MicroRNAs and genomic variations: from Proteus tricks to Prometheus gift

Muller Fabbri, Nicola Valeri and George A. Calin

Department of Molecular Virology, Immunology and Medical Genetics, Comprehensive Cancer Center, Ohio State University, Columbus, OH 43210, USA and 1Department of Experimental Therapeutics and Cancer Genetics, University of Texas M. D. Anderson Cancer Center, Houston, TX 77030, USA

*To whom correspondence should be addressed. Tel: +713 792 5461; Fax: +713 745 4528; Email: gcalin@mdanderson.org

MicroRNAs (miRNAs) are small non-coding RNAs with regulatory functions. MiRNAs are aberrantly expressed in almost all human cancers, leading to abnormal levels of target genes. Recently, an increasing number of studies have addressed whether genomic variations including germ line or somatic mutations and single-nucleotide polymorphisms can count for miRNA abnormal expression by altering their biogenesis and/or affect the ability of miRNAs to bind to target messenger RNAs. Here, we provide an extensive review of the studies that have investigated variations occurring both in miRNA genes and in target genes and we discuss the possible clinical implications of these findings. Furthermore, we propose that sequence variations in miRNAs or interactor sites located in miRNAs can be involved in cancer predisposition.

Introduction

MicroRNAs (miRNAs) are small (18–24 nucleotides (nts) in length) non-coding RNAs, which play important roles in regulating the expression of target messenger RNAs (mRNAs). MiRNAs are involved in a variety of biological processes such as cell proliferation and apoptosis, development, differentiation, senescence, metabolism and immunity (1–9). Initially identified in the nematode Caenorhabditis elegans (10), miRNAs are largely conserved in a variety of organisms (11) and share regulatory mechanisms similar to those of siRNAs (10). Recently, >650 miRNAs have been identified in the human genome (12), and it has been estimated that ~30% of the protein-coding genes might be regulated by miRNAs (13). Currently, we propose that sequence variations in miRNAs or interactor sites located in miRNAs can be involved in cancer predisposition.

Biogenesis and processing of miRNAs

MiRNA genes have been identified in all human chromosomes, either as independent non-coding RNAs or in the introns of protein-coding genes (25). In some cases, they are clustered in polycistronic transcripts (26,27). Interestingly, miRNA genes are more frequently located in cancer-associated genomic regions (16), which include minimal regions of amplification, loss of heterozygosity, common breakpoint regions in or near OGs and tumor suppressor genes and fragile sites, which are preferential sites of translocation, chromatid exchange, deletion, amplification or integration of plasmid DNA and tumor-associated viruses. This association represented the first evidence in support of miRNA involvement in cancer (16). Figure 1 schematically shows the main steps of miRNA biogenesis/processing. MiRNAs are initially transcribed by RNA polymerase II as long primary transcripts called primary-miRNAs (pri-miRNAs). A double-stranded RNA-specific ribonuclease called Drosophila, in conjunction with its binding partner DGCR8 (DGeorge syndrome critical region gene 8, or Pasha), processes pri-miRNAs into hairpin RNAs of 70–100 nts known as pre-miRNAs, which contain a two nt 3’ overhang characteristic of RNome III cleavage products (28). Transformed to the cytoplasm by the nuclear export factor Exportin 5, pre-miRNAs are processed by a ribonuclease complex, composed of a ribonuclease III (Dicer), and TRBP (HIV-1 transactivating response RNA-binding protein) in an 18–24 nt duplex. This duplex is incorporated into a large protein complex called RNA-induced silencing complex (RISC), which includes proteins of the Aragonate family (Ag01–4 in humans). One strand of the miRNA duplex remains stably associated with RISC and becomes the mature miRNA. The opposite strand, called passenger strand or miRNA*, is discarded through two different mechanisms. When the miRNA duplex is loaded into a RISC containing Ago2, the passenger strand may be cleaved. Alternatively, RISC containing any other Ago protein may remove the passenger strand through a bypass mechanism, which does not involve cleavage but presumably is based on duplex unwinding (2,29–33).

