A peer-reviewed version of this preprint was published in PeerJ on 13 November 2019.

View the peer-reviewed version (peerj.com/articles/8049), which is the preferred citable publication unless you specifically need to cite this preprint.

Aisina D, Niyazova R, Atambayeva S, Ivashchenko A. 2019. Prediction of clusters of miRNA binding sites in mRNA candidate genes of breast cancer subtypes. PeerJ 7:e8049 https://doi.org/10.7717/peerj.8049

Prediction of clusters of miRNA binding sites in mRNA candidate genes of breast cancer

Dana Aisina [Equal first author, 1], **Raigul Niyazova** [1], **Shara Atambayeva** [1], **Anatoliy Ivashchenko** [Corresp. Equal first author, 1]

[1] al-Farabi Kazakh National University, Almaty, Kazakhstan

Corresponding Author: Anatoliy Ivashchenko
Email address: a_ivashchenko@mail.ru

Distinct sets of candidate genes control the development of breast cancer subtypes. The expression of many genes is regulated by the binding of their mRNAs with miRNAs. The prediction of miRNA associations and target genes is essential in studying of breast cancer. The MirTarget program defines the following features of binding miRNA to mRNA: the start of the initiation of miRNA binding to mRNA; the localization of miRNA binding sites in 5'-untranslated regions (5'UTR), coding domain sequences (CDS) and 3'-untranslated regions (3'UTR); the free energy of binding of all miRNA nucleotides with mRNA; the schemes of interactions of all miRNAs nucleotides with mRNAs. The mRNAs of many genes have clusters (miRNA binding sites with overlapping nucleotide sequences) located in 5'UTR, CDS, or 3'UTR. There are clusters in 5'UTR of mRNA *EPOR, MAZ* and *NISCH* candidate genes of HER2 subtype. There are four clusters in CDS of mRNA *MAZ* gene, and in 3'UTR of mRNA *BRCA2* and *CDK6* genes. Candidate genes of triple-negative subtype are targets for multiple miRNAs. In 5'UTR of mRNA *CBL* gene, there are 11 sites; the mRNA for *MMP2* gene contains five sites; the mRNA of *RAB5A* gene contains two clusters each of three sites. In 3'UTR of mRNA *SFN* gene, there are two clusters, each of three sites, and one cluster of 21 sites. Candidate genes of luminal A and B subtypes are targets for miRNAs: there are 21 sites in 5'UTR of mRNA *FOXA1* gene and mRNA *HMGA2* gene contains 15 sites. There are clusters of five sites in CDS of mRNA *ITGB1* gene and five sites in 3'UTR of mRNA *SOX4* genes. Clusters of eight sites and ten sites are identified in 3'UTR of mRNA *SMAD3* and *TGFB1* genes, respectively. The organization of miRNA binding sites into clusters reduces the proportion of nucleotide binding sites in 5'UTR, CDS and 3'UTRs. This overlapping of miRNA binding sites creates a competition among miRNAs for the binding site. From 6,272 studied miRNAs only 29 miRNAs from miRBase and 88 novel miRNAs have binding sites in clusters of mRNA target genes of breast cancer.

Prediction of Clusters of miRNA Binding Sites in mRNA Candidate Genes of Breast Cancer

Dana Ye. Aisina[1], Raigul Ye. Niyazova[1], Shara A. Atambayeva[1], Anatoliy T. Ivashchenko[1]

[1] Department of biotechnology, SRI of Biology and Biotechnology Problems, al-Farabi Kazakh National University, Almaty, Kazakhstan

Corresponding Author:
Anatoliy Ivashchenko[1]
al-Farabi 71, Almaty, 050040, Kazakhstan
Email address: a_ivashchenko@mail.ru

Abstract

Distinct sets of candidate genes control the development of breast cancer subtypes. The expression of many genes is regulated by the binding of their mRNAs with miRNAs. The prediction of miRNA associations and target genes is essential in studying of breast cancer. The MirTarget program defines the following features of binding miRNA to mRNA: the start of the initiation of miRNA binding to mRNA; the localization of miRNA binding sites in 5'-untranslated regions (5'UTR), coding domain sequences (CDS) and 3'-untranslated regions (3'UTR); the free energy of binding of all miRNA nucleotides with mRNA; the schemes of interactions of all miRNAs nucleotides with mRNAs. The mRNAs of many genes have clusters (miRNA binding sites with overlapping nucleotide sequences) located in 5'UTR, CDS, or 3'UTR. There are clusters in 5'UTR of mRNA *EPOR, MAZ* and *NISCH* candidate genes of HER2 subtype. There are four clusters in CDS of mRNA *MAZ* gene, and in 3'UTR of mRNA *BRCA2* and *CDK6* genes. Candidate genes of triple-negative subtype are targets for multiple miRNAs. In 5'UTR of mRNA *CBL* gene, there are 11 sites; the mRNA for *MMP2* gene contains five sites; the mRNA of *RAB5A* gene contains two clusters each of three sites. In 3'UTR of mRNA *SFN* gene, there are two clusters, each of three sites, and one cluster of 21 sites. Candidate genes of luminal A and B subtypes are targets for miRNAs: there are 21 sites in 5'UTR of mRNA *FOXA1* gene and mRNA *HMGA2* gene contains 15 sites. There are clusters of five sites in CDS of mRNA *ITGB1* gene and five sites in 3'UTR of mRNA *SOX4* genes. Clusters of eight sites and ten sites are identified in 3'UTR of mRNA *SMAD3* and *TGFB1* genes, respectively. The organization of miRNA binding sites into clusters reduces the proportion of nucleotide binding sites in 5'UTR, CDS and 3'UTRs. This overlapping of miRNA binding sites creates a competition among miRNAs for the binding site. From 6,272 studied miRNAs only 29 miRNAs from miRBase and 88 novel miRNAs have binding sites in clusters of mRNA target genes of breast cancer.

Subjects Bioinformatics, Genomics, Computational Biology
Keywords miRNA, gene, binding site, cluster, breast cancer

Introduction

Breast cancer (BC) occupies one of the first places among all cancers in the world. These statistics demonstrate an intense, steady increase in the incidence and mortality from BC among women (*Benson & Jatoi, 2012*). More than 50% of patients with BC are detected for the first time in the late stages of the disease. Every year 1,400,000 new cases of the diseases are diagnosed in the world (*Jemal et al., 2010*). Establishing the interaction of miRNAs with mRNA genes involved in the development of BC (candidate genes) is one of the promising areas of research. miRNAs are found in tumors, blood, and may be potential biomarkers of BC (*Adhami et al., 2018*; *Hannafon et al., 2016*; *Kurozumi et al., 2017*; *Lagendijk et al., 2018*; *McDermott et al., 2014*; *Piasecka et al., 2018*; *Zhang et al., 2017*). The establishment of a correlation between the expression of miRNA and various BC subtypes is devoted to several publications (*Biagioni et al., 2012*; *Blenkiron et al., 2007*; *Enerly et al., 2011*; *Lee et al., 2013*; *Lowery et al., 2009*; *Mattie et al., 2006*; *Telonis et al., 2015*; *Yang et al., 2017*). Disruptions in the regulation of miRNA expression affect the development of a tumor, since they can regulate the expression of oncogenes and oncosuppressors. Increase or decrease in the expression of certain miRNAs influence the onset of a tumor and its progression (*Wang et al., 2013b*). miRNAs the expression of which varies with benign and malignant tumors have been revealed (*Tahiri et al., 2014*). It has been shown that many intron miRNAs are expressed together with host genes. Changes in miRNA expression may be associated with chromosomal mutations (*Qian et al., 2012*), epigenetic modifications (*Yu et al., 2013*) or defects in their biogenesis (*Sung et al., 2012*). miRNAs that inhibit translation of mRNA of tumor suppressors and apoptosis genes function as oncogenes, contributing to oncogenesis (*Wu et al., 2013*). Other miRNAs may be tumor suppressors if their target genes are oncogenes and cell cycle genes (*Nian et al., 2013*). Currently there is a little information about the interaction of miRNAs and genes associated with BC subtypes. Therefore, in the present work, the associations of miRNAs with mRNAs of the candidate genes of BC subtypes were revealed. According to the miRBase, more than 90% of the miRNA have a length in the range of 20-25 nucleotides (http://mirbase.org). This length of the nucleotide sequence of miRNA is necessary and sufficient for selective interaction with mRNA, like the length of primers in the polymerase chain reaction (*Huggett & O'Grady, 2014*). One miRNA can have binding sites in mRNA of many genes (*Atambayeva et al., 2017*; *Ivashchenko et al., 2016*; *Niyazova et al., 2015*) and mRNA of one gene can have binding sites for many miRNAs (*Kondybayeva et al., 2018*).

In this publication, on the example of studying the characteristics of the binding of miRNA with mRNA of BC candidate genes we show the advantage of the proposed changes in the perception of the interaction of miRNA with mRNA. The present article is devoted to ascertaining the interaction of miRNAs with mRNA candidate genes of BC, especially those that contain two and more miRNAs binding sites organized in clusters.

Materials & Methods

The nucleotide (nt) sequences of candidate genes of BC subtypes were downloaded from GenBank (http://www.ncbi.nlm.nih.gov). These candidate genes are specific for the development of triple-negative subtype, luminal A and B subtypes and HER2 subtype of BC (Table S1). Information about miRNAs that presumably bind to candidate genes of BC is provided in Table

S2. The table indicates that studied miRNAs are present in blood, serum, plasma, and cells in BC or other types of cancer. The nucleotide sequences of mRNAs genes of *Chlorocebus sabaeus* - Csa, *Gorilla gorilla* - Ggo, *Homo sapience* - Hsa, *Macaca mulatta* - Mml, *Mus musculus* - Mmu, *Pan paniscus* - Ppa, *Pan troglodytes* - Ptr, *Papio Anubis* - Pan, *Pongo abelii* - Pab, *Rattus norvegicus* - Rno were downloaded from NCBI GenBank (http://www.ncbi.nlm.nih.gov). The nucleotide sequences of 2565 miRNAs were taken from miRBase and 3707 miRNAs from the publication (*Londin et al., 2015*). RPKM value (*Mortazavi et al., 2008*) given in the Human Protein Atlas data (https://www.proteinatlas.org/ENSG00000150093-ITGB1/tissue/breast). Human Protein Atlas data were used as a quantitative measure of transcript expression in cerebral cortex.

