**INTRODUCTION** MicroRNAs (miRNAs) are small RNAs that play an important role in the regulation of gene expression. miRNA dysregulation has been associated with phenotypic changes, including cardiovascular diseases (CVDs).

**OBJECTIVES** The aim of the study was to obtain a list of single nucleotide polymorphisms (SNPs) related to CVDs, with computationally predicted effect on miRNA binding sites, which would verify the hypothesis that miRNA dysregulation can lead to the development of CVDs.

**MATERIALS AND METHODS** SNPs, CVDs, and miRNAs were the 3 factors subjected to analysis. Based on the publicly available databases, we created a set of SNPs associated with the phenotype of interest and of SNPs located in known miRNA binding sites. We then merged the records assigned by the same SNP, which allowed us to indicate miRNA target sites, whose variants may be associated with CVDs. The results were supplemented with the additional data such as miRNA and mRNA coexpression, differences in the expression between various tissues, and Expression Quantitative Trait Locus analysis. Only in-silico methods, on the basis of publically available information tools and databases, were used.

**RESULTS** We obtained a list of 47 entries, constituting unique miRNA–SNP allele–phenotype linkages.

**CONCLUSIONS** Computational approach supports the hypothesis of the linkage between alterations in miRNA function and numerous CVDs. Given the high frequency of SNP incidence, this pathomechanism may be common in the population. Although the obtained results need to be further experimentally validated, limiting the number of interactions to the most probable ones will facilitate the identification of clinically significant associations.

---

**KEY WORDS**
- cardiovascular diseases
- in silico
- microRNA
- polymorphisms

**ABSTRACT**
In-silico identification of cardiovascular disease-related SNPs affecting predicted microRNA target sites

Marcin J. Kamiński¹, Magdalena Kamińska¹, Iwona Skorupa¹, Remigiusz Kazimierczyk¹, Włodzimierz J. Musiał², Karol A. Kamiński²

1 Students’ Scientific Group of the Department of Cardiology, Medical University of Bialystok, Białystok, Poland
2 Department of Cardiology, Medical University of Białystok, Białystok, Poland

Correspondence to:
Marcin J. Kamiński, Studentkís
Kolo Naukowe przy Klinice
Kardiology, Uniwersytet
Medyczny w Białymstoku,
ul. M. Skłodowskiej-Curie 24a,
15-276 Białystok, Poland,
phone: +48-85-746-86-56,
fax: +48-85-746-86-04,
e-mail: marcinjakubkaminski@gmail.com

Received: March 4, 2013.
Accepted: March 15, 2013.
Published online: April 12, 2013.
Conflict of interest: none declared.
Copyright by Medycyna Praktyczna, Kraków 2013

* The authors won the first award of the Editor in Chief for the best student paper submitted to Pł Arch Med Wewn in 2013. For more information, go to www.pamw.pl.

The key location for miRNA activity is the so-called “seed region”, consisting of 2 to 7 nucleotides within the 5’ strand, which seems to be a crucial factor for determining miRNA functionality. This is confirmed by cases where single-point mutations within the miRNA seed region (or the corresponding target site) are able to significantly reduce the effectiveness of miRNA interaction or entirely deprive it of its regulatory capability, thereby altering the expression of the regulated gene.

The mature miRNA strand regulates gene expression by complementary binding with mRNA, which in the case of extensive complementarity
potential therapeutic agents for the treatment of CVDs has been proposed.

Although genome-wide association studies (GWAS) indicate numerous connections between single nucleotide polymorphisms (SNPs) and phenotypes, leading to the conclusion that these genetic variants are able to significantly affect the course of CVD, they do not provide direct information on the possible mechanisms by which these modulations occur. Given the association of multiple polymorphisms with CVDs, the presence of functional SNPs in noncoding intron sequences, and the effect of miRNAs on gene transcription, we hypothesize that SNPs within miRNA binding site, by destroying the existing or creating novel target sites or by changing miRNA binding strength, may change its effects on gene expression, which results in the onset or change in the course of CVD. In this paper, based on an in-silico analysis, we would like to determine whether there exist miRNAs associated with CVDs by the coexistence of SNPs interfering with their function, and, if so, to indicate the specific ones.

