The Role of MicroRNAs in Age-Related Disorders

From population-based genetic studies to experimental validation

Mohsen Ghanbari

ACKNOWLEDGMENTS

The research presented in this thesis was performed at Departments of Epidemiology and Immunology, Erasmus MC, Rotterdam, the Netherlands.

The research was supported by a scholarship from the Ministry of Health and Medical Education in IRAN and Mashhad University of Medical Sciences (MUMS), and by Utzchit grant awarded to M. Meester-Smoor. The contribution of the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists is gratefully acknowledged. The Rotterdam Study is supported by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam.

Publication of this thesis was kindly supported by the Department of Epidemiology and the Erasmus Medical Center Rotterdam, the Netherlands. Financial support was provided by Landelijke Stichting voor Blinden en Slechtzienden (LSBS), Alzheimer Nederland and ChipSoft. Additional financial support by the Dutch Heart Foundation for the publication of this thesis is gratefully acknowledged.

ISBN: 978-94-92863-49-6

Layout, cover design and printing by Optima Grafische Communicatie

The Role of MicroRNAs in Age-Related Disorders
From population-based genetic studies to experimental validation

De rol van microRNAs in leeftijdsgerelateerde ziekten
Van genetische populatiestudies naar experimentele validatie

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus
Prof.dr. H.A.P. Pols
en volgens besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op
woensdag 5 juli 2017 om 13:30 uur

Door

Mohsen Ghanbari
geboren te Mashhad, Iran

Erasmus University Rotterdam

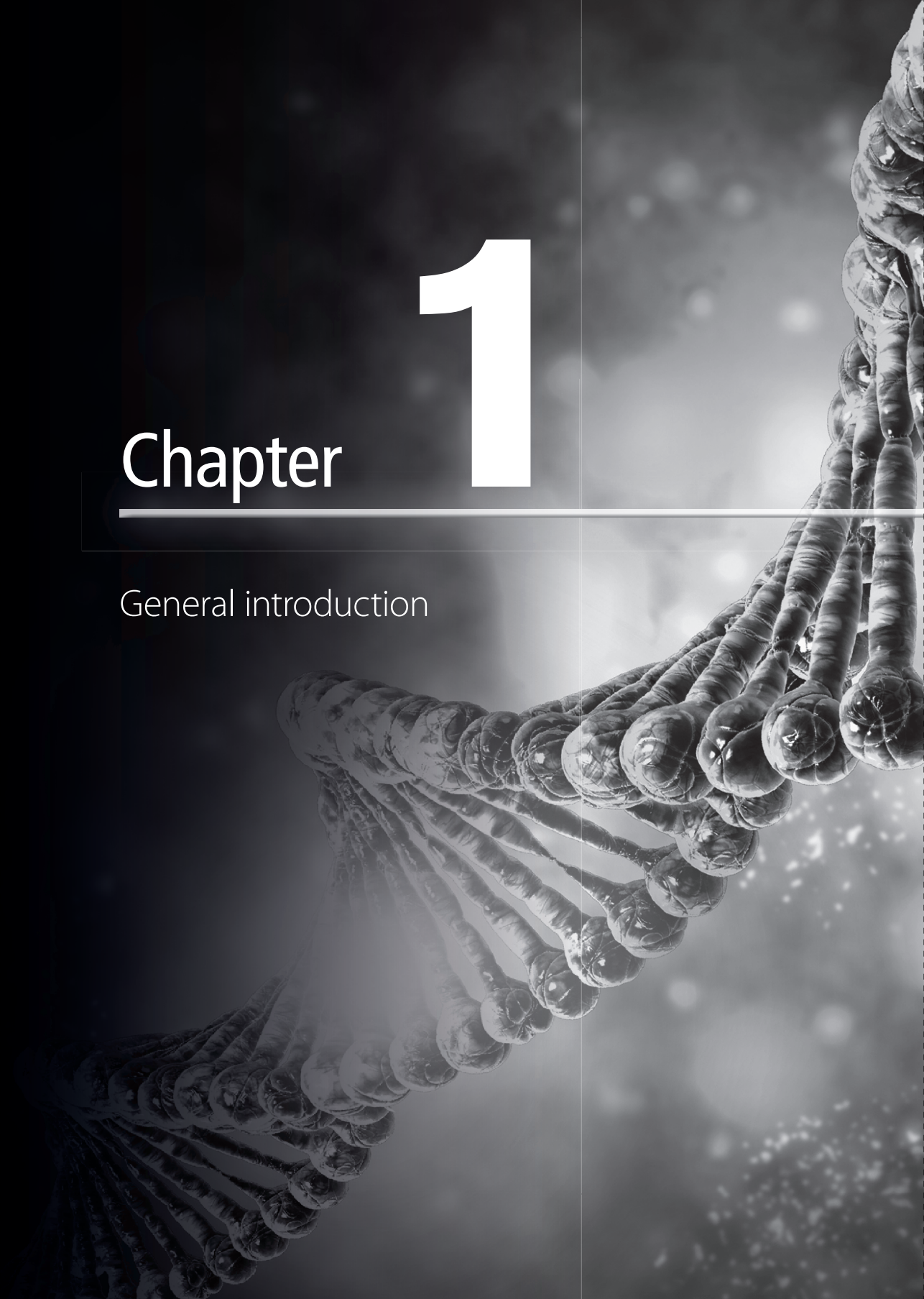
Promotiecommissie

Promotor:	Prof.dr. O.H. Franco
Overige leden:	Prof.dr. C.C.W. Klaver Prof.dr. A. Uitterlinden Prof.dr. L. Franke
Copromotors:	Dr. A. Dehghan Dr. S. Erkeland Dr. M. Meester-Smoor

Paranimfen:	E. Sam Shokrollahi Lea Jasmina Jabbarian

To my wife Sara and my son Shahrad
And to my parents

TABLE OF CONTENTS



Chapter 1

General introduction

Recent developments in next-generation sequencing technologies have led to extensive transcriptome analysis of many human tissues and cell lines at an unprecedented scale. These data display that nearly the entire human genome has been transcribed, however, only a very small proportion (2%) of the transcripts are translated into protein products [1]. The rest of the transcripts that do not possess any protein-coding capacity are annotated as non-coding RNAs (ncRNAs). Over the past few years, it has become increasingly evident that ncRNAs act as important players in the epigenetic, post-transcriptional and translational coordination of gene expression in developmental processes and human diseases [2]. The ncRNAs can be roughly categorized, based on their transcript size, into two main classes: small ncRNAs (< 200 nucleotides) and long ncRNAs (> 200 nucleotides) [2, 3]. By far the best-characterized ncRNAs are microRNAs (miRNAs), which are the main focus of this thesis.

MICRORNAS (MIRNAS)

MiRNAs are a conserved class of small ncRNAs, spanning 20-24 nucleotides, that post-transcriptionally regulate gene expression at both messenger RNA (mRNA) and protein levels [4]. MiRNAs were discovered in *C.elegans* in the early 1990s, when Ambros and colleagues observed that small RNA products from the heterochronic *lin*-4 gene regulated *lin*-14 gene expression by means of an RNA-RNA antisense pairing [5, 6]. However, not having homologs in other species, this RNA-based regulatory mechanism was thought to be unique to *C.elegans* until the second miRNA, *let*-7, was discovered with homologs in humans and *Drosophila melanogaster* [7, 8]. Since then, miRNAs have gained extensive attention as important modulators in diverse biological pathways, including stem cell differentiation, cell proliferation, apoptosis, and organ development, as well as being associated with a number of human diseases ranging from cancer to cardiovascular disease [9, 10]. The total number of confidently identified human miRNAs is now approximately 1,500 according to miRBase database (build 21), and around two-third of all coding genes in our genome are predicted to have at least one miRNA-binding site in their 3'UTR [11-14]. It was initially thought that human miRNAs are transcribed from genomic regions quite distant from previously annotated genes, suggesting that they are generated as independent transcriptional units. However, later analyses of miRNA gene positions demonstrated that approximately 50% of human miRNAs are hosted within and co-transcribed with a protein-coding mRNA transcription unit. Thus, the expression of intronic miRNAs largely coincides with the transcription of the host gene [15]. Nearly 50% of miRNAs are clustered with one or more other miRNAs and are believed to be co-transcribed as a single polycistron by RNA Pol II enzyme in the nucleus [15].

MIRNA BIOGENESIS

In the canonical biogenesis pathway, miRNA genes are initially transcribed in the nucleus by the normal cellular machinery to yield a large RNA transcript, approximately 1000 nucleotides, containing an extended stem-loop structure called a primary miRNA (pri-miRNA) [16]. Two subsequent processing events lead to the functional, mature miRNA in animals. First, the RNase-III endonuclease enzyme, Drosha, coupled with the dsRNA-binding protein DiGeorge Syndrome critical region 8 (*DGCR8*), cleaves the pri-miRNA transcript into a smaller, ~70 nucleotides, hairpin precursor miRNA (pre-miRNA) [17, 18]. The pre-miRNA is then shuttled out of the nucleus through the Exportin 5 (*XPO5*) nuclear transport receptor in cooperation with a Ran GTPase cofactor [19]. After being transported to the cytoplasm, a second RNase-III enzyme (*Dicer*) cleaves the loop structure of the pre-miRNA, resulting in an approximately 22bp double-stranded miRNA [20]. The duplex miRNA is then dissociated and the passenger strand is degraded. Subsequently, the strand with the less thermodynamically stable 5′ end (guide strand) gets incorporated into the RNA-induced silencing complex (RISC) [21]. This active RISC complex, consisting of the mature miRNA and Argonaute (*Ago*) proteins, interacts with the 3′-untranslated region (3′UTR) of target mRNA. This results in either translational repression or site-specific cleavage of target mRNAs, depending on the degree of complementarity to its target [14, 22]. An overall schematic of miRNA biogenesis and function is shown in **Figure 1**.

MIRNA-TARGET RECOGNITION AND REGULATORY FUNCTION

The core of a mature miRNA, called the "seed" sequence, includes nucleotides 2-7 (sometimes 2-8) from the 5′ end. This region of the miRNA is the most conserved among species and plays a critical role in target recognition and interaction [21]. The 3′-end of a miRNA modifies binding of the miRNA to its target genes and may compensate for seed mismatches [23]. Most miRNAs in animals are thought to control mRNA stability and translation through imperfect base-pairing to the 3′UTR of target mRNAs [16, 21]. The most common mechanism through which a miRNA regulates its target gene expression is the binding of the miRNA seed sequence to the 3′UTR sequence of the target mRNA (**Figure 2**) [13, 21]. Other less common mechanisms that have been described include non-canonical sites, which display good base-pairing at both the 5′ and 3′ ends, and 3′ compensatory sites, which show weak 5′ base-pairing and depend on a strong pairing at the 3′ end [24].

The biological role of individual miRNAs is dictated by regulating their target mRNAs expression [21]. Given that miRNAs need only minimum binding by a short seed sequence, it is not surprising that a single miRNA targets multiple genes [13]. As a consequence, the accurate prediction of the miRNA target genes has been a great

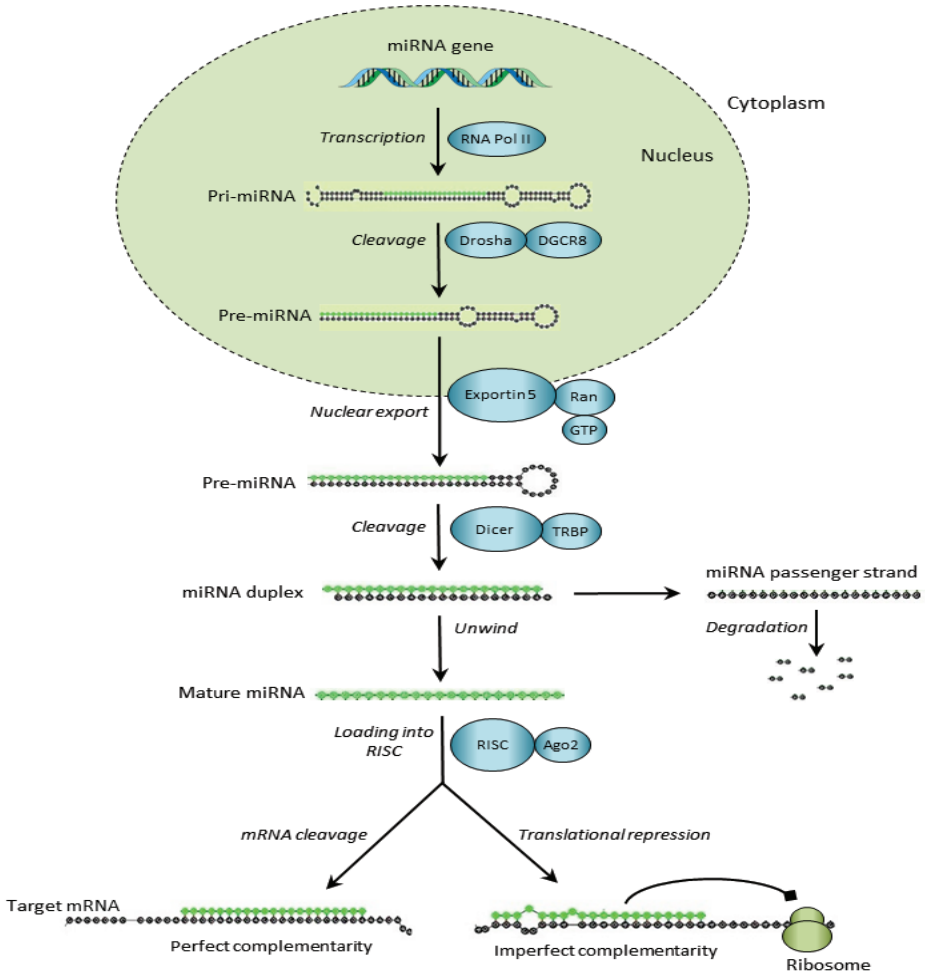


Figure 1. The biogenesis of miRNAs is a multistep coordinated process. After transcription, the Drosha enzyme processes the long primary transcripts (pri-miRNA), yielding a hairpin precursors (pre-miRNA) consisting of approximately 70 nucleotides. The pre-miRNA hairpins are exported by Exportin 5 (*XPO5*) to the cytoplasm where they are further processed into unstable, 19-25 nucleotides miRNA duplex structures by the Dicer enzyme. The 5′ end of the miRNA (guide strand) is incorporated into a multiple-protein nuclease complex, the RNA-induced silencing complex (RISC), which regulates gene expression by transcript destabilization, inhibition of protein translation and translational repression.

challenge. In recent years much effort has been put toward devising a genome-wide computational search that captures most of the regulatory targets without bringing in too many false-positive predictions [25]. Computational algorithms have been the major driving force in predicting miRNA targets [13, 25]. Current prediction methods are diverse, however a general agreement that has emerged on two important criteria includes: 1) Assessing the thermodynamic features of the miRNA:mRNA duplex by cal-

culating the free energy of the putative interaction, where a lower free energy indicating the binding of the miRNA to mRNA is stronger, and 2) Conservation of the miRNA target site among different species [13]. Several prediction programs are currently available that follow these major criteria, including TargetScan [13], miRTarBase [26], PicTar [27] and miRecords [28]. However, perfect seed pairing is not a generally reliable predictor for miRNA-target interactions, which may explain why some putative miRNA target sites were shown to be nonfunctional [29]. Therefore, in addition to the bioinformatic analysis of predicting an accessible miRNA-binding site, the functional significance of the predicted miRNA-mRNA interaction needs to be validated experimentally.

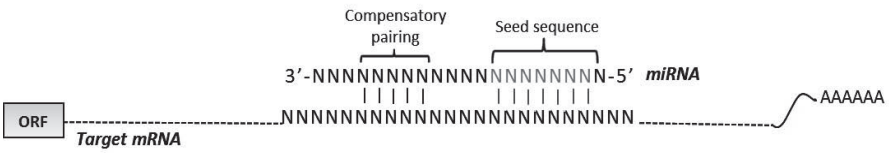


Figure 2. The miRNA seed sequence (nucleotides 2-7/8 from the 5′ end) binds to its complementary sequence in the 3′UTR of target mRNA.