The miRNAs binding sites in 5'UTRs, CDSs and 3'UTRs of several genes were predicted using the MirTarget program (*Ivashchenko et al., 2016*; *Ivashchenko, Issabekova & Berillo, 2013*). This program defines the following features of miRNA binding to mRNA: a) the start of the initiation of miRNA binding to mRNAs; b) the localization of miRNA binding sites in 5'UTRs, CDSs and 3'UTRs of the mRNAs; c) the free energy of interaction miRNA and the mRNA (ΔG, kJ/mole); d) the schemes of nucleotide interactions between miRNAs and mRNAs. The ratio ΔG/ΔGm (%) was determined for each site (ΔGm equals the free energy of miRNA binding with its fully complementary nucleotide sequence). The miRNA binding sites located in mRNAs had ΔG/ΔGm ratios of 87% or more. ΔG/ΔGm ratios were taken on the assumption that the members of the miRNA of one family generally differ by no more than 1-3 nucleotides, that with a miRNA length of 22 nt, the ΔG/ΔGm value was 96% (21 nt/22 nt = 96%) - 87% (19 nt/22 nt = 87%). With a larger difference in the number of mismatched nucleotides, the probability of two or more miRNAs to bind in one site increases, which excludes the natural property of the miRNA to interact selectively with the mRNA of the target gene. The MirTarget program identifies the positions of the binding sites on the mRNA, beginning from the first nucleotide of the mRNA's 5'UTR. The MirTarget program found hydrogen bonds between adenine (A) and uracil (U), guanine (G) and cytosine (C), G and U, A and C. The distances between A and C were equal 1.04 nanometers, between G and C, and A and U were equal 1.03 nanometers, between G and U equal to 1.02 nanometers (*Leontis, Stombaugh & Westhof, 2002*). The numbers of hydrogen bonds in the G-C, A-U, G-U and A-C interactions were found to be 3, 2, 1 and 1, respectively (*Kool, 2001*; *Lemieux & Major, 2002*; *Leontis, Stombaugh & Westhof, 2002*). The characteristics of miRNA interaction with mRNA reflect the intermolecular interactions of these molecules and are calculated for given parameters without their variation. Consequently, the results have no statistical scatter. Other factors that may influence these interactions have not been studied. The article did not address the issues of changing the concentration ratio of miRNA and mRNA, because this aspect is of independent interest and is not part of the objectives of this work. For any other pathology, other candidate genes should be used and other miRNAs binding sites will be determined.

The MirTarget program determines single miRNA binding sites in mRNA and miRNA binding sites which are in clusters (in series arranged with overlapping of nucleotide sequences of the same or several miRNAs). In the present work, we propose a hypothesis, that miRNA binding sites in mRNA are organized into clusters. MirTarget program does not work directly with miRBase and NCBI databases. The search for target genes from 17,508 human genes in a special format from NCBI for the known miRNAs from miRBase and novel miRNAs from other sources will be available on request at mirtarget8@gmail.com.

Results

The adequate prediction of miRNA binding sites in mRNA target genes is a key problem in the study of the biological role of miRNA in the regulation of gene expression. We have developed a MirTarget program that predicts the binding sites of miRNA with mRNA, thereby revealing fundamentally new properties of miRNA. Before presenting the results, we provide a few specific examples that demonstrate the features of the MirTarget program.

The schemes of miRNA nucleotides interaction with mRNA binding sites are shown in Fig. S1-3. It is shown the following advantages of the MirTarget program: that all miRNA nucleotides interaction with mRNA; the formation of non-canonical pairs G-U and A-C that do not change the double-stranded conformation of the miRNA complex with mRNA, since the distances between G-U and A-C are equal to the distances between G-C and A-U; the free energy of interaction is an important criterion for binding miRNA to mRNA; the localization of the miRNA binding site in mRNA (5'UTR, CDS and 3'UTR).

Several miRNAs bound of the entire nucleotide sequence to mRNA of candidate genes triple-negative subtype of BC (Fig. S1). For example, miR-5095, miR-5096, miR-619-5p, miR-1273g-3p, miR-1273f with the entire nucleotide sequence bound to mRNA of *ATM* gene. MiR-5095, miR-1273e, miR-1273f were bound to mRNA *IL11* gene by all nucleotides. Similarly, miR-1273c, miR-1285-3p bound to mRNA of *STMN1* gene and TJU_CMC_MD2.ID00436.3p-miR bound with mRNA of *SFN* gene.

TJU_CMC_MD2.ID01810.3p-miR had binding site in 5'UTR of mRNA *CBL* gene (Fig. S1). Of the 23 nucleotides of TJU_CMC_MD2.ID01810.3p-miR, only one nucleotide cannot form hydrogen bonds with mRNAs, and the other nucleotides form a double-stranded helical structure with mRNA. Three pairs of G-U and three pairs of A-C having each by one hydrogen bond contributed to the preservation of this structure due to stacking interactions between adjacent bases (*Yakovchuk et al., 2006*). The binding of TJU_CMC_MD2.ID01810.3p-miR to the mRNA of *CBL* gene, non-canonical pairs G-U, A-C formed and there was no interaction between A and G in the second position. Despite this, the interaction energy of TJU_CMC_MD2.ID01810.3p-miR with the mRNA of *CBL* gene was 87% of the maximum. The interaction schemes of TJU_CMC_MD2.ID01321.5p-miR with mRNA *RUNX1* gene, miR-3198 with mRNA *CBL* gene, miR-1273d with mRNA *IL11* gene, miR-5585-3p with mRNA *STMN1* gene are shown in Fig. S1. The free energy of interaction of these pairs of miRNAs and mRNAs was 87-98% of the maximum value of ΔGm (Table 1).

With mRNA of candidate genes luminal A and B subtypes of BC several miRNAs of the entire nucleotide sequence have been associated (Fig. S2). For example, TJU_CMC_MD2.ID01403.5p-miR and TJU_CMC_MD2.ID02428.3p-miR were bind to mRNA *HMGA2* gene. TJU_CMC_MD2.ID03332.3p-miR binds to mRNA *FOXA1* gene and TJU_CMC_MD2.ID01593.5p-miR was fully complementary bound to mRNA *ANGPTL4* gene.

MiR-3960, miR-7111-3p and TJU_CMC_MD2.ID01352.3p-miR were bind to mRNA *MAZ* gene by all nucleotides (Fig. S3). MiR-877-3p by all nucleotides bound to mRNA of *NISCH* and *MAZ* genes, and TJU_CMC_MD2.ID00436.3p-miR bound to mRNA of *CDK6* gene.

Subtype Triple-Negative Breast Cancer

The *CBL* gene is a target for six miRNAs, two of which had four binding sites (Table 1). A cluster of 12 binding sites for six miRNAs was located from 16 nt to 55 nt. All binding sites for miRNAs had a total length of 270 nt. The cluster size was 40 nt with a length of 5'UTR of mRNA *CBL* gene of 142 nt, so the need for cluster organization of miRNA binding sites is clear. The binding sites were compacted in 6.8 times. The average free energy of interaction of six miRNAs with the mRNA of *CBL* gene was -127 kJ/mole.

Results of supposed interactions of six miRNAs with the mRNA *CBL* gene can be represented as a diagram showing the location of miRNA binding sites relatively to cluster in mRNA (Fig. 1). A feature of TJU_CMC_MD2.ID03332.3p-miR is the location of beginning of repeating binding sites of them through three nucleotides. This miRNA interacts with mRNA with displacement of its binding sites coincides with open reading frame of mRNA *CBL* gene. Schemes of interaction of these 12 miRNAs with the mRNA of the *CBL* gene are shown in Fig. 2. It can be seen from the above schemes that the interaction of non-canonical pairs A-C and G-U increases the stability of the binding of miRNA to mRNA. From the data presented (Fig. 1) it can be seen that no more than one miRNA can bind with a cluster, which causes competition between the miRNAs for binding to mRNA of target gene. Twelve nucleotide sequences of miRNA with the indication of start of binding sites according to Table 1 located lower mRNA. Some genes expressed in the mammary gland with an RPKM value of less than 10 contain repeats of nucleotides that are targeted by several miRNAs. In mRNA *CBL* gene, with a RPKM value of 3.9, four binding sites were identified for TJU_CMC_MD2.ID03332.3p-miR and TJU_CMC_MD2.ID01310.3p-miR in a cluster located at 5'UTR from 16 nt to 54 nt. Another example of a target gene for miRNA with nucleotide repeats in 3'UTR is *SFN* gene with an RPKM value of 9.4 (Table 1). miR-466, TJU_CMC_MD2.ID01030.3.3p-miR and TJU_CMC_MD2.ID00436.3p-miR each had six binding sites in the cluster from 1190 nt to 1214 nt.

Five miRNAs with overlapping binding sites were found in the 5'UTR of mRNA *MMP2* gene with a length of cluster of 39 nt. The total length of miRNA of 114 nt, which is 2.9 times more than the total length of cluster. The average free energy of interaction of five miRNAs with the mRNA of *MMP2* gene was -122 kJ/mole.

The *RAB5A* gene was a target for six miRNAs, binding sites of which were formed into two clusters (Table 1). The length of binding sites cluster for TJU_CMC_MD2.ID02930.3p-miR, TJU_CMC_MD2.ID03445.3p-miR, TJU_CMC_MD2.ID01859.5p-miR located from 184 nt to 214 nt were 31 nt. The total length of the binding sites of these miRNAs, located arranged in succession was 71 nt. Therefore, due to the overlapping of nucleotide sequences of binding sites of these miRNAs, the total length of the binding sites decreased by 2.3 times. However, at the same time, only one miRNA can interact with mRNA in the 31 nt segment. Thus, there is a competition between three miRNAs for binding to the mRNA of target gene. It is more likely that miRNA will be bind with a greater free energy of interaction with mRNA at equal concentrations, or miRNA that is present in greater concentration at equal free energy of interaction with mRNA. The second cluster of miRNA binding sites was located from 325 nt to 356 nt and had a length of 32 nt. The total length of miRNA binding sites was 2.1 times the length of the cluster.

As a result, six binding sites of length 138 nt were compacted into clusters of 63 nt in length. This length is considerably smaller than the overall length (535 nt) of 5'UTR of mRNA *RAB5A* gene. The average free energy of interaction of six miRNAs with the mRNA of *RAB5A* was -128 kJ/mole.

There were six candidate genes that formed a cluster in 3'UTR of mRNA (Table 1). There were six miRNAs binding sites that formed two clusters of binding sites in the mRNA of *ATM* gene. The first cluster with a length of 37 nt began with 9778 nt and the second cluster with a length of 42 nt began with 11054 nt. The total miRNA length for the first and second cluster was 67 and 68 nt, respectively. The decrease in the total length of the miRNA binding sites at overlapping of their nucleotide sequences in the clusters was 1.6 - 1.8 times. The average free energy of interaction of six miRNAs with the mRNA of *ATM* gene was -115 kJ/mole.

The cluster of binding sites in the mRNA of the *IL11* gene is located from 1466 nt to 1497 nt in length 31 nt. The sum of the lengths of binding sites equal to 89 nt is 2.9 times the length of the cluster. The clusters of binding sites for three miRNAs were identified in 3'UTR of mRNA *RUNX1* and *CBL* genes. In 3'UTR of mRNA *CBL* gene, the cluster of three miRNA binding sites was 44 nt in length, and the total length of the binding sites was 70 nt. In 3'UTR of mRNA *RUNX1* gene, the cluster of three miRNA binding sites was 34 nt long, and the sum of the lengths of five binding sites was 115 nt. Compacting the length of the binding sites of these miRNAs led to the emergence of competition between them for the binding site in mRNA. The average free energy of miRNAs interaction with mRNA in *CBL* and *RUNX1* genes clusters were -116 kJ/mole and -109 kJ/mole, respectively.

