-premir16	Put-miR16-1 (MatureBayes 5'stem)	AGGCTTGTTTAAAAGCGAAGTA
	Put-miR16-2 (MatureBayes 3'stem)	CCGTGGGACGCGCCTTGGTCTG
	Put-miR16-3 (SSCpofiler putative mature miR)	CGAAGTAGCAGGGCCCCGTGGGACGC
Put-premir17	Put-miR17-1 (MatureBayes 5'stem)	GGGAGGAGGGCGGGAGTGGGCG
	Put-miR17-2 (MatureBayes 3'stem)	TGTGTTGCATTCCACTCCATGT
	Put-miR17-3 (SSCpofiler putative mature miR)	TGGGCGGCGAGGTGGGATGCGTTGTG
	Put-miR17-4 (UCSC putative mature miR)	GTTGTGTGTGTTATGTGTGTGTGTTG
Put-premir21	Put-miR21-1 (MatureBayes 5'stem)	CCGTCCTTTGGGGCCGCGGCTG
	Put-miR21-2 (MatureBayes 3'stem)	AGGCAGCTGCAGGGGGCGGAGG
	Put-miR21-3 (UCSC putative mature miR)	GGGGGCGGAGGCGGAGGAGGGA

miRNA: MicroRNA

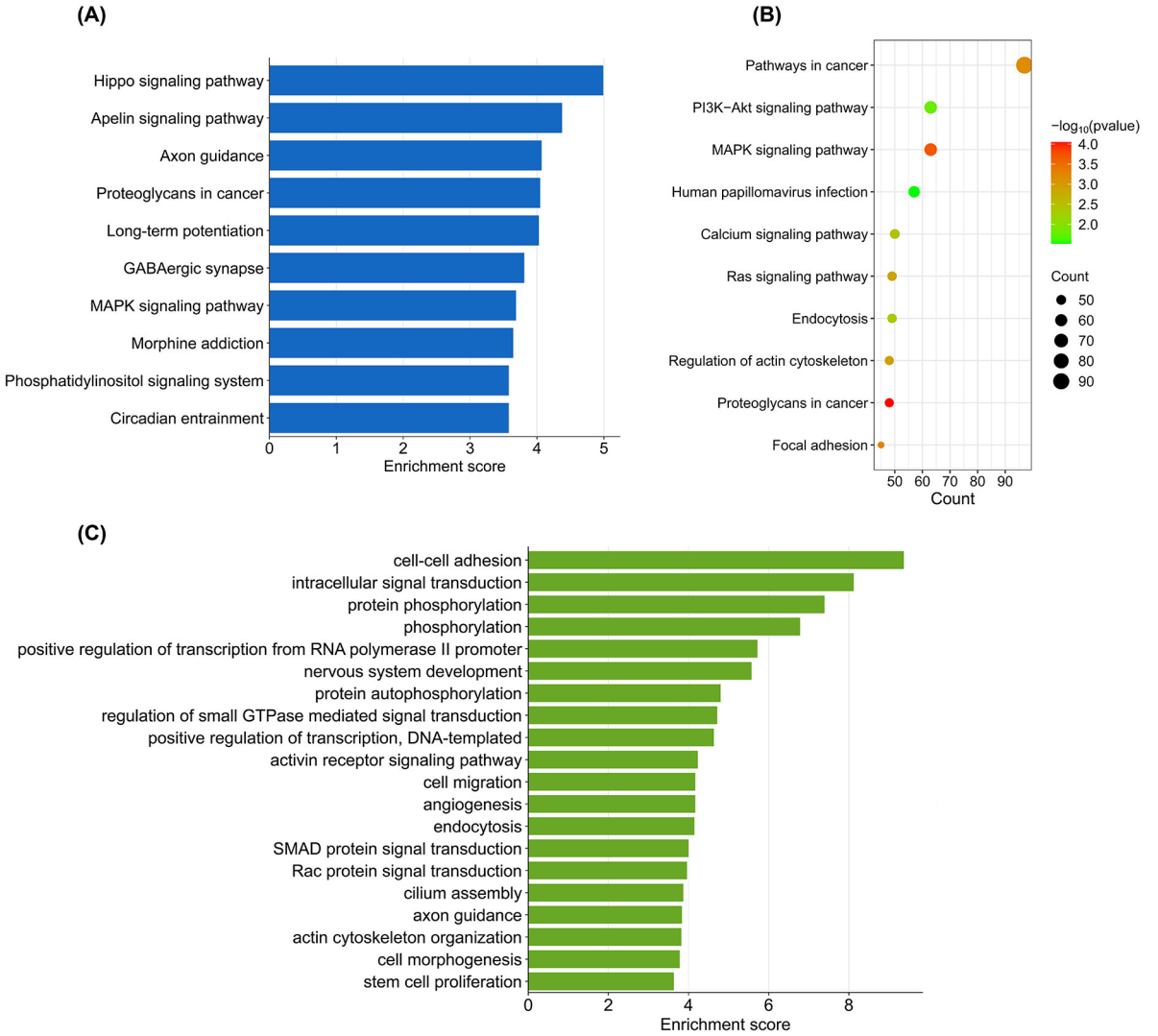


Figure 4: Significant KEGG pathways and biological processes (BP) associated with the candidate target genes are summarized. (A) This panel displays the top 10 most significant KEGG pathways of *FGFR1-miR1* candidate target genes. (B) The bubble plot illustrates the top 10 enriched KEGG pathways related to *FGFR1-miR1* that involve the highest count of associated target genes. (C) This graph highlights top 20 enriched biological processes for *FGFR1-miR1* target genes. Statistical significance for the enrichment analysis was determined using the EASE score (a modified Fisher's exact test) via the DAVID database (P<0.05). BP: Biological process; KEGG: Kyoto Encyclopedia of Genes and Genomes; DAVID: Database for annotation, visualization and integrated discovery

Functional Analysis of FGFR1-miR1 and Its Target Genes

The common target gene list was analyzed using the DAVID tool (supplementary files 2, 3, and 4). KEGG pathway enrichment analysis revealed that seven of the top ten most significant pathways were related to cancer (figure 4A and supplementary table S8). The top ten significant pathways