Copy number variation of microRNA genes in the human genome

Malgorzata Marcinkowska¹, Maciej Szymanski², Wlodzimierz J Krzyzosiak¹ and Piotr Kozlowski¹*

Abstract

Background: MicroRNAs (miRNAs) are important genetic elements that regulate the expression of thousands of human genes. Polymorphisms affecting miRNA biogenesis, dosage and target recognition may represent potentially functional variants. The functional consequences of single nucleotide polymorphisms (SNPs) within critical miRNA sequences and outside of miRNA genes were previously demonstrated using both experimental and computational methods. However, little is known about how copy number variations (CNVs) affect miRNA genes.

Results: In this study, we analyzed the co-localization of all miRNA loci with known CNV regions. Using bioinformatic tools we identified and validated 209 copy number variable miRNA genes (CNV-miRNAs) in CNV regions deposited in Database of Genomic Variations (DGV) and 11 CNV-miRNAs in two sets of CNVs defined as highly polymorphic. We propose potential mechanisms of CNV-mediated variation of functional copies of miRNAs (dosage) for different types of CNVs overlapping miRNA genes. We also showed that, consistent with their essential biological functions, miRNA loci are underrepresented in highly polymorphic and well-validated CNV regions.

Conclusion: We postulate that CNV-miRNAs are potential functional variants and should be considered high priority candidate variants in genotype-phenotype association studies.

Background
MicroRNAs (miRNAs) are a family of short (~20 nt), single-stranded, noncoding RNAs that are primarily involved in post-transcriptional down-regulation of gene expression in most eukaryotes [1]. Specific miRNAs are engaged in a variety of processes, including development, cell proliferation, differentiation and apoptosis [2]. Numerous studies have demonstrated that aberrant over-expression or down-regulation of certain miRNAs contribute to carcinogenesis and that these miRNAs can therefore be classified as either oncogenes (oncomirs) or tumor suppressors, respectively [3].

Mature, functional miRNAs are generated from primary precursors (pri-miRNA) encoded either by independent transcriptional units or within protein- or RNA-coding genes. In mammals, maturation of miRNAs involves two subsequent RNA cleavage steps. The first step takes place in the nucleus and is carried out by the Drosha nuclease to produce the secondary precursor (pre-miRNA) [4]. The pre-miRNAs (~60 nt) possess a hairpin structure, with the double-stranded portion interrupted by one or more mismatched nucleotides. Upon export to the cytoplasm, the pre-miRNA is further processed into an miRNA duplex by the RNAse III Dicer; [5] one of the duplex strands (passenger) is released, and the other serves as the mature miRNA [6]. The miRNA-induced silencing complex (miRISC) interacts with complementary target sequences, which are usually located within the 3' untranslated regions (3'UTRs) of mRNAs, causing mRNA degradation or inhibition of translation [7-9].

It is estimated that, in humans and other mammals, the expression of at least one-third of protein-coding genes is fine-tuned by approximately 1,000 miRNAs [10,11]. Currently, over 700 human miRNAs have been identified, and their sequences are deposited in miRBase (the microRNA database; http://www.mirbase.org).

Polymorphisms in miRNA genes can affect the expression of many downstream-regulated genes [12,13]. The most common form of polymorphism that affects the function of an miRNA (e.g., the structure of miRNA precursors, the efficiency of miRNA biogenesis and

* Correspondence: kozlowp@yahoo.com
¹Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland
Full list of author information is available at the end of the article

© 2011 Marcinkowska et al; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
miRNA-target recognition) is the single nucleotide polymorphism (SNP). Computational and experimental studies have revealed many SNPs located in different parts of pre-miRNA sequences [14-16]. The occurrence of SNPs (including INDELs) in pre-miRNA regions is significantly lower than that in the surrounding reference sequences [16]. While sequences of mature miRNAs are the most conserved, the sequences of anti-miRNAs and the stems (outside miRNA and anti-miRNA) and loops of pre-miRNAs are somewhat less conserved [16]. SNPs naturally occurring within pre-miRNA sequences may affect miRNA biogenesis and impair miRNA-mediated gene silencing, as demonstrated by functional assays [15,17]. Recently, large genome-wide association study has demonstrated that also SNPs located outside (>14 kb) of pre-miRNA sequences can modulate miRNA expression both as cis- and trans-regulators (miRNA-eQTLs). One of identified miRNA-eQTLs (rs1522653) was shown to correlate with expression of 5 different miRNAs [18].

MiRNA target sites are also conserved genetic elements. Bioinformatic analyses show that SNPs are underrepresented in both experimentally validated and computationally predicted miRNA target sites, [16,19] and SNPs have the potential to either disrupt or create new miRNA target sites [19]. It has also been proposed that target site polymorphisms may play a role in evolution by altering miRNA specificity and function.