There were two clusters of binding sites for three miRNAs in the 36 nt region from 826 nt to 861 nt and another 53 nt cluster from 1179 nt to 1231 nt in 3'UTR of mRNA *SFN* gene. The third cluster included 21 binding sites of five miRNAs. The sum of the lengths of all miRNA binding sites of two clusters was 619 nt. Due to the clustering of binding sites of these miRNAs, the actual binding site was only 89 nt, which is seven times less and amounts to 18% of the length of 3'UTR of mRNA *SFN* gene equal to 498 nt. The average free energy of miRNA binding at 27 sites was -108 kJ/mole.

The *STMN1* gene was the target of four miRNAs, the binding sites of which in 3'UTR occupied 43 nt, while the total miRNA length was 90 nt. The average free energy of miRNA binding at four sites was -110 kJ/mole.

The free energy value was higher than -125 kJ/mole for the interactions of TJU_CMC_MD2.ID03332.3p-miR, TJU_CMC_MD2.ID02430.3p-miR, TJU_CMC_MD2.ID02761.3p-miR, TJU_CMC_MD2.ID00278.3p-miR, TJU_CMC_MD2.ID03345.5p-miR, TJU_CMC_MD2.ID02930.3p-miR, TJU_CMC_MD2.ID03445.3p-miR, TJU_CMC_MD2.ID01804.3p-miR, TJU_CMC_MD2.ID00061.3p-miR, TJU_CMC_MD2.ID03006.5p-miR, miR-1273d, miR-6089 and TJU_CMC_MD2.ID01774.5p-miR with the mRNAs of *CBL, MMP2, RAB5A, ATM, IL11* and *SFN* genes.

Subtypes luminal A and B Breast Cancer

Eighteen miRNA binding sites with overlapping nucleotide sequences were identified in 5'UTR mRNA of *FOXA1* gene (Table 2). 20 binding sites formed a cluster with the length of 52 nt, from 99 nt to 150 nt. The total length of all 20 binding sites is 447 nt, which is longer than 5'UTR with length of 312 nt. All miRNA binding sites were located in the first half of 5'UTR. Since the cluster length is 52 nt, only two miRNAs can be contacted simultaneously, and other miRNAs will not affect the expression of the *FOXA1* gene.

The formation of a cluster of binding sites for the *FOXA1* gene in 5'UTR indicates a greater ability of this gene for compaction, which causes competition among miRNA for the binding site. Despite the fact that TJU_CMC_MD2.ID01099.5p-miR, TJU_CMC_MD2.ID01190.5p-miR and TJU_CMC_MD2.ID02457.3p-miR are fully complementary to mRNA gene, they had a free

energy interaction of -108 kJ/mole, which is significantly less than for other miRNAs. At equal concentrations of all miRNAs, TJU_CMC_MD2.ID00252.5p-miR, TJU_CMC_MD2.ID00296.3p-miR and TJU_CMC_MD2.ID01702.3p-miR had ΔG values equal to -140 kJ/mole having the advantage in binding to the mRNA of *FOXA1* gene. The average free energy of miRNA binding, without three miRNAs with a length of 17 nt, was -126 kJ/mole, which is characteristic of miRNA binding in 5'UTR.

The 5'UTR of mRNA *HMGA2* had 17 binding sites for 15 miRNAs. Binding sites of these miRNAs were in a 95 nt cluster from 512 nt to 606 nt. The total length of binding sites was equal to 407 nt and it was 4.3 times longer than the cluster. miRNA binding sites were located in the first half of 5'UTR and had a ΔG value more than -125 kJ/mole. TJU_CMC_MD2.ID00296.3p-miR and TJU_CMC_MD2.ID00296.3p-miR had a ΔG equal to -142 kJ/mole and -146 kJ/mole, respectively.

The *ITGB1* gene had no 5'UTR, but a cluster for five miRNA binding sites was located from 91 nt to 120 nt in the beginning of CDS with the length of 30 nt, which is 3.6 times less than the sum of the lengths of five miRNAs.

For *HMGA2* gene there was a cluster for four binding sites from 1255 nt to 1295 nt located in the beginning of 3'UTR. The cluster length was equal to 41 nt with total length of binding sites comprising 98 nt in length.

Apparently, the compaction of binding sites is due not only to the economy of gene length but also to the competition between miRNAs for interaction. For example, the cluster of eight binding sites with 3'UTR of mRNA *SMAD3* gene with the length of 35 nt was located from 2066 nt to 2101 nt. Therefore, only one miRNA can be bind in a cluster. At equal concentrations of all six miRNAs, TJU_CMC_MD2.ID02822.5p-miR and miR-6089 had free interaction energy of -127 kJ/mole to -136 kJ/mole will have an advantage in binding to cluster.

The 3'UTR of mRNA *SOX4* gene had four miRNA binding sites organized in a cluster of 29 nt. TJU_CMC_MD2.ID01282.3p-miR and TJU_CMC_MD2.ID03445.3p-miR bound to mRNA with a ΔG equal to -125 kJ/mole and -127 kJ/mole, respectively.

The mRNAs of the *TGFB1* gene had a cluster of binding sites for seven miRNAs with a length of 48 nt located from 2060 nt to 2107 nt. The length of 3'UTR was 146 nt with 10 miRNA binding sites equal to 230 nt, so the compacting of the binding sites was 4.8 times.

Fig. S2 shows the schemes of interaction of some miRNAs with mRNA of several candidate genes of the luminal A and B subtypes. The presented schemes clearly show the advantage of the MirTarget program in predicting the miRNA binding sites. For example, TJU_CMC_MD2.ID03367.5p-miR formed a non-canonical G-U pair in the mRNA *FOXA1* gene. But TJU_CMC_MD2.ID03367.5p-miR can bind to 19 nucleotide of mRNA and the free interaction energy was 93% of the maximum value. The TJU_CMC_MD2.ID02542.5p-miR interacted with 23 nucleotides of mRNA *FOXA1* gene, but had only one unpaired nucleotide. Such interaction between the miRNAs and their target genes is valid for the following pairs: TJU_CMC_MD2.ID00101.3p-miR and *HMGA2* gene, TJU_CMC_MD2.ID00849.3p-miR and *HMGA2* gene, miR-4507-3p and *SMAD3* gene, miR-937-5p and *TGFB1* gene, miR-937-5p and *SMAD3* gene, TJU_CMC_MD2.ID01403.5p-miR and *HMGA2* gene.

Subtype HER2 Breast Cancer

Twenty-three miRNAs were bound in 5'UTR mRNAs of three candidate genes of the breast cancer subtype HER2 (Table 3). The mRNA of *EPOR* gene had three miRNA binding sites, the nucleotide sequences of which overlapped. Three binding sites of TJU_CMC_MD2.ID01633.3p-miR, TJU_CMC_MD2.ID01599.3p-miR and TJU_CMC_MD2.ID01626.3p-miR comprised a 26

nt cluster located from 77 nt to 102 nt in 5'UTR of mRNA *EPOR* gene. Without overlapping sites, the length of three miRNAs would be 67 nt, which is half of the 135 nt length of 5'UTR. Consequently, the compacting of miRNA binding sites is useful in reducing the proportion of binding sites by 2.6 times in 5'UTR of mRNA *EPOR* gene.

In the mRNA of *MAZ* gene, the binding sites of TJU_CMC_MD2.ID00968.3p-miR, TJU_CMC_MD2.ID01476.3p-miR, miR-1470, and TJU_CMC_MD2.ID00620.3p-miR were located in a cluster with length of 34 nt from 16 nt to 49 nt. The total length of the four miRNAs was equal to 87 nt. Another cluster in *MAZ* mRNA with a length of 44 nt was formed by miR-6850-5p, miR-4466, miR-762, TJU_CMC_MD2.ID00915.3p-miR and TJU_CMC_MD2.ID02979.5p-miR binding sites. Both clusters occupied only 78 nt, and the total length of binding sites of nine miRNAs was 196 nt.

In the mRNA of *NISCH* gene, the binding sites of TJU_CMC_MD2.ID03445.3p-miR, TJU_CMC_MD2.ID01560.3p-miR and TJU_CMC_MD2.ID03119.5p-miR formed a cluster with the length of 35 nt from 31 nt to 64 nt. With cluster formation, the length of these binding sites was 71 nt, i.e. 52 % of the length of 5'UTR equal to 134 nt.

There were 24 miRNAs for which the mRNA was targeted in CDS. The mitogen-activated protein kinase three (*MAPK3*) gene was a target of three miRNAs, the binding sites of which were located in a cluster with the length of 26 nt. The mRNA of *MAZ* gene had miRNA binding sites with overlapping of nucleotide sequences into four different clusters (Table 3). The first cluster with the length of 33 nt included binding sites of miR-6729-5p, TJU_CMC_MD2.ID02623.3p-miR, TJU_CMC_MD2.ID02460.5p-miR and miR-2861. Nucleotide sequences of binding sites cluster encoded the oligopeptide APAPPPTPQA which was conservative in orthologous proteins MAZ of *Hsa, Pab, Ptr, Csa* (Fig. 3A). The second cluster with the length of 74 nt was located from 457 nt to 530 nt. The total length of all binding sites of miRNAs of this cluster was 302 nt. This length requires binding site compaction, since all nucleotides participate in the coding of functionally important amino acids in CDS. Despite the large length of the cluster of nine miRNA binding sites encoding the oligopeptide AAAAAAAAAAAAVAAAPPAPAAA, it was conservative in the MAZ orthologs (Fig. 3B). The third cluster consisted of miR-4706, TJU_CMC_MD2.ID01641.3p-miR, TJU_CMC_MD2.ID01705.3p-miR, and miR-3960 binding sites with the length of 30 nt. The cluster encoded the conservative oligopeptide APPASAAT (Fig. 3C). The fourth cluster with a length of 30 nt included binding sites for three miRNAs from 893 nt to 922 nt. The encoded by cluster oligopeptide GAGGGGGEAG was also conservative (Fig. 3D).

All binding sites for miRNAs that interact with *MAZ* mRNA had a total length of 472 nt, which is approximately 33% of the total CDS length. Clustered binding sites for miRNA occupied only 12% of CDS length equal to 1434 nt. *MAZ* gene was the most vulnerable target for miRNA, so its expression should be monitored as a matter of priority.

Fifteen miRNAs were bound within the mRNA of candidate breast cancer subtype HER2 genes with a free energy of -125 kJ/mole or greater (Table 3). For example, TJU_CMC_MD2.ID01626.3p-miR had a competitive advantage over TJU_CMC_MD2.ID01633.3p-miR and TJU_CMC_MD2.ID01599.3p-miR for binding in the mRNA cluster *EPOR*. In two clusters of *MAZ* mRNA, TJU_CMC_MD2.ID01476.3p-miR and TJU_CMC_MD2.ID00915.3p-miR will predominantly bind. The translation of mRNA *MAZ* gene will be significantly suppressed if TJU_CMC_MD2.ID02294.5p-miR (which had three binding sites) is present, and TJU_CMC_MD2.ID01804.3p-miR and

TJU_CMC_MD2.ID01641.3p-miR had two sites with a free energy of binding of -132 kJ/mole or greater.