with the highest number of genes are listed in figure 4B. Furthermore, pathways such as the MAPK signaling pathway, Ras signaling pathway, PI3K-Akt signaling pathway, Breast cancer, and Rap1 signaling pathway, which are also related to the host gene (*FGFR1*), were detected by KEGG analysis. GO annotation indicated the involvement of target genes in cancer-related biological processes such as cell-cell adhesion, intracellular signal transduction, angiogenesis, and cell migration (figure 4C).

To validate the regulatory effect of putative *FGFR1-miR1* on the selected targets *BRCA1* and *SMAD2,* real-time qPCR was performed. The analysis indicated significant down-regulation of the *BRCA1* gene expression with a 59% decrease (fold change=0.41) following the overexpression of the *FGFR1-premir1* in the HEK293T cells compared to the un-transfected cells. The *BRCA1* expression level between mock and un-transfected control cells revealed no significant change (figure 1D). In contrast, no significant alteration in the *SMAD2* expression level was detected in the *FGFR1-miR1*

overexpressed HEK293T cells compared to the controls (figure 1E). Data were normalized using *GAPDH* and *β-actin* as reference genes.

Pearson correlation coefficient (r) test showed a significant negative correlation between *FGFR1-miR1* and *BRCA1* expression levels across overexpressed, mock, and un-transfected cells with a correlation coefficient of r=-0.8994 (P=0.0010) (figure 1F).

Discussion

In the current study, we aimed to uncover novel intronic miRNAs within the *FGFR1* gene, which is frequently amplified in different cancers.[6] Our findings demonstrated the successful identification and initial experimental validation of a novel miRNA (designated as *FGFR1-miR1*) that potentially downregulates *BRCA1* expression.