GENETIC VARIANTS IN MIRNA GENES AND THEIR BINDING SITES

MiRNA transcripts need to fulfill structural and sequences prerequisites in order to result in expression of the right mature miRNAs [16, 30]. In addition, the expression level of a miRNA is critical for its activity on the target mRNA. Genetic polymorphisms in miRNA genes can affect the expression level as well as the functionality of miRNAs (**Figure 3**) [31-33]. In this context, variants located in pri- or pre-miRNA sequences can change the binding affinity of the miRNA hairpin to biogenesis enzymes (Drosha and Dicer) or accessory proteins [34, 35]. These variants can lead to changed processing accuracy, alter frequencies of alternative cleavage sites of the biogenesis enzymes or change miRNA strand loading bias into the RISC complex [31]. These changes in maturation can all result in altered expression of the miRNA, resulting in deregulation of target genes [32]. Variants in the miRNA seed sequence can further affect the miRNA-target gene interaction [36]. In addition to variants in miRNA-encoding sequences, variants within miRNA promoter regions, and other regulatory regions, may result in an altered transcription rate of the pri-miRNA [37]. Given the potentially significant effect of genetic variants on the tightly regulated miRNA repertoire and the importance of miRNA-mediated gene regulation, it is not surprising that variants in miRNA genes have been found to be associated with human diseases [31, 36, 38, 39].

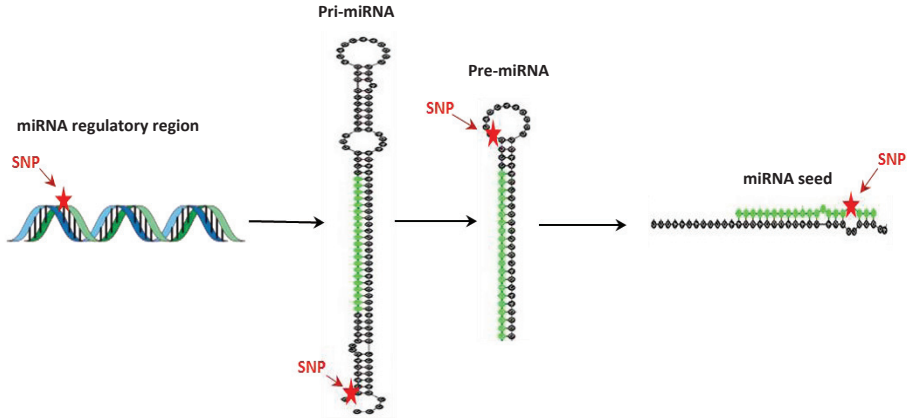
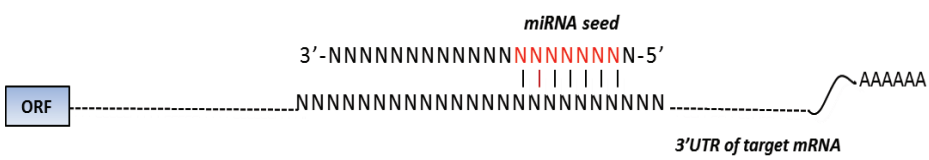


Figure 3. Genetic variants in miRNA regulatory regions, pri- and pre-miRNA sequences may affect the miRNA expression level. Variants in the miRNA seed sequences can further affect the interaction between miRNAs and their target mRNAs.

In addition to variants in miRNA genes, genetic polymorphisms located in miRNA-binding sites within the 3'UTRs of target genes can lead to impaired miRNA regulation (**Figure 4**) [40]. Polymorphisms in miRNA-binding sites are likely to have a smaller impact than variants in miRNA genes, because only the targeting of one gene by one miRNA would be disrupted in case the target gene does not contain multiple binding sites for that miRNA. Nevertheless, the loss of a miRNA-binding site can have important biological implications. Previous computational and experimental studies have shown that miRNA-binding site variants can modulate the risk for disease [38, 41, 42].

A. Disruption of a miRNA-binding site within a gene 3'UTR



B. Creation of a new miRNA-binding site within a gene 3'UTR

Figure 4. Genetic variants within gene 3'UTRs can disrupt an existing miRNA-binding site or create new binding sites for miRNAs.

GENETICS OF COMPLEX DISORDERS

The human genome is thought to consist of approximately 20,000 protein-coding genes [43]. The general positions of these genes within the genome are referred to as loci (singular, locus). Each locus is surrounded by or embedded with DNA sequences dictating when, where and how each gene should be expressed per its interaction with DNA-regulatory factors. The human variation comes in many forms and is influenced by multiple factors including evolutionary forces. The origin of this variation is likely to consist of errors in DNA replication, DNA recombination and the insertion or deletion of mobile transposon elements [44]. The most common form of human variation is the single-nucleotide polymorphism (SNP), which includes variation at a single base position [45]. It has been estimated that SNPs occur at a frequency of 1/1000 base pairs, on average, and constitute approximately 90% of the isolated variations in the human genome [46]. The number of SNPs in human identified currently exceeds 50 million [47, 48]. SNPs in protein coding sequences can result in an amino acid change, which may modify the peptide or the entire protein, and play a role in disease pathophysiology. SNPs in non-coding regions of the genome could change the gene expression through different mechanisms, for example by affecting binding sites of transcription factors or miRNAs [49, 50]. Genetic studies have traditionally used candidate gene studies and family-based linkage studies to identify novel genes for human disease. Candidate gene studies rely on our partial understanding of genes with known biological function regulating the developmental processes of the investigated traits [51]. The practicability of candidate gene approach is largely limited by its reliance on the current knowledge about the known or presumed biology of the trait under study [51]. Family-based linkage analysis has been a traditional means for identifying regions of the genome that contribute to a disease which will guide the search for the causative gene (Mendelian disease) [52]. However, this approach has only achieved little success in identifying genes for complex disorders because of low power and lack of replication [53].

Genome-wide association study (GWAS) is a hypothesis-free approach that investigates the genetic component of common diseases without relying on prior knowledge [54]. In the recent decade, with the development of high-throughput genotyping and next-generation sequencing platforms, GWAS have been successfully conducted to identify genetic variants associated with many complex traits and diseases [55-60]. The completion of the Human Genome project and the International HapMap project provided the possibility to select a set of genetic variants that are representative of all human haplotypes, groups of alleles that are co-inherited based on linkage disequilibrium (LD) maps [60, 61]. This allowed researchers to investigate the genetic variants that nearly cover the entire genome variation by genotyping hundreds of thousands to one million SNPs using SNP array technologies. GWAS survey individual human genetic

variations in thousands of individuals and test for association between those variants and a given trait or condition (**Figure 5**) [62, 63]. For GWAS to be successful, there must be multiple common loci (>1% of the population) that influence a trait or disease risk, by which each locus contributes a small proportion of the overall trait variance or disease susceptibility [64]. Although large-scale GWAS have led us to the discovery of a large number of genetic variants at multiple loci associated with complex disorders, majority of associated variants (~90%) map to non-coding regions of the genome. These include variants in non-coding RNAs and gene 3'UTRs that are responsible for fine-tuning gene expression regulation. Investigation of the functional consequences of non-coding variants and elucidating the biological mechanisms by which these variants and related genes act is a major challenge facing the biological sciences today.

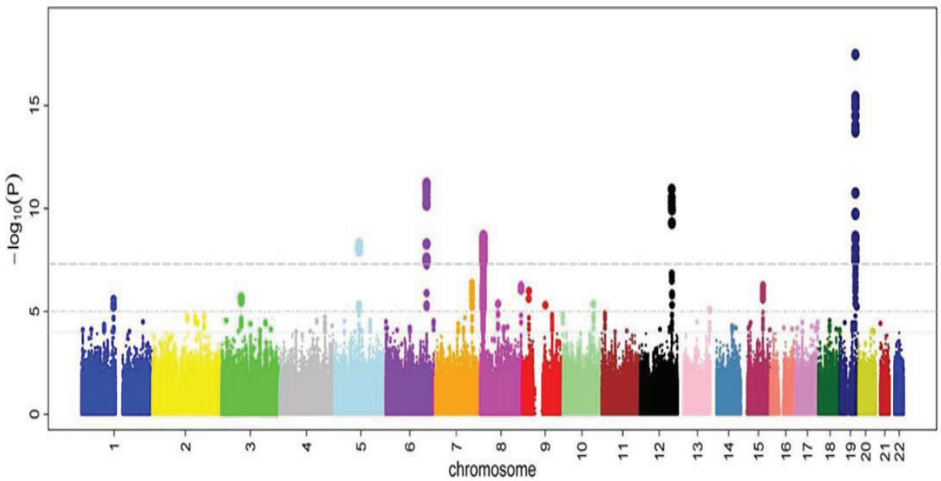


Figure 5. Manhattan plot showing the statistical association between SNPs and a trait of interest, each dot on this plot signifies a SNP. Genomic coordinates (Chr 1-22) are displayed along the X-axis and the negative logarithm of the association p-value for each SNP displayed on the Y-axis. Dash line shows significance threshold.

STUDYING MIRNAS INVOLVED IN HUMAN DISEASES

MiRNAs have become a source of great excitement as regulators of complex diseases because they provide new insights into disease mechanisms and could potentially be therapeutically targeted. Elucidating the associations between miRNAs and diseases in a systematic approach will extend our understanding of the molecular mechanisms underlying human diseases. To study whether miRNAs are involved in the pathogenesis of human diseases, researchers usually perform miRNA expression profiling to conduct a genome-wide analysis of miRNA expression in normal and diseased samples, or to

distinguish miRNA expression signatures associated with diagnosis and therapeutic interventions (Figure 6). The expression of miRNAs can be tested by different technologies such as microarray based methods or RNA sequencing. Microarray technology is a powerful high-throughput tool capable of monitoring the expression of hundreds of miRNAs at once processed in parallel in a single experiment [65]. The miRNA expression profiles are compared between the normal and diseased samples in order to determine differentially expressed miRNAs. The next step would be to identify the target genes and pathways that are regulated by the identified miRNAs. To determine deregulated target genes of miRNAs, gene expression profiling should be done in the same samples. The interactions between deregulated miRNAs and their putative target genes need to be experimentally verified. One possible experiment is to clone the 3'UTR of the target gene immediately downstream of the luciferase (or Renilla) or green fluorescent protein (GFP) open reading frame sequence contained in the reporter plasmid [66]. The recombinant plasmid and the miRNA are then transiently transfected into a host cell, preferably one that does not endogenously express the miRNA under study, and luciferase activity or fluorescence is measured 48 hours after transfection [66]. Alternatively, just the reporter construct can be transfected into cells which express the relevant miRNAs, along with vectors which express mutant versions of the miRNA-binding sites. In this experimental system, the wild-type reporters should have less activity than their respective mutants [67]. To fully understand how the miRNA contributes to the pathology of disease, further experimental studies are needed to be done such as inhibition and overexpression of miRNAs in the right cellular context or animal models [67, 68].

Genome-wide miRNA expression profiling can provide a global insight in which miRNAs may play a role in a disease by generating disease-specific miRNA expression signatures, however, it has some limitations. The main limitation is that differential expression of miRNAs in itself does not provide an indication of causality because the altered expression can be a causal factor, a consequence, or could even be unrelated to the disease pathogenesis [37]. Expression profiling studies are therefore subject to confounding bias or reverse causation. Another constraint is that the affected tissue has to be accessible which is especially problematic in the study of neurodegenerative diseases such as Alzheimer's and Parkinson's diseases. In addition, these type of studies are mainly based on differentially expressed miRNAs in a small number of samples, which makes it difficult to extrapolate the results to the general population. Expression profiling of miRNAs in large cohort studies with validated clinical data is also costly, time-consuming and labor intensive.

The availability of high-throughput omics data (e.g. genomics and transcriptomics) provide a valuable resource for genetic epidemiologic studies. These data encourage researchers to develop novel integration approaches to understand the complexity of the genomic mechanisms behind complex diseases. In this thesis we intended to use

a new method to identify miRNAs associated with age-related disorders using genetic evidence. In this approach, the starting point is an association analysis between genetic variants in miRNA-related sequences and the trait of interest. When a genetic variant in a miRNA or in a miRNA-binding site is associated with the trait, functional validation studies can be initiated to investigate the effect of the variant on the miRNA-mediated gene regulation. The genetic approach has some advantages in comparison with miRNA expression profiling approach. The main advantage is that when a miRNA variant is associated with the disease risk, it supports the idea that the miRNA may play a role in the disease pathomechanism. In addition, since genetic information is constant over the life course, reverse causation and confounding bias, two matters of using miRNA expression profiling approach, are refuted. Another advantage is that diseases with no accessible affected tissue (e.g. Alzheimer's disease) can still be studied because genomic DNA can be isolated from easy-accessible patient tissues such as blood.

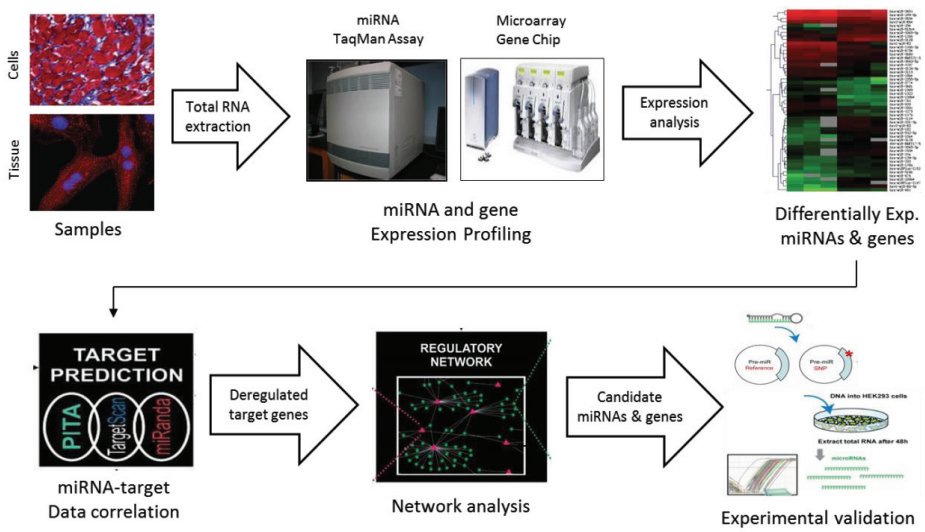


Figure 6. A schematic of studying miRNAs involved in human disease using miRNA profiling approach.

AIM OF THIS THESIS

The overall objective of this thesis was to identify miRNAs involved in age-related disorders (including cardiovascular risk factors and disease, and a number of common neurodegenerative and ophthalmic diseases). The second aim was to ascertain target genes of the identified miRNAs that may mediate the miRNA downstream effects in relation to

the associated traits and diseases. To accomplish these aims, we used a new approach in which we investigated the association of miRNAs with these complex disorders using genetic data followed by various *in silico* and *in vitro* validation studies.

OUTLINE OF THIS THESIS

Chapter 2 of this thesis is devoted to the association of genetic variants in non-coding RNAs (miRNAs and lncRNAs) with cardiovascular risk factors and disease. In **Chapter 2.1** we make a dataset of genetic variants located in miRNA-encoding sequences. We examine the association between variants in miRNA seed sequences with cardiometabolic disorders using data from the largest available GWAS. We perform *in silico* and *in vitro* analyses to examine the functionality of a miRNA seed sequence variant associated with cardiovascular traits. In **Chapter 2.2** we scan for the association between variants located in miRNA-encoding sequences, beyond the seed sequence, and cardiovascular risk factors and disease followed by experimental validation studies. In **Chapter 2.3** we make a comprehensive dataset of variants located in miRNA-binding sites and investigate their associations with different cardiometabolic traits and diseases. We apply a prioritization method based on a set of predefined criteria to highlight the most likely functional candidate variants for functional studies. We experimentally test whether the candidate variants affect miRNA-mediated regulation of their genes. In **Chapter 2.4** we investigate the association of genetic variants in lncRNAs with ten cardiometabolic disorders using the publicly available GWAS data. We further examine the association of expression and DNA methylation of hosting lncRNAs with cardiometabolic disorders using the Rotterdam Study data. We underline a number of potentially functional variants associated with cardiometabolic disorders that may act through lncRNAs. We further highlight a number of loci for which the role of lncRNAs may have been overlooked in the current literature.

Chapter 3 focuses on the association between non-coding RNAs (miRNAs and lncRNAs) and neurodegenerative diseases. In **Chapter 3.1** we investigate the association of genetic variants in miRNAs as well as miRNA-binding sites with Alzheimer's disease (AD) using the largest available GWAS. Subsequently, we perform several computational and experimental studies to determine the impact of a candidate miRNA variant associated with AD. In **Chapter 3.2** we scan for the association of miRNA-related variants with Parkinson's disease (PD). Performing subsequent *in silico* and *in vitro* studies, we test whether the associated variants in miRNAs and their target sites may affect miRNA-mediated regulation of PD genes. In **Chapter 3.3** we examine the association between genetic variants in lncRNAs and AD. We perform various *in silico* and *in vitro* analyses to provide evidence for the functionality of a non-coding variant associated with AD.