However, little is known about copy number variation (CNV) of miRNA genes. CNVs are segments of genomic DNA (roughly 1 kb to 1 Mb in length) that show variable numbers of copies in the genome due to deletions or duplications. CNVs recurrently occurring in a population are often called copy number polymorphisms (CNPs). Only a few CNV discovery studies report the presence of miRNAs in detected CNV regions and recognize their potential consequences [20-22]. Indeed, it was suggested that a comprehensive analysis of the co-localization of miRNAs and CNVs is needed [12].

Numerous studies show that CNVs can influence the expression of protein-coding genes in a copy number-dependent manner [23-25]. Recent results of genome-wide association study has confirmed such association for dozens of protein-coding genes and showed that CNVs capture at least 18% of the total detected genetic variation in gene expression [26]. It seems obvious that the expression of miRNA genes can also be modified by CNVs. This notion is supported by results from cancer genetics studies. For instance, there is a correlation between somatic copy number variation and the expression of miRNA genes, and miRNA genes recurrently amplified or lost in cancer genomes can serve as oncogenes or cancer suppressor genes, respectively [27-31].

In this study, by comparing the coordinates of human miRNAs with different sets of CNV regions (DGV-deposited and highly polymorphic), we identified over 200 human copy number variable miRNA loci. By comparing fractions of miRNAs and the genome that are covered by differentially validated CNV regions, we showed that miRNA loci are underrepresented in highly polymorphic CNVs, but not in CNVs deposited in the DGV database. We discuss the potential functional relevance of identified copy number variable miRNAs and propose models of how different types of CNVs can affect miRNA dosage.

Results and Discussion

Prior to bioinformatic identification of copy number variable miRNA genes (CNV-miRNAs), we compared the frequency of SNPs in annotated pre-miRNA sequences (3.7 SNPs/1,000 bp) and in reference human genome (4.8 SNPs/1,000 bp). Significantly lower number of SNPs in the pre-miRNA sequences (Fisher’s exact test; p < 0.0001) most likely results from SNP purification effect and confirms general conservation of the analyzed pre-miRNA sequences. These analyses confirmed a SNP purification effect in pre-miRNA sequences reported previously [16]. The much higher number of SNPs identified in annotated pre-miRNA sequences in our study (N = 229; Additional file 1) versus N = 65 reported previously [16] results from the increased number of both SNPs (dbSNP - build 130; Apr 30, 2009; only annotated as 'single'; ~14 million SNPs) and miRNAs (miRBase - v 13.0), available in versions of databases used in this study.

To identify CNV-miRNAs, we compared the positions of miRNA loci with three sets of CNVs: ‘DGV-deposited’ (N = 29133; 30% genome coverage), ‘polymorphic-SMC’ (N = 1319; 1.2% genome coverage) [32] and ‘polymorphic-DC’ (N = 5037; 2.3% genome coverage) [22]. CNVs, ‘DGV-deposited’ CNVs include all 29133 CNVs deposited in the Database of Genomic Variants (DGV update Aug 05, 2009 - http://projects.tcag.ca/variation). Two sets of ‘polymorphic’ CNV (’polymorphic-SMC’ [32] and ‘polymorphic-DC’ [22]) include highly polymorphic CNVs (minor allele frequency >0.01) validated by high-quality genotyping in two recent CNV-discovery studies using CNV-dedicated high-density hybrid arrays (combining traditional SNP probes and probes targeting CNVs) [22,32]. In both of these studies, precise breakpoints and unambiguous copy numbers were determined for each analyzed sample. All ‘DGV-deposited’ CNV-miRNA regions were further characterized by the following validation factors: (i) number of publications reporting CNVs (references), (ii) number of overlapping CNVs (DGV records) and (iii) number of observations in discovery studies (frequency) (Additional file 2). Since
the exact boundaries of miRNA genes (including regulatory elements) are difficult to determine, we used the genomic coordinates of all pre-miRNA loci deposited in miRBase (v 13.0; N = 715) as a proxy of miRNA gene sequences (three pre-miRNA loci located in the mitochondrial genome were excluded from our analysis) [33,34]. We realize, however, that CNVs overlapping other functional regions of miRNA coding genes (e.g., promoters) can also affect miRNA biogenesis and functionality, and those CNVs will be missed in our analysis.

The CNV-miRNAs identified in ‘DGV-deposited’ CNVs (N = 209) and in two sets of ‘polymorphic’ CNVs (N = 4 and N = 8) are shown in Additional file 2 and Table 1, respectively. Top-validated ‘DGV-deposited’ CNV-miRNAs are also shown in Table 2. Most miRNA loci identified in ‘polymorphic’ CNVs also overlapped with top-validated ‘DGV-deposited’ CNV regions (Table 1 and Table 2). All ‘polymorphic’ CNV-miRNAs were relatively frequent (combined minor genotype frequency >0.1 in at least one HapMap population). Among the identified miRNA-CNVs, we found deletions (e.g., hsa-mir-384 and hsa-mir-1324), duplications (e.g., hsa-mir-1972 and hsa-mir-1977), and multiple duplications (multiallelic polymorphisms; e.g., hsa-mir-1233 and hsa-mir-1268). The number of observed copies ranged from 0 (e.g., hsa-mir-384 and hsa-mir-650) to 6 (e.g., hsa-mir-1268).