Through comprehensive bioinformatics prediction tools, we identified several pre-miRNA candidates within the *FGFR1* gene. Among these, candidates including put-premir11, 12, 16, 17, and 20 demonstrated the most promising characteristics for further investigation. Based on the results of multiple bioinformatics tools, put-premir12, which exhibited the strongest *in silico* evidence, was selected for experimental validation. We successfully confirmed its processing into a mature miRNA (*FGFR1-miR1*) through overexpression in HEK293T cells, followed by reverse transcription-quantitative polymerase chain reaction (RT-qPCR) and sequencing. Furthermore, we observed the downregulation of *BRCA1*, one of its predicted target genes, following *FGFR1-miR1* overexpression. These findings provide initial evidence for the existence of a novel miRNA hosted within the *FGFR1* gene and suggest a possible regulatory link with *BRCA1*.

Identifying a highly conserved miRNA (*FGFR1-miR1*) within the *FGFR1* intronic region suggests its potential co-regulation with the host gene. The *FGFR1* gene is a well-known oncogene whose amplification drives tumor growth through pathways such as MAPK and PI3K-Akt signaling.[6] Our pathway analysis of FGFR1-miR1's predicted target genes showed significant enrichment in cancer-related pathways, including MAPK, PI3K-Akt, and Ras signaling. The overlap between these pathways and the known downstream signaling of *FGFR1* suggests a potential functional link between the host gene and its intronic miRNA. This finding is consistent with studies on other intronic miRNAs that often share regulatory elements and functional relationships with their host genes and can regulate the same or related biological pathways.[35] For example, intronic *miR-4728-5p* is co-transcribed with its host gene, erb-b2 receptor tyrosine kinase 2 (*HER2*; also known as *ERBB2*), and enhances HER2 oncogenic function by targeting ErbB3-binding protein 1 (*EBP1*), forming a positive feedback loop that promotes *HER2* expression and breast cancer progression.[36] These findings suggest further investigation into *FGFR1-miR1*'s role in *FGFR1*-driven cancers. However, whether *FGFR1-miR1* is actually co-expressed in cancer tissues remains to be determined experimentally.

Investigation of next-generation sequencing (NGS) data from the NCBI-SRA database provided preliminary support for the existence and identity of this miRNA. The identification of slightly truncated reads at the 3′ end may reflect the biological processing of this sequence into isomiRs. These length variations could arise from the imprecise cleavage by DROSHA and DICER enzymes during miRNA biogenesis. Furthermore, the action of specific exonucleases that remove nucleotides from the ends of the sequence can lead to the generation of shorter mature miRNAs.[37]

The identification of several reads aligning with the predicted *FGFR1-miR1* sequence in breast cancer tumor samples and the MDA-MB-468 cell line strengthens the possibility of its endogenous expression in these tissues. Additionally, the identification of 19 reads (22nt in length) from MDA-MB-231 breast cancer cells that aligned to our sequence, but differed by two nucleotides compared to the precursor, possibly indicates a single nucleotide polymorphism (SNP) event and the presence of an isomiR. While this may suggest the potential endogenous presence of *FGFR1-miR1* or its isomiRs in this cell line, linking this directly to the aggressive form of the breast cancer cell line requires further investigation.

Although similar studies have successfully identified novel miRNAs within other host genes, such as cadherin 1 (*CDH4*) and *HER2*,[31, 38] to the best of our knowledge, no prior research has reported the detection of miRNAs encoded within the *FGFR1* gene. Furthermore, while previous literature has primarily focused on miRNAs that target the *FGFR1* transcript to regulate its expression (such as *miR-136* and *miR-214-3p*),[11, 12] the identification of a miRNA originating from within the *FGFR1* gene itself adds a novel layer of regulatory complexity.