Chapter 4 focuses on the association between miRNA-related variants and ophthalmic diseases. In **Chapter 4.1** we perform a genome-wide scan for miRNAs contributing to Age-related macular degeneration (AMD). We demonstrate that three variants in miRNAs and a number of variants in miRNA-binding sites may affect miRNA-mediated regulation of genes involved in AMD. Finally, in **Chapter 4.2** we investigate the association of miRNA-related variants with Primary Open-Angle Glaucoma (POAG) endophenotypes, aiming to identify miRNAs and target genes that may play a role in the pathogenesis of POAG.

REFERENCES

1. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, et al: Landscape of transcription in human cells. *Nature* 2012, 489:101-108.
2. Pennisi E: Genomics. ENCODE project writes eulogy for junk DNA. *Science* 2012, 337:1159, 1161.
3. Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermuller J, Hofacker IL, et al: RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* 2007, 316:1484-1488.
4. He L, Hannon GJ: MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet* 2004, 5:522-531.
5. Lee RC, Feinbaum RL, Ambros V: The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 1993, 75:843-854.
6. Wightman B, Ha I, Ruvkun G: Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell* 1993, 75:855-862.
7. Lee RC, Ambros V: An extensive class of small RNAs in Caenorhabditis elegans. *Science* 2001, 294:862-864.
8. Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, Horvitz HR, Ruvkun G: The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature* 2000, 403:901-906.
9. Esquela-Kerscher A, Slack FJ: Oncomirs - microRNAs with a role in cancer. *Nat Rev Cancer* 2006, 6:259-269.
10. van Rooij E, Olson EN: MicroRNA therapeutics for cardiovascular disease: opportunities and obstacles. *Nat Rev Drug Discov* 2012, 11:860-872.
11. Griffiths-Jones S: miRBase: microRNA sequences and annotation. *Curr Protoc Bioinformatics* 2010, Chapter 12:Unit 12 19 11-10.
12. Kozomara A, Griffiths-Jones S: miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* 2014, 42:D68-73.
13. Agarwal V, Bell GW, Nam JW, Bartel DP: Predicting effective microRNA target sites in mammalian mRNAs. *Elife* 2015, 4.
14. 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.
15. 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.
16. Shukla GC, Singh J, Barik S: MicroRNAs: Processing, Maturation, Target Recognition and Regulatory Functions. *Mol Cell Pharmacol* 2011, 3:83-92.
17. Gregory RI, Yan KP, Amuthan G, Chendrimada T, Doratotaj B, Cooch N, Shiekhattar R: The Microprocessor complex mediates the genesis of microRNAs. *Nature* 2004, 432:235-240.
18. Han J, Lee Y, Yeom KH, Nam JW, Heo I, Rhee JK, Sohn SY, Cho Y, Zhang BT, Kim VN: Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell* 2006, 125:887-901.
19. Bohnsack MT, Czaplinski K, Gorlich D: Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA* 2004, 10:185-191.
20. Macrae IJ, Zhou K, Li F, Repic A, Brooks AN, Cande WZ, Adams PD, Doudna JA: Structural basis for double-stranded RNA processing by Dicer. *Science* 2006, 311:195-198.
21. Bartel DP: MicroRNAs: target recognition and regulatory functions. *Cell* 2009, 136:215-233.

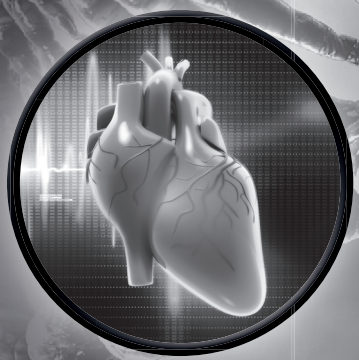
22. Azuma-Mukai A, Oguri H, Mituyama T, Qian ZR, Asai K, Siomi H, Siomi MC: Characterization of endogenous human Argonautes and their miRNA partners in RNA silencing. *Proc Natl Acad Sci U S A* 2008, 105:7964-7969.
23. Yekta S, Shih IH, Bartel DP: MicroRNA-directed cleavage of HOXB8 mRNA. *Science* 2004, 304:594-596.
24. Shin C, Nam JW, Farh KK, Chiang HR, Shkumatava A, Bartel DP: Expanding the microRNA targeting code: functional sites with centered pairing. *Mol Cell* 2010, 38:789-802.
25. Akhtar MM, Micolucci L, Islam MS, Olivieri F, Procopio AD: Bioinformatic tools for microRNA dissection. *Nucleic Acids Res* 2016, 44:24-44.
26. Chou CH, Chang NW, Shrestha S, Hsu SD, Lin YL, Lee WH, Yang CD, Hong HC, Wei TY, Tu SJ, et al: miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. *Nucleic Acids Res* 2016, 44:D239-247.
27. Krek A, Grun D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, MacMenamin P, da Piedade I, Gunsalus KC, Stoffel M, Rajewsky N: Combinatorial microRNA target predictions. *Nat Genet* 2005, 37:495-500.
28. Xiao F, Zuo Z, Cai G, Kang S, Gao X, Li T: miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Res* 2009, 37:D105-110.
29. Didiano D, Hobert O: Perfect seed pairing is not a generally reliable predictor for miRNA-target interactions. *Nat Struct Mol Biol* 2006, 13:849-851.
30. Krol J, Sobczak K, Wilczynska U, Drath M, Jasinska A, Kaczynska D, Krzyzosiak WJ: Structural features of microRNA (miRNA) precursors and their relevance to miRNA biogenesis and small interfering RNA/short hairpin RNA design. *J Biol Chem* 2004, 279:42230-42239.
31. Ryan BM, Robles AI, Harris CC: Genetic variation in microRNA networks: the implications for cancer research. *Nat Rev Cancer* 2010, 10:389-402.
32. Ghanbari M, de Vries PS, de Looper H, Peters MJ, Schurmann C, Yaghootkar H, Dorr M, Frayling TM, Uitterlinden AG, Hofman A, et al: A genetic variant in the seed region of miR-4513 shows pleiotropic effects on lipid and glucose homeostasis, blood pressure, and coronary artery disease. *Hum Mutat* 2014, 35:1524-1531.
33. Ghanbari M, Sedaghat S, de Looper HW, Hofman A, Erkeland SJ, Franco OH, Dehghan A: The association of common polymorphisms in miR-196a2 with waist to hip ratio and miR-1908 with serum lipid and glucose. *Obesity (Silver Spring)* 2015, 23:495-503.
34. Ghanbari M, Darweesh SK, de Looper HW, van Luijn MM, Hofman A, Ikram MA, Franco OH, Erkeland SJ, Dehghan A: Genetic Variants in MicroRNAs and Their Binding Sites Are Associated with the Risk of Parkinson Disease. *Hum Mutat* 2016, 37:292-300.
35. Ghanbari M, Ikram MA, de Looper HW, Hofman A, Erkeland SJ, Franco OH, Dehghan A: Genome-wide identification of microRNA-related variants associated with risk of Alzheimer's disease. *Sci Rep* 2016, 6:28387.
36. Mencia A, Modamio-Hoybjor S, Redshaw N, Morin M, Mayo-Merino F, Olavarrieta L, Aguirre LA, del Castillo I, Steel KP, Dalmay T, et al: Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss. *Nat Genet* 2009, 41:609-613.
37. Cammaerts S, Strazisar M, De Rijk P, Del Favero J: Genetic variants in microRNA genes: impact on microRNA expression, function, and disease. *Front Genet* 2015, 6:186.
38. Gong J, Tong Y, Zhang HM, Wang K, Hu T, Shan G, Sun J, Guo AY: Genome-wide identification of SNPs in microRNA genes and the SNP effects on microRNA target binding and biogenesis. *Hum Mutat* 2012, 33:254-263.

39. Hughes AE, Bradley DT, Campbell M, Lechner J, Dash DP, Simpson DA, Willoughby CE: Mutation altering the miR-184 seed region causes familial keratoconus with cataract. *Am J Hum Genet* 2011, 89:628-633.
40. Ghanbari M, Franco OH, de Looper HW, Hofman A, Erkeland SJ, Dehghan A: Genetic Variations in MicroRNA-Binding Sites Affect MicroRNA-Mediated Regulation of Several Genes Associated With Cardio-metabolic Phenotypes. *Circ Cardiovasc Genet* 2015, 8:473-486.
41. Bhattacharya A, Ziebarth JD, Cui Y: PolymiRTS Database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. *Nucleic Acids Res* 2014, 42:D86-91.
42. Richardson K, Lai CQ, Parnell LD, Lee YC, Ordovas JM: A genome-wide survey for SNPs altering microRNA seed sites identifies functional candidates in GWAS. *BMC Genomics* 2011, 12:504.
43. Ezkurdia I, Juan D, Rodriguez JM, Frankish A, Diekhans M, Harrow J, Vazquez J, Valencia A, Tress ML: Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes. *Hum Mol Genet* 2014, 23:5866-5878.
44. Hartman JLt, Garvik B, Hartwell L: Principles for the buffering of genetic variation. *Science* 2001, 291:1001-1004.
45. Collins FS, Brooks LD, Chakravarti A: A DNA polymorphism discovery resource for research on human genetic variation. *Genome Res* 1998, 8:1229-1231.
46. Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF, Denisov G, et al: The diploid genome sequence of an individual human. *PLoS Biol* 2007, 5:e254.
47. Stark AL, Zhang W, Zhou T, O'Donnell PH, Beiswanger CM, Huang RS, Cox NJ, Dolan ME: Population differences in the rate of proliferation of international HapMap cell lines. *Am J Hum Genet* 2010, 87:829-833.
48. Thorisson GA, Smith AV, Krishnan L, Stein LD: The International HapMap Project Web site. *Genome Res* 2005, 15:1592-1593.
49. Khurana E, Fu Y, Chakravarty D, Demichelis F, Rubin MA, Gerstein M: Role of non-coding sequence variants in cancer. *Nat Rev Genet* 2016, 17:93-108.
50. Dastani Z, Hivert MF, Timpson N, Perry JR, Yuan X, Scott RA, Henneman P, Heid IM, Kizer JR, Lyytikainen LP, et al: Novel loci for adiponectin levels and their influence on type 2 diabetes and metabolic traits: a multi-ethnic meta-analysis of 45,891 individuals. *PLoS Genet* 2012, 8:e1002607.
51. Zhu M, Zhao S: Candidate gene identification approach: progress and challenges. *Int J Biol Sci* 2007, 3:420-427.
52. Risch N: Linkage strategies for genetically complex traits. I. Multilocus models. *Am J Hum Genet* 1990, 46:222-228.
53. Altmuller J, Palmer LJ, Fischer G, Scherb H, Wjst M: Genomewide scans of complex human diseases: true linkage is hard to find. *Am J Hum Genet* 2001, 69:936-950.
54. Hirschhorn JN, Daly MJ: Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* 2005, 6:95-108.
55. Global Lipids Genetics C, Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, Ganna A, Chen J, Buchkovich ML, et al: Discovery and refinement of loci associated with lipid levels. *Nat Genet* 2013, 45:1274-1283.
56. International Consortium for Blood Pressure Genome-Wide Association S, Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, Chasman DI, Smith AV, Tobin MD, Verwoert GC, et al: Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* 2011, 478:103-109.

57. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, DeStafano AL, Bis JC, Beecham GW, Grenier-Boley B, et al: Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* 2013, 45:1452-1458.
58. Nalls MA, Pankratz N, Lill CM, Do CB, Hernandez DG, Saad M, DeStefano AL, Kara E, Bras J, Sharma M, et al: Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat Genet* 2014, 46:989-993.
59. Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, Saleheen D, Kyriakou T, Nelson CP, Hopewell JC, et al: A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet* 2015, 47:1121-1130.
60. International HapMap C: The International HapMap Project. *Nature* 2003, 426:789-796.
61. International HapMap C, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, et al: A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007, 449:851-861.
62. Collins FS, Guyer MS, Charkravarti A: Variations on a theme: cataloging human DNA sequence variation. *Science* 1997, 278:1580-1581.
63. Weiss KM, Clark AG: Linkage disequilibrium and the mapping of complex human traits. *Trends Genet* 2002, 18:19-24.
64. Stranger BE, Stahl EA, Raj T: Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics* 2011, 187:367-383.
65. Chen Y, Gelfond JA, McManus LM, Shireman PK: Reproducibility of quantitative RT-PCR array in miRNA expression profiling and comparison with microarray analysis. *BMC Genomics* 2009, 10:407.
66. Meenhuis A, van Veelen PA, de Looper H, van Boxtel N, van den Berge IJ, Sun SM, Taskesen E, Stern P, de Ru AH, van Adrichem AJ, et al: MiR-17/20/93/106 promote hematopoietic cell expansion by targeting sequestosome 1-regulated pathways in mice. *Blood* 2011, 118:916-925.
67. Thomson DW, Bracken CP, Goodall GJ: Experimental strategies for microRNA target identification. *Nucleic Acids Res* 2011, 39:6845-6853.
68. Kuhn DE, Martin MM, Feldman DS, Terry AV, Jr., Nuovo GJ, Elton TS: Experimental validation of miRNA targets. Methods 2008, 44:47-54.
69. upta SK, Bang C, Thum T: Circulating microRNAs as biomarkers and potential paracrine mediators of cardiovascular disease. Circ Cardiovasc Genet 2010, 3:484-488.
70. Olson EN: MicroRNAs as therapeutic targets and biomarkers of cardiovascular disease. Sci Transl Med 2014, 6:239ps233.

Chapter 2

Contribution of microRNAs to cardiovascular risk factors and disease



CHAPTER 2.1

miR-4513 associates with multiple cardiometabolic traits and diseases

Mohsen Ghanbari, Paul S. de Vries, Hans de Looper, Marjolein J. Peters, Claudia Schurmann, Hanieh Yaghootkar, Marcus Dörr, Timothy M. Frayling, Andre G. Uitterlinden, Albert Hofman, Joyce B.J. van Meurs, Stefan J. Erkeland, Oscar H. Franco, Abbas Dehghan

Manuscript based on this chapter

A genetic variant in the seed region of miR-4513 shows pleiotropic effects on lipid and glucose homeostasis, blood pressure and coronary artery disease.

Human Mutation. 2014; 35(12): 1524-31.

ABSTRACT

Background. MicroRNAs (miRNA) play a crucial role in the regulation of diverse biological processes by post-transcriptional modulation of gene expression. Genetic polymorphisms in miRNA genes can potentially contribute to a wide range of phenotypes. The effect of such variants on cardiometabolic diseases has not yet been defined.

Methods and Results. We systematically investigated the association of genetic variants in the seed region of miRNAs with cardiometabolic phenotypes, using data from the available genome wide association studies (GWAS) on 17 cardiometabolic traits and diseases. We found that rs2168518:G>A, in the seed region of miR-4513, associates with fasting glucose, LDL-cholesterol and total cholesterol, systolic and diastolic blood pressure and risk of coronary artery disease. We experimentally showed that miR-4513 expression is significantly reduced in presence of the rs2168518 minor allele. We sought to identify miR-4513 target genes that may mediate these associations and revealed five genes (*PCSK1, BNC2, MTMR3, ANK3* and *GOSR2*) through which these effects might be taking place. Using luciferase reporter assays we validated *GOSR2* as a target of miR-4513 and demonstrated that the miRNA mediated regulation of this gene is reduced by rs2168518.

Conclusion. Our findings suggest a pleiotropic role of miR-4513 in cardiometabolic traits and disease, which may improve our understanding of the pathophysiology of cardiovascular disease.

INTRODUCTION

MicroRNAs (miRNAs) are a class of small non-coding RNAs, spanning 20-24 nucleotides, function as crucial regulators in a broad range of biological processes [1]. Since the first miRNA was discovered in the early 1990s, over 1500 miRNAs have been identified with confidence in humans [2, 3]. This number of miRNAs together are predicted to regulate expression levels of approximately 60% of all protein-coding genes [4]. In recent years, miRNAs have been widely studied as potential diagnostic biomarkers and therapeutic targets in complex disorders [5]. Furthermore, miRNAs have gained attention as important modulators of cardiovascular diseases such as myocardial infarction [6, 7], cardiac hypertrophy [8] and heart failure [9], as well as various metabolic processes such as insulin production [10], glucose homeostasis [11], lipid metabolism [12] and obesity [13].