In 3'UTR of mRNA *BRCA2* gene, three miRNA binding sites were identified with overlapping of nucleotide sequences (Table 3). The *CDK6* gene was a target for nine miRNAs. miR-548h-3p, miR-548z, miR-548aq-3p, miR-548az-3p, TJU_CMC_MD2.ID03264.3p-miR formed a cluster from 1677 nt to 1699 nt. The mRNA of *BRCA2* and *CDK6* genes had binding sites for miRNA in 3'UTR with a low free energy of binding: from -98 kJ/mole to -117 kJ/mole. In the 3'UTR of mRNA *CDK6* gene with an RPKM value of 2.2, there are ten binding sites of miR-466, nine binding sites of TJU_CMC_MD2.ID00436.3p-miR, seven binding sites of TJU_CMC_MD2.ID01030.3p-miR formed a cluster from 1896 nt to 1948 nt (Table 3). Multiple binding sites for these miRNAs allow them to bind with mRNA and significantly increase the probability of translation inhibition of the mRNA *CDK6* gene.

Compacting of miRNA binding sites is difficult to explain if its sole purpose is saving the length of 3'UTR. Apparently, there are other reasons for compacting binding sites. For example, the binding of one miRNA precludes other miRNAs binding with their site. If this miRNA is a signal of the host gene (gene encoding miRNA), the target gene will not perceive this signal. That is, there is competition between different miRNAs for the binding site and for the ability to regulate the expression of the target gene.

It should be noted that most miRNA binding sites were located at the beginning of 5'UTR and CDS mRNA regions of *MAZ* gene (Table 3). This localization of miRNA binding sites allows protein synthesis to be stopped earlier in the case of the formation of abortive proteins. For example, the first three clusters of miRNA binding sites were located in CDS of mRNA *MAZ* gene comprise an area from the 158 nt to the 477 nt. All binding sites of nine miRNAs in 5'UTR of mRNA *MAZ* were located from 16 nt to 114 nt of the 168 nt length of 5'UTR.

The miRNA binding sites in 3'UTR of mRNA *BRCA2* and *CDK6* genes were also located at the beginning of 3'UTR (Table 3).

Fig. S3 shows the schemes of interaction of miRNAs with mRNAs of candidate genes of the subtype HER2. Several miRNAs and their target genes: TJU_CMC_MD2.ID02998.3p-miRand *MAZ* gene, miR-5008-5p and *MAZ* gene, TJU_CMC_MD2.ID02499.3p-miR and *MAZ* gene, miR-6805-3p and *MAPK3* gene, miR-3960 and *MAZ* gene. miR-877-3p, miR-7111-3p, TJU_CMC_MD2.ID01352.3p-miR had binding sites in the same region of the mRNA *MAZ* gene from 2273 nt to 2774 nt. All miRNA nucleotides form hydrogen bonds with this region of mRNA.

After discovery of clusters of miRNA binding sites with mRNA of candidate genes of breast cancer subtypes, the question arises as to how stable these structural forms are. It is known that some miRNAs arose in early stages of evolution and are stable for tens of millions of years of species divergence (*Kondybayeva et al., 2018*). Other miRNA associations with mRNA have appeared recently, and they are not observed even in closely related species. In this regard, we checked the variability of nucleotide sequences of binding sites in clusters identified by us. In Tables S3-5 of are given results of analysis of nucleotide sequences of clusters in mRNA candidate breast cancer subtypes. The data obtained show that in most cases the nucleotide sequences of clusters are identical. Observed differences in single nucleotides slightly change the degree of interaction of miRNA with binding sites. Consequently, established bindings between miRNAs and binding sites organized in clusters are stable in genomes of objects studied which have diverged for millions of years. The evolutionary conservatism of the associations of

miRNAs and mRNAs allows the choice of adequate animal models for studying associations of miRNAs and mRNAs.

Discussion

Dysregulation of gene expression by miRNAs is one of the causes of oncogenesis (*Adhami et al., 2018*). The information about the participation of miRNAs in oncogenesis testifies to their important role in this process (*Persson et al., 2011*). Many studies have been devoted to the study of miRNA in various diseases, including triple-negative BC (*Bar et al., 2017*; *Buschmann et al., 2018*; *Yao et al., 2018*), luminal A and B subtypes of BC (Aure *et al.*, 2017; *Hannafon et al., 2016*; *Wang & Luo, 2015*) and HER2 subtype of BC (*Halvorsen et al., 2017*; *Patel et al., 2016*; *Wang & Lin, 2013a*). However, there are few reliably established associations of miRNA and target genes. In this paper, we identified associations of the miRNAs and their target genes that may be responsible for the development of breast cancer, taking into account the subtypes of the disease. The selectivity of miRNA and mRNA interaction, as well as the facts of preservation of the entire miRNA nucleotide sequence and the corresponding binding sites in the mRNA of target genes, demonstrates the stability of these interactions over tens of millions of years of evolution.

We applied new bioinformatics approaches to the assessment of these relationships, which allowed us to reveal important characteristics of the binding between the miRNAs and target genes (Tables 1-3). Previously experimentally established miRNA binding sites with mRNA of target *RTL1* genes (*Davis et al., 2005*) were verified using the MirTarget program. *RTL1* is the host gene for 10 miRNA and through them participates in oncogenesis. For example, miR-127-5p can suppress the expression of a gene that is a tumorsupressor (*Wang et al., 2011a*), miR-136-5p can regulate the expression of *CLDN15, ENAH* genes involved in a tumor invasion (*Forse et al., 2015*; *Takehara et al., 2009*), miR-432-3p can affect the gene *IL2RB* involved in the development of breast cancer (*Garcia-Tunon et al., 2004*). We confirmed with the MirTarget program the experimentally established seven binding sites for miR-127, miR-136, miR-431, miR-433-3p/5p, miR-434-3p/5p with the mRNA of *RTL1* gene, and predicted other three binding sites for miRNAs (*Davis et al., 2005*). All ten miRNA binding sites are located in the CDS of mRNA gene and the nucleotide sequences of miRNA and mRNA are fully complementary. Note that the nucleotide sequences miR-127-5p and miR-127-3p are conservative in species diverted tens of millions of years ago, according to miRBase (http://www.mirbase.org/cgi-bin/query.pl?terms=mir-127&submit=Search). Using the MirTarget program, we predicted miRNA binding sites in the CDS, 5'UTR and 3'UTR of many genes (*Ivashchenko et al., 2016*; *Ivashchenko et al., 2014a*; *Ivashchenko et al., 2014b*; *Ivashchenko, Issabekova & Berillo, 2013*). The nucleotide sequences of miRNA and their binding sites have been conserved in the mRNA genes of animals and plant organisms over tens of millions of years of evolution (*Bari, Orazova & Ivashchenko, 2013*; *Ivashchenko et al., 2016*; *Ivashchenko et al., 2014a*; *Ivashchenko et al., 2014b*; *Ivashchenko, Issabekova & Berillo, 2013*).

On the basis of the MirTarget program, the organization of binding sites was established in arranged located sites with overlapping nucleotide sequences. In this miRNA binding sites cluster of mRNA, several miRNAs can competitively interact. The organization of binding sites into clusters has two consequences: a) compacts binding sites to reduce their share in the total length of mRNA; b) competitive between miRNA for binding to mRNA is created taking into account the free energy of miRNA interaction with mRNA.

The competition between miRNA complicates the interpretation of the expected effects of changing the miRNA concentration. In diseases, the concentration of miRNA can increase and decrease dozens of times (*Lu et al., 2016; Yang, Sui & Liang G, 2017*). For example, a decrease in the concentration of one miRNA may not cause an effect, since other miRNAs will inhibit protein translation. In most cases, the binding of miRNA with mRNA is unlikely to lead in a complete suppression of translation, because a fully complementary interaction of miRNA with mRNA is not observed (Davis *et al*., 2005). Even a few miRNAs will not cause complete suppression of translation if their concentration is less than the concentration of mRNA.

It has been found that some miRNAs may bind to mRNA of more than one candidate gene of different subtypes. For example, TJU_CMC_MD2.ID03445.3p-miR may interact with the mRNA of *NISCH* gene of HER2 subtype, mRNA of *RAB5A* gene of triple-negative subtype, and mRNA of *SOX4* gene of luminal A and B subtype (Tables 1-3). TJU_CMC_MD2.ID00061.3p-miR can bind to mRNA of *RAB5A* gene of triple-negative subtype and mRNA of *FOXA1* gene of luminal A and B (Tables 1 and 2). TJU_CMC_MD2.ID01641.3p-miR can interact with mRNA of *FOXA1, HMGA2, TGFB1* genes of luminal A and B subtypes, and with mRNA of *MAZ* gene of HER2 subtype (Tables 2 and 3). miR-466 could bind to mRNA of *RUNX1, SFN* and *CDK6* genes. One miRNA can interact with mRNA of two or more candidate genes of the same subtype. For example, miR-6089 can bind to mRNA of *SMAD3* and *TGFB1* genes, while it has two and four consecutive sites (Table 2), and in mRNA of *SFN* gene one binding site (Table 1). The TJU_CMC_MD2.ID00367.5p-miR and miR-1273g-3p binding sites were located through seven nucleotides in mRNA of *ATM* gene and *STMN1* gene (Table 1). Binding sites of these miRNAs were part of corresponding clusters.

A distinctive feature of candidate genes of triple-negative subtype is the presence in mRNA of several genes of miR-1273 family binding sites (Table 1). The mRNA of *IL11* gene in the cluster included binding sites of miR-1273d, miR-1273e and miR-1273f. The mRNA genes of *ATM, CBL, STMN1* included binding sites of miR-1273a and miR-1273g-3p, the distance between the origins of their binding sites was 22 nt. These characteristics of interaction of miR-1273 family with candidate genes of triple-negative subtype are obviously not random, as is the fact that mRNA candidate genes of luminal A and B, and HER2 subtypes lack the binding sites of the miR-1273 family (Tables 2 and 3). For the first time we predicted that several consecutive binding sites (multiple sites) of a single miRNA. For example, miR-466 had two binding sites in mRNA of *RUNX1* gene, six binding sites in mRNA of *SFN* gene, and 10 binding sites in mRNA of *CDK6* gene (Tables 1 and 3). Binding sites of TJU_CMC_MD2.ID01030.3p-miR and TJU_CMC_MD2.ID00436.3p-miR were included in clusters of mRNA binding sites of *ATM* and *SFN* genes (Table 1). The mRNA of *SFN* gene had six binding sites with TJU_CMC_MD2.ID01030.3p-miR and TJU_CMC_MD2.ID00436.3p-miR. It is known that single nucleotide, dinucleotide, trinucleotide repeats are found in mRNA genes. We have shown which repeats are targets for miRNA (*Kondybayeva et al., 2018*; *Niyazova et al., 2015*). Position of binding sites in the same cluster suggests competition between miRNA for binding to mRNA of target gene. Competition among miRNAs for binding sites also occurs when they are linked to mRNA of different genes, if they are expressed both in a single cell and in different cells of the body, since miRNA is transported through the body with blood (*Hannafon et al., 2016*; *Lagendijk et al., 2018*; *Zhang et al., 2017*). It is necessary to take into account the concentration of miRNA and mRNA to explain the effectiveness of their interaction.