Abbreviations: DHFR, dihydrofolate reductase; mRNA, messenger RNA; miRNA, MicroRNA; NSCLC, non-small cell lung cancer; nts, nucleotide; OG, oncogene; pri-miRNA, primary-miRNA; PTC, papillary thyroid carcinoma; RISC, RNA-induced silencing complex; SNP, single-nucleotide polymorphism; UTR, untranslated region.
When the miRNA–mRNA base pair complementarity is perfect, the target mRNA is cleaved (this mechanism has been described as predominant in plants). Imperfect complementarity between miRNA and target mRNA (predominant in \textit{C.elegans}, \textit{Drosophila melanogaster} and mammals) leads to translational silencing of the target, albeit also in case of imperfect base pairing a reduction of the target mRNA has been described (3). As a result, miRNAs negatively regulate the expression of target mRNAs. Recently, it has been experimentally demonstrated that by binding to the 3\#-UTR of tumor necrosis factor alpha mRNA, miR-369-3p can actually increase the translation of the gene (34), suggesting a more complex level of regulation of miRNAs on their target genes. Despite the miRNA–mRNA binding mechanism has not been fully elucidated yet, nt 2–8 from the 5\# end on the miRNA (the so called ‘seed region’) appear critical in the match (35). A weaker complementarity of the seed region can be compensated by a stronger pairing at the 3\# end of the mature miRNA, but extensive 3\# pairing is not enough to confer targeting (30). Finally, nt 13–16 of the mature miRNA seem to contribute to enhance miRNA–mRNA binding (36). Nonetheless, the classical Watsonian match between mature miRNA and a mRNA sequence does not define a target (37). The secondary structure of the miRNAs affects the accessibility of miRNAs to their binding sequence (38), and the energy required to free the site that binds to miRNA seed sequence is affected by intracellular microenvironmental factors [such as concentration of the miRNA and of other miRNAs competing to target the same mRNA (39)], cellular growth condition or cellular stressing agents (40). All of these factors could explain why the same miRNA can regulate different genes in different cells.

A better comprehension of the physiology of miRNA–mRNA interaction is knowledge of utmost importance in our understanding of miRNA targeting and gene regulatory functions. Despite several levels of complexity are being described, the first step in the attempt to decode the rules of this interaction relies on interpreting whether, and to what extent, variations in the plane miRNA–mRNA matching sequences impact the final outcome of the interaction.

