Functional genetic variation of human miRNAs and phenotypic consequences

BOREL, Christelle, ANTONARAKIS, Stylianos

Abstract

A large number of human protein-coding genes are finely regulated by one or more microRNAs. Members of this small noncoding RNA family have emerged as important post-transcriptional regulators of gene expression and are involved in a number of disease phenotypes. Variability in the human genome is extensive and includes the common and rare single nucleotide polymorphisms (SNPs) and copy number variations (CNVs). The functional significance of the genome's variability is under intense investigation. In this article we review the emerging literature on how human genomic variation influences the outcome of microRNA targeting and the associated phenotypic effects. Illustrative examples are discussed that demonstrate the biological importance of functional polymorphisms affecting miRNA-mediated gene regulation.

Reference


DOI : 10.1007/s00335-008-9137-6
PMID : 18787897
Functional genetic variation of human miRNAs and phenotypic consequences

Christelle Borel · Stylianos E. Antonarakis

Received: 28 April 2008 / Accepted: 28 July 2008 / Published online: 12 September 2008 © Springer Science+Business Media, LLC 2008

Abstract A large number of human protein-coding genes are finely regulated by one or more microRNAs. Members of this small noncoding RNA family have emerged as important post-transcriptional regulators of gene expression and are involved in a number of disease phenotypes. Variability in the human genome is extensive and includes the common and rare single nucleotide polymorphisms (SNPs) and copy number variations (CNVs). The functional significance of the genome’s variability is under intense investigation. In this article we review the emerging literature on how human genomic variation influences the outcome of microRNA targeting and the associated phenotypic effects. Illustrative examples are discussed that demonstrate the biological importance of functional polymorphisms affecting miRNA-mediated gene regulation.

Introduction

The human genome contains many forms of genetic variation. One class of variants comprises the millions of single nucleotide polymorphisms (SNPs) first identified more than 30 years ago in the β-globin gene cluster (Kan and Dozy 1978), and recently cataloged in different populations (Hinds et al. 2005; The International HapMap Consortium 2003). A second common class of variation are copy number variations (CNVs), also identified about 30 years ago in the α-globin gene cluster (Goossens et al. 1980; Kan et al. 1975). A considerable effort is now devoted toward understanding the extent and magnitude of CNVs in human populations (Wheeler et al. 2008). Here we review the emerging evidence of the effect of human genetic variation on microRNA-mediated gene regulation and discuss how human interindividual variability can influence the outcome of microRNA targeting.

miRNAs and their functions

MicroRNAs (miRNAs, miR), which are short, noncoding RNA molecules (~21 nt), have revealed a new dimension in the complexity of translation control. Our current understanding of the functions of miRNAs is incomplete (Bushati and Cohen 2007). There is considerable evidence that miRNAs function by binding to complementary sequences that usually lie in the 3′ untranslated regions (3′UTR) of target mRNAs to induce cleavage, repression of productive translation, and regulation of mRNA stability in the cytoplasm (Ambros 2004; Bartel 2004; Bushati and Cohen 2007; Carrington and Ambros 2003; Filipowicz et al. 2008; Jackson and Standart 2007; Nilsen 2007). Recently, studies revealed additional functions of miRNAs such as upregulation of target mRNA translation upon cell-cycle arrest (Vasudevan et al. 2007), import into the nucleus (Hwang et al. 2007), or even the secretion of the miRNAs from the cell (Valadi et al. 2007). A novel aspect of miRNA-mediated repression with a noncoding transcript in Arabidopsis thaliana that can regulate the activity of an miRNA by mimicking its target site has been demonstrated (Franco-Zorrilla et al. 2007). Some authors suggest that some RNAs have a nonproductive interaction with a complementary miRNA to inhibit and regulate the activity of the miRNA. These examples indicate that miRNA
molecules have very diverse functions that remain to be explored.

To date, the miRNA registry (MiRBase) contains 6397 mature miRNAs, among which 678 are human miRNAs (April 2008, release 11.0). miRNAs can be encoded in independent transcription units, in polycistronic clusters, or within the introns of protein-coding genes (Bartel 2004). Thus far, most registered miRNAs are widely expressed and highly conserved. Deep sequencing of small RNA libraries and cell-type–specific analyses are currently valuable approaches to uncovering miRNAs with less abundant expression; thus, the catalog of miRNAs is expected to grow substantially in the future (Bar et al. 2008; Glazov et al. 2008; Sunkar et al. 2008).