Identifying miRNA target genes is a key step in understanding the role of miRNAs in various cellular functions.[34] To investigate its function experimentally, we assessed the

effect of *FGFR1-miR1* overexpression on two predicted targets: *BRCA1* and *SMAD2*. We observed a significant downregulation of *BRCA1*, whereas the *SMAD2* expression level remained unchanged. The lack of a significant change in *SMAD2* expression highlights the complexity of miRNA-target interactions and the inherent limitations of computational predictions. Various miRNA target prediction tools rely on sequence specificity, thermodynamic stability, and evolutionary conservation, but often fail to account for the limited effects miRNAs exert under physiological conditions, frequently leading to false predictions.[34, 39] Therefore, although *SMAD2* was identified as a high-confidence *in silico* target, the lack of experimental downregulation suggests that its regulation may depend on distinct regulatory timing, specific cellular contexts, or the tissue-specific nature of miRNA function. Furthermore, the uncertainty regarding the exact 5′ seed sequence of *FGFR1-miR1* demonstrates that *in silico* predictions, even those with high scores, do not always guarantee biological validation *in vitro*.

The observed significant downregulation of *BRCA1* expression provides preliminary experimental support for our bioinformatics predictions. *BRCA1* is a critical tumor suppressor gene involved in numerous cellular processes, including cell cycle regulation, DNA repair, and gene transcription modulation, and its dysregulation is observed in breast cancer and mammary tumors.[40]

This study provides the first evidence for a novel miRNA in the *FGFR1* locus, but several limitations must be acknowledged. First, the experimental validation was performed in a single, non-cancerous cell line (HEK293T) using an overexpression system, which may not fully reflect the endogenous processing and function of *FGFR1-miR1* in a cancer context. Second, computational target predictions have inherent limitations, including the potential for false positives. Although our *in silico* analyses predicted *BRCA1* as a target across various hypothetical lengths (18-27 nt) of the miRNA, and *FGFR1-miR1* overexpression demonstrated a significant downregulation of *BRCA1*, we cannot definitively conclude that *BRCA1* is a direct target. The observed downregulation could be a secondary effect or specific to the HEK293T cellular context. Further experiments, such as a dual-luciferase reporter assay, are essential to confirm whether *BRCA1* is a direct target of *FGFR1-miR1*. Third, the precise mature sequence of the miRNA remains unconfirmed. Techniques such as rapid amplification of cDNA ends (RACE) or small RNA sequencing of the transfected cells are required to determine the exact 5′ and 3′ ends of *FGFR1-miR1*. Additionally, evaluating the endogenous expression and functional impact of *FGFR1-miR1* across different cancer and normal cell lines, followed by *in vivo* studies, is necessary to elucidate its potential oncogenic or tumor-suppressive role.

Conclusion

In conclusion, our findings demonstrated the existence of *FGFR1-miR1*, a novel highly conserved miRNA located in the second intron of the *FGFR1* gene, revealing a new potential layer of regulation within this well-known oncogene. Furthermore, the overexpression of *FGFR1-miR1* resulted in the downregulation of the *BRCA1* gene. Although this downregulation suggests a potential regulatory role, its direct implication in cancer progression remains uncertain. Due to current limitations, including reliance on a single non-cancerous cell line and ambiguity regarding the mature miRNA's exact length, future investigations, such as RACE for precise sequence determination, dual-luciferase assays to confirm direct targets, and functional studies, are essential to elucidate its biological role.

Acknowledgment

This research was carried out at the University of Isfahan, Iran. The present study is extracted from the PhD thesis of Maryam Parvaresh and was funded by the Graduate Office of the University of Isfahan. We gratefully acknowledge the Molecular Biology and Microbiology Department of the University of Isfahan for their support. In addition, this study was financially supported by the Biotechnology Development Council of the Islamic Republic of Iran [grant number: 960606].

Authors' Contribution

M.P: Conceptualization, study design, Methodology, Investigation, data analysis, data interpretation, and drafting the manuscript; M.MB: Conceptualization, study design, Supervision, Project administration, and reviewing the manuscript; S.D: Conceptualization, study design, Investigation, data interpretation, and drafting the manuscript. All authors have read and approved the final manuscript and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Declaration of AI

The authors used Grammarly only for grammar correction and to improve the English readability of the manuscript. The AI tool was not used to generate content or interpret results. The authors thoroughly reviewed and edited sentences and take full responsibility for the final content of the manuscript.