MiRNAs are post-transcriptional regulators of gene expression by interacting with the 3' untranslated region (3'UTR) of target mRNAs [1]. Thereby they repress translation and to a lesser extent accelerate the decay of target transcripts [14]. Given the central role of miRNAs in gene expression, genetic polymorphisms in the corresponding sequences of a miRNA may contribute to a phenotypic variation and disease susceptibility [15, 16]. The core of a mature miRNA, called the "seed region", includes nucleotides 2-7/8 from the 5' end, and plays a critical role in target gene recognition and interaction [17]. Genetic variation within this critical region of miRNA may both disrupt the interaction of a miRNA with target transcripts and create illegitimate miRNA targets [18, 19]. Therefore, miRNA seed polymorphisms are expected to alter the expression profile of target genes and subsequently affect corresponding phenotypes. However, so far only very few pathogenic variants in miRNAs have been evidenced in cardiovascular disease and metabolic syndrome.

In the present study, we aimed to systematically investigate the association between miRNA seed variants and cardiometabolic disorders. In addition, we sought to determine whether any of the target genes of the identified miRNAs may mediate the miRNA effects in the associated phenotypes.

MATERIAL AND METHODS

Identification of miRNA seed polymorphisms

A flow chart of our approach to retrieve single nucleotide polymorphisms (SNPs) in miRNA seed regions is shown in **Figure 1**. We systematically screened all known human miRNAs to identify variants in their seed regions, by reviewing the literature and searching the following online databases: microSNiPer [20], PolymiRTS [21], Patrocles [22] and miRvar [23]. We included variants with minor allele frequency (MAF) > 0.01.

Since previous GWAS meta-analysis on cardiometabolic traits and diseases have been performed using HapMap imputed data, we focused on SNPs that were present in the international HapMap project (release 22) (http://www.hapmap.org/) [24]. For SNPs that were not present in the HapMap, we used the SNAP web tool to find proxy SNPs in high Linkage Disequilibrium (LD) ($R^2 > 0.8$ and distance < 200 kb) (http://www.broadinstitute.org/mpg/snap/id) [25].

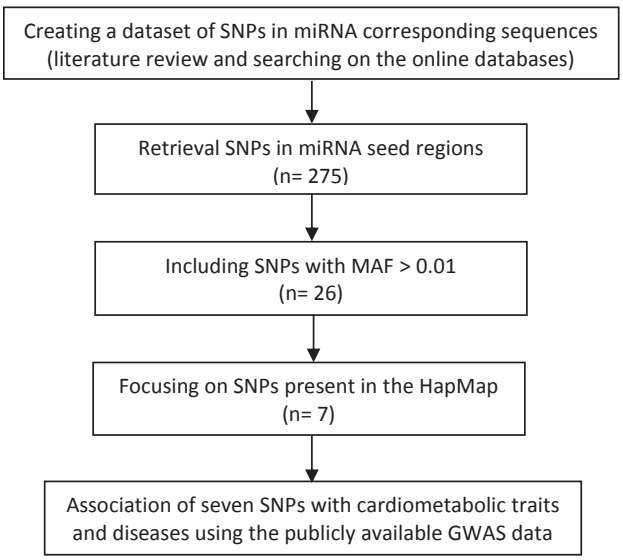


Figure 1. This flow chart describes our method to retrieve SNPs in miRNA seed sequences.

Association of miRNA seed variants with cardiometabolic phenotypes

We examined the association of miRNA seed SNPs with cardiometabolic phenotypes using data from the available GWAS meta-analyses on 17 cardiometabolic traits and diseases. **Table 1** shows a description of the studied phenotypes and consortia that we used in this study.

Data on glycemic traits have been contributed by the Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC) investigators, including fasting glucose, serum glucose after 2hr, fasting insulin, fasting pro-insulin, HbA1c, HOMA-B, and HOMA-IR from up to 133,000 individuals (http://www.magicinvestigators.org) [26-30]. The DIAbetes Genetics Replication and Meta-analysis (DIAGRAMv3) consortium has done a GWAS meta-analysis in 12,171 type 2 diabetes (T2D) cases and 56,862 controls [31]. The Global Lipid Genetics Consortium (GLGC) has carried out GWAS on plasma concentrations of total cholesterol, low-density lipoprotein cholesterol (LDL), high-density lipoprotein cholesterol (HDL) and triglyceride for approximately 100,000 individuals [32]. The Genetic Investigation of ANthropometric Traits (GIANT) consortium has performed

Table 1. Description of 17 cardiometabolic phenotypes and publicly available GWAS data used in this study

Phenotype	Consortium sharing GWAS data	GWAS sample Size	Reference
Glycemic indices			
Fasting glucose	MAGIC	133010	26
Fasting insulin	MAGIC	108557	26
Glucose after2h	MAGIC	42854	27
Pro-insulin	MAGIC	10701	28
HbA1c	MAGIC	46368	29
HOMA-B	MAGIC	46186	30
HOMA-IR	MAGIC	46186	30
Type 2 diabetes	DIAGRAM	12,171 cases/ 56,862 controls	31
Lipid traits			
Total Cholesterol	GLGC	100184	32
Tri -Glyceride	GLGC	96598	32
HDL Cholesterol	GLGC	99900	32
LDL Cholesterol	GLGC	95454	32
Anthropometric measures			
BMI	GIANT	123764	33
WHR	GIANT	77105	34
Blood pressure			
Systolic BP	Global BPGen	71225	35
Diastolic BP	Global BPGen	71225	35
Coronary artery disease	CARDIoGRAMplusC4D	63,746 cases/ 130,681 controls	36

GWAS on anthropometric traits including body mass index (BMI) of over 120,000 and waist/hip ratio (WHR) adjusted for BMI of 77,000 individuals [33, 34]. The Global BPgen consortium has done GWAS on systolic and diastolic blood pressure in over 71,000 individuals. Individuals under treatment for hypertension were imputed to have 15 mm Hg higher systolic blood pressure and 10 mm Hg higher diastolic blood pressure than the observed measurements [35]. The CARDIoGRAMplusC4D consortium conducted GWAS in 63,746 coronary artery disease (CAD) cases and 130,681 controls. In this study, they have assessed 79,138 important related SNPs with CAD on the Metabochip [36].

Effect of miRNA seed variant on the miRNA secondary structure and expression

When a miRNA seed variant was associated with cardiometabolic phenotypes, we used the Vienna RNAfold algorithm to predict the effect of variant on the secondary structure of the primary miRNA sequence [37]. Furthermore, we examined whether the SNP affect mature miRNA expression. We cloned the pre-miRNA sequence containing either major (wild-type) or minor (mutant) allele behind the gene encoding green fluorescent

protein (GFP) in the expression plasmid MSCV-BC [38], resulting in GFP-miRNA fusion transcripts. We thus had two separate expression vectors containing either pre-miRNA (Wt) or pre-miRNA (Mut). HEK293 cell transfection, total RNA isolation and quantitative PCRs were performed as previously described [38]. Mature miRNA levels in cells transfected with pre-miRNA (Wt) compared with cells transfected with pre-miRNA (Mut).

Association of miRNA target genes with cardiometabolic phenotypes

To explore the putative mediatory role of the target genes of a miRNA associated with cardiometabolic phenotypes, we investigated the association of genetic variants in these target genes with the associated phenotypes. The significance threshold for this analysis was set by using a Bonferroni correction based on the number of independent SNPs. We calculated the number of independent SNPs using of Linkage disequilibrium based SNP pruning in PLINK with $R^2 > 0.5$ (http://pngu.mgh.harvard.edu/purcell/plink/). The TargetScan database was used to identify target gene information, including their context score, and evolutionary conserved sites of miRNAs (release 6.2) (http://www.TargetScan.org/).

Expression Quantitative Trait Loci (eQTL)

We examined the association of miRNA seed variant with expression levels of the miRNA target genes using whole blood *trans*-eQTL, and on the host genes expression using *cis*-eQTL data from the Rotterdam Study (n=762). We further replicated the eQTL analyses in two other cohorts; SHIP-TREND (n=963) and InCHIANTI (n=611). The designs of these cohorts have been described in detail elsewhere [39-41]. Association of SNPs or their proxies, based on an $R^2 > 0.7$, were assessed with gene expression levels in whole blood cells. Whole-blood cells were collected in PAXgene-tubes (Becton Dickinson, Breda, the Netherlands). Total RNA was isolated using PAXgene Blood RNA kits (Qiagen, Benelux BV, Venlo, the Netherlands), and to ensure a constant high quality of the RNA preparations, all RNA samples were analyzed using the Labchip GX (Calliper) according to the manufacturer's instructions. Samples with an RNA Quality Score>=7 were amplified and labelled (Ambion Total Prep RNA, Life Technologies, Bleiswijk, the Netherlands), and hybridized to the Illumina Whole-Genome Expression Beadchips (HumanHT-12 v4). Processing of the samples was performed at the Genetic Laboratory of Internal Medicine, Erasmus University Medical Center Rotterdam. The RS-III expression dataset is available at GEO (Gene Expression Omnibus) public repository under the accession GSE 33828. For normalization, raw intensity data generated with the expression arrays were exported from Illumina's Genome Studio V 2010.1 Gene Expression Module to the R environment and quantile normalized and log2-transformed, as well as probe-centered, and sample-standardized. We used the eQTL mapping pipeline called MegaQTL. eQTLs were deemed cis when the distance between the SNP chromosomal position and the probe midpoint

was less than 250 kb; eQTLs were deemed trans when the distance between the SNP chromosomal position and the probe midpoint position was larger than 5 Mbp. eQTLs were mapped using Spearman's rank correlation, using the imputation dosage values as genotypes. Resultant correlations were then converted to p-values and respective z-scores weighted with the square root of the sample size. The model was adjusted for 40 principal components [42].

Luciferase reporter assay

Luciferase reporter assay system was used to validate the interaction of a miRNA with its identified target genes and also to determine the functional consequence of miRNA seed SNP on the binding of the miRNA to its target genes. To amplify the mature miR-4513 sequence, we used a forward primer containing *XhoI* restriction site (AACTCGAGAGGATGTGGTCTTTGCATCTTC) and a reverse primer containing *EcoRI* restriction site (AAGAATTCCCTCCAGTCTCCCCACCTAG). The miRNA sequences with major or minor alleles were cloned in the MSCV-BC vector. In addition, 3'UTR sequence of *GOSR2* was amplified with the forward primer (AATCTAGAGTGATCCCAGCGACTCTTCA) containing the restriction enzyme site *XbaI* and the reverse primer (AAGGGCCCCCGTAGAGATGGCAG GGACT) containing an *ApaI* restriction site. The 3'UTR fragment of *GOSR2*, containing the putative target site of miR-4513, was cloned in the pGL3 Luciferase reporter vector [31]. All constructs were confirmed by Sanger sequencing. HEK293 cells were plated into 12-well plates and co-transfected with MSCV-wild type miR-4513 (contain major allele) or MSCV-mutant miR-4513 (contain minor allele) and pGL3 containing the 3'UTR fragment of *GOSR2*. Luciferase activity was measured with the Dual-Glo Luciferase Assay System according to manufacturer's protocol (Promega, Madison, Wisconsin). Renilla luciferase activity was normalize to the corresponding firefly luciferase activity and plotted as a percentage of the control. The experiments were performed in triplicate.

Pathway analysis for the identified miRNA target genes

To explore the pathways and networks in which the identified miRNA's target genes may play a role, we performed Ingenuity Pathway Analysis (IPA). IPA is a knowledge database generated from peer-reviewed scientific publications that enables the discovery of highly represented biological mechanisms, pathways or functions most relevant to the genes of interest from large, quantitative datasets (http://www.ingenuity.com/products/ipa/). We uploaded list of the miRNA target genes associated with cardiometabolic phenotypes, and performed a core analysis with the default settings in IPA. We mapped these target genes to biological functions or canonical pathways. We looked at each gene separately to identify the associated pathways and biological networks. We further sought to determine whether the highlighted target genes of a miRNA that were found to be associated with cardiometabolic phenotypes are correlated together.

RESULTS

Genetic variants in miRNA seed regions

We retrieved all possible SNPs in miRNA corresponding sequences, of which a total number of 275 SNPs were located in the miRNA seed regions (**Table S1**). We included SNPs with MAF > 0.01 (n=26) and focused on the SNPs present in the HapMap project. Using the SNAP web tool, we found 2 proxy SNPs in high LD (R2 > 0.8 and distance < 200 kb) with 2 further miRNA seed variants that were not present in HapMap (**Figure 1**). In total, we examined 7 SNPs pertaining to 7 different miRNAs, including miR-146a-3p, miR-548a, miR-1178-5p, miR-1269b, miR-4513, miR-4741, and miR-6499-5p (**Table 2**).

Table 2. Characteristics of seven miRNA seed variants with MAF> 0.01 and present in the HapMap project

SNP ID	Chr.	Coded allele	Other allele	MAF	SNP proxy	miRNA ID	miRNA location
rs2910164	5	C	G	0.24	-	miR-146a-3p	Intergenic
rs3734050	5	T	C	0.098	-	miR-6499-5p	*FAT2*
rs7210937	4	C	G	0.074	-	miR-1269b	*ARHGAP44*
rs7311975	12	C	T	0.028	-	miR-1178-5p	*CIT*
rs515924	6	G	A	0.15	rs676103[a]	miR-548a	Intergenic
rs2168518	15	A	G	0.31	rs1378942[b]	miR-4513	*CSK*
rs7227168	18	T	C	0.12	rs7239066[c]	miR-4741	*RBBP8*

Shown are seven miRNA seed SNPs with minor allele frequency (MAF) > 0.01 and available in the HapMap project (release 22). For those SNPs that were not present in HapMap imputed data, we used the proxy in high linkage disequilibrium (LD): including a, R^2= 1.0, distance= 189bp (A/G); b, R^2= 1.0, distance= 928bp (A/C); c, R^2= 1.0, distance= 1351bp (A/G).

A miR-4513 seed variant associates with multiple cardiometabolic phenotypes

The genetic association analysis of 7 miRNA seed SNPs with 17 cardiometabolic traits/diseases are shown in **Table S2**. We used a Bonferroni correction to compensate for 119 tests (7×17= 119), resulting in a p-value of 4.2×10^{-4} as a threshold of study-wide significance. We found rs1378942:G>T, a proxy in full LD (R^2= 1.0) with rs2168518:C>T in the seed region of miR-4513 (**Figure S1**), to be significantly associated with multiple cardiometabolic phenotypes. Among glycemic traits, rs1378942 was significantly associated with increased levels of fasting glucose (effective allele: A, p-value= 2.5×10^{-4}, β= 1.2×10^{-2}). For lipid traits, the A allele of rs1378942 was significantly associated with higher LDL (p-value= 5.6×10^{-5}, Z.score= 4.03) and total cholesterol (p-value= 5.7×10^{-5}, Z.score= 4.02). This allele was also significantly associated with higher systolic (p-value= 3.4×10^{-10}) and diastolic blood pressure (p-value= 3.5×10^{-12}). Moreover, the minor allele A of rs1378942 showed a suggestive association with increased risk of CAD (p-value= 9.2×10^{-4}). Additionally, we generated regional association plots of the related genomic

region of this SNP for the identified traits using LocusZoom web tool (Version 1.1.) [15]. **Figure S2** illustrates the association of rs1378942 with these traits in regional association plots, showing that the SNP either has the strongest association with the trait in the given genomic region or is one of the strongest ones.

Rs2168518 affects the miR-4513 processing and expression

We observed 0.49 kcal/mol difference in the minimum free energy (MFE) of the thermodynamic ensemble of the mutant versus the wild type primary miR-4513 sequence, which may affect the processing of the primary miRNA (**Figure S3**). We cloned the pre-miR-4513 sequence (containing the wild type or mutant alleles) behind the GFP in the expression plasmid to examine the effect of rs2168518 on the level of mature miR-4513 expression. Transient transfection experiments in HEK293 cells showed a significant reduced level of miR-4513 from the mutant allele relative to GFP compared to the wild type allele (p-value= 0.0048) (**Figure 2**).