miRNA expression profiles indicated that some miRNAs are under complex control during development and in a variety of tissues (Farh et al. 2005; Lim et al. 2005; Stark et al. 2005). These studies suggest that miRNAs define highly specific cell identities and have a critical role in cell differentiation. In C. elegans, the miRNA lousy-6 (lsy-6) is specifically activated by the transcription factor die-1 in ASE left cells, whereas lsy-6 is downregulated by miR-273 and then activated by the transcription factor cog-1 in ASE right cells (Chang et al. 2004; Johnston and Hobert 2003; Johnston et al. 2005). Genetic data demonstrated that lsy-6, cog-1, miR-273, and die-1 act in a regulatory double-negative feedback loop (Johnston et al. 2005). This inverse expression of lsy-6 controls laterality of the nematode chemosensory system and is an excellent illustration of the role of miRNAs in controlling terminally differentiated cellular states.

Although the proteins mediating miRNA biogenesis and function have been clarified, the precise mechanism by which miRNAs regulate the expression of target mRNAs remains unclear. Mature miRNAs preferentially form nonperfect duplexes with the target mRNA (usually in the mRNA 3’UTR) and recruit a repressing complex termed the RNA-induced silencing complex (RISC). This regulation can be highly pleiotropic, with one miRNA able to target several hundred different transcripts. In fact, bioinformatic prediction indicates that 30% of animal genes may be miRNA targets.

Principles of miRNA biogenesis

miRNAs are transcribed by RNA polymerase II (Bartel 2004) as long primary transcripts (pri-miRNAs, 100 nt-10 kb) and are then processed within the nucleus into hairpin-shaped precursor miRNAs (pre-miRNAs, ~70 nt) by the RNase III enzyme Drosha (Lee et al. 2003) and the double-stranded RNA-binding domain cofactor Dgcr8/Pasha (DiGeorge syndrome critical region gene 8) (Denli et al. 2004; Gregory et al. 2004; Han et al. 2004; Landthaler et al. 2004; Lee et al. 2003). An alternative miRNA processing pathway has recently been described that uses splicing processes to bypass Drosha cleavage and to generate miRNA precursors from short intronic sequences (Berezikov et al. 2007; Okamura et al. 2007; Ruby et al. 2007). These small intronic sequences of noncoding RNA are called mirtrons.

In the cytoplasm, Dicer (RNase III) processes the premiRNAs to generate ~22-nt miRNA duplexes (Lund et al. 2004). One strand remains a mature miRNA and is then assembled into the effector complex RISC (Khvorova et al. 2003; Schwarz et al. 2003).

Principles of miRNA—mRNA interactions

To specify repression, miRNAs seem to require only short stretches of complementarity to a mRNA, following a broad set of principles that have been identified experimentally and by bioinformatic studies (Brennecke et al. 2005; Doench and Sharp 2004; Grimson et al. 2007; Lewis et al. 2005; Lytle et al. 2007; Nielsen et al. 2007). However, the mechanistic details of miRNA–mRNA interactions are poorly understood and there are always exceptions that make the predictions of miRNA targets difficult. Many studies have confirmed that the interaction is initiated by a continuous and perfect base pairing of seven contiguous nucleotides at positions 2-8 from the 5’ end of the miRNA, representing the “seed region” (Lewis et al. 2003, 2005). Mismatches, a guanine–uracil pairing (GU), or bulges in the seed sequences disturb the binding of miRNA with this target site and consequently affect function (Brennecke et al. 2005). However, it has been observed that an A on position 1 of the miRNA (5’) and either A or U at position 9 can boost miRNA targeting, although these do not need to base pair with target nucleotides (Grimson et al. 2007). Other characteristics specify targeting: mismatches and bulges are present in the central part of the miRNA-target site duplex, a core region of 3’ pairing to residues 13-16 within the miRNA 3’ end. However, the degree of repression is also related to UTR contexts and factors that can improve site efficiency (Grimson et al. 2007): multiple, closely spaced target sites act cooperatively, high local AU content in the vicinity of the miRNA target site improves site efficiency, and effective sites are located preferentially near both the poly(A) tail and the stop codon. As part of the different models proposed for miRNA–mRNA interactions, miRNA target sites can be categorized into two distinct classes: 5’-dominant and 3’-compensatory. 5’-Dominant target sites have perfect base pairing with at least 7-8 nt at the 5’ end of the miRNA; the base pairing to the rest of the miRNA is considered irrelevant (Chen and Rajewsky 2006; Lewis et al. 2003). 3’-Compensatory target sites have extended base pairing to the 3’ end of the miRNA in order to
compensate for an imperfect or shorter stretch of base pairing with the 5′ end of the miRNA (Gupta et al. 2006; Kiriakidou et al. 2004; Schratt et al. 2006; Stark et al. 2005).