The sequences of miRNA deposited in miRBase are derived from discovery studies in which many strict miRNA verification criteria were applied (e.g. hairpin forming potential, evolutionary conservation, presence in multiple clones/sequence reads or homogeneity of the 5’end). The SNP frequency analysis presented in this study also confirmed global conservation of annotated pre-miRNA sequences. However, there is still a possibility that some of the miRNAs in the miRBase represent experimental artifacts of false positive discoveries [35]. To provide additional data that can further validate miRNAs identified in CNVs we have conducted bioinformatic analysis of their expression and conservation. Table 1 and Table 2 show that according to different miRNA expression resources summarized in mimiRNA

| miRNA ID | miRNA position | dupl. | CNV region position | genotypes | CNV ID | functional relevance | expression (mimiRNA/[18]) | conservation |
|----------|----------------|-------|---------------------|-----------|--------|----------------------|---------------------------|-------------|
| mir-1268 | chr15:20014593-20014644 | chr15:19803370-20089386 | 2,3,4,5,6 | 2057 | 1) recurrently deleted in classical Hodgkin’s lymphoma [47] | not reported/NA | primates |
| mir-1233 | chr15:32607783-32607864 | chr15:32487975-32617680 | 0,1,2,3 | 2082 | 1) | not reported/NA | primates |
| mir-1972 | chr16:15011679-15011755 | chr16:14807364-15016088 | 2,3,4 | 2141 | not reported/NA | primates |
| mir-384 | chrX:76056092-76056179 | chrX:76053855-76057477 | 0,1,2 | 2648 | in several tissues/NA | mammals |

| miRNA ID | miRNA position | dupl. | CNV region position | genotypes | CNV ID | functional relevance | expression (mimiRNA/[18]) | conservation |
|----------|----------------|-------|---------------------|-----------|--------|----------------------|---------------------------|-------------|
| mir-1977 | chr1:556050-556128 | chrM | chr1:554403-560267 | 2,3,4 | 3.1 | not reported/NA | primates |
| mir-1324 | chr3:75762604-75762699 | chr3:75464498-75782745 | 1.2 | 1432.2 | not reported/NA | primates |
| mir-548i-2 | chr4:9166887-9167035 | chr4:9117494-9354801 | 1.2 | 1815.3 | not reported/NA | primates |
| mir-1275 | chr6:34075277-34075806 | chr6:34071086-34077139 | 1.2 | 2853.1 | 2) upregulated in blood cells of MS patients [41] | not reported/NA | primates |
| mir-1302-2 | chr9:20144-20281 | chr9:485-38531 | 2.3 | 4134.1 | | not reported/NA | primates |
| mir-1323 | chr15:32461562-32461643 | chr15:32450046-32662643 | 2,3,4 | 6351.3 | 1) | not reported/NA | primates |
| mir-650 | chr22:219459320-21495365 | chr22:20711019-21578090 | 0,1,2 | 8103.1 | | in several tissues (mostly ovary and ovary-derived cancers)/high | primates |

dupl. - localization of duplicated copies; mimiRNA/[18] - miRNA expression according to database mimiRNA/and according to resent result of expression analysis in primary fibroblast cells (high - high expression, absent - low or undetectable expression in fibroblast cells, NA - not analyzed).
### Table 2 miRNA loci localized in CNV regions validated by multiple overlapping CNVs