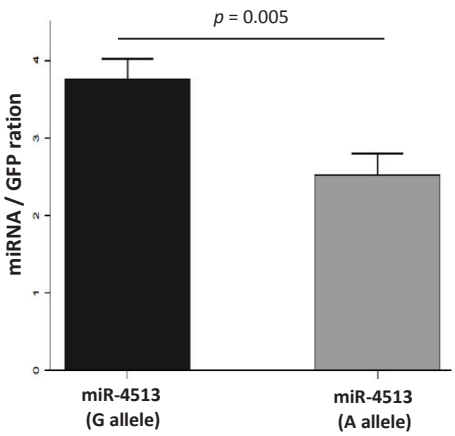


Figure 2. The effect of rs2168518 on miR-4513 expression containing the major allele G and the minor allele A. The pre-miR-4513 sequences containing either G or A alleles were cloned behind the GFP in the expression plasmid MSCV-BC. HEK293 cells were vector transfected, and total RNA was isolated for subsequent analysis by qPCR to determine the levels of mature miR-4513 in these expression vectors. This experiment showed a significant reduced level of mature miR-4513 from the minor allele A relative to GFP compared with the major allele G.

miR-4513 target genes are associated with cardiometaboenotyp

We examined the association of 109 predicted target genes of miR-4513 with the cardiometabolic traits to identify their putative mediatory roles in our findings (**Table S3**). After applying a Bonferroni correction to compensate for the multiple testing, we found five target genes to be significantly associated with the identified traits, including *PSCK1* with fasting glucose (p-value= 8.1×10^{-6}), *BNC2* with LDL (p-value= 7.6×10^{-6}) and total cholesterol (p-value= 6.6×10^{-6}) , *MTMR3* with total cholesterol (p-value= 3.6×10^{-5}), *GOSR2* with systolic blood pressure (p-value= 7.3×10^{-7}) and CAD (p-value= 1.5×10^{-6}) , and *ANK3* with systolic blood pressure (p-value= 3.9×10^{-5}) (**Figure 3**).

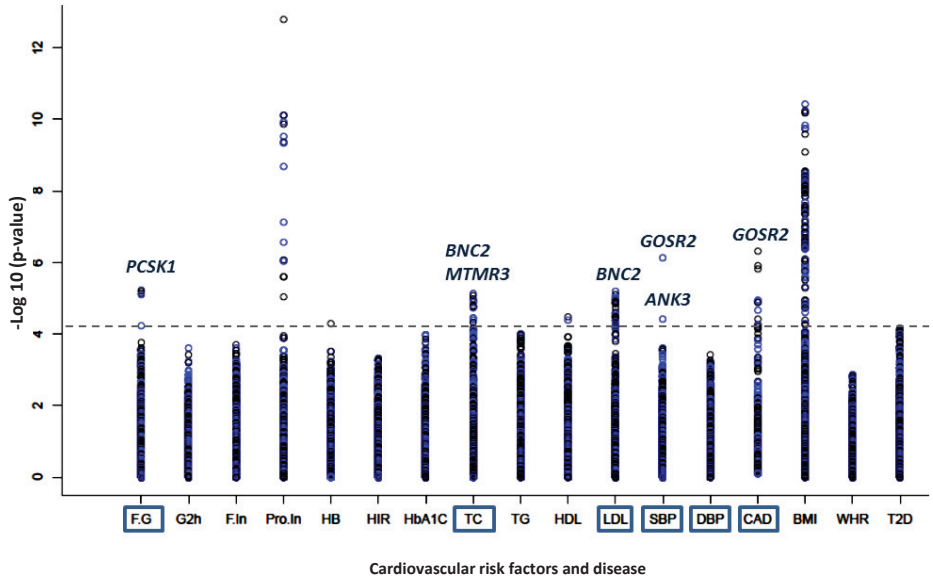


Figure 3. This figure shows the association of 2,261 SNPs in 109 predicted target genes of miR-4513 with 17 cardiovascular risk factors and disease. Dashed line indicates the significance threshold set at p-value< 2.2×10^{-5}. We highlighted the target genes which are most suspected to be influenced by the significantly associated SNPs. F.G, Fasting glucose; G2H, Glucose after 2hours; F.In, Fasting insulin; Pro.In, Pro-Insulin; HB, Homa-B; HIR, Homa-IR; HbA1c; TC, Total cholesterol; TG, Triglyceride; HDL, High-density lipoprotein; LDL, Low-density lipoprotein; SBP, Systolic blood pressure; DBP, Diastolic blood pressure; T2D, Type 2 diabetes; CAD, Coronary artery disease; BMI, Body mass index; WHR, Waist to hip ratio.

Association of rs2168518 with miR-4513 target genes expression

We examined the association of rs2168518 with expression levels of the five identified target genes of miR-4513 using blood *trans*-eQTL data in 2,336 individuals. We did not find a statistically significant difference in the expression levels of target genes *PSCK1, BNC2, MTMR3, GOSR2*, and *ANK3* across different alleles of rs2168518. However, there was a positive trend in the mean RNA-expression levels of *GOSR2* in individuals carrying the minor allele of rs2168518 (**Table S4**). Our *cis*-eQTL analysis showed a significant association between rs2168518 (G) and expression of the miR-4513's host gene *CSK* (Z.score= 16.2, p-value= 5.12×10^{-59}).

Rs2168518 affects miR-4513 controlled expression of GOSR2

Next, we investigated whether rs2168518 effects on the expression level of *GOSR2 in vitro*. We generated expression vectors with miR-4513 sequences containing either the major or the minor allele and co-transfected these constructs with Luciferase reporters containing the 3' UTR of *GOSR2*. Overexpression of miR-4513 significantly decreased the Luciferase activity of *GOSR2* 3'UTR fragment by 45% (p-value= 0.04), indicating *GOSR2*

is a direct target of miR-4513 (**Figure 4**). In addition, the rs2168518 mutant allele causes a reduced miR-4513 activity compared to the wild type miRNA, when the miRNA was overexpressed at lower levels, suggesting that the target repression efficiency, but not the specificity is changed by this SNP (**Figure 4**).

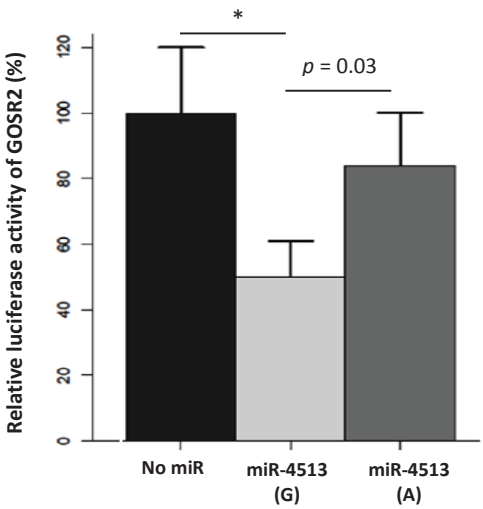


Figure 4. Luciferase reporter activity of the *GOSR2* 3'UTR in presence of miR-4513 containing the rs2168518 major allele (G) or minor allele (A). A significance differences of the mean relative luciferase activity between cells transfected with pGL3 vector coupled to the 3'UTR region of *GOSR2* with or without miR-4513 marked by * sign. The figure also shows that the rs2168518 minor allele A in miR-4513 affects the repression of *GOSR2*.

Potential role of the identified miR-4513 target genes in cardiometabolic disorders

The IPA core analysis was performed to determine the canonical pathways and networks that link the five identified miR-4513 target genes with the associated phenotypes. In agreement with our findings in the association study, there was a link between *PCSK1* and the insulin biosynthesis pathway and hyperglycemia. In addition, *MTMR3* and *BNC2* were correlated with lipid metabolism, *GOSR2* was associated with CAD and myocardial infarction and *ANK3* was linked with pulmonary and renal hypertension (**Figure S4**). We further generated interaction networks between these five target genes of miR-4513 and their associated phenotypes. **Figure S5** illustrates a potential pleiotropic effect of miR-4513 on cardiometabolic traits and diseases.

DISCUSSION

We found that rs2168518, a variant in the seed region of miR-4513, associates with fasting glucose, LDL and total cholesterol, and systolic and diastolic blood pressure. We identified five miR-4513 target genes, *GOSR2, ANK3, PCSK1*, *BNC2*, and *MTMR3*, as

potential mediators of these associations. We experimentally showed two mechanisms through which rs2168518 affects miR-4513 function. First, the rs2168518 minor allele A decreases miR-4513 expression. Second, the minor allele reduces the ability of miR-4513 to repress its target genes (*GOSR2*) expression compared to the major allele.

In recent years numerous studies have provided strong evidence showing miRNAs as major players in complex disorders [43, 44]. In addition, large advances have been made to identify miRNAs regulatory role in the pathophysiology of cardiometabolic diseases [4, 45, 46]. Since each miRNA regulates the expression of a large number of genes, genetic polymorphisms in miRNA corresponding sequences are expected to contribute to phenotypic variation and subsequently disease susceptibility [47, 48]. Previous studies have reported an appreciable level of variation at miRNA binding sites and associated some of them with complex disorders [49]. However, since genetic variation in miRNA seed regions has important phenotypic consequences, they are expected to be at a low frequency. Polymorphisms in the seed of miRNAs render a strong effect on miRNA interaction with its target genes. For instance mutation in the miR-96 seed results in a non-syndromic progressive hearing loss and variants in the seed regions of miR-146a-3p and miR-499a-3p are associated with increased risk of cancer [19, 50, 51]. Although variants on the miRNA target sites have previously linked with metabolic disorders [52], the association of miRNA seed polymorphisms with cardiometabolic phenotypes were not defined yet. Here we applied a systematic approach to investigate the association of miRNA seed SNPs with different cardiometabolic phenotypes. In agreement with previous studies, we show that the occurrence of common variants in the seed region of miRNAs is relatively rare and because of that many of the SNPs are not present in HapMap imputed data and are of negligible population genetic importance [49].

We found that the SNP rs2168518 in miR-4513 is associated with fasting glucose, LDL and total cholesterol, blood pressure, and CAD. This is the first finding concerning the role of miR-4513 in disease since its discovery by deep sequencing in 2010 [53]. We showed that the mature miR-4513 expression from the minor allele of rs2168518 is significantly reduced. The lower miR-4513 levels may be explained two possible mechanisms, which are not mutually exclusive. First, this variant could affect the expression of mature miRNA by interfering with miRNA processing efficiency and components such as the RNA-induced silencing complex (RISC) assembly and Dicer cleavage [54, 55]. Second, the stability of rs2168518 containing miR-4513 may be reduced due to aberrant RISC loading and RNA degradation mechanisms [56].

We highlighted five predicted target genes of miR-4513, *PCSK1, BNC2*, *MTMR3, ANK3* and *GOSR2*, as potential mediators of this effect on cardiometabolic phenotypes. We revealed a significant association between *PCSK1* and fasting glucose. This gene is previously associated with obesity [57], glucose metabolism, insulin secretion and risk of T2D [58]. *BNC2* is associated with HbA1c and glucose in type 1 diabetes [58, 59]. Our

results here indicate that this gene is also a regulator of cholesterol metabolism. *MTMR3* is an inositol lipid 3-phosphatase which is involved in lipid metabolism [60]. In agreement with our study, a recent large-scale meta-analysis of GWA studies of lipid traits has reported *MTMR3* to be associated with LDL cholesterol [61]. Our findings further showed an association between *ANK3* and higher systolic blood pressure. This gene has been previously highlighted to be involved in cardiac arrhythmia [62] and psychological disorders like bipolar disorders [63]. In addition, our pathway analysis using Ingenuity showed *ANK3* to be linked to pulmonary and renal hypertension. Finally, we report *GOSR2* to be associated with blood pressure and CAD by use of the GWAS data. Previous studies of other investigators have also shown it to be associated with increased hypertension [64] and pulse pressure [65]. These findings indicate that our approach is valid to identify miRNA target genes that may mediate the effect of a miRNA on the studied traits. Since each miRNA regulate a large number of target genes, miRNAs have this potential to play a pleiotropic role in biological pathways. We demonstrate the pleiotropic effect of miR-4513 on cardiometabolic traits may be through its highlighted target genes.

Gene expression patterns are highly variable across tissues. Therefore, although we did not find an association between rs2168518 and blood expression levels of the highlighted target genes in the *trans*-eQTL data, this does not rule out an effect in other tissues. Accordingly, previous studies have shown that *trans*-regulatory effects of gene expression are highly complex and with small effect size [66]. However, we identified a positive trend in the RNA-expression levels of *GOSR2* in individuals carrying the risk allele of rs2168518 in blood. Therefore, to have higher priority about the functional effect of rs2168518 on the expression of *GOSR2,* we employed the luciferase reporter assay system. We experimentally validated *GOSR2* as target genes of miR-4513 which is the first report of a validated target gene for this miRNA. We then showed that miR-4513 mediated regulation of *GOSR2* was significantly affected by SNP rs2168518 at lower concentration. This dose-dependent effect of the miRNA concentrations can be explained by the minimal concentration that is necessary for a miRNA to regulate the target gene [67]. Alternatively, this may further indicates that rs2168518 changes the expression levels of mature miR-4513 rather than impairing the targeting. We found an association between rs2168518 and the expression of its host gene *CSK* in blood. Several reports demonstrate that the expression profiles of intragenic miRNAs are highly correlated with their corresponding host genes [68-70]. Therefore, it is possible to use the miRNA host gene expression as a proxy to monitor the expression of its embedded miRNA [71]. The identified association of rs2168518 with expression levels of *CSK* may subsequently indicate an altered expression of miR-4513 in individual carrying the mutant allele.

Previous GWAS reported rs1378942, the SNP we used as a proxy for rs2168518, to be significantly associated with systolic and diastolic blood pressure and annotated that to *CSK* [72]. Our results indicate that rs1378942 may tag the altered function of miR-4513

caused by rs216518, and may resulting in up-regulation of *GOSR2*. Furthermore, *GOSR2* has been robustly associated with blood pressure traits: in our study with systolic blood pressure, and previously with hypertension and pulse pressure [64, 65]. Our findings further indicate that *GOSR2* is significantly associated with CAD. This may suggests miR-4513 as a candidate miRNA for blood pressure and CAD. Further research on miR-4513 are needed to investigate the expression levels of the miRNA in hypertensive and CAD patients.

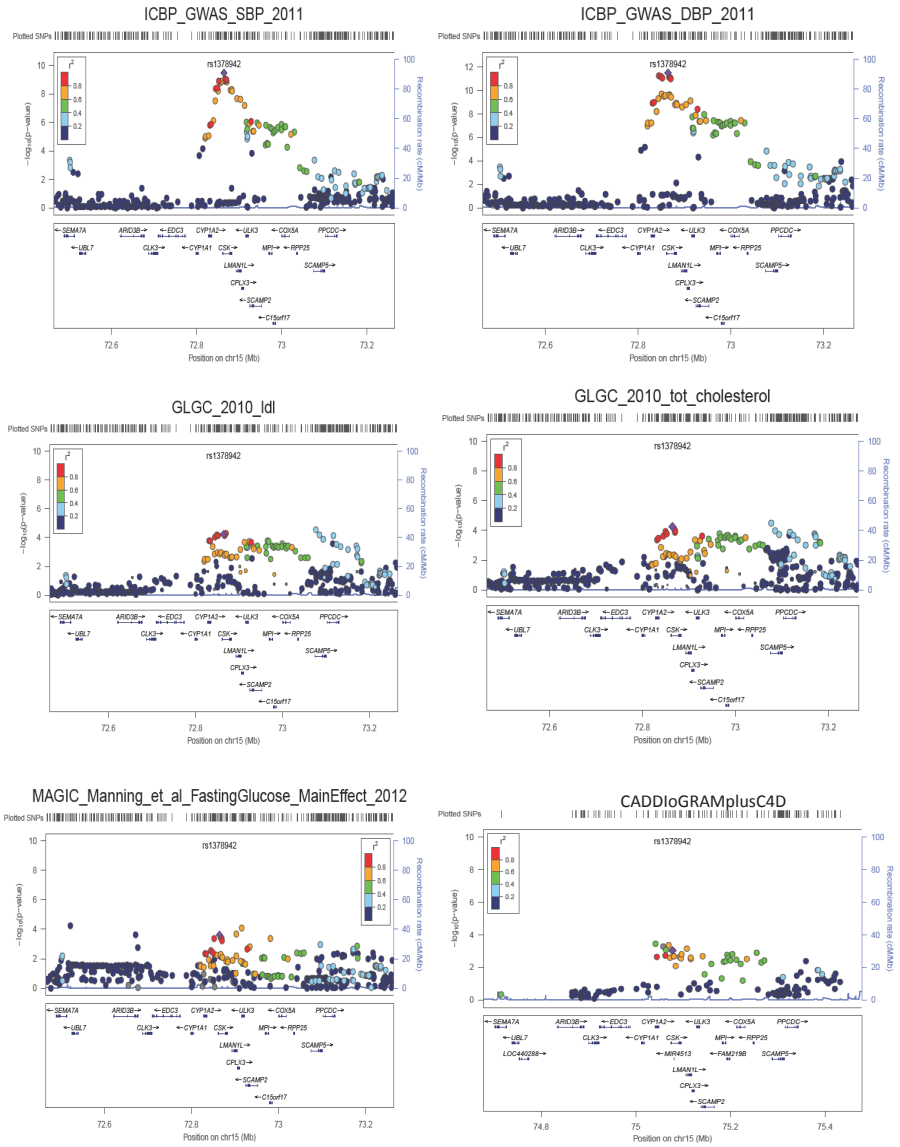
To the best of our knowledge, this is the first study to systematically investigate the association of genetic variations in the seed regions of miRNAs with cardiometabolic phenotypes. We identified a cardiometabolic-associated variant in the miR-4513 region seed that affects the miRNA expression and activity. We provide data supporting a pleiotropic role for miR-4513 in cardiometabolic traits and highlight a number of its target genes including *GOSR2* as potential mediators. This may improve our understanding of the pathophysiology of cardiometabolic disorders. Moreover, our work introduces the investigation of miRNA variants as a novel approach to study the putative role of miRNAs in complex disorders. Given the completion of the first phase of GWA studies and providing valuable information on the association of millions of SNPs with complex disorders, the time is ripe to apply this kind of approach to a wide range of traits and diseases to detect miRNA involved in complex disorders.