Genetic variation

Genetic variations range from large chromosome anomalies to single-nucleotide changes. Recently, multiple studies have identified a large number of submicroscopic CNVs of DNA segments ranging from kilobases (kb) to megabases (Mb) in size (Beckmann et al. 2007). CNVs consist of deletions, insertions, duplications, and complex multisite variants in comparison with a reference genome. All those DNA sequence variations could potentially affect the maturation of miRNAs, the silencing machinery, the structure or the expression level of mature miRNA, and the base pairing at the target site, and have functional role for miRNA-mediated gene regulation. We focus on human miRNAs and review all the evidence for human regulatory polymorphisms perturbing miRNA function and thus in turn causing phenotypic variations and disorder.

Polymorphisms and heterogeneity of miRNA sequences

Polymorphism in pri-, pre-, and mature miRNA sequences

A representative mammalian pri-miRNA comprises a stem (≈33 bp) with a terminal loop and flanking sequences (≈100 bp). Experiments show that a part of the terminal loop (Zeng and Cullen 2005; Zeng et al. 2005) and sequences flanking the stem loop (Han et al. 2006; Lee et al. 2003; Zeng and Cullen 2003) are critical for processing by Drosha and Dicer. It has been hypothesized that polymorphisms in pre-miRNA may influence the miRNA maturation and thereby modulate miRNA expression.

A total of 173 human pre-miRNAs in 96 Japanese individuals have been sequenced (Iwai and Naraba 2005). The study identified ten SNPs in ten pre-miRNA hairpins among which only one SNP in the mature sequence of miR-30c-2 is likely to affect stem integrity (this prediction has not yet been confirmed experimentally).

In another study, a bioinformatic search selected 323 known SNPs located within 227 human pre-miRNA sequences (Duan et al. 2007). Twelve of these SNPs were within pre-miRNA sequences and one, a G/U polymorphism, located at the eighth nucleotide within the mature sequence of miR-125a, has been functionally characterized. This SNP (rs12975333) blocks in vitro the processing of pri-miRNA into pre-miRNA and alters the translation suppression by miR-125a on a well-known Lin-28 target mRNA. The biological consequences of these miRNA SNPs remain to be determined.

In a similar study, the occurrence of common SNPs in 474 human pre-miRNAs found by screening the dbSNP database was studied (Saunders et al. 2007). Sixty-five SNPs (including indel polymorphisms) in 49 pre-miRNAs were found, thus exhibiting a SNP density of ≈1.3 SNPs per kilobase pair. Only three miRNAs (hsa-mir-125a, hsa-mir-627, and hsa-mir-662) were found to have SNPs within the seed region (rs12975333, rs2620381, and rs9745376, respectively). The conclusion of the study was that miRNA genes have low polymorphism density and that most identified polymorphisms are not within the seed region, indicative of strong selective constraint on human pre-miRNAs. None of these polymorphisms has been tested experimentally to determine whether these SNPs might affect the structure or processing of miRNAs.

A germline DNA variant (C/T) in the primary sequence of the miRNA cluster encoding for miR-16-1 and miR-15a (7 bp in the 3′ direction after the precursor) has been identified (Calin et al. 2005). This DNA variant was found in 11 of 75 patients with chronic lymphocytic leukemia, but not seen in 160 subjects without cancer. This nucleotide variation results in reduced mature expression of miR-16-1 and miR-15 in vitro and in vivo and is associated with deletion of the normal allele containing this miR cluster.

A recent example illustrates how a common polymorphism in a miR precursor can consequently affect the mature miRNA expression level. Jazdzewski et al. (2008) reported a G/C (rs 2910164) SNP in pre-miR-146a that results in vitro in reduced amounts of mature miR-146a and contributes to the genetic predisposition to papillary thyroid carcinoma.

A recent genome-wide analysis in healthy individuals revealed 43 miRNAs mapping in CNV regions in one study (Redon et al. 2006), while another study identified 14 CNVs encompassing 21 known miRNAs (Wong et al. 2007). It is not known if the variation in copy number of the miRNAs alters their function and contributes to the phenotypic diversity in humans.