In silico is, apart from in vivo or in vitro, one of the experimental techniques. It represents a modern approach to research, based on the use of computing power to perform mathematical analyses of a large amount of data and the creation of complex databases.

**MATERIALS AND METHODS** Based on the in-silico method, 3 factors – SNPs, CVDs, and miRNAs – were subjected to analysis. Databases available in public domains, based on the previous experiments and containing a set of SNPs associated with the phenotype of interest (CVD) and SNPs located in known miRNA binding sites, were obtained. By transforming the structure of both database types and by merging the records assigned by the same SNP, we were able to indicate the miRNA target site, the SNPs of which may be associated with CVDs. These results were further analyzed with the use of additional computational tools (FIGURE 1).

To obtain a list of SNPs associated with CVDs, an “Open Access GWAS Database”, built on the basis of 118 GWAS articles, was used. Only the SNPs related to CVDs, such as MI, coronary artery disease (CAD), subclinical atherosclerosis (SA), and arterial hypertension (HT), were extracted.

Because of the low number of entries associated with MI, we decided to extend the list by additional entries, obtained from the available medical literature. Via a PubMed search, 225 additional SNPs affecting the risk of MI were identified and included in the downloaded database.

By downloading databases available online, we obtained a list of SNPs with a computationally demonstrated effect on known miRNA binding sites. Due to the computational nature, and hence the need for experimental verification, 2 different databases were downloaded, created on the basis.
of different algorithms, allowing both to expand the range of data subjected to the study and indicate hypothetically more likely interactions. The databases used were PolymiRTS Database 2.0\cite{10} and mirSNPscore.\cite{11} Due to the different origins and methods of data storage, both underwent modification. The databases were exported to.xls filetype with MS Excel 2010 and numerous scripts and macros were used, allowing to merge the cells and rows, automatic search, content comparison, and conditional formatting.

As the final step allowing to identify the association between miRNAs and the phenotype included in the databases, we searched the content of integrated databases for GWAS entries using an SNP ID as the common variable for both databases. As a result, a list of miRNA–SNP allele–phenotype triplets was obtained (\textbf{Table 1}).

miRNAs identified in the present study were analyzed for the changes in expression between the various tissues. This was possible through the use of \textit{miRNA body map} database.\cite{12} Only the records for which the change of expression in the tissues of the cardiovascular system was at least 1.5-fold were gathered. A similar search was performed for miRNAs within the same family.

By using the mimiRNA tool,\cite{13} it was possible to analyze the collected miRNA–mRNA pairs for coexpression. The results were entered into a spreadsheet or saved in a graphic form. Due to the limited resource of the mimiRNA database, only a fraction of the obtained mRNA–miRNA pairs could have been tested.

A list of circulating miRNAs was gathered based on miRandola database resources. The data-base was searched for each miRNA obtained in this study, and each positive result was included in the table.

To compare the changes in mRNA expression caused by the difference in an SNP allele, an online application – Genevar – was used.\cite{15} The Expression Quantitative Trait Locus (eQTL) analysis of SNP–gene relations was based on the HapMap3 study.\cite{16}