Supplement available online at:
http://onlinelibrary.wiley.com/doi/10.1002/humu.22706/suppinfo

Supplemental Table S2. Association of seven miRNA seed sequence SNPs with cardiometabolic disorders

Phenotype	rs2910164	rs515924	rs3734050	rs2168518 *	rs7227168 *	rs7210937	rs7311975
Proxy SNP	-	rs676103	-	rs1378942	rs7239066	-	-
Effect Allele	C	A	T	A	G	C	C
MAGIC							
Fasting glucose	2.5×10^{-2}	0.97	0.87	**2.5×10^{-4}**	0.90	0.18	0.92
Glucose after2h	0.49	0.15*	5.5×10^{-2}	0.78	0.2	0.88	NA
Fasting insulin	0.85	0.53	0.37	0.96	0.44	0.77	0.35
Pro-insulin	0.75	9.8×10^{-2}	0.5	0.84	0.31	0.62	0.62
HOMA-B	0.37	0.15	0.29	0.42	0.71	0.42	0.37
HOMA-IR	0.36	0.65	0.42	0.98	0.17	0.81	0.87
HbA1c	0.47	1.8×10^{-2}	0.83	0.94	0.98	0.8	0.31
DIAGRAM							
Type 2 Diabetes	0.52	0.51*	0.40	8.1×10^{-2}	9.1×10^{-2}	8.6×10^{-2}	NA
GLGC							
Total-Chol	0.90	0.55	0.17	**5.6×10^{-5}**	0.31	0.18	0.33
Tri Glyceride	0.82	0.86	0.41	0.74	0.78	4.5×10^{-2}	0.3
HDL- Chol	0.77	0.74	0.74	0.29	6.0×10^{-2}	0.95	0.75
LDL- Chol	0.82	0.86	0.60	**5.7×10^{-5}**	0.82	0.34	0.52
GIANT							
BMI	0.33	0.56	0.49	1.6×10^{-2}	0.19	0.43	0.46
WHR	0.36	0.93 *	0.16	0.13	0.59	0.93	NA
Global BPgen	0.24		0.94		0.43	0.96	
Systolic BP		0.51*		**3.4×10^{-10}**			NA
Diastolic BP	0.34	0.65*	0.56	**3.5×10^{-12}**	0.46	0.71	NA
CARDIoGRAMC4D							
CAD	NA	8.3×10^{-3}	NA	**9.2×10^{-4}**	NA	NA	NA

Shown are association between seven miRNA seed variants and different cardiovascular risk factors and disease. In this table, numbers are p-values that significant ones are shown in bold. Those associations that we used proxy SNP are marked by * sign.



Supplemental Figure S2. Regional plots showing the association of rs1378942 with different cardiometabolic disorders.

REFERENCES

1. He L, Hannon GJ: MicroRNAs: small RNAs with a big role in gene regulation. Nat Rev Genet 2004, 5:522-531.
2. Lee RC, Feinbaum RL, Ambros V: The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 1993, 75:843-854.
3. Kozomara A, Griffiths-Jones S: miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res 2014, 42:D68-73.
4. Friedman RC, Farh KK, Burge CB, Bartel DP: Most mammalian mRNAs are conserved targets of microRNAs. Genome Res 2009, 19:92-105.
5. Kong YW, Ferland-McCollough D, Jackson TJ, Bushell M: microRNAs in cancer management. Lancet Oncol 2012, 13:e249-258.
6. Corsten MF, Dennert R, Jochems S, Kuznetsova T, Devaux Y, Hofstra L, Wagner DR, Staessen JA, Heymans S, Schroen B: Circulating MicroRNA-208b and MicroRNA-499 reflect myocardial damage in cardiovascular disease. Circ Cardiovasc Genet 2010, 3:499-506.
7. Eulalio A, Mano M, Dal Ferro M, Zentilin L, Sinagra G, Zacchigna S, Giacca M: Functional screening identifies miRNAs inducing cardiac regeneration. Nature 2012, 492:376-381.
8. Song DW, Ryu JY, Kim JO, Kwon EJ, Kim DH: MicroRNA-19a/b family positively regulates cardiomyocyte hypertrophy by targeting atrogin-1 and MuRF-1. Biochem J 2013.
9. Stather PW, Sylvius N, Wild JB, Choke E, Sayers RD, Bown MJ: Differential MicroRNA Expression Profiles in Peripheral Arterial Disease. Circ Cardiovasc Genet 2013, 6:490-497.
10. Trajkovski M, Hausser J, Soutschek J, Bhat B, Akin A, Zavolan M, Heim MH, Stoffel M: MicroRNAs 103 and 107 regulate insulin sensitivity. Nature 2011, 474:649-653.
11. Gauthier BR, Wollheim CB: MicroRNAs: 'ribo-regulators' of glucose homeostasis. Nat Med 2006, 12:36-38.
12. Chen L, Song J, Cui J, Hou J, Zheng X, Li C, Liu L: microRNAs regulate adipocyte differentiation. Cell Biol Int 2013.
13. Kornfeld JW, Baitzel C, Konner AC, Nicholls HT, Vogt MC, Herrmanns K, Scheja L, Haumaitre C, Wolf AM, Knippschild U, et al: Obesity-induced overexpression of miR-802 impairs glucose metabolism through silencing of Hnf1b. Nature 2013, 494:111-115.
14. Bartel DP: MicroRNAs: target recognition and regulatory functions. Cell 2009, 136:215-233.
15. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ: LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 2010, 26:2336-2337.
16. Lin E, Pei D, Huang YJ, Hsieh CH, Wu LS: Gene-gene interactions among genetic variants from obesity candidate genes for nonobese and obese populations in type 2 diabetes. Genet Test Mol Biomarkers 2009, 13:485-493.
17. Petersen CP, Bordeleau ME, Pelletier J, Sharp PA: Short RNAs repress translation after initiation in mammalian cells. Mol Cell 2006, 21:533-542.
18. Richardson K, Nettleton JA, Rotllan N, Tanaka T, Smith CE, Lai CQ, Parnell LD, Lee YC, Lahti J, Lemaitre RN, et al: Gain-of-function lipoprotein lipase variant rs13702 modulates lipid traits through disruption of a microRNA-410 seed site. Am J Hum Genet 2013, 92:5-14.
19. Mencia A, Modamio-Hoybjor S, Redshaw N, Morin M, Mayo-Merino F, Olavarrieta L, Aguirre LA, del Castillo I, Steel KP, Dalmay T, et al: Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss. Nat Genet 2009, 41:609-613.

20. Barenboim M, Zoltick BJ, Guo Y, Weinberger DR: MicroSNiPer: a web tool for prediction of SNP effects on putative microRNA targets. Hum Mutat 2010, 31:1223-1232.
21. Bao L, Zhou M, Wu L, Lu L, Goldowitz D, Williams RW, Cui Y: PolymiRTS Database: linking polymorphisms in microRNA target sites with complex traits. Nucleic Acids Res 2007, 35:D51-54.
22. Hiard S, Charlier C, Coppieters W, Georges M, Baurain D: Patrocles: a database of polymorphic miRNA-mediated gene regulation in vertebrates. Nucleic Acids Res 2010, 38:D640-651.
23. Bhartiya D, Laddha SV, Mukhopadhyay A, Scaria V: miRvar: A comprehensive database for genomic variations in microRNAs. Hum Mutat 2011, 32:E2226-2245.
24. International HapMap C, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, et al: A second generation human haplotype map of over 3.1 million SNPs. Nature 2007, 449:851-861.
25. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI: SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics 2008, 24:2938-2939.
26. Manning AK, Hivert MF, Scott RA, Grimsby JL, Bouatia-Naji N, Chen H, et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. Nat Genet. 2012;44:659-669.
27. Saxena R, Hivert MF, Langenberg C, Tanaka T, Pankow JS, Vollenweider P, et al. Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. Nat Genet. 2010;42:142-148.
28. Strawbridge RJ, Dupuis J, Prokopenko I, Barker A, Ahlqvist E, Rybin D, et al. Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes. Diabetes. 2011;60:2624-2634.
29. Soranzo N, Sanna S, Wheeler E, Gieger C, Radke D, Dupuis J, et al Common variants at 10 genomic loci influence hemoglobin A(1)(C) levels via glycemic and nonglycemic pathways. Diabetes. 2010;59:3229-3239.
30. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet. 2010;42:105-116.
31. Scott RA, Lagou V, Welch RP, et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. Nat Genet 2012, 44:991-1005.
32. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Ripatti S, Chasman DI, Willer CJ, et al: Biological, clinical and population relevance of 95 loci for blood lipids. Nature 2010, 466:707-713.
33. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, Lango Allen H, Lindgren CM, Luan J, Magi R, et al: Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet 2010, 42:937-948.
34. Heid IM, Jackson AU, Randall JC, et al. Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. Nat Genet 2010, 42:949-960.
35. International Consortium for Blood Pressure Genome-Wide Association S, Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature. 2011;478:103-109.

36. Consortium CAD, Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, Thompson JR, Ingelsson E, Saleheen D, Erdmann J, et al: Large-scale association analysis identifies new risk loci for coronary artery disease. Nat Genet 2013, 45:25-33.
37. Lorenz R, Bernhart SH, Honer Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL: ViennaRNA Package 2.0. Algorithms Mol Biol 2011, 6:26.
38. Meenhuis A, van Veelen PA, de Looper H, van Boxtel N, van den Berge IJ, Sun SM, Taskesen E, Stern P, de Ru AH, van Adrichem AJ, et al: MiR-17/20/93/106 promote hematopoietic cell expansion by targeting sequestosome 1-regulated pathways in mice. Blood 2011, 118:916-925.
39. Ferrucci L, Bandinelli S, Benvenuti E, Di Iorio A, Macchi C, Harris TB, Guralnik JM: Subsystems contributing to the decline in ability to walk: bridging the gap between epidemiology and geriatric practice in the InCHIANTI study. J Am Geriatr Soc 2000, 48:1618-1625.
40. Volzke H, Alte D, Schmidt CO, Radke D, Lorbeer R, Friedrich N, Aumann N, Lau K, Piontek M, Born G, et al: Cohort profile: the study of health in Pomerania. Int J Epidemiol 2011, 40:294-307.
41. Hofman A, Darwish Murad S, van Duijn CM, Franco OH, Goedegebure A, Ikram MA, Klaver CC, Nijsten TE, Peeters RP, Stricker BH, et al: The Rotterdam Study: 2014 objectives and design update. Eur J Epidemiol 2013, 28:889-926.
42. Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, Christiansen MW, Fairfax BP, Schramm K, Powell JE, et al: Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet 2013.
43. Croce CM: Causes and consequences of microRNA dysregulation in cancer. Nat Rev Genet 2009, 10:704-714.
44. van Rooij E, Olson EN: MicroRNA therapeutics for cardiovascular disease: opportunities and obstacles. Nat Rev Drug Discov 2012, 11:860-872.
45. Fichtlscherer S, Zeiher AM, Dimmeler S: Circulating microRNAs: biomarkers or mediators of cardiovascular diseases? Arterioscler Thromb Vasc Biol 2011, 31:2383-2390.
46. Kumar M, Nath S, Prasad HK, Sharma GD, Li Y: MicroRNAs: a new ray of hope for diabetes mellitus. Protein Cell 2012, 3:726-738.
47. Gong J, Tong Y, Zhang HM, Wang K, Hu T, Shan G, Sun J, Guo AY: Genome-wide identification of SNPs in microRNA genes and the SNP effects on microRNA target binding and biogenesis. Hum Mutat 2012, 33:254-263.
48. Abelson JF, Kwan KY, O'Roak BJ, Baek DY, Stillman AA, Morgan TM, Mathews CA, Pauls DL, Rasin MR, Gunel M, et al: Sequence variants in SLITRK1 are associated with Tourette's syndrome. Science 2005, 310:317-320.
49. Saunders MA, Liang H, Li WH: Human polymorphism at microRNAs and microRNA target sites. Proc Natl Acad Sci U S A 2007, 104:3300-3305.
50. Dorn GW, 2nd, Matkovich SJ, Eschenbacher WH, Zhang Y: A human 3' miR-499 mutation alters cardiac mRNA targeting and function. Circ Res 2012, 110:958-967.
51. Wang J, Wang Q, Liu H, Shao N, Tan B, Zhang G, Wang K, Jia Y, Ma W, Wang N, Cheng Y: The association of miR-146a rs2910164 and miR-196a2 rs11614913 polymorphisms with cancer risk: a meta-analysis of 32 studies. Mutagenesis 2012, 27:779-788.
52. Zhao X, Ye Q, Xu K, Cheng J, Gao Y, Li Q, Du J, Shi H, Zhou L: Single-nucleotide polymorphisms inside microRNA target sites influence the susceptibility to type 2 diabetes. J Hum Genet 2013, 58:135-141.
53. Jima DD, Zhang J, Jacobs C, Richards KL, Dunphy CH, Choi WW, Au WY, Srivastava G, Czader MB, Rizzieri DA, et al: Deep sequencing of the small RNA transcriptome of normal and malignant human B cells identifies hundreds of novel microRNAs. Blood 2010, 116:e118-127.

54. Khvorova A, Reynolds A, Jayasena SD: Functional siRNAs and miRNAs exhibit strand bias. Cell 2003, 115:209-216.
55. Krol J, Sobczak K, Wilczynska U, Drath M, Jasinska A, Kaczynska D, Krzyzosiak WJ: Structural features of microRNA (miRNA) precursors and their relevance to miRNA biogenesis and small interfering RNA/short hairpin RNA design. J Biol Chem 2004, 279:42230-42239.
56. Kawamata T, Seitz H, Tomari Y: Structural determinants of miRNAs for RISC loading and slicer-independent unwinding. Nat Struct Mol Biol 2009, 16:953-960.
57. Heni M, Haupt A, Schafer SA, Ketterer C, Thamer C, Machicao F, Stefan N, Staiger H, Haring HU, Fritsche A: Association of obesity risk SNPs in PCSK1 with insulin sensitivity and proinsulin conversion. BMC Med Genet 2010, 11:86.
58. Paterson AD, Waggott D, Boright AP, Hosseini SM, Shen E, Sylvestre MP, Wong I, Bharaj B, Cleary PA, Lachin JM, et al: A genome-wide association study identifies a novel major locus for glycemic control in type 1 diabetes, as measured by both A1C and glucose. Diabetes 2010, 59:539-549.
59. Hertel JK, Johansson S, Raeder H, Platou CG, Midthjell K, Hveem K, Molven A, Njolstad PR: Evaluation of four novel genetic variants affecting hemoglobin A1c levels in a population-based type 2 diabetes cohort (the HUNT2 study). BMC Med Genet 2011, 12:20.
60. Walker DM, Urbe S, Dove SK, Tenza D, Raposo G, Clague MJ: Characterization of MTMR3. an inositol lipid 3-phosphatase with novel substrate specificity. Curr Biol 2001, 11:1600-1605.
61. Global Lipids Genetics C, Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, Ganna A, Chen J, Buchkovich ML, et al: Discovery and refinement of loci associated with lipid levels. Nat Genet 2013.
62. Mohler PJ, Rivolta I, Napolitano C, LeMaillet G, Lambert S, Priori SG, Bennett V: Nav1.5 E1053K mutation causing Brugada syndrome blocks binding to ankyrin-G and expression of Nav1.5 on the surface of cardiomyocytes. Proc Natl Acad Sci U S A 2004, 101:17533-17538.
63. Ferreira MA, O'Donovan MC, Meng YA, Jones IR, Ruderfer DM, Jones L, Fan J, Kirov G, Perlis RH, Green EK, et al: Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. Nat Genet 2008, 40:1056-1058.
64. International Consortium for Blood Pressure Genome-Wide Association S, Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, Chasman DI, Smith AV, Tobin MD, Verwoert GC, et al: Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature 2011, 478:103-109.
65. Wain LV, Verwoert GC, O'Reilly PF, Shi G, Johnson T, Johnson AD, Bochud M, Rice KM, Henneman P, Smith AV, et al: Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. Nat Genet 2011, 43:1005-1011.
66. Grundberg E, Small KS, Hedman AK, Nica AC, Buil A, Keildson S, Bell JT, Yang TP, Meduri E, Barrett A, et al: Mapping cis- and trans-regulatory effects across multiple tissues in twins. Nat Genet 2012, 44:1084-1089.
67. Shu J, Xia Z, Li L, Liang ET, Slipek N, Shen D, Foo J, Subramanian S, Steer CJ: Dose-dependent differential mRNA target selection and regulation by let-7a-7f and miR-17-92 cluster microRNAs. RNA Biol 2012, 9:1275-1287.
68. Baskerville S, Bartel DP: Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. RNA 2005, 11:241-247.
69. Karali M, Peluso I, Marigo V, Banfi S: Identification and characterization of microRNAs expressed in the mouse eye. Invest Ophthalmol Vis Sci 2007, 48:509-515.
70. Kim YK, Kim VN: Processing of intronic microRNAs. EMBO J 2007, 26:775-783.