Some miRNAs can bind with high free energy to mRNA of targets genes of different subtypes: TJU_CMC_MD2.ID.01804.3p-miR, TJU_CMC_MD2.ID.00252.5p-miR, miR-6089,

TJU_CMC_MD2.ID.02294.5p-miR, TJU_CMC_MD2.ID.00296.3p-miR, TJU_CMC_MD2.ID.01641.3p-miR (Tables 1-3). Such associations of miRNA and target gene can be used as markers of two BC subtypes, since the expression of these genes will be significantly suppressed by the corresponding miRNA. Simultaneously with controlling the expression of these miRNA and gene associations, it is necessary to control the expression of specific associations for each subtype. For example, such associations may be TJU_CMC_MD2.ID.03332.3p-miR, TJU_CMC_MD2.ID.02761.3p-miR and the *CBL* gene, TJU_CMC_MD2.ID.02930.3p-miR and the *RAB5A* gene for triple-negative subtype (Table 1) for the triple-negative subtype, TJU_CMC_MD2.ID.01702.3p-miR and the *FOXA1* gene for the luminal subtype A and B (Table 2), TJU_CMC_MD2.ID.1476.3p-miR, TJU_CMC_MD2.ID.02294.5p-miR and the *MAZ* gene for the subtype HER2 (Table 3).

Tables 1-3 provide information (RPKM) on the normal expression of candidate genes in the mammary gland. The most strongly expressed genes are *MMP2* (Table 1), *ITGB1* (Table 2), *NISCH, MAPK3* (Table 3). The mRNA of *MMP2* and *ITGB1* genes contain clusters of binding sites for five miRNAs, and the mRNA of *NISCH* and *MAPK3* genes for three miRNAs. Consequently, the expression of these candidate genes and miRNAs binding in respective clusters can be used to develop methods for diagnosing BC subtypes. The *HMGA2* gene is not normally expressed (Table 2); however, its mRNA has two binding clusters for 18 miRNAs, and some miRNAs can bind with the large free energy to mRNA, which suggests suppression of its possible expression. Several studies have shown that the gene can be expressed in tumor cells and its increased expression leads to the development of oncogenesis (*Chen et al., 2019*; *Niu et al., 2019*; *Pearlman et al., 2019*; *Piscuoglio et al., 2012*; *Sun et al., 2014*; *Wang et al., 2011b*).

The proposed associations of miRNA and target genes should be analyzed taking into account the following factors: a) miRNA and their target genes perform the limiting stages of key biological processes involved in the development of diseases; b) these binding events have a large free energy of miRNA interaction with mRNA; c) there is a greater number of miRNAs that bind to mRNA; and d) included miRNAs have more target genes. Depending on the circumstances, the adequacy and significance of the listed miRNAs association with mRNA may vary.

Conclusions

The associations of miRNAs and their targets genes have been identified for a set of candidate genes for breast cancer subtypes. The clustering of miRNA binding sites decreases the fraction of nucleotide sequence comprising binding sites in mRNA. The average free energy of miRNA binding in mRNA sites decreases in the order: 5'UTR > CDS > 3'UTR. The cluster organization of miRNA binding sites is mainly manifested in 5'UTR and 3'UTR. In the CDS, the share of miRNA binding sites organized into clusters is less than that of single miRNA binding sites. The cluster organization of miRNA binding sites together with the free energy of miRNA interaction with mRNA causes competition between miRNA for binding to mRNA. This phenomenon demonstrates the competitive relationship of miRNA in the regulation of the expression of target genes. The number of miRNA binding sites in clusters indicates the degree of dependence of the expression of target genes on the expression of other genes generating miRNAs. Some associations of miRNAs and their target genes can be used to develop methods for diagnosing BC subtypes.

Acknowledgements

The work was carried out with the financial support of the Ministry of Education and Science of the Republic of Kazakhstan within the framework of the grant №AP05132460. We are grateful to Pyrkova A.Yu. to performing calculations on the program MirTarget.

References

Adhami M, Haghdoost AA, Sadeghi B, Malekpour AR. 2018. Candidate miRNAs in human breast cancer biomarkers: a systematic review. *Breast Cancer* **25**: 198-205 DOI 10.1007/s12282-017-0814-8.

Atambayeva S, Niyazova R, Ivashchenko A, Pyrkova A, Pinsky I, Akimniyazova A, Labeit S. 2017. The binding sites of miR-619-5p in the mRNAs of human and orthologous genes. *BMC Genomics* **18**: 428 DOI 10.1186/s12864-017-3811-6.

Aure MR, Vitelli V, Jernström S, Kumar S, Krohn M, Due EU, Haukaas TH, Leivonen SK, Vollan HK, Lüders T, Rødland E, Vaske CJ, Zhao W, Møller EK, Nord S, Giskeødegård GF, Bathen TF, Caldas C, Tramm T, Alsner J, Overgaard J, Geisler J, Bukholm IR, Naume B, Schlichting E, Sauer T, Mills GB, Kåresen R, Mælandsmo GM, Lingjærde OC, Frigessi A, Kristensen VN, Børresen-Dale AL, Sahlberg KK. 2017. Integrative clustering reveals a novel split in the luminal A subtype of breast cancer with impact on outcome. *Breast Cancer Research* **19:** 44 DOI 10.1186/s13058-017-0812-y.

Bar I, Merhi A, Abdel-Sater F, Ben Addi A, Sollennita S, Canon JL, Delrée P. 2017. The MicroRNA miR-210 is expressed by cancer cells but also by the tumor microenvironment in triple-negative breast cancer. *J Histochem Cytochem*. **65**: 335-346 DOI 10.1369/0022155417702849.

Bari A, Orazova S, Ivashchenko A. 2013. miR156- and miR171-binding sites in the protein-coding sequences of several plant genes. *Bio.Med.Research International* **2013:** 1-7 DOI 10.1155/2013/307145

Benson JR, Jatoi I. 2012. The global breast cancer burden. *Future Oncology* **8:** 697-702 DOI 10.2217/fon.12.61

Biagioni F, Bossel BMN, Fontemaggi G, Canu V, Mori F, Antoniani B, Di Benedetto A, Santoro R, Germoni S, De Angelis F, Cambria A, Avraham R, Grasso G, Strano S, Muti P, Mottolese M, Yarden Y, Domany E, Blandino G. 2012. miR-10b*, a master inhibitor of the cell cycle, is down-regulated in human breast tumours. *EMBO Mol Med*. **4**: 1214-1229 DOI 10.1002/emmm.201201483.

Blenkiron C, Goldstein LD, Thorne NP, Spiteri I, Chin SF, Dunning MJ, Barbosa-Morais NL, Teschendorff AE, Green AR, Ellis IO, Tavaré S, Caldas C, Miska EA. 2007. MicroRNA expression profiling of human breast cancer identifies new markers of tumor subtype. *Genome Biol*. **8**: R214 DOI 10.1186/gb-2007-8-10-r214

Buschmann D, González R, Kirchner B, Mazzone C, Pfaffl MW, Schelling G, Steinlein O, Reithmair M. 2018. Glucocorticoid receptor overexpression slightly shifts microRNA expression patterns in triple-negative breast cancer. *Int J Oncol*. **52:** 1765-1776 DOI 10.3892/ijo.2018.4336.

Chen X, Zeng K, Xu M, Liu X, Hu X, Xu T, He B, Pan Y, Sun H, Wang S. 2019. P53-induced miR-1249 inhibits tumor growth, metastasis, and angiogenesis by targeting VEGFA and HMGA2. *Cell Death Dis*. **10**: 131 DOI 10.1038/s41419-018-1188-3.

Davis E, Caiment F, Tordoir X, Cavaillé J, Ferguson-Smith A, Cockett N, Georges M, Charlier C. 2005. RNAi-mediated allelic trans-interaction at the imprinted Rtl1/Peg11 locus. *Current Biology* **15**: 743-749 DOI 10.1016/j.cub.2005.02.060

Enerly E, Steinfeld I, Kleivi K, Leivonen SK, Aure MR, Russnes HG, Rønneberg JA, Johnsen H, Navon R, Rødland E, Mäkelä R, Naume B, Perälä M, Kallioniemi O, Kristensen VN, Yakhini Z, Børresen-Dale AL. 2011. miRNA-mRNA integrated analysis reveals roles for miRNAs in primary breast tumors. *PLoS One* **6:** e16915 DOI 10.1371/journal.pone.0016915.

Forse CL, Agarwal S, Pinnaduwage D, Gertler F, Condeelis JS, Lin J, Xue X, Johung K, Mulligan AM, Rohan TE, Bull SB, Andrulis IL. 2015. Menacalc, a quantitative method of metastasis assessment, as a prognostic marker for axillary node-negative breast cancer. *BMC Cancer* **15**: 483 DOI 10.1186/s12885-015-1468-6.

García-Tuñón I, Ricote M, Ruiz A, Fraile B, Paniagua R, Royuela M. 2004. Interleukin-2 and its receptor complex (alpha, beta and gamma chains) in in situ and infiltrative human breast cancer: an immunohistochemical comparative study. *Breast Cancer Res*. **6:** R1-7.

Halvorsen AR, Helland A, Gromov P, Wielenga VT, Talman MM, Brunner *N*, Sandhu V, Børresen-Dale AL, Gromova I, Haakensen VD. 2017. Profiling of microRNAs in tumor interstitial fluid of breast tumors – a novel resource to identify biomarkers for prognostic classification and detection of cancer. *Mol Oncology* **11:** 220-234 DOI 10.1002/1878-0261.12025

Hannafon BN, Trigoso YD, Calloway CL, Zhao YD, Lum DH, Welm AL, Zhao ZJ, Blick KE, Dooley WC, Ding WQ. 2016. Plasma exosome microRNAs are indicative of breast cancer. *Breast Cancer Research* **18**: 90 DOI 10.1186/s13058-016-0753-x

Huggett J, O'Grady J. 2014. Molecular diagnostics: current research and applications. *Caister Academic Press*, 248.

Ivashchenko A, Issabekova AS, Berillo OA. 2013. miR-1279, miR-548j, miR-548m, and miR-548d-5p binding sites in CDSs of paralogous and orthologous PTPN12, MSH6, and ZEB1 Genes. *Biomed Res Int*. **2013:** 902467 DOI 10.1155/2013/902467.

Ivashchenko A, Berillo O, Pyrkova A, Niyazova R, Atambayeva Sh. 2014a. The properties of binding sites of miR-619-5p, miR-5095, miR-5096 and miR-5585-3p in the mRNAs of human genes. *Biomed Research International* **2014a:** e8 DOI 10.1155/2014/720715

Ivashchenko A, Berillo O, Pyrkova A, Niyazova R. 2014b. Binding sites of miR-1273 family on the mRNA of target genes. *Biomed Research International* **2014**: e11 DOI 10.1155/2014/620530

Ivashchenko AT, Pyrkova AY, Niyazova RY, Alybayeva A, Baskakov K. 2016. Prediction of miRNA binding sites in mRNA. *Bioinformation* **12:** 237-240.