| miRNA ID | miRNA position | dupl. | minimal CNV region | # CNVs | functional relevance | expression (miRNA/18) | conservation |
|----------|----------------|-------|--------------------|--------|----------------------|-----------------------|--------------|
| mir-1977 | chr1:556050-556128 | chrM | chr1:554340-569354 | 6      | in multiple tissues/ primates |
| mir-149  | chr2:24104491-24104479 |   | chr2:24103969-241051687 | 6      | in several tissues/ primates |
| mir-566  | chr3:50185763-50185856 |   | chr3:50173490-50214015 | 7      | in multiple tissues/ high priamtes |
| mir-1324 | chr3:75762604-75762699 |   | chr3:75761737-75839337 | 6      | in multiple tissues/ primates |
| mir-570  | chr3:196911452-196911548 |   | chr3:196905807-196918722 | 9      | in multiple tissues/ primates |
| mir-548-2| chr4:9166887-9167035 |   | chr4:9152768-9182838 | 9      | not reported/NA primates |
| mir-548-3| chr8:7983873-7984021 |   | chr8:7965981-8024983 | 14     | in multiple tissues/ high vertebrates |
| mir-383  | chr14:755318-14755390 |   | chr14:4741501-14763659 | 8      | in multiple tissues/ high vertebrates |
| mir-661  | chr18:5091347-145091435 |   | chr18:45090343-145104971 | 8      | in multiple tissues/ high vertebrates |
| mir-1299 | chr9:68292059-68292141 |   | chr9:68291272-68298205 | 7      | not reported/NA primates |
| mir-126  | chr13:86864875-138684959 |   | chr13:8680837-13888363 | 14     | high, in multiple tissues/high vertebrates |
| mir-202  | chr10:134911006-134911115 |   | chr10:13490301-134918923 | 10     | in multiple tissues/ high vertebrates |
| mir-1268 | chr15:20014593-20014614 |   | chr15:19975453-20004635 | 37     | not reported/NA primates |
| mir-1233 | chr15:32461562-32461643 |   | chr15:32461525-32469857 | 9      | not reported/NA primates |
| mir-1233 | chr15:32607783-32607864 |   | chr15:32599966-32615283 | 17     | not reported/NA primates |
| mir-662  | chr16:760184-760278 |   | chr16:750040-764098 | 6      | in multiple tissues/ high vertebrates |
| mir-1972 | chr16:68621750-68621826 |   | chr16:68621490-68653097 | 6      | not reported/NA primates |
| mir-142  | chr17:3373592-53763687 |   | chr17:33735160-53767652 | 11     | high, in multiple tissues/absent |
| mir-1270 | chr19:20371080-20371162 |   | chr19:20370872-20383238 | 9      | not reported/NA primates |
| mir-663  | chr20:26136822-26136914 |   | chr20:26136626-26139184 | 6      | in several tissues/NA primates |
| mir-650  | chr22:21495270-21495365 |   | chr22:21494381-21502189 | 38     | in several tissues/ high primates |
| mir-514-2| chrX:146171153-146171240 |   | chrX:146168796-146174575 | 6      | in several tissues/ NA mammals |
| mir-514-3| chrX:146173851-146173938 |   | chrX:146168796-146174575 | 6      | in several tissues/ NA mammals |

dupl. - localization of duplicated copies; miRNA/18 - miRNA expression according to database miRNA/and according to resent result of expression analysis in primary fibroblast cells (high - high expression, absent - low or undetectable expression in fibroblast cells, NA - not analyzed).
database [36] over half (14/26) of top-validated CNV-miRNAs (Table 1 and Table 2) were shown to be expressed in at least several tissues/cell lines (detailed expression profiles are shown in Additional file 3). MiRNA whose expression is not reported in miRBase were either not analyzed for expression or did not show expression in the analyzed tissues. Additionally, three out of ten (30%) top-validated CNV-miRNAs (Table 1 and Table 2) which expression in primary fibroblast cell lines was analyzed by the micro-fluidics-based TaqMan Human MiRNA Array show high level of expression [18]. Based on the currently available sequence data for miRNAs deposited in miRBase and blast searches of the vertebrate genomic sequences we also determined evolutionary conservation of the miRNAs found in top-validated CNV regions. Most of these miRNAs seem to be specific only for primates. There are, however, 8 miRNAs that are conserved across mammals or vertebrates (Table 1 and Table 2).

The functional relevance of several of the CNV-miRNAs identified in this survey was previously reported in the literature (manual screening; Table 1 and Table 2). CNV-miRNAs are involved in many processes and phenotypes (diseases), including organ development [37], angiogenesis [38], male infertility [39], transplant rejection [40], multiple sclerosis [41] and cancer. Many CNV-miRNAs are specifically deleted, amplified or expressed in different types of cancers [42-47] and can regulate the expression of important cancer-related genes [37,48]. The copy number variation of those functionally relevant miRNAs can modulate or predispose one to the aforementioned phenotypes.

In the next step, we determined whether the overlap of CNVs and miRNA loci was random (null hypothesis) or whether the CNVs were underrepresented at these loci (alternative hypothesis). To test this hypothesis, we compared fractions of miRNA loci and fractions of the genome covered by differentially defined CNV regions. Figure 1A shows that the fraction of miRNA loci covered by two sets of ‘polymorphic’ CNVs is approximately two times lower than expected (fraction of the covered genome). Although this effect was only marginally significant (Figure 1A), it suggested that at least highly polymorphic CNVs are under negative (purifying) selection at miRNA genes. Conversely, the fraction of miRNAs (0.292) covered by ‘DGV-deposited’ CNVs corresponded almost exactly to the fraction of the genome covered by those CNVs (0.299). The CNV purification effect was not observed, even after narrowing ‘DGV-deposited’ CNV regions by different validation factors defined above (Figure 1B and 1C). The fact that the purifying effect did not apply to the ‘DGV-deposited’ CNVs suggested that a significant portion of these CNVs are very rare, private, or significantly oversized or represents false positive artifacts. This observation is consistent with the conclusions from other recently published results [32,49].