71. Tsang J, Zhu J, van Oudenaarden A: MicroRNA-mediated feedback and feedforward loops are recurrent network motifs in mammals. Mol Cell 2007, 26:753-767.
72. Newton-Cheh C, Johnson T, Gateva V, Tobin MD, Bochud M, Coin L, Najjar SS, Zhao JH, Heath SC, Eyheramendy S, et al: Genome-wide association study identifies eight loci associated with blood pressure. Nat Genet 2009, 41:666-676.

CHAPTER 2.2

Association of miR-196a2 with waist to hip ratio and miR-1908 with serum lipid and glucose

Mohsen Ghanbari, Sanaz Sedaghat, Hans de Looper, Albert Hofman, Stefan J. Erkeland, Oscar H. Franco, Abbas Dehghan

Manuscript based on this chapter

The association of common polymorphisms in miR-196a2 with waist to hip ratio and miR-1908 with serum lipid and glucose.

Obesity. 2015, 23(2):495-503.

ABSTRACT

Background. MicroRNAs (miRNAs) have been implicated in the regulation of cardiometabolic disorders. Given the crucial role of miRNAs in gene expression, genetic variation within miRNA genes are expected to affect miRNA function and contribute to disease risk.

Methods and Results. We retrieved 2,320 genetic variants in miRNA-encoding sequences and investigated their associations with 17 cardiovascular risk factors and disease, using genome-wide association studies (GWAS) on glycemic indices, anthropometric measures, lipid traits, blood pressure, coronary artery disease and type2 diabetes. We then examined whether target genes of the associated miRNAs may mediate their effects in the related phenotypes. Furthermore, we performed *trans*-eQTL analysis and Luciferase reporter assays to provide functional evidence for our findings. We found that rs11614913:C>T in miR-196a2 is associated with waist to hip ratio (p-value= 1.7×10^{-5}, β= 0.023). We identified two target genes, *SFMBT1* and *HOXC8*, which may mediate the effect of miR-196a2 and experimentally showed that they are direct targets of this miRNA. Moreover, we found rs174561:T>C in miR-1908 to be associated with total-cholesterol (p-value= 6.5×10^{-16}, β= 0.044), LDL-cholesterol (p-value= 4.3×10^{-18}, β= 0.049), HDL-cholesterol (p-value= 1.7×10^{-6}, β= 0.026), triglyceride (p-value= 7.8×10^{-14}, β= 0.038) and fasting glucose (p-value= 4.3×10^{-10}, β= 0.02). We further highlighted a number of putative target genes of miR-1908 as potential mediators.

Conclusion. Our results support the miRNA-dependent regulation of fat distribution by miR-196a2 and lipid metabolism by miR-1908.

INTRODUCTION

Cardiometabolic disorders including cardiovascular disease (CVD) and type 2 diabetes mellitus (T2D) remain the leading cause of mortality and morbidity worldwide [1]. Although a number of risk factors are identified for these disorders, the quest for further risk factors and new diagnostic markers remains to be a research priority. Over the past two decades, a large number of studies have provided evidence for the association between microRNAs (miRNA) and complex disorders [2, 3]. In addition, miRNAs have been highlighted as potential diagnostic biomarkers as well as therapeutic targets in cardiometabolic disorders [3, 4]. Consequently, miRNAs have recently gained extensive attention as important players in the pathophysiology of cardiovascular disease and metabolic syndrome [5].

MiRNAs are a class of small non-coding RNAs with a pivotal role in the regulation of a broad range of biological processes [6]. Since the discovery of the first miRNA in the early 1990s, more than 1500 miRNAs have been identified in human, that they are predicted to regulate the expression of approximately 60% of all protein-coding genes [7, 8]. MiRNAs regulate gene expression by binding to the 3' untranslated regions (3'UTR) of their target messenger RNAs (mRNA). They inhibit translation of target mRNAs and to a lesser extent decay of the transcripts [9]. MiRNAs are initially transcribed as long primary-miRNA (pri-miRNA) transcripts in the nucleus which are cleaved by the enzyme Drosha to liberate precursor miRNAs (pre-miRNA) of 60-70 nt long. Following this initial processing, pre-miRNAs are transported to the cytoplasm to be further processed by the enzyme Dicer into double-stranded miRNA of approximately 21 nt in length. The mature single-stranded miRNA then integrate into the RNA-induced silencing complex (RISC) and interact with the target mRNAs [6, 9]. Genetic variation in miRNA genes, including pre- or mature miRNA sequences, are expected to affect the expression of miRNAs and their target genes and subsequently contribute to phenotypic variations [10, 11]. Previous studies have reported a few pathogenic variants within miRNA genes in cardiometabolic phenotypes [4, 12].

Here, we aimed to systematically investigate the association of genetic variants in miRNA-encoding sequences with different cardiometabolic traits and diseases. Furthermore, we examined whether target genes of the identified miRNAs harboring the variants may mediate their effects on the associated phenotypes.

MATERIALS AND METHODS

Identification of genetic variants in miRNA-encoding sequences

We screened known human miRNAs report in the miRBase database (release 21) (http://www. mirbase.org), to identify genetic variants in the miRNA corresponding sequences. Reviewing the literature and searching two online databases: PolymiRTS and miRNASNP [13, 14], we made a dataset of single nucleotide polymorphisms (SNPs) in human pre- and mature miRNAs, beyond the seed region (the nucleotides 2-8 of a mature miRNA) that we have been previously studied their associations with cardiometbolic phenotypes [15]. We included SNPs with minor allele frequency (MAF)> 0.01, since variants with smaller MAF are more likely to have a low imputation quality and in most cases we are under powered to study them. Given that previous GWAS meta-analyses were conducted on the basis of the international HapMap imputed data, we focused on the SNPs present in the HapMap project (http://www.hapmap.org/). For miRNA-SNPs that were not present in HapMap, we used the SNAP web tool to find proxy SNPs in high Linkage Disequilibrium (LD) ($R^2 > 0.8$ and distance < 200 kb) (http://www.broadinstitute.org/mpg/snap/id) (**Figure 1**).

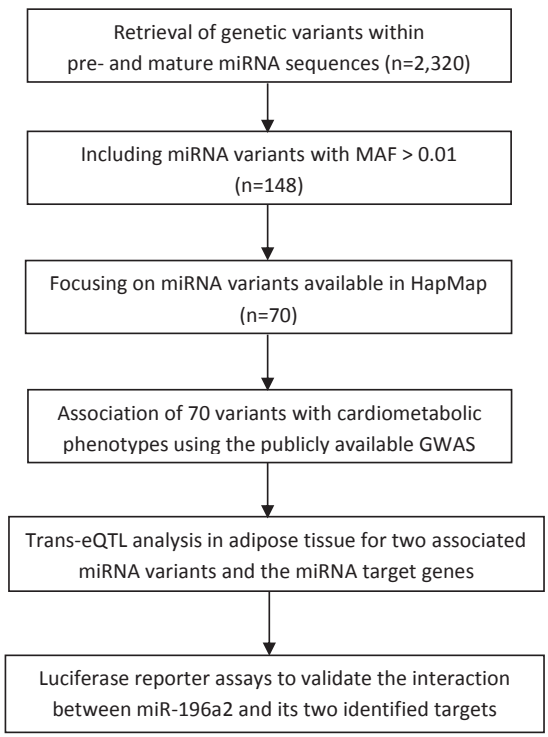


Figure 1. This figure shows our approach to examine the association between genetic variants in pre- and mature miRNA sequences and cardiometabolic phenotypes.

Association between miRNA-SNPs and cardiometabolic phenotypes

We studied the association of SNPs within the miRNA-encoding sequences with 17 cardiometabolic phenotypes using the publicly available GWAS meta-analyses on glycemic indices, anthropometric measures, lipid traits, systolic and diastolic blood pressure, coronary artery disease (CAD), and type 2 diabetes (T2D). A description of GWAS consortia and cardiometabolic phenotypes are shown in **Table S1**. To compensate for the multiple testing in the association studies, we used the Bonferroni correction, in which the number of SNPs multiplied by the number of phenotypes.

Data on glycemic traits have been contributed by MAGIC investigators and have been downloaded from www.magicinvestigators.org. We included the following traits: fasting glucose, serum glucose after 2hr, fasting insulin, fasting pro-insulin, HbA1c, HOMA-B, and HOMA-IR from up to 133,000 individuals. The Genetic Investigation of ANthropometric Traits (GIANT) consortium has performed GWAS meta-analysis on anthropometric measures including body mass index (BMI) of over 120,000 and waist to hip ratio (WHR) adjusted for BMI of 77,000 individuals. We also used the sex-stratified GWAS including 270,000 individuals released by the GIANT [15, 16]. The Global Lipid Genetics Consortium (GLGC) has provided a GWAS meta-analysis on plasma concentrations of total cholesterol, LDL-cholesterol, HDL-cholesterol and triglyceride for 187,000 individuals [17]. The Global BPgen consortium has done GWAS meta-analysis on systolic and diastolic blood pressure in over 71,000 individuals [18]. The CARDIoGRAM GWAS is a meta-analysis of 22 GWAS of European descent imputed to HapMap 2 involving 22,233 cases and 64,762 controls [19]. The DIAbetes Genetics Replication and Meta-analysis (DIAGRAMv3) consortium has done a GWAS meta-analysis in 12,171 T2D cases and 56,862 controls [20].

Effect of miRNA variants on the processing of pre-miRNAs

For those variants in miRNA corresponding sequences that were associated with cardiometabolic phenotypes, we used the Vienna RNAfold algorithm (ViennaRNA package 2.0) to predict the effect of the variants on the secondary structure and processing of the pre-miRNA sequences [21].

Association of miRNA target genes with cardiometabolic phenotypes

When a miRNA-variant was significantly associated with cardiometabolic phenotypes, we sought to identify the miRNA target genes that may mediate the effect of miRNA on the associated phenotypes. We used GWAS data on the identified phenotypes to detect target genes that are more likely to be involved in these phenotypes, in a candidate gene approach. We retrieved all genetic variants in the target genes of miRNA and assessed their associations with cardiometabolic phenotypes. The significance threshold for this analysis was set by using the Bonferroni correction based on the number of in-

dependent SNPs in the miRNA target genes. To obtain the number of independent SNPs, we used the linkage disequilibrium (LD) based SNP pruning in PLINK (http://pngu.mgh.harvard.edu/~purcell/plink/), where we excluded the SNPs with $R^2 > 0.5$. The TargetScan database (v6.2) was used to identify target genes information, including their context score and conserved sites of miRNAs, (http://www.TargetScan.org/).

Expression quantitative trait loci (eQTL)

Polymorphisms in miRNA genes are expected to alter expression level of the miRNAs and subsequently affect their target genes expression. We performed *trans*-eQTL analysis, showing the downstream consequences of variants, to examine the association between miRNA-SNPs and expression levels of the miRNA target genes. We scanned *trans*-eQTL data in adipose tissue with 856 healthy female twins using data from the MuTHER Study. The design of the MuTHER Study has been described in detail elsewhere [22].

Luciferase reporter assay

Primers were designed to amplify the pre-miR-196a2 sequence included in the restriction enzyme sites *XhoI* for the forward primer and *EcoRI* for the reverse primer. The miR-196a2 expression vector was cloned in the MSCV-BC vector. In addition, primers were designed to clone the 3'UTR sequences of *SFMBT1* and *HOXC8* included in the restriction enzyme sites XbaI for the forward primer and ApaI for the reverse. All primers are shown in **Table S2**. The 3'UTR sequences (wild-type and mutated), containing the putative binding site of miR-196a2, were cloned downstream of the Luciferase gene in the pGL3 vectors [23]. The inserts of all constructs were validated by Sanger sequencing. HEK293 cells were plated into 12-well plates and co-transfected with MSCV-miR-196a2, pGL3 containing the different 3'UTR and a plasmid expressing the Renilla transfection control. Luciferase activity was measured with the Dual-Glo Luciferase Assay System according to manufacturer's protocol. Renilla luciferase activity was used for normalization. The experiments were performed in triplicate.

RESULTS

Genetic variants within miRNA-encoding sequences

We retrieved a total of 2,320 SNPs in human pre- and mature miRNA sequences. We included SNPs with MAF > 0.01 (n=148). Of these, 68 SNPs were present in the HapMap project, using GWAS on cardiometabolic phenotypes that were conducted on the basis of HapMap imputed data. By use of the SNAP web tool, we found 2 proxy SNPs in high LD with 2 other miRNA polymorphisms that were not present in HapMap. A total of 70 SNPs within 66 different miRNAs were candidates for the association studies (**Table 1**). In

this study, we did not investigate variants in pri-miRNA transcripts, as they have different length and in many cases are comprised of more than one miRNA [24]. Thus it was challenging to disentangle the specific effect of such variants in each individual miRNA.