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|c|c|c|c|}
\hline
\textbf{Chr} & \textbf{Gene} & \textbf{SNP} & \textbf{Phen} & \textbf{Anc} & \textbf{MAF} & \textbf{All} & \textbf{miRNA} & \textbf{Scr} | \textbf{Cons} | \textbf{p} \\
\hline
\textbf{9} & \textit{CDKN2B} & rs3217992\textsuperscript{a} & MI & G & 0.41 & T & miR-655\textsuperscript{b} & 0.26 | 9 | 0.62\textsuperscript{d}  \\
& & & & & & T & miR-374\textsuperscript{c} & 0.22 | 9 | &  \\
& & & & & & C & miR-205\textsuperscript{a,c} & 0.18 | 3 | –0.18 |  \\
& & & & & & C & miR-138-2\textsuperscript{a,c} & – | 3 | –0.21 |  \\
& rs3217989\textsuperscript{a} & CAD & A & 0.07 & A & miR-1305\textsuperscript{a} & – | 5 | &  \\
\hline
\textbf{11} & \textit{TPCN2} & rs1060435 & CAD & A & 0.45 & G & miR-483-5p\textsuperscript{a,c} & 0.60 | 3 | –0.88 |  \\
& & & & & & A & miR-3935 & 0.31 | 3 | &  \\
& & & & & & A & miR-323b-5p\textsuperscript{a,c} & 0.30 | 4 | &  \\
& & & & & & G & miR-323b-5p\textsuperscript{a,c} & 0.30 | – | &  \\
& & & & & & A & miR-218-2\textsuperscript{a} & 0.18 | – | –0.07 |  \\
\hline
\textbf{3} & \textit{GATA2} & rs3803\textsuperscript{a} & CAD & G & 0.23 & G & miR-3150b-3p & 0.40 | 0 |  \\
& & & & & & G & miR-3916 & 0.32 | – |  \\
& & & & & & A & miR-3123 & 0.30 | – |  \\
& & & & & & A & miR-3925 & 0.30 | – |  \\
& & & & & & G & miR-3125 & 0.30 | – |  \\
& & & & & & G & miR-4516 & – | 10 |  \\
& & & & & & G & miR-4784 & – | 0 |  \\
\hline
\textbf{6} & \textit{ESR1} & rs2813563 & SA & C & 0.19 & C & miR-3188 & – | 0 |  \\
\hline
\end{tabular}
\caption{Computationally predicted phenotype–SNP allele–miRNA triplets with coexisting cis-regulatory effects}
\end{table}

\textbf{Abbreviations:} All – allele, Anc – ancestral allele, CAD – coronary artery disease, Chr – chromosome, Cons – miRNA sequence conservation, MAF – mean allele frequency, MI – myocardial infarction, Phen – phenotype, SA – subclinical atherosclerosis, Scr – mirSNPscore database score, p – miRNA–mRNA coexpression coefficient, others – see Figure 1

---

\textbf{RESULTS} On the basis of the present in-silico analysis, the computationally identified variants that may affect the regulation of gene expression, and thus the existence of disease through changes within miRNA binding sites, are herein specified. From the "open-access GWAS database" containing 56,411 significant SNP–phenotype associations, after filtering out entries associated with CVDs (MI, CAD, SA, HT), 1331 results (2.35%) were merged with 225 results for MI derived from PubMed. This gave a total of 1556, of which 241 (15.49%) were related to MI, 1059 (68.06%) to CAD, 97 (6.23%) to SA, and 13 (0.84%) to HT.
Another analysis referred to the comparison of the modulated gene expression level depending on the SNP within it. It was possible based on the results of the previous cis-eQTL studies (expression Quantitative Trait Loci) that showed a statistically significant relationship between the gene transcript abundance and SNP allele present in the immediate vicinity of the gene. Using the Genevar expression dataset to analyze 13 SNPs, 4 were identified (rs3803, rs3217989, rs3217992, rs2813563), for which the difference in gene expression reached statistical significance in one of the studied populations (FIGurE 2).

Three of the tested miRNAs (miR-181a-2*, miR-628-5p, miR-378*) featured a variable...
differences in miRNA coexpression, tissue-dependent miRNA expression, and significant eQTL effect for the indicated SNP–mRNA pairs, substantiated on the basis of earlier experimental studies, makes this hypothesis even more likely. It has to be acknowledged that the accuracy of the above statement is based on the assumption that the databases used in the study, built on the basis of computational algorithms, are characterized by high prediction accuracy, indicate interactions on the SNP–miRNA line actually existing in vivo, and correctly identify miR-NAs’ targets. For obvious reasons, the applied mechanisms are based on a simplified model of the miRNA function, focusing the analysis on the seed region and minimizing the effect of interactions affecting the rest of the strand.