3′- and 5′-end heterogeneity of miRNAs—isomiRs

Detailed analysis of mature miRNA sequences revealed an extensive degree of variation at the terminal nucleotides compared to the current miRBase reference sequences (Cummins et al. 2006; Lagos-Quintana et al. 2002; Landgraf et al. 2007; Morin et al. 2008; Ruby et al. 2006; Wu et al. 2007). This population of variants in known miRNAs is collectively termed isomiRs. All cited studies observed in most cases nonrandom and conserved 3′-end
heterogeneity with 3′-end deletions and extensions of one to three nucleotides. The most prevalent type of modifications generated single-nucleotide 3′ extensions that differ from the genomic sequence. In contrast, 5′ ends were more homogeneous, with only a few cases of variation (Landgraf et al. 2007; Morin et al. 2008). These results predict that a particular miRNA hairpin may generate numerous mature miRNAs.

Although the biological consequence of this observation remains to be determined, such variants might act by influencing the miRNA half-life, subcellular localization, and miRNA target specificity (particularly 5′-end variation). In addition, the most abundant isomiRs’ sequences need to be determined for each miRNA, and it is essential to verify whether any of these variants associate within RISC and function as gene silencers.

**Polymorphisms affecting miRNA–mRNA interactions**

**Polymorphism in miRNA target sites**

The thermodynamics of RNA–RNA interaction plays an essential role in the binding of a miRNA with its target mRNA; it is proposed that SNPs at miRNA binding sites may alter the expression of target genes. Polymorphism density is significantly lower in conserved target site regions that are complementary to the seed region of the miRNA (Chen and Rajewsky 2006). SNPs in this region are expected to be functional and they are candidates for causal variants of human disease. Several studies scanned for the occurrence of SNPs in human 3′ UTRs with the purpose of identifying SNPs that may modulate expression of computationally predicted miRNA target sites (Bao et al. 2007; Chen and Rajewsky 2006; Georges et al. 2007; Saunders et al. 2007). Among the 120,000 known 3′ UTR SNPs, 19,913 modified putative miRNA targets either by destroying predicted conserved target sites (785) or predicted nonconserved target sites (9470) or by creating 10,283 novel predicted target sites (http://www.patrocles.org; Georges et al. 2007).

The first evidence that mutations in a miRNA target site might affect the phenotype came from a study of the gene SLITRK1, considered a strong candidate for Tourette’s syndrome (Abelson et al. 2005). Mutation screening in 174 unrelated individuals with Tourette’s syndrome revealed a G-to-A variant that was not present in 1800 normal individuals. This variant is predicted to stabilize the interaction of 3′ UTR of SLITRK1 with miR-189 by changing a G:U wobble pair to an A:U Watson-Crick pair at position 9 of the miRNA. It was demonstrated in vitro that the 3′ UTR of SLITRK1 indeed facilitates miR-189-mediated downregulation of a luciferase reporter in Neuro2a cells, and the G-to-A variant resulted in a repression of luciferase expression compared to the wild type. There is a developmentally regulated and overlapping expression pattern of SLITRK1 mRNA and miR-189 in brain regions previously implicated in Tourette’s syndrome (Abelson et al. 2005). Moreover, SLITRK1 promotes dendritic growth while the mutation likely results in a loss of function. These results suggest an association of rare SLITRK1 sequence variants with Tourette’s syndrome (Abelson et al. 2005).

Another example of a functional SNP in the target miRNA sequence is related to hypertension. A SNP rs5186 (1166A/C) in the 3′ UTR for the angiotensin receptor-1 (AGTR1) gene has been characterized (Sethupathy et al. 2007). The 1166C allele of AGTR1 has been shown to increase the risk of essential hypertension in several association studies in various populations (EHT [MIM 145500]). Moreover, it has been shown independently that increased levels of AGTR1 contribute to cardiovascular disease (Song and White 2002; Van Geel et al. 2000) and antagonists of AGTR1 are now widely used in the treatment of hypertension (Burnier and Brunner 2000). The likely mechanism through which SNP rs5186 regulates AGTR1 protein levels has recently been elucidated (Sethupathy et al. 2007). This SNP has been localized in a target site for human miR-155 within the 3′ UTR of AGTR1 (Martin et al. 2006). Using reporter silencing assays, it was shown that miR-155 downregulates the expression of only the A and not the C allele of rs5186. Thus, the C allele may be functionally associated with hypertension by abrogating regulation by miR-155, thereby increasing AGTR1 levels (Sethupathy et al. 2007). Moreover, miR-155 maps to human chromosome 21 and it was observed by quantitative real-time PCR that miR-155 is overexpressed in trisomy 21 individuals. Fibroblasts from monzygotic twins discordant for trisomy 21 showed that the AGTR1 protein is lower in trisomy 21 fibroblasts compared to normal. Remarkably, it has been reported that individuals with trisomy 21 have lower diastolic and systolic blood pressure levels than age- and gender-matched controls (Draheim et al. 2002; Morrison et al. 1996). Overexpression of miR-155 in trisomy 21 is likely to excessively suppress the AGTR1 common allele; this may contribute to the lower blood pressure in trisomy 21 individuals.