Although copy number variation can influence gene expression through different mechanisms (e.g., position effect and deletion or duplication of regulatory elements
that control transcription or splicing), the most obvious mechanism is in the variability of dosage (number of functional copies). All of these mechanisms can affect both protein-coding and miRNA genes. However, mechanisms of dosage variation may be different for protein-coding and miRNA genes. In Figure 2, potential consequences of different CNV types overlapping different parts of miRNA genes are proposed. Not only whole gene amplification but also certain partial gene duplications (multiple duplications) can increase the dosage of miRNAs. Conversely, partial gene deletions may not always result in decreased miRNA dosage. This contrasts with the situation observed for protein-coding genes, in which only duplication of the entire gene (including the promoter and regulatory sequences) can lead to an increased number of functional copies, and almost every (even partial) gene deletion is deleterious.

Analysis of 11 miRNAs located in CNVs with well defined breakpoints (Table 1) showed that (i) 3 of these miRNAs are located in the protein coding genes which are entirely positioned within CNVs, (ii) 4 of the miRNAs are located in intergenic regions and are flanked by at least 20 kb of CNV sequences, (iii) 3 miRNAs are located in intergenic regions flanked by short CNV sequences (< 5 kb) and (iv) 1 miRNA is located in a gene of which the 3'end extends beyond CNV (Additional file 4). Taking into account the average size of a human gene (~30 kb) one can expect that miRNAs located in large CNVs (groups (i) and (ii)) will be expressed from genes entirely embedded within the CNV regions. According to the model presented in Figure 2A the expression of such miRNAs very likely will correlate with expression (number of copies) of genes from which these miRNAs are generated (no matter whether generated from protein-coding or non-coding transcripts). MiRNA located in short CNVs (group (iii)) most likely will form the tandem copies transcribed from one promoter. A number of such copies may modulate the number of miRNA precursors (pre-miRNAs) present in one primary transcript (pri-miRNA) and thus may modulate expression of miRNA (Figure 2D). Expression of miRNA whose gene only partially is embedded in CNV (iii) may be modified according to the model shown in Figure 2B and will depend on expression and stability of the transcript truncated at the 3'end. Moreover, it should be noted that some pre-miRNA sequences occur in the genome in multiple copies. Although the functionality of such copies is still mostly unknown, the duplicated copies of miRNA genes may mask the effect of copy number variations that usually affect only one copy.

Finally, not only common CNVs, but also CNVs implicated in specific diseases can affect miRNA loci and thus can play important role in pathogenesis. We

---

Figure 2: Potential mechanism of CNV-mediated variation of miRNA dosage. Schematic representation of an miRNA gene and its primary transcript (solid or dotted arrow-lines). The position of the pre-miRNA sequence is indicated as a hairpin-loop structure in the miRNA primary transcript. Dotted lines represent transcripts unlikely to be produced due to the lack of promoter and transcriptional start sequences. Orange boxes represent CNV regions (deletions, duplications and dispersed duplications). The following panels show a CNV spanning different parts of the miRNA gene: (A) whole gene, (B) 5'-portion, (C) 3'-portion and (D) intragenic region of the gene. +, - and 0 indicate potential increase, decrease and no change of miRNA dosage, respectively.
have identified 38 loci of miRNAs located in chromosomal regions implicated in microdeletion/microduplication syndromes (DECYPHER v5.0 [50]) (Additional file 5). For example, six miRNA loci (hsa-mir-185, hsa-mir-1306, hsa-mir-1286, hsa-mir-649, hsa-mir-301b and hsa-mir-130b) are located within genomic region implicated in DiGeorge syndrome. The role of somatic copy number variation of miRNA genes in cancer is extensively investigated in multiple studies (e.g. [27-31]) and was recently summarized in several review articles [51-53].

Conclusions
Although ‘polymorphic’ CNVs showed some purifying effects at miRNA loci, there were still many miRNA loci that overlapped with known CNV regions (Additional file 2 and Table 2), including those that are highly validated and confirmed by high-quality genotyping (Table 1). Taking into account the CNV genome coverage (1.2% ‘polymorphic-SMC’ and 2.3% ‘polymorphic-DC’) and the relatively small overlapping fractions (0.39 and 0.20, respectively) between the two sets of ‘polymorphic’ CNVs analyzed in this study, we estimated that up to 10% of the human genome is covered by highly polymorphic CNVs. This fraction corresponds to approximately 30 highly polymorphic CNV-miRNAs in the human genome (extrapolation of the fraction of miRNA loci covered by highly polymorphic CNVs analyzed in this study). It is likely that at least some of these loci are among the CNV-miRNAs identified from the top-validated ‘DGV-deposited’ CNVs (Table 2 and Additional file 2).