Genomic variations in miRNA genes

Genomic variation in miRNA genes can affect their processing and ultimately their ability to properly regulate the expression of target genes. Figure 1 summarizes the main functionally relevant genetic variations described in miRNA genes. The first polymorphisms in pre-miRNAs were described by Iwai \textit{et al.} (41). By sequencing 173 human pre-miRNA genome regions in 96 subjects, 10 polymorphisms in 10 pre-miRNA hairpin regions were identified, one of which (a C to A SNP) in the mature miR-30c-2. None of these SNPs had any effect on miRNA processing (41). The fact that most of the polymorphisms are neutral may reflect the evolutionary importance of the miRNA system. Saunders \textit{et al.} (42) established that the occurrence of SNPs in pre-miRNA sequences is relatively rare (\(\sim 10\%\)) and \(< 1\%\) of miRNAs have SNPs in the functional seed region. Since the SNPs at the stem and loop regions of pre-miRNAs are more frequent, it can be postulated that changes occurring outside the seed region undergo less stringent functional constraint. Duan \textit{et al.} have systematically identified SNPs associated with 227 known human miRNAs. They
identified one SNP located at the eighth nt of the mature miR-125a, which effectively blocks the processing of pri-miR-125a to pre-miRNA and reduces miRNA-mediated translational suppression (43). By systematically screening sequence variations in several hundred of human miRNAs from >100 human tumor tissues and 20 cancer cell lines, Wu et al. (44) identified eight new SNPs and 14 novel mutations in several pri-, pre- and mature miRNAs. Despite most of the mutations were not functional, a G–A mutation at 19 nt downstream miRNA let-7e led to a significant reduction of its expression in vivo and could contribute to carcinogenesis. Recently, Hu et al. (45) have provided evidence that common SNPs in miRNA genes might predict survival in non-small cell lung cancer (NSCLC) patients. The CC homozygous SNP rs1614913 located in miR-196a-2 was associated with statistically significant increase in the mature miR-196a-2, with no changes in its pre-miRNA, and with a worse prognosis in NSCLC (45). Yang et al. (46) assessed the effects of 41 SNPs in genes of the miRNA biogenesis pathway, pre-miRNAs and pri-miRNAs on bladder cancer predisposition and found that a non-synonymous SNP in GEMIN3 and a common haplotype in GEMIN4 (two genes part of RISC) were associated with a significantly increased risk of bladder cancer. Moreover, they identified a signature of combined unfavorable genotypes of selected SNPs, which might be used jointly to predict bladder cancer risk (46). Jazdzewski et al. (47) identified a common G–C SNP (rs2910164) within the pre-miR-146a sequence in papillary thyroid carcinoma (PTC). The C allele reduces the amount of pre- and mature miR-146a and decreases the amount of each pre-miR generated from the corresponding pri-miR-146a in an in vitro processing reaction (47). This study also showed that in addition to impairing miR-146a processing, this SNP contributes to genetic predisposition to PTC and plays a role in tumorigenesis by reducing the targeting effect of this miRNA on experimentally validated targets of the Toll-like receptor and cytokine signaling pathways (47). Shen et al. (48) studied the same SNP (rs2910164) in familial breast/ovarian cancers and concluded that the C allele increases mature miR-146a levels and create a stronger match with the 3′-UTR of BRCA1, leading to an earlier age onset of familial breast and ovarian cancers. In a cancer association study, Xu et al. (49) described a 2-fold increased risk of hepatocellular carcinoma for male individuals with the GG phenotype of the SNP rs2910164. Moreover, in accordance with the PTC study, they showed that the C allele reduces the levels of mature miR-146a (49). The opposite effects of SNP rs2910164 on pre-miR-146a processing described in studies of Jazdzewski and Xu, with respect to Shen’s study, might reflect tissue-specific differences. Further investigations are warranted to clarify whether other factors in the cell microenvironment might affect the outcome of this and other SNPs in miRNA biogenesis and processing.

In addition to SNPs, functional germ line mutations have been described also in the pri-miRNA sequence. In the initial report of sequence variations in miRNAs, we reported in two patients diagnosed with chronic lymphocytic leukemia, one of which has a family history of chronic lymphocytic leukemia and breast cancer, a C–T homozygous substitution in the pri-miR-16-1, seven nt in the 3′ direction after the end of the pre-miRNA (22). This substitution was associated with lower levels of mature miR-16 production revealing a functional impact on the processing of this miRNA. Mutation in a nearly identical location in the 3′ flanking region of miR-16-1 was also described in the New Zealand black mouse, a model for human chronic lymphocytic leukemia that spontaneously develops the disease when it ages (50). Recently, Shen et al. (51) have screened genetic variants in 17 selected miRNA genes, which are predicted to regulate key breast cancer genes, in 42 patients with familial breast cancer. They identified seven new variants, two of which in pre-miRNAs (pre-miR-30c-1 and pre-miR-21) and five in pri-miRNA transcripts (pri-miR-17, pri-miR-24-1, pri-miR-125a, pri-miR-191 and pri-miR-125b-1). Interestingly, the rare variants in pre-miR-30c-1 and in pri-miR-17 were only observed in non-carriers of BRCA1/2 mutations. Since miR-17 can target BRCA1, and the described variant affects the processing of miR-17 (51), miRNA genomic variations can potentially alter the regulation of key breast cancer genes. Changes in the promoter sequence of miRNAs could also significantly affect the expression of miRNA transcripts and have important phenotypical implications. Sevignani et al. (52) found that five of seven (71.4%) miRNAs with sequence differences among tumor-susceptible and tumor-resistant mouse strains (Figure 1) had changes within their predicted promoter regions, revealing a possible profound impact of miRNA promoter sequence variations in mouse carcinogenesis. In summary, SNP/mutations in miRNA genes are very rare in the seed regions, most of the times are neutral but in some cases can affect miRNA processing and have a functional role in human tumorigenesis. Considering that each miRNA can target hundreds of miRNAs, variations in miRNA sequence, especially at the seed level, may affect several biologic pathways and contribute to vast wide deleterious effects (41). Therefore, it can be postulated that cells carrying the variant miRNA undergo negative selection, explaining the very low occurrence of SNP/mutation in the miRNA gene, especially in the seed sequence.