There are four additional examples of loss of miRNA target site sequences due to SNPs. The presence of SNP 829C/T (rs34764978) in the 3′ UTR of DHFR led to a loss of miR-24 function and resulted in dihydrofolate reductase overexpression and methotrexate resistance (Mishra et al. 2007). Furthermore, SNP 3142C/G (rs1063320) in the HLA-G 3′ UTR affects the targeting of miR-148a, miR-148b, and miR-152 and is associated with risk of asthma in
children (Tan et al. 2007). Wang et al. (2008) demonstrated a direct link between miR-433 and a common SNP (rs12720208, T/C) located in the 3′ UTR of fibroblast growth factor 20 (FGF20), already associated with an increased risk for Parkinson’s disease. The risk allele for this SNP abolished the miR-433 functional activity and increased the translation activity of FGF20 and α-synuclein expression.

Another recent example of loss of miRNA target site sequences due to SNPs is that type 2 diabetes implicated miR-657 and the IGF-II receptor gene (IGF2R). IGF2R is located in a region that has been shown to be related to insulin resistance and obesity-related metabolic phenotypes (Duggirala et al. 2001). A case control association study concluded that the ACAA insertion/deletion (144/140 bp) polymorphism at the 3′ UTR of the IGF2R might participate in the pathogenesis of type 2 diabetes (Villuendas et al. 2006). Recently, new evidence showed that hsa-miR-657 can post-transcriptionally regulate the IGF2R expression levels in Hep G2 cells by targeting its 3′ UTR. Moreover, it has been demonstrated that the ACAA insertion/deletion polymorphism alters the interaction between IGF2R and miR-657 (Lv et al. 2008).

On the other hand, a DNA variant can create a gain of a miRNA target site. QTL mapping and genetic analysis reveal a G-to-A substitution in the 3′ UTR of the GDF8 gene implicated in the polygenic hypermuscularity of Texel sheep (Clop et al. 2006). This point mutation creates two novel target sites for miR-1 and miR-206 (sharing a common seed sequence), which are highly expressed in skeletal muscle. In the presence of the mutation, there was a threefold reduction of circulating myostatin and 1.5-fold reduction in the GDF8 transcript level. The novel miRNA target sites were also experimentally validated.

**Polymorphisms affecting miRNA expression**

To date there are no SNPs described that affect miRNA expression levels. It is anticipated that such variation will be discovered in the near future as the miRNAs are better characterized, and more genomic variants, common or rare, will be functionally analyzed and be associated with phenotypes.

**Conclusions**

The study of SNPs has revealed that any two randomly selected human genomes differ by 0.1% (Frazer et al. 2007). A fraction of this genetic variation would affect functional elements directly and will have a causative role in the phenotype. It is therefore expected that genetic variability would also affect the function of miRNAs. This review points out the emergence of a new class of regulatory variability that might substantially contribute to the heritability of complex traits. Some illustrative examples have been discussed here. However, the majority of this functional variability (particularly rare private variants) may remain unknown. The miRNA function provides excellent opportunities for the discovery of potential epistatic interactions between polymorphisms in miRNA genes and their targets as important risk allele combinations. Because miRNAs affect the fine-tuning of gene expression, we predict that functional variants related to miRNA biology will likely be involved in the complex multifactorial phenotypes.

**Acknowledgments**

The authors’ lab is supported by the Swiss National Science Foundation, the NCCR Frontiers in Genetics, The European Union, the NIH, and the Lejeune and ChildCare Foundations. We thank our laboratory colleagues for discussions, critical reading, and numerous suggestions.

**Web Resources**

- SeattleSNPS database: http://pga.gs.washington.edu
- International HapMap project: http://www.hapmap.org
- Perlegen Sciences: http://genome.perlegen.com
- Patrocles: http://www.patrocles.org
- PolymiRTS: http://compbio.utmem.edu/miRSNP/
- Tarbase: http://www.diana.pcbi.upenn.edu/tarbase.html

**References**

Lytle JR, Yario TA, Steitz JA (2007) Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR. Proc Natl Acad Sci USA 104:9667–9672