CNV-miRNAs are potential functional variants and should be considered high priority candidate variants in genotype-phenotype association studies, especially when they are located in regions implicated by linkage or association studies. As indicated in Table 1, only a small fraction of CNV-miRNAs were genotyped in three HapMap populations, which provides precise information about their polymorphisms. This is mostly due to the lack of appropriate methods for precise characterization of CNV polymorphisms. Although several genome-wide approaches that substantially fulfill the above requirement were proposed recently, a simple and inexpensive method that enables accurate characterization of several CNVs of interest in a large number of samples is still needed. The lack of such a method significantly hampers the analyses of CNVs and their correlation with the phenotype. To verify and characterize the polymorphisms of all CNV-miRNAs, we are developing several medium-throughput assays suited for large scale population studies that are focused on selected CNVs of potential functional effect. These assays will take advantage of the MLPAbased strategy proposed previously [54-56].
Additional file 5: miRNAs located in chromosomal regions implicated in microdeletion/microduplication syndromes. Excel table containing list of miRNAs located in chromosomal regions implicated in microdeletion/microduplication syndromes (DECIPHER v5.0 [50]).

Acknowledgements
This work was supported by the Ministry of Science and Higher Education [N N302 279937, N N302 260938]. The authors have declared no conflict of interest.

Author details
1Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland. 2Computational Genomics Laboratory, Institute of Molecular Biology and Biotechnology, Adam Mickiewicz University, Poznan, Poland.

Authors' contributions
MM performed the computational analysis, literature screening, participated in the manuscript preparation. MS participated in the computational analysis (sequence conservation analysis) and the manuscript preparation. WJK participated in the design and coordination. All authors have read and approved the final manuscript.