Jemal A, Center MM, DeSantis C, Ward EM. 2010. Global patterns of cancer incidence and mortality rates and trends. *Cancer Epidemiology, Biomarkers & Prevention* **10:** 1893-1907 DOI 10.1158/1055-9965.EPI-10-0437

Kondybayeva AM, Akimniyazova AN, Kamenova SU, Ivashchenko AT. 2018. The characteristics of miRNA binding sites in mRNA of *ZFHX3* gene and its orthologs. *Vavilov journal of Genetics and Breeding* **22:** 438-444 DOI 10.18699/VJ18.380

Kool ET. 2001. Hydrogen bonding, base stacking, and steric effects in DNA replication. *Annual Review of Biophysics and Biomolecular Structure* **30:** 1-22 DOI 10.1146/annurev.biophys.30.1.1

Kurozumi S, Yamaguchi Y, Kurosumi M, Ohira M, Matsumoto H, Horiguchi J. 2017. In microRNA research into breast cancer with particular focus on the associations between microRNAs and intrinsic subtypes. *J Hum Genet*. **62:** 15-24 DOI 10.1038/jhg.2016.89

Lagendijk M, Sadaatmand S, Koppert LB, Tilanus-Linthorst MMA, de Weerd V, Ramírez-Moreno R, Smid M, Sieuwerts AM, Martens JWM. 2018. MicroRNA expression in pre-treatment plasma of patients with benign breast diseases and breast cancer. *Oncotarget* **9:** 24335-24346 DOI 10.18632/oncotarget.25262

Lee CH, Kuo WH, Lin CC, Oyang YJ, Huang HC, Juan HF. 2013. MicroRNA-regulated protein-protein interaction networks and their functions inbreast cancer. *Int J Mol Sci*. **14:** 11560-11606 DOI 10.3390/ijms140611560.

Lemieux S, Major F. 2002. RNA canonical and non-canonical base pairing types: a recognition method and complete repertoire. *Nucleic Acids Res*. **30:** 4250-4263 PMID: 12364604

Leontis NB, Stombaugh J, Westhof E. 2002. The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Research* **30:** 3497-3531 PMID: 12177293

Londin E, Loher P, Telonis AG, Quann K, Clark P, Jing Y, Hatzimichael E, Kirino Y, Honda S, Lally M, Ramratnam B, Comstock CE, Knudsen KE, Gomella L, Spaeth GL, Hark L, Katz LJ, Witkiewicz A, Rostami A, Jimenez SA, Hollingsworth MA, Yeh JJ, Shaw CA, McKenzie SE, Bray P, Nelson PT, Zupo S, Van Roosbroeck K, Keating MJ, Calin GA, Yeo C, Jimbo M, Cozzitorto J, Brody JR, Delgrosso K, Mattick JS, Fortina P, Rigoutsos I. 2015. Analysis of 13 cell types reveals evidence for the expression of numerous novel primate- and tissue-specific microRNAs. *PNAS USA* **112:** 1106-1115 DOI 10.1073/pnas.1420955112

Lowery AJ, Miller N, Devaney A, McNeill RE, Davoren PA, Lemetre C, Benes V, Schmidt S, Blake J, Ball G, Kerin MJ. 2009. MicroRNA signatures predict oestrogen receptor, progesterone receptor andHER2/neu receptor status in breast cancer. *Breast Cancer Res*. **11:** R27 DOI 10.1186/bcr2257.

Lu Y, Qin B, Hu H, Zhang J, Wang Y, Wang Q, Wang S. 2016. Integrative microRNA-gene expression network analysis in genetic hypercalciuric stone-forming rat kidney. *PeerJ*. **4:** e1884 DOI 10.7717/peerj.1884.

Mattie MD, Benz CC, Bowers J, Sensinger K, Wong L, Scott GK, Fedele V, Ginzinger D, Getts R, Haqq C. 2006. Optimized high-throughput microRNA expression profiling provides novel biomarker assessment of clinical prostate and breast cancer biopsies. *Mol Cancer* **5:** 24. DOI 10.1186/1476-4598-5-24

McDermott AM, Miller N, Wall D, Martyn LM, Ball G, Sweeney KJ, Kerin MJ. 2014. Identification and validation of oncologic miRNA biomarkers for luminal A-like breast cancer. *PLoS One* **9:** e87032 DOI 10.1371/journal.pone.0087032.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5:** 621-628 DOI 10.1038/nmeth.1226

Nian W, Ao X, Wu Y, Huang Y, Shao J, Wang Y, Chen Z, Chen F, Wang D. 2013. miR-223 functions as a potent tumor suppressor of the Lewis lung carcinoma cell line by targeting

insulin-like growth factor-1 receptor and cyclin-dependent kinase 2. *Oncol. Lett*. **6**: 359-366. DOI 10.3892/ol.2013.1375

Niu Y, Zhou H, Liu Y, Wang Y, Xie J, Feng C, An N. 2019. miR-16 regulates proliferation and apoptosis of pituitary adenoma cells by inhibiting HMGA2. *Oncol Lett*. **17:** 2491-2497 DOI 10.3892/ol.2018.9872.

Niyazova R, Berillo O, Atambayeva Sh, Pyrkova A, Alybayeva A, Ivashchenko A. 2015. miR-1322 binding sites in paralogous and orthologous genes. *BioMed Research Int*. **1**: 7 DOI 10.1155/2015/962637

Patel Y, Shah N, Lee JS, Markoutsa E, Jie C, Liu S, Botbyl R, Reisman D, Xu P, Chen H. 2016. A novel double-negative feedback loop between miR-489 and the HER2-SHP2-MAPK signaling axis regulates breast cancer cell proliferation and tumor growth. *Oncotarget* **7:** 18295-18308 DOI 10.18632/oncotarget.7577

Pearlman A, Rahman M, Upadhyay K, Loke J, Ostrer H. 2019. Ectopic Otoconin 90 expression in triple negative breast cancer cell lines is associated with metastasis functions. *PLoS One* **14:** e0211737 DOI 10.1371/journal.pone.0211737.

Persson H, Kvist A, Rego N, Staaf J, Vallon-Christersson J, Luts L, Loman N, Jonsson G, Naya H, Hoglund M, Borg A, Rovira C. 2011. Identification of new microRNAs in paired normal and tumor breast tissue suggests a dual role for the ERBB2/Her2 gene. *Cancer Res*. **71:** 78-86 DOI 10.1158/0008-5472.CAN-10-1869.

Piasecka D, Braun M, Kordek R, Sadej R, Romanska H. 2018. MicroRNAs in regulation of triple-negative breast cancer progression. *J Cancer Res Clin Oncol*. **144:** 1401-1411 DOI 10.1007/s00432-018-2689-2.

Piscuoglio S, Zlobec I, Pallante P, Sepe R, Esposito F, Zimmermann A, Diamantis I, Terracciano L, Fusco A, Karamitopoulou E. 2012. HMGA1 and HMGA2 protein expression correlates with advanced tumour grade and lymph node metastasis in pancreatic adenocarcinoma. *Histopathology* **60:** 397-404 DOI 10.1111/j.1365-2559.2011.04121.x

Qian P, Banerjee A, Wu ZS, Zhang X, Wang H, Pandey V, Zhang WJ, Lv XF, Tan S, Lobie PE, Zhu T. 2012. Loss of SNAIL regulated miR-128-2 on chromosome 3p22.3 targets multiple stem cell factors to promote transformation of mammary epithelial cells. *Cancer Res*. **72:** 6036-6050 DOI 10.1158/0008-5472.CAN-12-1507

Sun M, Gomes S, Chen P, Frankenberger CA, Sankarasharma D, Chung CH, Chada KK, Rosner MR. 2014. RKIP and HMGA2 regulate breast tumor survival and metastasis through lysyl oxidase and syndecan-2. *Oncogene* **33**: 3528-3537 DOI 10.1038/onc.2013.328

Sung H, Jeon S, Lee KM, Han S, Song M, Choi JY, Park SK, Yoo KY, Noh DY, Ahn SH, Kang D. 2012. Common genetic polymorphisms of microRNA biogenesis pathway genes and breast cancer survival. *BMC Cancer* **12:** 195 DOI 10.1186/1471-2407-12-195.

Tahiri A, Leivonen SK, Lüders T, Steinfeld I, Ragle Aure M, Geisler J, Mäkelä R, Nord S, Riis ML, Yakhini Z, Kleivi Sahlberg K, Børresen-Dale AL, Perälä M, Bukholm IR, Kristensen VN. 2014. Deregulation of cancer-related miRNAs is a common event in both benign and malignant human breast tumors. *Carcinogenesis* **35:** 76-85 DOI 10.1093/carcin/bgt333

Takehara M, Nishimura T, Mima S, Hoshino T, Mizushima T. 2009. Effect of claudin expression on paracellular permeability, migration and invasion of colonic cancer cells. *Biol Pharm Bull*. **32:** 825-831 PMID: 19420749

Telonis AG, Loher P, Jing Y, Londin E, Rigoutsos I. 2015. Beyond the one-locus-one-miRNA paradigm: microRNA isoforms enable deeper insights into breast cancer heterogeneity. *Nucleic Acids Res*. **43:** 9158-9175 DOI 10.1093/nar/gkv922.

Wang S, Huang X, Li Y, Lao H, Zhang Y, Dong H, Xu W, Li JL, Li M. 2011a. RN181 suppresses hepatocellular carcinoma growth by inhibition of the ERK/MAPK pathway. *Hepatology* **53:** 1932-1942 DOI 10.1002/hep.24291.

Wang SE, Lin RJ. 2013a. MicroRNA and HER2-overexpressing cancer. *Microrna* **2:** 137-147 DOI 10.2174/22115366113029990011

Wang S, Li H, Wang J, Wang D. 2013b. Expression of microRNA-497 and its prognostic significance in human breast cancer. *Diagn. Pathol*. **8**: 172 DOI 10.1186/1746-1596-8-172.

Wang X, Liu X, Li AY, Chen L, Lai L, Lin HH, Hu S, Yao L, Peng J, Loera S, Xue L, Zhou B, Zhou L, Zheng S, Chu P, Zhang S, Ann DK, Yen Y. 2011b. Overexpression of HMGA2 promotes metastasis and impacts survival of colorectal cancers. *Clin Cancer Res*. **17:** 2570-2580 DOI 10.1158/1078-0432.CCR-10-2542

Wang W, Luo YP. 2015. MicroRNAs in breast cancer: oncogene and tumor suppressors with clinical potential. *J Zhejiang Univ Sci B*. **16:** 18-31 DOI 10.1631/jzus.B1400184

Wu ZB, Cai L, Lin SJ, Lu JL, Yao Y, Zhou LF. 2013. The miR-92b functions as a potential oncogene by targeting on Smad3 in glioblastomas. *Brain Res*. **1529:** 16-25 DOI 10.1016/j.brainres.2013.07.031.

Yakovchuk P, Protozanova E, Frank-Kamenetskii MD. 2006. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Research* **34:** 564-574 DOI 10.1093/nar/gkj454

Yang S, Sui J, Liang G. 2017. Diagnosis value of aberrantly expressed microRNA profiles in lung squamous cell carcinoma: a study based on the Cancer Genome Atlas. *PeerJ*. **5:** e4101. DOI 10.7717/peerj.4101.