Conflicts of Interest: None declared.

References

1 Bagban M, Sharma K, Saifi S, Ilangovan I, Sultana S, Numanoğlu EN, et al. miR-96 and its versatile role in cancer. Advances in Cancer Biology - Metastasis. 2023;7:100082. doi: https://doi.org/10.1016/j.adcanc.2022.100082.

2 Treiber T, Treiber N, Meister G. Regulation of microRNA biogenesis and its crosstalk with other cellular pathways. Nat Rev Mol Cell Biol. 2019;20:5-20. doi: 10.1038/s41580-018-0059-1. PubMed PMID: 30228348.

3 Chen L, Heikkinen L, Wang C, Yang Y, Sun H, Wong G. Trends in the development of miRNA bioinformatics tools. Brief Bioinform. 2019;20:1836-52. doi: 10.1093/bib/bby054. PubMed PMID: 29982332; PubMed Central PMCID: PMC7414524.

4 Hussen BM, Hidayat HJ, Salihi A, Sabir DK, Taheri M, Ghafouri-Fard S. MicroRNA: A signature for cancer progression. Biomed Pharmacother. 2021;138:111528. doi: 10.1016/j.biopha.2021.111528. PubMed PMID: 33770669.

5 Soh J, Cho H, Choi CH, Lee H. Identification and Characterization of MicroRNAs Associated with Somatic Copy Number Alterations in Cancer. Cancers (Basel). 2018;10. doi: 10.3390/cancers10120475. PubMed PMID: 30501131; PubMed Central PMCID: PMC6315597.

6 Fan S, Chen Y, Wang W, Xu W, Tian M, Liu Y, et al. Pharmacological and Biological Targeting of FGFR1 in Cancer. Curr Issues Mol Biol. 2024;46:13131-50. doi: 10.3390/cimb46110783. PubMed PMID: 39590377; PubMed Central PMCID: PMC11593329.

7 Erber R, Rubner M, Davenport S, Hauke S, Beckmann MW, Hartmann A, et al. Impact of fibroblast growth factor receptor 1 (FGFR1) amplification on the prognosis of breast cancer patients. Breast Cancer Res Treat. 2020;184:311-24. doi: 10.1007/s10549-020-05865-2. PubMed PMID: 32852708; PubMed Central PMCID: PMC7599145.

8 Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences to function. Nucleic Acids Res. 2019;47:D155-D62. doi: 10.1093/nar/gky1141. PubMed PMID: 30423142; PubMed Central PMCID: PMC6323917.

9 Alles J, Fehlmann T, Fischer U, Backes C, Galata V, Minet M, et al. An estimate of the total number of true human miRNAs. Nucleic Acids Res. 2019;47:3353-64. doi: 10.1093/nar/gkz097. Pub

10 Çakmak E. Computational and experimental tools of miRNAs in cancer. Middle East J Cancer. 2020;11:381-9.

11 Zhou L, Yang J, Ma S, Gao H. Effect of miRNA-136-targeted regulation of FGFR1 on proliferation and apoptosis of triple-negative breast cancer cells. Am J Transl Res. 2021;13:7723-9. PubMed PMID: 34377249; PubMed Central PMCID: PMC8340197.

12 Yang Y, Li Z, Yuan H, Ji W, Wang K, Lu T, et al. Reciprocal regulatory mechanism between miR-214-3p and FGFR1 in FGFR1-amplified lung cancer. Oncogenesis. 2019;8:50. doi: 10.1038/s41389-019-0151-1. PubMed PMID: 31492847; PubMed Central PMCID: PMC6731303.

13 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;37:3276-87. doi: 10.1093/nar/gkp120. PubMed PMID: 19324892; PubMed Central PMCID: PMC2691815.

14 Gao D, Middleton R, Rasko JE, Ritchie W. miREval 2.0: a web tool for simple microRNA prediction in genome sequences. Bioinformatics. 2013;29:3225-6. doi: 10.1093/bioinformatics/btt545. PubMed PMID: 24048357; PubMed Central PMCID: PMC5994938.