The heterogeneity of both databases used in the analysis may come from the use of different algorithms. On the other hand, overlapping of the results seems to justify the use of those sources as well as allow for a more accurate prediction.

The possibility of using computational methods as an approximation only is a significant limitation. For this reason, further investigation involving experimental laboratory methods is required to verify the presented interdependencies. The verification procedure may include examination of coexpression for the remaining miRNA–mRNA transcript pairs, SNPs in coexistence with gene expression changes, and miRNA expression profiles impaired in patients with the phenotype of interest. The necessity to
indicated in this paper lays within such a region. The locus identified in this case is 9p21.3, correlated with numerous CVDs. Changes in gene expression were associated with CAD and confirmation for these experiments are provided by GWAS, in which CDKN2B SNPs have been associated with atherosclerosis and CAD.

Of note, the ANRIL gene, antisense to CDKN2B, is situated in the same location. Although all of its transcripts are noncoding RNA, it has been assigned a number of key functions such as involvement in epigenetic changes and modulation of the expression of genes responsible for the process of atherogenesis, which is further confirmed by its altered expression in haplotypes associated with atherosclerosis and CAD.

The GWAS conclusions for ANRIL are consistent with the results provided for CDKN2B, suggesting the involvement of ANRIL SNPs in the pathogenesis of CAD and atherosclerosis. The involvement of noncoding RNA in many complex processes suggests the need to search for causative factors not only among the coding genes but also within other loci.

Particularly noteworthy is the characteristic of the rs3217992 polymorphism, which hypothetically affects miRNA-205 binding, specified for the CDKN2B gene.

rs3217992 is associated with the occurrence of MI, and its presence has a negative effect (demonstrates cis-acting effects) on the expression of CDKN2B and ANRIL, which was confirmed by the tools used in the present analysis.
miR-205, for which the coexpression with the CDKN2B gene has been demonstrated, was found to be overexpressed in ischemic striated muscle of the rat and correlated with metabolism of the failing heart.

Still not much is known on the effects of miRNA on the cardiovascular system, and this is the main issue addressed in the present publication. Based on the previous studies, it can be concluded that associations with CVDs have been shown for some of herein identified miRNAs (as is the case for miR-483-5p showing the relationship with angiogenesis or miR-224 showing altered expression profile in myocardial hypertrophy).

Moreover, miRNA function is often associated with the effect on the course of neoplasia. Although similar association studies are also conducted for miRNAs relation to phenotype and pathophysiology of the cardiovascular system, they are far less numerous. Given the positive results of studies linking tumor development with miRNA expression, a strong effect of genetic factors on cardiovascular risk, and the possibility that miRNA affects the function of genes associated with CVDs as discussed in this study, we conclude that the studies on miRNAs are extensive, untapped niche, which allows to show their relationship to the pathology of the cardiovascular system. Moreover, not in all cases a functional miRNA must demonstrate mutual expression with its target mRNA within the examined tissue. The existence of circulating miRNAs subjected to exocytosis and transported in serum in the form of exosomes or complexes with high-density lipoprotein and Ago2, showing an intracellular activity, has been proved. The discovery of novel candidates brings the opportunity for developing new diagnostic methods based on serum miRNA expression profile and use of appropriate miRNA analogues or inhibitors as therapeutic agents in the future.