Yang Z, Wu L, Wang A, Tang W, Zhao Y, Zhao H, Teschendorff AE. 2017. dbDEMC 2.0: updated database of differentially expressed miRNAs in human cancers. *Nucleic Acids Res*. **45:** D812-D818 DOI 10.1093/nar/gkw1079.

Yao L, Liu Y, Cao Z, Li J, Huang Y, Hu X, Shao Z. 2018. MicroRNA-493 is a prognostic factor in triple-negative breast cancer. *Cancer Sci*. **109:** 2294-2301 DOI 10.1111/cas.13644.

Yu Y, Wu J, Guan L, Qi L, Tang Y, Ma B, Zhan J, Wang Y, Fang W, Zhang H. 2013. Kindlin 2 promotes breast cancer invasion via epigenetic silencing of the microRNA200 gene family. *Int. J. Cancer* **133:** 1368-1379 DOI 10.1002/ijc.28151.

Zhang K, Wang YW, Wang YY, Song Y, Zhu J, Si PC, Ma R. 2017. Identification of microRNA biomarkers in the blood of breast cancer patients based on microRNA profiling. *Gene* **619:** 10-20 DOI 10.1016/j.gene.2017.03.038.

Table 1(on next page)

Characteristics of miRNA interaction in the mRNA of BC subtype triple-negative

Table 1 Characteristics of miRNA interaction in the mRNA of BC subtype triple-negative

Gene, RPKM	miRNA	Start of binding site, nt	ΔG, kJ/mole	ΔG/ΔGm, %	Length, nt
CBL	TJU_CMC_MD2.ID03332.3p-miR (4)	16 ÷ 25	-134 ÷ -140	90 ÷ 94	24
	TJU_CMC_MD2.ID01310.3p-miR (4)	17 ÷ 26	-121	92	22
	TJU_CMC_MD2.ID02761.3p-miR	28	-138	93	24
3.9	miR-1908-3p	30	-121	92	21
	TJU_CMC_MD2.ID00278.3p-miR	32	-125	91	23
	TJU_CMC_MD2.ID02430.3p-miR	34	-110	98	18
MMP2	TJU_CMC_MD2.ID00278.3p-miR	110	-123	89	23
	TJU_CMC_MD2.ID01310.3p-miR	113	-121	92	22
192.4	TJU_CMC_MD2.ID03037.3p-miR	115	-121	90	22
	TJU_CMC_MD2.ID03345.5p-miR	124	-127	90	24
	TJU_CMC_MD2.ID03368.3p-miR	125	-117	89	23
RAB5A	TJU_CMC_MD2.ID02930.3p-miR	184	-132	89	24
	TJU_CMC_MD2.ID03445.3p-miR	189	-127	90	24
	TJU_CMC_MD2.ID01859.5p-miR	191	-121	89	23
16.1	TJU_CMC_MD2.ID01804.3p-miR	325	-140	88	25
	TJU_CMC_MD2.ID03367.5p-miR	328	-121	97	20
	TJU_CMC_MD2.ID00061.3p-miR	334	-127	92	22
*ATM***	TJU_CMC_MD2.ID03006.5p-miR	9778	-121	89	24
	miR-5095	9787	-108	93	21
	miR-619-5p	9793	-119	98	22
3.9	miR-1273a	11054	-119	90	25
	TJU_CMC_MD2.ID00367.5p-miR	11069	-110	90	22
	miR-1273g-3p	11076	-113	96	21
*CBL***	miR-1273a	7727	-117	89	25
3.9	TJU_CMC_MD2.ID01838.5p-miR	7728	-117	93	24
	miR-1273g-3p	7749	-115	98	21
*IL11***	miR-1273f	1466	-102	98	19
	miR-1273d	1467	-121	89	25
0.1	TJU_CMC_MD2.ID01404.5p-miR	1470	-113	91	23
	miR-1273e	1476	-113	96	22
*RUNX1***	TJU_CMC_MD2.ID01030.3p-miR (2)	5454 ÷ 5464	-108 ÷ -113	89 ÷ 93	23
9.0	miR-466 (2)	5456 ÷ 5460	-106 ÷ -110	91 ÷ 95	23
	TJU_CMC_MD2.ID00436.3p-miR	5464	-108	93	23
*SFN***	miR-6089	826	-129	87	24
	TJU_CMC_MD2.ID01774.5p-miR	835	-129	90	23
	miR-6846-5p	839	-113	91	22
	TJU_CMC_MD2.ID00790.3p-miR	1179	-104	89	23
	TJU_CMC_MD2.ID02868.3p-miR	1188	-113	90	23
	TJU_CMC_MD2.ID00436.3p-miR	1190	-104	89	23
9.4	miR-466 (6)	1190 ÷ 1200	-106	91	23
	TJU_CMC_MD2.ID01030.3p-miR (6)	1190 ÷ 1200	-108	89	23
	TJU_CMC_MD2.ID00436.3p-miR (6)	1192 ÷ 1202	-104	89	23
	TJU_CMC_MD2.ID01727.5p-miR (2)	1203 ÷ 1205	-104 ÷ -106	89 ÷ 91	23
	TJU_CMC_MD2.ID02882.3p-miR	1210	-108	91	21
*STMN1***	miR-1273a	1729	-115	87	25
	TJU_CMC_MD2.ID03011.5p-miR	1730	-106	91	22
6.6	TJU_CMC_MD2.ID00367.5p-miR	1744	-113	91	22
	miR-1273g-3p	1751	-108	93	21

Note. In Tables 1, 2 and 3 shown: Genes without * - miRNA binding sites are in the 5'UTR, genes with * - miRNA

binding sites are in the CDS, ** - miRNA binding sites are in the 3'UTR; in parentheses indicates the number of binding sites; ÷ - the change of the parameter in the interval.

2

Table 2(on next page)

Characteristics of miRNAs interaction in mRNA of BC subtype luminal A and B

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27583v1 | CC BY 4.0 Open Access | rec: 13 Mar 2019, publ: 13 Mar 2019

Table 2 Characteristics of miRNAs interaction in mRNA of BC subtype luminal A and B

Gene, RPKM	miRNA	Start of binding site, nt	ΔG, kJ/mole	ΔG/ΔGm, %	Length, nt
FOXA1	TJU_CMC_MD2.ID00297.5p-miR	99	-123	89	24
	TJU_CMC_MD2.ID02106.3p-miR	110	-123	89	23
	TJU_CMC_MD2.ID00252.5p-miR	111	-140	94	24
	TJU_CMC_MD2.ID02769.5p-miR	112	-127	92	22
	TJU_CMC_MD2.ID00296.3p-miR	115	-140	89	25
	TJU_CMC_MD2.ID01099.5p-miR	116	-108	100	17
	TJU_CMC_MD2.ID00071.3p-miR (2)	118 ÷ 121	-117 ÷ -121	93 ÷ 97	20
	TJU_CMC_MD2.ID01190.5p-miR	118	-108	100	17
	TJU_CMC_MD2.ID02457.3p-miR	118	-108	100	17
10.2	TJU_CMC_MD2.ID02595.5p-miR	118	-115	92	20
	TJU_CMC_MD2.ID01403.5p-miR	120	-123	91	23
	TJU_CMC_MD2.ID01702.3p-miR	120	-140	89	25
	miR-3960	120	-115	92	20
	TJU_CMC_MD2.ID03367.5p-miR (2)	121 ÷ 122	-117	93	20
	TJU_CMC_MD2.ID01641.3p-miR	122	-134	90	24
	TJU_CMC_MD2.ID00457.3p-miR	124	-123	91	22
	TJU_CMC_MD2.ID00061.3p-miR	127	-129	94	22
	TJU_CMC_MD2.ID02499.3p-miR (2)	127 ÷ 130	-119 ÷ -121	92 ÷ 93	21
HMGA2	miR-3960	549	-117	93	20
	miR-6756-5p	529	-117	87	23
	TJU_CMC_MD2.ID01737.3p-miR	539	-119	93	21
	TJU_CMC_MD2.ID01041.5p-miR (2)	541 ÷ 544	-129 ÷ -134	88 ÷ 91	24
	TJU_CMC_MD2.ID00089.3p-miR	542	-125	91	22
	TJU_CMC_MD2.ID01323.3p-miR	542	-117	95	20
	TJU_CMC_MD2.ID02296.5p-miR	542	-115	93	20
0.0	TJU_CMC_MD2.ID00296.3p-miR	544	-146	93	25
	TJU_CMC_MD2.ID01641.3p-miR	544	-142	96	24
	TJU_CMC_MD2.ID01403.5p-miR	547	-119	88	23
	TJU_CMC_MD2.ID00061.3p-miR	550	-132	95	22
	TJU_CMC_MD2.ID03367.5p-miR	550	-115	92	20
	miR-4739	573	-123	87	25
	TJU_CMC_MD2.ID00425.5p-miR	575	-121	88	24
	TJU_CMC_MD2.ID00564.5p-miR	585	-110	90	22
*ITGB1**	TJU_CMC_MD2.ID02187.5p-miR	91	-127	92	23
	miR-4787-5p	92	-123	92	22
63.6	TJU_CMC_MD2.ID00457.3p-miR	95	-123	91	22
	TJU_CMC_MD2.ID02770.5p-miR	98	-117	93	20
	TJU_CMC_MD2.ID01184.3p-miR	101	-117	93	20
*HMGA2***	TJU_CMC_MD2.ID01970.3p-miR	1255	-113	90	23
0.0	TJU_CMC_MD2.ID00849.3p-miR (2)	1261 ÷ 1268	-117	90	22
	TJU_CMC_MD2.ID01545.3p-miR	1275	-115	95	21
*SMAD3***	miR-4690-5p	2066	-115	92	22
	miR-3620-5p (2)	2069 ÷ 2074	-117 ÷ -115	87 ÷ 89	22
	TJU_CMC_MD2.ID02822.5p-miR	2070	-127	91	23
14.0	TJU_CMC_MD2.ID00978.5p-miR	2072	-119	90	22
	miR-6089 (2)	2073 ÷ 2078	-132 ÷ -136	89 ÷ 91	24
	TJU_CMC_MD2.ID01382.3p-miR	2075	-113	93	20
*SOX4***	TJU_CMC_MD2.ID01839.3p-miR	2994	-123	89	23
	TJU_CMC_MD2.ID01282.3p-miR	3000	-125	95	23
13.2	TJU_CMC_MD2.ID03445.3p-miR	3000	-127	90	24

	TJU_CMC_MD2.ID00101.3p-miR	3001	-115	92	22
*TGFB1***	TJU_CMC_MD2.ID03306.3p-miR	2060	-123	94	21
	miR-6089 (4)	2060 ÷ 2095	-132 ÷ -136	89 ÷ 91	24
	TJU_CMC_MD2.ID01382.3p-miR	2062	-113	93	20
19.5	TJU_CMC_MD2.ID03208.5p-miR	2066	-125	88	24
	miR-3620-5p	2086	-115	87	22
	TJU_CMC_MD2.ID00978.5p-miR	2089	-119	90	22
	TJU_CMC_MD2.ID00296.3p-miR	2093	-140	89	25

3

4

Table 3(on next page)