### Genomic variations in miRNA targets

Polymorphisms and genomic variations can also occur in the target mRNAs and affect the miRNA–mRNA complementarity by destroying their binding or by creating new illegitimate interactions. Table 1 lists the published genomic variations of possible biologic relevance in the miRNA-binding sites on target mRNAs. A combined analysis of SNP database at the coordinates of ~29 000 computationally predicted sites and 706 experimentally validated sites identified 17 SNPs that disrupt target sequences with experimental evidence (42). In particular, the T–A SNP rs1042538, which involves the binding site of miR-124 on the 3′-UTR of the IQGAP1 (IQ motif containing GTPase-activating protein) gene (53), and the C–T SNP rs17168525 on the binding site for let-7 on the MTPN (myotrophin) gene (54) have a relatively high population frequency (42), although further functional assays are needed to clarify the biological significance of these SNPs. Interestingly, Saunders et al. (42) also reported that the SNP newly created target sites for miRNAs in several human genes. Among them, the rs7284767 (G–A) on the gene TUG1 (taurine upregulated gene 1), in which the A allele creates a new binding site for miR-20, and the rs11755 (A–G) on the gene ARPC3 (actin related protein 2/3 complex, subunit 5), in which the G allele creates an illegitimate target site for miR-34ac (42). By using SNP genotype data and techniques from population genetics to study an entire layout of short, cis-regulatory sites in the human genome, Chen et al. (55) observed that negative selection in humans occurs more strongly on bioinformatically predicted miRNA-binding sites than on other conserved motifs in the 3′-UTR region. An increasing number of studies are investigating how SNPs in miRNA target genes affect human carcinogenesis. Yu et al. (56) conducted a genome-wide search for SNPs located in miRNA-binding sites of target mRNAs using the SNP database and comprehensively defined the display of each SNP in cancers versus normal tissues by mining the expressed sequence tags (EST) database. The study showed that some miRNA-binding SNPs exhibit significant different allele frequencies between the human cancer EST libraries and the SNP database. Moreover, matching human cancer specimens against the SNP database for case–control association studies, they identified 12 miRNA-binding SNPs that indeed display an aberrant allele frequency in human cancers; therefore, revealing that SNPs located in miRNA-binding sites might contribute to human tumorigenesis (56). By selecting the 3′-UTRs of 129 genes involved in pathways commonly recognized as important to cancer, and studying the effects of SNPs on the Gibbs free energy of the miRNA–mRNA binding, Landi et al. (57) identified 23 SNPs of potential relevance for cancer, which deserve further investigation in case–control association studies. A C–T SNP in the 3′-UTR of the estrogen receptor alpha gene (ESR1), the rs93410170, impacts the ability of miR-206 to target ESR1 mRNA (58). In particular, the T allele enhances miR-206 ability to repress ESR1 expression. Since ~70% of primary breast cancers overexpress ESR1, the higher frequency of the T allele in Hispanic and European populations might be responsible for their lower incidence of breast cancer. Despite this
conclusion is rational and appealing, future genome-wide association studies, which include SNP rs93410170, are warranted in order to confirm this association. More recently, a Swedish study examined the effect of SNPs in predicted miRNA target sites of six integrin genes on breast cancer risk and clinical outcome (59). Despite none of the SNPs was significantly associated with breast cancer risk; several SNPs were associated with the hormone receptor status (59). Among them, the A allele of the SNP rs743554 in the ITGB4 (integrin, beta 4) gene, occurring at the binding site of miR-34a, is significantly associated with estrogen receptor negative tumors and with a worse survival with respect to the wild-type genotype carriers (59). Hence, the ITGB4 SNP could harbor prognostic implications for breast cancer patients. Landi et al. identified a G–C SNP (rs17281995) in the 3′-UTR of the cluster of differentiation 86 (CD86) gene, which was significantly associated with colorectal cancer. This SNP involved the targeted site of five different miRNAs, resulting in a reduced binding for three of them (miR-337, -582 and -200a), and in an increased matching for the other two (miR-184 and -212) (60). Further experimental validation for SNP-mediated differential miRNA binding and function is necessary, such as already done by Chiu et al. It has been experimentally documented that the rat sarcoma (RAS) OG is regulated by the let-7 family of miRNAs in lung cancer (61). Recently, Chiu et al. (62) identified a new SNP in the let-7 complementarity site in v-ki-RAS2 Kirsten rat sarcoma viral oncogene homolog (KRAS) 3′-UTR (namely let-7 complementarity site 6). Functionally, the SNP induces KRAS overexpression in vitro and is associated with significantly increased risk for NSCLC among moderately smokers (62). Moreover, in tumors from patients harboring the SNP, let-7s were lower than in non-variant allele tumors, suggesting that the SNP might be associated with NSCLC with severe clinical outcome (62). He et al. (63) sequenced the regions harboring two binding domains, one for miR-221 and miR-222 and another for miR-146a and miR-146b, in genomic DNA samples from the PTC patients and found a SNP in each recognition site. The G–A SNP rs17084733 is located within the v-kit Hardy–Zuckerman 4 feline sarcoma viral oncogene homolog (KIT) 3′-UTR at the miR-221/222-binding site and leads to a conformation change with an increase of free energy. The synonymous G–C SNP rs3733542 in exon 18 of the KIT miRNA affects the binding of miR-146a and miR-146b by changing miRNA–mRNA duplex conformation and results in hybridization with a different region (63). In 5 of 10 PTC cases, marked KIT downregulation was associated with germ line single-nt changes in the two recognition sequences in KIT for these miRNAs (63).