15 Ahmed F, Kaundal R, Raghava GP. PHDcleav: a SVM based method for predicting human Dicer cleavage sites using sequence and secondary structure of miRNA precursors. BMC Bioinformatics. 2013;14:S9. doi: 10.1186/1471-2105-14-S14-S9. PubMed PMID: 24267009; PubMed Central PMCID: PMC3851333.

16 Jiang P, Wu H, Wang W, Ma W, Sun X, Lu Z. MiPred: classification of real and pseudo microRNA precursors using random forest prediction model with combined features. Nucleic Acids Res. 2007;35:W339-44. doi: 10.1093/nar/gkm368. PubMed PMID: 17553836; PubMed Central PMCID: PMC1933124.

17 Liu B, Fang L, Liu F, Wang X, Chen J, Chou KC. Identification of real microRNA precursors

with a pseudo structure status composition approach. PLoS One. 2015;10:e0121501. doi: 10.1371/journal.pone.0121501. PubMed PMID: 25821974; PubMed Central PMCID: PMC4378912

18 Liu B, Fang L, Liu F, Wang X, Chou KC. iMiRNA-PseDPC: microRNA precursor identification with a pseudo distance-pair composition approach. J Biomol Struct Dyn. 2016;34:223-35. doi: 10.1080/07391102.2015.1014422. PubMed PMID: 25645238.

19 Liu B, Fang L, Chen J, Liu F, Wang X. miRNA-dis: microRNA precursor identification based on distance structure status pairs. Mol Biosyst. 2015;11:1194-204. doi: 10.1039/c5mb00050e. PubMed PMID: 25715848.

20 Chen J, Wang X, Liu B. iMiRNA-SSF: Improving the Identification of MicroRNA Precursors by Combining Negative Sets with Different Distributions. Sci Rep. 2016;6:19062. doi: 10.1038/srep19062. PubMed PMID: 26753561; PubMed Central PMCID: PMC4709562.

21 Liu B, Fang L, Wang S, Wang X, Li H, Chou KC. Identification of microRNA precursor with the degenerate K-tuple or Kmer strategy. J Theor Biol. 2015;385:153-9. doi: 10.1016/j.jtbi.2015.08.025. PubMed PMID: 26362104.

22 Gkirtzou K, Tsamardinos I, Tsakalides P, Poirazi P. MatureBayes: a probabilistic algorithm for identifying the mature miRNA within novel precursors. PLoS One. 2010;5:e11843. doi: 10.1371/journal.pone.0011843. PubMed PMID: 20700506; PubMed Central PMCID: PMC2917354.

23 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:e48236. doi: 10.1371/journal.pone.0048236. PubMed PMID: 23118959; PubMed Central PMCID: PMC3484136.

24 Chen Y, Wang X. miRDB: an online database for prediction of functional microRNA targets. Nucleic Acids Res. 2020;48:D127-D31. doi: 10.1093/nar/gkz757. PubMed PMID: 31504780; PubMed Central PMCID: PMC6943051.

25 Reczko M, Maragkakis M, Alexiou P, Grosse I, Hatzigeorgiou AG. Functional microRNA targets in protein coding sequences. Bioinformatics. 2012;28:771-6. doi: 10.1093/bioinformatics/bts043. PubMed PMID: 22285563.

26 Sherman BT, Hao M, Qiu J, Jiao X, Baseler MW, Lane HC, et al. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). Nucleic Acids Res. 2022;50:W216-W21. doi: 10.1093/nar/gkac194. PubMed PMID: 35325185; PubMed Central PMCID: PMC9252805.

27 Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. 1988;16:1215. doi: 10.1093/nar/16.3.1215. PubMed PMID: 3344216; PubMed Central PMCID: PMC334765.

28 Xu N, Papagiannakopoulos T, Pan G, Thomson JA, Kosik KS. MicroRNA-145 regulates OCT4, SOX2, and KLF4 and represses pluripotency in human embryonic stem cells. Cell. 2009;137:647-58. doi: 10.1016/j.cell.2009.02.038. PubMed PMID: 19409607.