Characteristics of miRNA interaction in mRNA genes of BC subtype HER2

Table 3 Characteristics of miRNA interaction in mRNA genes of BC subtype HER2

Gene, RPKM	miRNA	Start of binding site, nt	ΔG, kJ/mole	ΔG/ΔGm, %	Length, nt
EPOR 8.1	TJU_CMC_MD2.ID01633.3p-miR	77	-108	91	21
	TJU_CMC_MD2.ID01599.3p-miR	79	-119	89	23
	TJU_CMC_MD2.ID01626.3p-miR	80	-129	90	23
MAZ 9.5	TJU_CMC_MD2.ID00968.3p-miR	16	-117	93	20
	TJU_CMC_MD2.ID01476.3p-miR	16	-134	91	23
	miR-1470	18	-123	97	21
	TJU_CMC_MD2.ID00620.3p-miR	27	-127	91	23
	miR-6850-5p	92	-115	87	22
	miR-4466	107	-110	98	18
	miR-762	111	-123	91	22
	TJU_CMC_MD2.ID00915.3p-miR	112	-127	88	24
	TJU_CMC_MD2.ID02979.5p-miR	114	-121	92	22
NISCH 32.2	TJU_CMC_MD2.ID03445.3p-miR	31	-125	88	24
	TJU_CMC_MD2.ID01560.3p-miR	38	-123	89	23
	TJU_CMC_MD2.ID03119.5p-miR	41	-125	88	24
*MAPK3** 32.6	TJU_CMC_MD2.ID00149.3p-miR	1144	-117	93	22
	TJU_CMC_MD2.ID01748.3p-miR	1144	-110	91	21
	miR-6805-3p	1145	-117	87	23
*MAZ** 9.5	miR-6729-5p	361	-115	87	22
	TJU_CMC_MD2.ID02623.3p-miR	363	-125	89	23
	TJU_CMC_MD2.ID02460.5p-miR	372	-119	92	22
	miR-2861	375	-110	95	19
	TJU_CMC_MD2.ID02294.5p-miR (3)	457 ÷ 469	-134 ÷ -138	91 ÷ 94	24
	TJU_CMC_MD2.ID02986.5p-miR	459	-119	93	21
	TJU_CMC_MD2.ID01819.5p-miR	461	-125	87	25
	TJU_CMC_MD2.ID01804.3p-miR (2)	464 ÷ 467	-140	88	25
	TJU_CMC_MD2.ID02064.5p-miR	489	-121	92	21
	TJU_CMC_MD2.ID02538.3p-miR	489	-125	94	22
	TJU_CMC_MD2.ID00296.3p-miR	500	-138	88	25
	miR-3960	505	-119	95	20
	TJU_CMC_MD2.ID01641.3p-miR	506	-132	89	24
	miR-4706	605	-123	87	25
	TJU_CMC_MD2.ID01641.3p-miR	608	-134	90	24
	TJU_CMC_MD2.ID01705.3p-miR	608	-117	92	21
	miR-3960	612	-117	93	20
	TJU_CMC_MD2.ID01768.3p-miR	893	-113	90	22
	TJU_CMC_MD2.ID01911.5p-miR	900	-123	89	23
	TJU_CMC_MD2.ID00849.3p-miR	901	-125	97	22
*BRCA2*** 0.1	TJU_CMC_MD2.ID00112.5p-miR	10722	-102	91	21
	TJU_CMC_MD2.ID02744.3p-miR	10738	-104	92	22
	miR-619-5p	10746	-117	96	22
*CDK6*** 2.2	miR-548h-3p	1677	-104	91	23
	miR-548z	1677	-104	91	23
	miR-548aq-3p	1678	-102	94	22
	miR-548az-3p	1678	-98	94	21
	TJU_CMC_MD2.ID03264.3p-miR	1678	-98	90	22
	miR-466 (10)	1908 ÷ 1926	-104 ÷ -108	90 ÷ 93	23
	TJU_CMC_MD2.ID00436.3p-miR (9)	1896 ÷ 1920	-104 ÷ -106	89 ÷ 91	23
	TJU_CMC_MD2.ID01030.3p-miR (7)	1900 ÷ 1918	-108 ÷ -115	89 ÷ 95	23
	TJU_CMC_MD2.ID02513.5p-miR	1901	-102	91	22

Figure 1(on next page)

Location of nucleotide sequences of miRNA binding sites cluster in mRNA *CBL* gene.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27583v1 | CC BY 4.0 Open Access | rec: 13 Mar 2019, publ: 13 Mar 2019

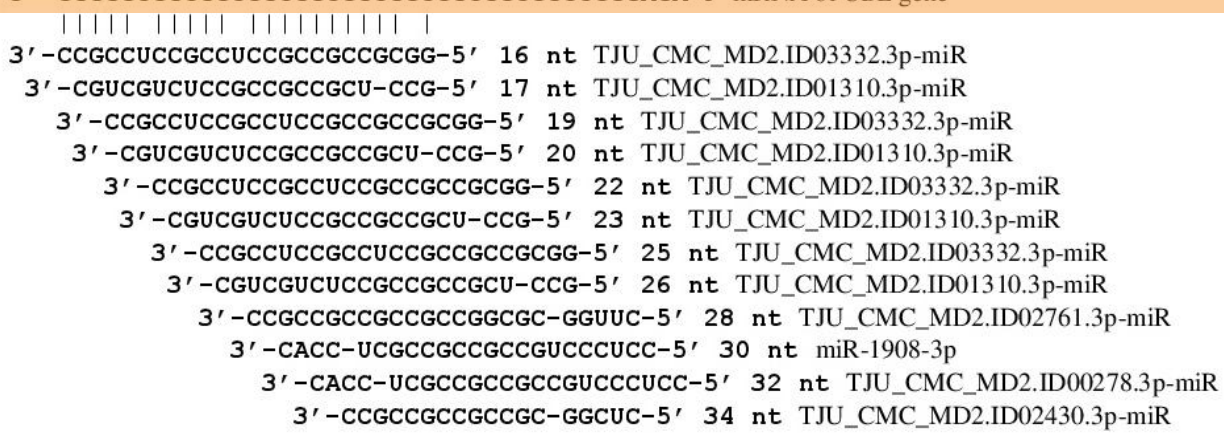

Figure 1 Location of nucleotide sequences of miRNA binding sites cluster in mRNA *CBL* gene.

Figure 2(on next page)

Schemes of interactions of miRNAs with mRNA of *CBL* gene in cluster of binding sites.

TJU_CMC_MD2.ID03332.3p-miR; 5'UTR; 16; -134; 90; 24 5'-GGCGGCGGCGGCGGCGGCGGCGGC-3' ||||| ||||| |||||||||| | 3'-CCGCCUCCGCCUCCGCCGCCGCGG-5'	TJU_CMC_MD2.ID03332.3p-miR; 5'UTR; 19; -134; 90; 24 5'-GGCGGCGGCGGCGGCGGCGGCGGC-3' ||||| ||||| |||||||||| | 3'-CCGCCUCCGCCUCCGCCGCCGCGG-5'
TJU_CMC_MD2.ID03332.3p-miR; 5'UTR; 22; -134; 90; 24 5'-GGCGGCGGCGGCGGCGGCGGCGGC-3' ||||| ||||| |||||||||| | 3'-CCGCCUCCGCCUCCGCCGCCGCGG-5'	TJU_CMC_MD2.ID03332.3p-miR; 5'UTR; 25; -140; 94; 24 5'-GGCGGCGGCGGCGGCGGCGGCGGCC-3' ||||| ||||| |||||||||| || 3'-CCGCCUCCGCCUCCGCCGCCGC-GG-5'
TJU_CMC_MD2.ID01310.3p-miR; 5'UTR; 17; -121; 92; 22 5'-GC**G**GC**G**GCGGCGGCGGCG**G**CGGC-3' |||||| ||||||||| ||| 3'-CG**U**CG**U**CUCCGCCGCCGC**U**-CCG-5'	TJU_CMC_MD2.ID01310.3p-miR; 5'UTR; 20; -121; 92; 22 5'-GC**G**GC**G**GCGGCGGCGGCG**G**CGGC-3' |||||| ||||||||| ||| 3'-CG**U**CG**U**CUCCGCCGCCGC**U**-CCG-5'
TJU_CMC_MD2.ID01310.3p-miR; 5'UTR; 23; -121; 92; 22 5'-GC**G**GC**G**GCGGCGGCGGCG**G**CGGC-3' |||||| ||||||||| ||| 3'-CG**U**CG**U**CUCCGCCGCCGC**U**-CCG-5'	TJU_CMC_MD2.ID01310.3p-miR; 5'UTR; 26; -121; 92; 22 5'-GC**G**GC**G**GCGGCGGCGGCG**G**CGGC-3' |||||| ||||||||| ||| 3'-CG**U**CG**U**CUCCGCCGCCGC**U**-CCG-5'
TJU_CMC_MD2.ID02761.3p-miR; 5'UTR; 28; -138; 93; 24 5'-GGCGGCGGCGGCGGCGGCGGCC**GG**G-3' |||||||||||||| ||| |||| 3'-CCGCCGCCGCCGCCGGCGC-GG**UU**C-5'	miR-1908-3p; 5'UTR; 30; -121; 92; 21 5'-CGGCGGCGGCGGCGGCGGCCGG-3' ||| ||||| | |||||||||| 3'-GCC-CCGCCUCGGCCGCCGGCC-5'
TJU_CMC_MD2.ID00278.3p-miR; 5'UTR; 32; -125; 91; 23 5'-G**C**GGC**G**GCGGCGGCGGCCGGGAG**A**-3' |||| ||||||||||| |||||| 3'-C**A**CC-**U**CGCCGCCGCCGUCCCUC**C**-5'	TJU_CMC_MD2.ID02430.3p-miR; 5'UTR; 34; -110; 98; 18 5'-GGCGGCGGCGGCGGCGCCG**G**G-3' ||||||||||||| ||||| 3'-CCGCCGCCGCCGC-GGC**U**C-5'
Note: miRNA; the miRNA region; start of binding site (nt); the free energy, ΔG (kJ/mole); the ΔG/ΔGm (%); length of miRNA (nt). The upper and lower nucleotide sequences of mRNA and miRNA, respectively. The nucleotides of non-canonical pairs G-U and A-C highlighted in bold type.	

Figure 2 Schemes of interactions of miRNAs with mRNA of *CBL* gene in cluster of binding sites.

Figure 3(on next page)

Logo plots of variation of amino acids in the region of orthologous MAZ protein.

They containing: APAPPPTPQA oligopeptide (A), AAAAAAAAAAAAAVAAAPPAPAAA oligopeptide (B), APPASAAT oligopeptide (C), and GAGGGGGEAG oligopeptide (D) of *Hsa, Pab, Ptr, Csa.* Conservative oligopeptides are highlighted in blue.

Figure 3 Logo plots of variation of amino acids in the region of orthologous MAZ proteins containing: APAPPPTPQA oligopeptide (A), AAAAAAAAAAAAAVAAAPPAPAAA oligopeptide (B), APPASAAT oligopeptide (C), and GAGGGGGEAG oligopeptide (D) of *Hsa, Pab, Ptr, Csa.* Conservative oligopeptides are highlighted in blue.