Received: 24 March 2010 Accepted: 12 April 2011
Published: 12 April 2011

References
1. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 2004, 116: 281-297.
2. Kim VN, Nam JW. Genomics of microRNA. Trends Genet 2006; 22: 165-173.
3. Esquela-Kerscher A, Slack FJ. Oncomir - microRNAs with a role in cancer. Nat Rev Cancer 2006, 6: 259-269.
4. Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S, Kim VN. The nuclear RNAs III Drosha initiates microRNA processing. Nature 2003, 425: 415-419.
5. Bernstein E, Caudy AA, Hammond SM, Hannon GJ. Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature 2001, 409: 363-366.
6. Hammond SM, Bernstein E, Beach D, Baltimore D, Hannon GJ. An RNA-directed nuclelease mediates post-transcriptional gene silencing in Drosophila cells. Nature 2000, 404: 296-299.
7. Guo H, Ingolia NT, Weissman JS, Bartel DP. Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature 2010, 466: 835-840.
8. Pillai RS, Bhattacharyya SN, Artus CG, Zoller T, Cougot N, Basyuk E, Bertrand E, Filipowicz W. Inhibition of translational initiation by Let-7 MicroRNA in human cells. Science 2005, 309: 1573-1576.
9. Yekta S, Shih IH, Bartel DP. MicroRNA-directed cleavage of HOXB8 mRNA. Science 2004, 304: 594-596.
10. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 2005, 120: 15-20.
11. Rajewsky N. microRNA target predictions in animals. Nat Genet 2006, 38(Suppl): S8-13.
12. Borel C, Antonarakis SE. Functional genetic variation of human miRNAs and phenotypic consequences. Mamm Genome 2008, 19: 503-509.
13. Georges M, Coppersmith W, Charlier C. Polymorphic miRNA-mediated gene regulation: contribution to phenotypic variation and disease. Curr Opin Genet Dev 2007, 17: 166-176.
14. Iwai N, Naraba H. Polymorphisms in human pre-miRNAs. Biochem Biophys Res Commun 2005, 331: 1439-1444.
15. Duan R, Pal C, Jin P. Single nucleotide polymorphism associated with mature mir-125a alters the processing of pri-miRNA. Hum Mol Genet 2007, 16: 1124-1131.
16. Saunders MA, Li WH. Human polymorphism at miRNAs and microRNA target sites. Proc Natl Acad Sci USA 2007, 104: 3300-3305.
17. Sun G, Yan J, Nolten K, Feng J, Li H, Sarkis DA, Sommer SS, Rossi JJ. SNPs in human miRNA genes affect biogenesis and function. RNA 2009, 15: 1640-1651.
18. Borel C, Deutsch S, Letoumeau A, Migliavacca E, Montgomery SB, Dimas AS, Veijar CE, Attar H, Gagnebin M, Gehrig C, et al. Identification of cis- and trans-regulatory variation modulating microRNA expression levels in human fibroblasts. Genome Res 2011, 21: 68-73.
19. Chen K, Rajewsky N. Natural selection on human microRNA binding sites inferred from SNP data. Nat Genet 2006, 38: 1452-1456.
20. Wong KK, deLeeuw RI, Dosanjh NS, Kimm LR, Zheng C, Horsman DE, Macaulay C, Ng RT, Brown CJ, Eckler EE, Lam WL. A comprehensive analysis of common copy-number variations in the human genome. Am J Hum Genet 2007, 80: 91-104.
21. Lin CH, Li LH, Ho SF, Chuang TP, Wu YJ, Chen YT, Fann CS. A large-scale survey of genetic copy number variations among Han Chinese residing in Taiwan. BMC Genet 2008, 9: 92.
22. Conrad DF, Pinto D, Redon R, Feuk L, Gokcenmen O, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P, et al. Origins and functional impact of copy number variation in the human genome. Nature 2010.
23. Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, Werner J, Villaneau FA, Mountain JL, Misra R, et al. Diet and the evolution of human amylase gene copy number variation. Nat Genet 2007, 39: 1256-1260.
24. González E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, Nibbs RJ, Freedman BI, Quinones MP, Banshd MJ, et al. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science 2005, 307: 1434-1440.
25. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Q Y, Scherer SW, Lee C. Detection of large-scale variation in the human genome. Nat Genet 2004, 36: 949-951.
26. Stranger BE, Forrest MS, Dunning M, Ingle CE, Beadley C, Thorne N, Redon R, Bird CP, de Grassi A, Lee C, et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science 2007, 315: 848-853.
27. Bottomo A, Picon D, Tagliati F, Luchin A, Zatelli MC, degli Uberti EC. mir15a and mir16-1 down-regulation in pituitary adenomas. J Cell Physiol 2005, 204: 283-285.
28. Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch S, Oldridge N, Rattan S, Keating M, Rai K, et al. Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proc Natl Acad Sci USA 2002, 99: 15524-15529.
29. Zhang L, Huang J, Yang N, Gershlick J, Meigraw MS, Giannakakis A, Liang S, Naylor TL, Barchetti A, Ward MR, et al. microRNAs exhibit high frequency genomic alterations in human cancer. Proc Natl Acad Sci USA 2006, 103: 9136-9141.
30. Ota A, Tagawa H, Kaman S, Tsuzuki S, Karpas A, Ito K, Seto M. Identification and characterization of a novel gene, C13orf25, as a target for 13q31-32 amplification in malignant lymphoma. Cancer Res 2004, 64: 3087-3095.
31. He L, Thomson JM, Hernant MT, Hernando-Monge E, Mu D, Goodson S, Powers S, Cordon-Cardo C, Lowe SW, Hannon GJ, Hammond SM. A microRNA polycistron as a potential human oncogene. Nature 2005, 435: 628-633.
32. McCarroll SA, Kuruvilla FG, Kom JCM, Cawley S, Nemesh J, Wysocka J, Shapero MH, de Bakker PI, Maller JB, Kirby A, et al. Integrated detection and population-genetic analysis of SNPs and copy number variation. Nat Genet 2008, 40: 1166-1174.
33. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. miRBase: tools for microRNA genomics. Nucleic Acids Res 2008, 36: D154-158.
34. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res 2006, 34: D140-144.
35. Chiang HK, Bronsfield LW, Ruby JG, Aveyung VC, Nipper S, Aerts D, Johnstone WK, Russ C, Luo S, Babiarz JE, et al. Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. Genes Dev 2010, 24: 992-1009.
36. Ritchie W, Flamant S, Rasko JE. miRNA: a microRNA expression profiler and classification resource designed to identify functional correlations between microRNAs and their targets. Bioinformatics 2010, 26: 223-227.
37. Shen WF, Hu YL, Uttarwar L, Passegue E, Largman C. MicroRNA-126 regulates HOXA9 by binding to the homeobox. Mol Cell Biol 2008, 28: 4609-4619.
38. Fish JE, Santoro MW, Morton SU, Yu S, Yeh RF, Wythe JD, Ivey KN, Bruneau BG, Stainier DY, Stavrintsa D: mir-126 regulates angiogenic signaling and vascular integrity. Dev Cell 2008, 15: 272-284.

39. Lian J, Zhang X, Tian H, Liang N, Wang Y, Liang C, Li X, Sun F: Altered microRNA expression in patients with non-obstructive azoospermia. Reprod Biol Endocrinol 2009, 7: 13.

40. Anglicheau D, Sharma VK, Ding R, Hummel A, Snopkowski C, Dadhania D, Seshan SV, Suthanthiran M: MicroRNA expression profiles predictive of human renal allograft status. Proc Natl Acad Sci USA 2009, 106: 5330-5335.

41. Keller A, Leidinger P, Lange J, Bomes A, Schreers H, Scheffler M, Lenhof HP, Ruepecht K, Meese E: Multiple sclerosis: microRNA expression profiles accurately differentiate patients with relapsing-remitting disease from healthy controls. PLoS One 2009, 4: e7440.

42. Zhang H, Luo XQ, Zhang P, Huang LB, Zheng YS, Wu J, Zhou H, Qu LH, Xu L, Chen YQ: MicroRNA patterns associated with clinical prognostic parameters and CNS relapse prediction in pediatric acute leukemia. PLoS One 2009, 4: e7626.