SNPs in miRNA target genes can also affect patient sensitivity to antiblastic drugs. Goto et al. (64) suggested that the T allele in a C–T SNP in the 3′-UTR of the dihydrofolate reductase (DHFR) gene, the rs34764978, was associated with higher DHFR protein levels. Mishra et al. (65) demonstrated that the T allele of this SNP disrupts miR-24 ability to target DHFR; therefore, conferring 4-fold more methotrexate chemoresistance. Notably, the rs34764978 is not located in miR-24-binding site on the DHFR gene 3′-UTR; therefore, raising the question whether the SNP actually affects miR-24 binding to this target. It can be postulated that rs34764978 modifies the secondary structure of DHFR 3′-UTR and reduces miR-24 accessibility to its binding site, although this hypothesis still needs to be experimentally proven. Moreover, a direct association between rs34764978 and methotrexate resistance has still to be shown. In summary, there is increasing evidence that genomic variations/SNPs in miRNA target genes occur and can be functional. In theory, the impact of such variations should be limited to the target gene or its pathway and should not be as wide as that of variations in miRNA genes. Nevertheless, functional genomic variations in miRNA target genes can release miRNAs from their ‘physiological’ miRNA control and in some cases put them under different miRNA regulation, varying the spectrum of targets of miRNAs, and ultimately harboring the potential for a greater regulatory turmoil.