Acknowledgments Karol Kamiński was supported by the Foundation for Polish Science.
7 Kathiresan S, Voight BF, Purcell S, et al.; Myocardial Infarction Genet-
ics Consortium. Genome-wide association of early-onset myocardial infar-
tion with single nucleotide polymorphisms and copy number variants. Nat
8 Broadbent HM, Peden JF, Lorkowski S, et al.; PROCARDIS consortium. Susceptibility to coronary artery disease and diabetes is encoded by dis-
tinct, tightly linked SNPs in the ANRIL locus on chromosome 9q. Hum Mol
10 Ziebarth JD, Bhattacharya A, Chen A, Cui Y. PolymiRTS Database 2.0:
11 Thomas LF, Saito T, Saetrom P. Inferring causative variants in microR-
12 Mestdagh P, Lefever S, Pattyn F, et al.; PROCARDIS consortium. Susceptibility to coronary artery disease and diabetes is encoded by dis-
tinct, tightly linked SNPs in the ANRIL locus on chromosome 9p. Hum Mol
14 Broadbent HM, Peden JF, Lorkowski S, et al.; PROCARDIS consortium. Susceptibility to coronary artery disease and diabetes is encoded by dis-
tinct, tightly linked SNPs in the ANRIL locus on chromosome 9q. Hum Mol
16 Ziebarth JD, Bhattacharya A, Chen A, Cui Y. PolymiRTS Database 2.0:
11 Thomas LF, Saito T, Saetrom P. Inferring causative variants in microR-
12 Mestdagh P, Lefever S, Pattyn F, et al.; PROCARDIS consortium. Susceptibility to coronary artery disease and diabetes is encoded by dis-
tinct, tightly linked SNPs in the ANRIL locus on chromosome 9p. Hum Mol
ARTYKUŁ ORYGINALNY – KONKURS STUDENCKI 2013*

Identyfikacja *in silico* SNP i powiązanych z nimi microRNA, wpływających na występowanie chorób sercowo-naczyniowych

Marcin J. Kamiński¹, Magdalena Kamińska¹, Iwona Skorupa¹,
Remigiusz Kazimierczyk¹, Włodzimierz J. Musiał², Karol A. Kamiński²

1 Studenckie Kolo Naukowe przy Klinice Kardiologii, Uniwersytet Medyczny w Białymstoku, Białystok
2 Klinika Kardiologii, Uniwersytet Medyczny w Białymstoku, Białystok

SŁOWA KLUCZOWE
choroby sercowo-naczyniowe, *in silico*, microRNA, polimorfizmy

STRESZCZENIE

**WProwadzenie** MicroRNA (miRNA) to małe RNA odgrywające ważną rolę w regulacji ekspresji genów. Zaburzenia miRNA zostały powiązane z istnieniem zmian fenotypowych, włączając w to choroby sercowo-naczyniowe (*cardiovascular diseases* – CVD).

**CELE** Uzyskanie listy polimorfizmów pojedynczego nukleotydu (*single nucleotide polymorphisms* – SNP) związanych z występowaniem CVD, dla których obliczeniowo wykazano wpływ na miejsca wiązania miRNA, przez co zostanie zweryfikowana hipoteza, że wywołana przez SNP zmiana funkcji miRNA może wpływać na wystąpienie CVD.

**MATERIAL I METODY** SNP, CVD i miRNA były trzema elementami poddanyanalizie. Na podstawie informacji dostępnych w bazach danych stworzono listy SNP powiązanych z konkretną chorobą sercowo-naczyniową oraz SNP zlokalizowanych w znanych miejscach wiązania miRNA. Następnie połączono wpisy odpowiadające temu samemu SNP, co pozwoliło na wskazanie miejsc wiązania miRNA, których warianty mogą być związane z CVD. Wyniki uzupełniono o dodatkowe informacje, obejmujące koekspresję miRNA i mRNA, różnice ekspresji pomiędzy różnymi tkankami oraz analizę Expression Quantitative Trait Locus. Wykorzystano wyłącznie techniki *in silico*, przy użyciu baz znajdujących się w domenie publicznej.

**WYNIKI** Otrzymaliśmy listę 47 wpisów będących unikalnymi powiązaniami miRNA – allel SNP – fenotyp. Nie zgłoszono sprzeczności interesów.

**WNIOSEKI** Podejście obliczeniowe wspiera hipotezę istnienia związku między zmianami funkcji miRNA i licznymi CVD. Podczas gdy obserwacją częstotliwość występowania SNP, patomechanizm ten może występować pospolito w obrębie populacji. Dalszym etapem badań musi być eksperymentalna weryfikacja uzyskanych wyników, jednak ograniczenie liczby interakcji do najbardziej prawdopodobnych w oparciu o prawdopodobieństwo predykcji interakcji ułatwi znalezienie współzależności istotnych klinicznie.