29 da Conceicao Braga L, Goncalves BOP, Coelho PL, da Silva Filho AL, Silva LM. Identification of best housekeeping genes for the normalization of RT-qPCR in human cell lines. Acta Histochem. 2022;124:151821. doi: 10.1016/j.acthis.2021.151821. PubMed PMID: 34861601.

30 Bignotti E, Calza S, Tassi RA, Zanotti L, Bandiera E, Sartori E, et al. Identification of stably expressed reference small non-coding RNAs for microRNA quantification in high-grade serous ovarian carcinoma tissues. J Cell Mol Med. 2016;20:2341-8. doi: 10.1111/jcmm.12927. PubMed PMID: 27419385; PubMed Central PMCID: PMC5134371.

31 Shabaninejad Z, Mowla SJ, Yousefi F, Soltani BM. A Novel miRNA Located in the HER2 Gene Shows an Inhibitory Effect on Wnt Signaling and Cell Cycle Progression. Biomed Res Int. 2022;2022:7216758. doi: 10.1155/2022/7216758. PubMed PMID: 35747498; PubMed Central PMCID: PMC9213177.

32 Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C(T) method. Nat Protoc. 2008;3:1101-8. doi: 10.1038/nprot.2008.73. PubMed PMID: 18546601.

33 Steiman-Shimony A, Shtrikman O, Margalit H. Assessing the functional association of intronic miRNAs with their host genes. RNA. 2018;24:991-1004. doi: 10.1261/rna.064386.117. PubMed PMID: 29752351; PubMed Central PMCID: PMC6049507.

34 Riolo G, Cantara S, Marzocchi C, Ricci C. miRNA Targets: From Prediction Tools to Experimental Validation. Methods Protoc. 2020;4. doi: 10.3390/mps4010001. PubMed PMID: 33374478; PubMed Central PMCID: PMC7839038.

35 Calin GA. Being Small and Intronic: miRNAs That Count! Cancer Res. 2021;81:1212-3. doi: 10.1158/0008-5472.CAN-20-4172. PubMed PMID: 33822742.

36 Zhou Y, Yuan Y, Li L, Wang X, Quan Y, Liu C, et al. HER2-intronic miR-4728-5p facilitates HER2 expression and accelerates cell proliferation and migration by targeting EBP1 in breast cancer. PLoS One. 2021;16:e0245832. doi: 10.1371/journal.pone.0245832. PubMed PMID: 33529238; PubMed Central PMCID: PMC7853520.
37 Tomasello L, Distefano R, Nigita G, Croce CM. The MicroRNA Family Gets Wider: The IsomiRs Classification and Role. Front Cell Dev Biol. 2021;9:668648. doi: 10.3389/fcell.2021.668648. PubMed PMID: 34178993; PubMed Central PMCID: PMC8220208.
38 Rahimi A, Sedighi R, Emadi-Baygi M, Honardoost MA, Mowla SJ, Khanahmad H, et al. Bioinformatics prediction and experimental validation of a novel microRNA: hsa-miR-B43 within human CDH4 gene with a potential metastasis-related function in breast cancer. J Cell Biochem. 2020;121:1307-16. doi: 10.1002/jcb.29367. PubMed PMID: 31489987.
39 Oh M, Rhee S, Moon JH, Chae H, Lee S, Kang J, et al. Literature-based condition-specific miRNA-mRNA target prediction. PLoS One. 2017;12:e0174999. doi: 10.1371/journal.pone.0174999. PubMed PMID: 28362846; PubMed Central PMCID: PMC5376335.
40 Fu X, Tan W, Song Q, Pei H, Li J. BRCA1 and Breast Cancer: Molecular Mechanisms and Therapeutic Strategies. Front Cell Dev Biol. 2022;10:813457. doi: 10.3389/fcell.2022.813457. PubMed PMID: 35300412; PubMed Central PMCID: PMC8921524.