Conclusion remarks and clinical implications

The increasing evidence of genomic variations affecting miRNAs and their target genes, and the disclosure of their functional implications, raise at least three ways of possible clinical applications. First, several studies indicate that some SNPs in both miRNA genes and miRNA target genes increase the risk of specific types of cancers. Despite some of these studies are limited by the lack of information on the population frequency of the SNP, the inclusion of at least some miRNA-associated SNPs in genetic counseling programs can be foreseen. Abnormal expression of miRNAs due to sequence variations could represent a new form of cancer predisposition. Since each miRNA has numerous targets, inherited minor variations in miRNA expression could have important consequences on the expression of various protein-coding oncogenes and tumor suppressors involved in malignant transformation. Accumulation of additional somatic events occurring in protein-coding genes or in non-coding RNAs, including miRNAs, is necessary for the full development of the malignant phenotype. Secondly, some miRNA genomic variations have been associated with prognosis and response to treatment, and a signature of combined unfavorable genotypes of selected SNPs can predict the risk of bladder cancer (46). Finally, decoding how these genomic variations affect miRNA biogenesis/processing and the range of targeted miRNAs will open new possible therapeutic scenarios. For instance, restoration of wild-type miRNAs or miRNAs by gene therapy could revert an aberrant phenotype by restoring a physiologic miRNA–mRNA interaction. Although this kind of therapy is still a long way to come, since many other issues need to be addressed before it can be successfully pursued, there is no doubt that we are at the dawn of a better understanding of key regulatory mechanisms in human gene regulation. Deciphering the ‘tricks’ that genomic variations can trigger in the world of miRNAs, builds up the fundamental background of knowledge that is needed to achieve the ‘gift’ of understanding and possibly curing the aberrant miRNA gene regulation that occurs in human carcinogenesis.

**Funding**

University of Texas M. D. Anderson Research Trust; University of Texas System Regents Research Scholar Fund; Ladjevardian Regents

---

**Table 1. Genomic variations in miRNA Target genes of biological relevance**

<table>
<thead>
<tr>
<th>Target gene</th>
<th>miRNA</th>
<th>Variation</th>
<th>Alleles</th>
<th>Type of cancer</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>IQGAP1</td>
<td>miR-124</td>
<td>rs1042538</td>
<td>A/T</td>
<td>Not reported</td>
<td>(42)</td>
</tr>
<tr>
<td>MTPN</td>
<td>let-7</td>
<td>rs17168525</td>
<td>C/T</td>
<td>Not reported</td>
<td>(42)</td>
</tr>
<tr>
<td>TUG1</td>
<td>miR-20</td>
<td>rs7284767</td>
<td>G/A</td>
<td>Not reported</td>
<td>(42)</td>
</tr>
<tr>
<td>ARPC5</td>
<td>miR-34a</td>
<td>rs17155</td>
<td>A/G</td>
<td>Not reported</td>
<td>(42)</td>
</tr>
<tr>
<td>ESR1</td>
<td>miR-206</td>
<td>rs93410170</td>
<td>C/T</td>
<td>Breast</td>
<td>(58)</td>
</tr>
<tr>
<td>ITGB4</td>
<td>miR-34a</td>
<td>rs743554</td>
<td>G/A</td>
<td>Breast</td>
<td>(59)</td>
</tr>
<tr>
<td>CD86</td>
<td>miR-337, -582, -200a', -184, -212</td>
<td>rs17281995</td>
<td>G/C</td>
<td>Colorectal</td>
<td>(60)</td>
</tr>
<tr>
<td>KRAS</td>
<td>let-7</td>
<td>SNP</td>
<td></td>
<td>NSCLC</td>
<td>(62)</td>
</tr>
<tr>
<td>KIT</td>
<td>miR-221, -222</td>
<td>rs17084733</td>
<td>G/A</td>
<td>PTC</td>
<td>(63)</td>
</tr>
<tr>
<td>KIT</td>
<td>miR-146a, -146b</td>
<td>rs3733542</td>
<td>G/C</td>
<td>PTC</td>
<td>(63)</td>
</tr>
<tr>
<td>DHRF</td>
<td>miR-24</td>
<td>rs34764978</td>
<td>C/T</td>
<td>Not reported</td>
<td>(65)</td>
</tr>
</tbody>
</table>
References


Received January 28, 2009; revised March 9, 2009; accepted March 10, 2009