43. Guo C, Sah JF, Beard L, Willson JK, Markowitz SD, Guda K: The noncoding RNA, miR-126, suppresses the growth of neoplastic cells by targeting phosphatidylinositol 3-kinase signaling and is frequently lost in colon cancers. Genes Chromosomes Cancer 2008, 47: 939-946.

44. Wong TS, Liu XB, Wang Y, Ng RW, Tuen AP, Wei W: Mature miR-184 as Potential Oncogenic microRNA of Squamous Cell Carcinoma of Tongue. Clin Cancer Res 2008, 14: 2588-2592.

45. Rossi S, Sevignani C, Nnadi SC, Siracusa LD, Calin GA: Cancer-associated genomic regions (CAGRs) and noncoding RNAs: bioinformatics and therapeutic implications. Mamm Genome 2008, 19: 526-540.

46. Ju X, Li D, Shi Q, Hou H, Sun N, Shen B: Differential microRNA expression in childhood B-cell precursor acute lymphoblastic leukemia. Pediatr Hematol Oncol 2009, 26: 1-10.

47. Hartmann S, Martin-Subero JI, Gesk S, Husken J, Giebing M, Nagel I, Remke J, Chott A, Klapper W, Parens M, et al: Detection of genomic imbalances in microdissected Hodgkin and Reed-Sternberg cells of classical Hodgkin’s lymphoma by array-based comparative genomic hybridization, Hematol Oncol 2008, 26: 1318-1326.

48. Reddy SD, Pakala SB, Ohchrbo K, Rayala SK, Kumar R: MicroRNA-661, a cEBPalpha target, inhibits metastatic tumor antigen 1 and regulates its functions. Cancer Res 2009, 69: 5639-5642.

49. Itsara A, Cooper GM, Baker C, Girirajan S, Li J, Absher D, Krauss RM, Myers RM, Ridker PM, Chasman DI, et al: Population analysis of large copy number variants and hotspots of human genetic disease. Am J Hum Genet 2009, 84: 148-161.

50. Firth HV, Richards SM, Bevan AP, Clayton S, Corbiss M, Rajan D, Van Vooren S, Moreau Y, Pettett RM, Carter NP: DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. Am J Hum Genet 2009, 84: 524-533.

51. Deng S, Calin GA, Croce CM, Coukos G, Zhang L: Mechanisms of microRNA deregulation in human cancer. Cell Cycle 2008, 7: 2643-2646.

52. Di Leva G, Croce CM: Roles of small RNAs in tumor formation. Trends Mol Med 2010, 16: 257-267.

53. Ruan K, Fang X, Ouyang G: MicroRNAs: novel regulators in the hallmarks of human cancer. Cancer Lett 2009, 285: 116-126.

54. Kozlowska K, Jasinska AJ, Kwiatkowski DJ: New applications and developments in the use of multiplex ligation-dependent probe amplification. Electrophoresis 2008, 29: 4627-4636.

55. Kozlowska K, Roberts P, Dabowska S, Franz D, Bisler J, Northrup H, Au KS, Lazarus R, Domanska-Piakala D, Kotulka K, et al: Identification of 54 large deletions/duplications in TSC1 and TSC2 using MLPA, and genotype-phenotype correlations. Hum Genet 2007, 121: 389-400.

56. Marcinkowska M, Wong WK, Kwiatkowski DJ, Kozlowski P: Design and generation of MLPA probe sets for combined copy number and small-mutation analysis of human genes: EGFR as an example. ScientificWorldJournal 2010, 10: 2003-2018.

57. Taylor J, Schenck I, Blankenberg D, Nekrutenko A: Using galaxy to perform large-scale interactive data analyses. Cyber Bioinf Bioinformatics 2007, Chapter 10: Unit 10.15.

58. Landgraf P, Rusu M, Sheridan R, Sewer A, Loviso N, Araven A, Pfeffer S, Rice A, Kamphorst AO, Landthaler M, et al: A mammalian microRNA expression atlas based on small RNA library sequencing. Cell 2007, 129: 1401-1414.

59. Gaur A, Jewell DA, Liang Y, Ridzon D, Moore JH, Chen C, Ambros VR, Israel MA: Characterization of microRNA expression levels and their biological correlates in human cancer cell lines. Cancer Res 2007, 67: 2456-2468.

60. Lee EJ, Baek M, Gusev Y, Brackett DJ, Nuovo GJ, Schmittgen TD: Systematic evaluation of microRNA processing patterns in tissues, cell lines, and tumors. RNA 2008, 14: 35-42.

61. Barrett T, Edgar R: Gene expression omnibus: microarray data storage, submission, retrieval, and analysis. Methods Enzymol 2006, 411: 352-369.

62. Agosti A: A Survey of Exact Inference for Contingency Tables. Statist Sci 1992, 7: 131-153.