Table 1. Characteristics of 70 genetic variants in pre- and mature miRNA sequences

SNP ID	Chr.	Position	Effect allele	Other allele	MAF	miRNA ID	miR-SNP location
rs10406069	19	40528370	a	g	0.212	miR-5196	Pre-miR
rs10505168	8	113724928	t	c	0.292	miR-2053	Pre-miR
rs1077020	4	4023371	t	c	0.205	miR-943	Pre-miR
rs10878362	12	64703760	a	c	0.358	miR-6074	Pre-miR
rs10934682	3	125934002	t	g	0.15	miR-544b	Pre-miR
rs11014002	10	24604659	t	c	0.064	miR-603	Pre-miR
rs11032942	11	34920035	t	c	0.125	miR-1343	Pre-miR
rs11176006	12	64703749	a	g	0.312	miR-6074	Pre-miR
rs11237828	11	78810868	t	c	0.225	miR-5579	Mature-miR
rs11259096	10	14518624	t	c	0.075	miR-1265	Pre-miR
rs11614913	12	52671866	t	c	0.442	miR-196a-2	Mature-miR
rs11713052	3	126353066	c	g	0.025	miR-5092	Pre-miR
rs11907020	20	18399325	t	c	0.028	miR-3192	Pre-miR
rs12314280	12	93479716	t	c	0.085	miR-5700	Mature-miR
rs12456845	18	44830056	t	c	0.04	miR-4744	Pre-miR
rs12473206	2	64421420	c	g	0.254	miR-4433	Mature-miR
rs12523324	5	1761983	a	g	0.314	miR-4277	Pre-miR
rs12780876	10	14465210	a	t	0.288	miR-4293	Pre-miR
rs13186787	5	153706962	a	g	0.017	miR-1294	Pre-miR
rs13299349	9	18563360	a	g	0.275	miR-3152	Mature-miR
rs1683709	12	127344656	a	g	0.164	miR-3612	Pre-miR
rs17022749	12	93479734	t	c	0.068	miR-5700	Pre-miR
rs17091403	10	115923895	t	c	0.098	miR-2110	Pre-miR
rs174561	11	61815236	t	c	0.35	miR-1908	Pre-miR
rs17797090	12	102848396	a	g	0.111	miR-3652	Pre-miR
rs2042253	5	143039626	t	c	0.225	miR-5197	Pre-miR
rs2043556	10	52729412	t	c	0.186	miR-605	Pre-miR
rs2060455	15	63798684	a	g	0.181	miR-4511	Pre-miR
rs2114358	8	129090361	a	g	0.417	miR-1206	Pre-miR
rs215383	12	46044299	a	g	0.125	miR-4494	Pre-miR
rs2289030	12	93752417	c	g	0.082	miR-492	Pre-miR
rs2291418	5	179157930	a	g	0.05	miR-1229	Mature-miR
rs2292832	2	241044176	t	c	0.268	miR-149	Pre-miR
rs2368392	10	29874009	a	g	0.246	miR-604	Pre-miR

Table 1. (continued)

SNP ID	Chr.	Position	Effect allele	Other allele	MAF	miRNA ID	miR-SNP location
rs243080	2	60468076	a	g	0.458	miR-4432	Pre-miR
rs257095	5	9106945	t	c	0.119	miR-4636	Pre-miR
rs2663345	17	872514	a	g	0.322	miR-3183	Pre-miR
rs2910164	5	159844996	c	g	0.235	miR-146a	Seed-miR
rs356125	9	96612065	a	g	0.067	miR-2278	Pre-miR
rs3734050	5	150881892	t	c	0.111	miR-6499	Seed-miR
rs4112253	19	54282167	c	g	0.35	miR-4752	Pre-miR
rs4351242	6	32825700	t	c	0.075	miR-3135b	Pre-miR
rs4414449	15	84169902	a	g	0.325	miR-548ap	Mature-miR
rs4577031	15	84169963	a	t	0.325	miR-548ap	Pre-miR
rs4674470	2	220479467	t	c	0.158	miR-4268	Pre-miR
rs4822739	22	26555219	c	g	0.045	miR-548j	Pre-miR
rs4919510*	10	102724768	c	g	0.177	miR-608	Mature-miR
rs515924	11	74399308	a	g	0.177	miR-548al	Seed-miR
rs521188	1	65296107	a	g	0.025	miR-3671	Pre-miR
rs5997893	22	29886103	a	g	0.354	miR-3928	Pre-miR
rs6505162*	17	25468309	a	c	0.406	miR-423	Pre-miR
rs653667	1	12174395	t	g	0.354	miR-4632	Pre-miR
rs701213	1	233624172	t	c	0.237	miR-4427	Pre-miR
rs701214	1	233624193	t	c	0.107	miR-4427	Pre-miR
rs702742	5	154189217	a	g	0.111	miR-378h	Pre-miR
rs7207008	17	43444845	a	t	0.438	miR-2117	Pre-miR
rs7210937	17	12761371	c	g	0.095	miR-1269b	Seed-miR
rs7311975	12	119713688	t	c	0.017	miR-1178	Seed-miR
rs7522956	1	222652581	a	c	0.248	miR-4742	Pre-miR
rs7896283	10	12735183	a	g	0.5	miR-4481	Pre-miR
rs7911488	10	105144079	a	g	0.341	miR-1307	Pre-miR
rs8078913	17	6499492	t	c	0.451	miR-4520a	Mature-miR
rs832733	12	45867896	t	c	0.333	miR-4698	Pre-miR
rs8667	19	55128183	a	g	0.368	miR-4751	Pre-miR
rs877722	1	232508880	a	t	0.127	miR-4671	Pre-miR
rs895819	19	13836478	t	c	0.297	miR-27a	Pre-miR
rs936581	5	153955769	a	g	0.221	miR-3141	Pre-miR
rs9842591	3	152766381	a	c	0.434	miR-5186	Pre-miR
rs9877402	3	40528370	a	g	0.017	miR-5682	Pre-miR

Shown are 70 miRNA-variants that are present in the HapMap project (release 22). For those SNPs that were not present in the HapMap, marked them by * sign, we used their proxies in strong linkage disequilibrium (LD): including rs58078477 proxy for rs4919510 with R^2=1 and distance=0, and rs61093106 proxy for rs6505162 with R^2=1 and distance=0.

Two variants in miR-196a2 and miR-1908 were associated with cardiometabolic phenotypes

Figure 1 shows our approach to examine the association between miRNA variants and cardio-metabolic phenotypes. We performed 1,190 tests (17×70) resulting in a *p*-value threshold of 4.2×10^{-5} (Bonferroni correction). **Table S3** shows the associations between 70 miRNA-SNPs and 17 cardiometabolic phenotypes. We found that rs11614913:C>T (chr12.hg19:54385599) in miR-196a2 is associated with WHR adjusted BMI (*p*-value= 1.7×10^{-5}, β= 0.023) in up to 73,014 individuals, suggesting that individuals who carrying the rs11614913 minor allele have a bigger WHR index. This SNP was associated with WHR adjusted BMI in 32,706 men (*p*-value= 9.4×10^{-4}, β= 0.025) and in 40,308 women (*p*-value= 5.0×10^{-3}, β= 0.020). Furthermore, rs174561:T>C (chr11.gh19:61582708) in miR-1908 was significantly associated with multiple phenotypes, including total cholesterol (*p*-value= 6.5×10^{-16}, β= 0.044), LDL-cholesterol (*p*-value= 4.3×10^{-18}, β= 0.049), HDL-cholesterol (*p*-value=1.7×10^{-6}, β= 0.026), triglyceride (*p*-value= 7.8×10^{-14}, β= 0.038), fasting glucose (*p*-value= 4.35×10^{-10}, β= 0.02) and HOMA-B (*p*-value= 3.8×10^{-5}, β= -0.014). These findings indicate that carriers of the minor allele T of rs174561 have increased serum lipid and fasting glucose (**Table 2**). **Figure S1** illustrates regional association plots of rs11614913 and rs174561 and their close by SNPs with the identified phenotypes, using LocusZoom web tool, Version 1.1.

Table 2. Two miRNA variants associated with cardiometabolic phenotypes

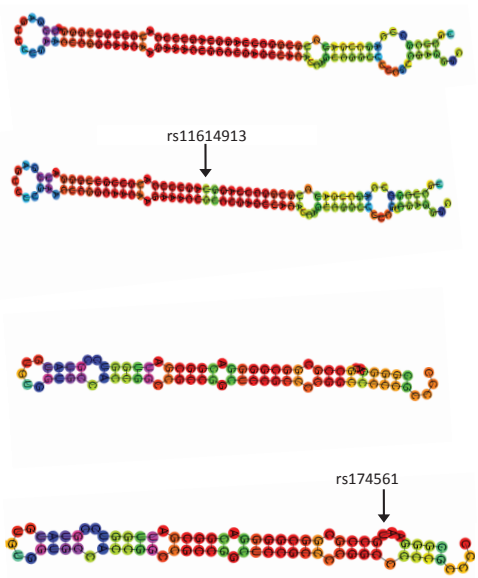
SNP ID	miRNA ID	MFE change	Associated phenotype	Effect estimate	*P*-value	Associated target genes
rs11614913 (C/T)	miR-196a2	-4.6 kcal/mol	Waist to hip ratio	0.023	1.7×10^{-5}	*SFMBT1*, *HOXC8*
rs174561 (C/T)	miR-1908	-2.8 kcal/mol	LDL-cholesterol	0.049	4.3×10^{-18}	O*TX1*
			Total cholesterol	0.044	6.5×10^{-16}	*NEUROD2*
			Triglyceride	0.038	7.8×10^{-14}	*PRRT1*, *SLC12A5*, *VPS37D*
			Fasting glucose	0.020	4.35×10^{-10}	-
			HDL-cholesterol	0.026	1.7×10^{-6}	*NEUROD2*
			HOMA-B	-0.014	3.8×10^{-5}	*EPHB3*

Shown are the associations between two miRNA-variants and cardiometabolic phenotypes. MFE; minimum free energy. Associated target genes are the miRNA putative targets that show the most significant associations with the related phenotype using GWAS data.

rs11614913 and rs174561 with potential to affect the pre-miRNA secondary structure

Genetic variants in pre-miRNA sequences may affect the folding of RNA. Using the Vienna RNAfold algorithm, we found differences in the RNAfold predicted secondary structure of

miR-196a2 and miR-1908 containing the wild type and mutant alleles, **Figure 2** [21]. As shown in this figure, there is a 4.6 kcal/mol difference in the minimum free energy (MFE) of the thermodynamic ensemble of pre-miR-196a2 containing the mutant versus the wild type. Accordingly, we observed a 2.8 kcal/mol difference between the wild type and mutant pre-miR-1908 sequence. This suggests that these variants may affect the processing of the pre-miRNAs, the RNA structures with lower MFE are thermodynamically more stable.



A. Pre-miR-196a2 with major allele C
ΔG= -52.40 kcal/mol

B. Pre-miR-196a2 with minor allele U
ΔG= -46.90 kcal/mol

A. Pre-miR-1908 with major allele C
ΔG= -49.40 kcal/mol

B. Pre-miR-1908 with minor allele U
ΔG= -46.60 kcal/mol

Figure 2. Schematic view of the RNAfold predicted secondary structures of the two identified miRNAs with their wild type and mutant alleles. **A.** RNAfold predicted secondary structure of pre-miR-196a2 and pre-miR-1908 with major alleles (wild type) and their MFE of the thermodynamic ensemble (ΔG). **B.** The secondary structure of miRNAs with the mutant alleles. The pre-miRNA structures with lower MFE are thermodynamically more stable. The position of SNPs in the pre-miRNA sequences are shown by arrow.

miR-196a2 and miR-1908 target genes associated with the identified phenotypes

MiRNAs are implicated in phenotypic variation and human disease through regulation of their target genes expression. We assessed the putative mediatory roles of the target genes of miR-196a2 on WHR and miR-1908 target genes on serum lipid and fasting glucose. A list of the conserved target genes of these two miRNAs are shown in **Table S4.** The association of SNPs in 295 putative target genes of miR-196a2 with WHR were examined. After excluding SNPs in high LD ($R^2 > 0.8$), we extracted 2,656 independent SNPs in all targets genes of miR-196a2 and using Bonferroni correction the significance threshold was set to 1.88×10^{-5}. We identified one target gene of miR-196a2, *SFMBT1*, to be significantly associated with WHR (*p*-value= 2.8×10^{-6}). Accordingly, we examined

the association of SNPs in 68 targets genes of miR-1908 with lipid traits, fasting glucose and HOMA-B. We retrieved 890 independent SNPs in all targets genes of miR-1908 and therefore a significance threshold was set to 5.6×10^{-5}. We found six target genes to be significantly associated with the identified phenotypes, including *NEUROD2* with total cholesterol (*p*-value= 3.9×10^{-7}) and HDL-cholesterol (*p*-value= 1.8×10^{-19}), O*TX1* with LDL-cholesterol (*p*-value= 2.6×10^{-7}), *PRRT1* (*p*-value= 2.3×10^{-6}), *SLC12A5* (*p*-value= 1.4×10^{-5}), and *VPS37D* (*p*-value= 2.1×10^{-5}) with triglyceride, and finally *EPHB3* with HOMA-B (*p*-value= 4.5×10^{-5}). These target genes of miR-196a2 and miR-1908 are more likely to play a role in mediating WHR and serum lipid and fasting glucose respectively (**Table 2**).

Association of rs11614913 and rs174561 with the miRNA target genes expression

We explored the effect of rs11614913 and rs174561 on the miRNA target genes expression using *trans*-eQTL data in adipose tissue, which is a relevant tissue for WHR and lipid. The SNP rs11614913 in miR-196a2 was associated with an increased expression levels of *SFMBT1* (*p*-value= 2.2×10^{-3}) and *HOXC8* (*p*-value= 1.6×10^{-3}) in individuals who carried the risk allele. For SNP rs174561 in miR-1908, we did not find significant association with the highlighted target genes. However, this SNP was significantly associated with its host gene, *FADS1*, across different genotypes (*p*-value= 1.6×10^{-5}).

miR-196a2-mediated repression of *SFMBT1* and *HOXC8*

Next, we investigated whether miR-196a2 control expression level of *SFMBT1 and HOXC8* in-vitro. We generated expression vectors containing the pre-miR-196a2 sequence and co-transfected the constructs with Luciferase reporters containing the 3' UTR of *SFMBT1* and *HOXC8* (wild-type and mutated). Overexpression of miR-196a2 significantly decreased the Luciferase activity of wild type construct reporters of *SFMBT1* (*P*-value= 0.01) and *HOXC8* (*P*-value= 4×10^{-5}), while the miRNA did not significantly regulate the mutated reporters (**Figure 3**). These data indicate that *SFMBT1* and *HOXC8* are direct targets of miR-196a2.

The LD structure evaluation of rs11614913 and rs174561

We investigated the LD structure of proxy SNPs in high LD with rs11614913 and rs174561 in the related genomic regions. Using the HaploReg and SNPinfo websites we found only 4 SNPs in strong LD with rs11614913 that are located in *HOXC5* and *HOXC-AS1*. Among them, rs11614913 has some predicted functions and is located within a miRNA, thus is more likely to be a functional SNP in this locus (**Table S5**). In addition, we found 23 proxies for rs174561 that are in *FADS1, C11orf9* and *C11orf10*. As shown in **Table S5**, there are predicted functions for rs174561 and some of these proxies that can be functional SNPs and may contribute to explain the observed GWAS associations.

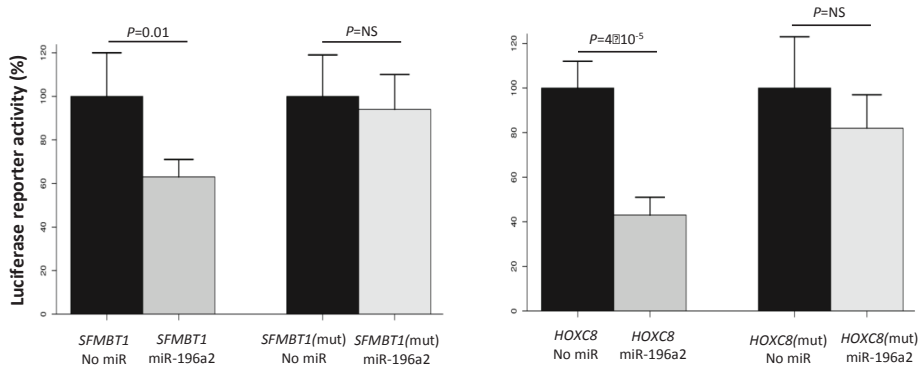


Figure 3. miR-196a2-dependent regulation of *SFMBT1* and *HOXC8*. This figure illustrates Luciferase reporter activity of cells transfected with pGL3 vector coupled to the miR-196a2 and the 3'UTR regions of *SFMBT1* and *HOXC8*. A significant difference of the mean relative luciferase activity between cells transfected with pGL3 vector coupled to 3'UTR region of these genes with or without miR-196a2 was observed. The figure further shows that the miR-196a2-mediated regulation of the genes are abolished when the reporter constructs containing mutation.

DISCUSSION

We found that rs11614913, a common variant in miR-196a2 associates with WHR and rs174561 in miR-1908 associates with LDL-cholesterol, HDL-cholesterol, total cholesterol, triglyceride and fasting glucose. We identified that these variants may affect processing of the pre-miRNAs to the mature forms. Furthermore, we identified a number of target genes of these two miRNAs that may mediate their effect on the associated phenotypes. We performed *trans*-eQTL analysis in adipose tissue and found positive association between rs11614913 in miR-196a2 with expression levels of its two target genes, *SFMBT1* and *HOXC8*. We then demonstrated that these genes are direct targets of miR-196a2 *in vitro*.

In recent years large advances have been made in the discovering of the regulatory role of miRNAs in complex disorders [2, 9]. MiRNAs comprise a key regulatory layer that intersect with transcriptional and translational control mechanisms to maintain the metabolic homeostasis. Deregulation of miRNA activities is implicated in cardiovascular disease [3-5]. Accordingly, numerous studies have provided strong evidences for a crucial role of miRNAs in fat cell development and obesity [25]. Recently, genome-wide computational analyses and experimental studies have suggested that miRNA-SNPs contribute to phenotypic variations [11, 13, 26]. Variants in pre-miRNAs are expected to alter miRNA processing and expression [10]. This may affect the expression of miRNA target genes and subsequently play a role in disease [27]. In this study, we applied a systematic approach to identify miRNAs involved in cardiometabolic disorders by investigating the correlation of miRNA polymorphisms with phenotypes. We examined the association

of miRNA variants with cardiometabolic phenotypes using the available GWAS data to find the associations deserve sufficient power. Our approach is statistically powerful since it benefits from extremely large GWA studies together with a liberal significance threshold. By implementing this approach, we found two miRNA variants, rs11614913 and rs174561, to be significantly associated with cardiometabolic phenotypes.

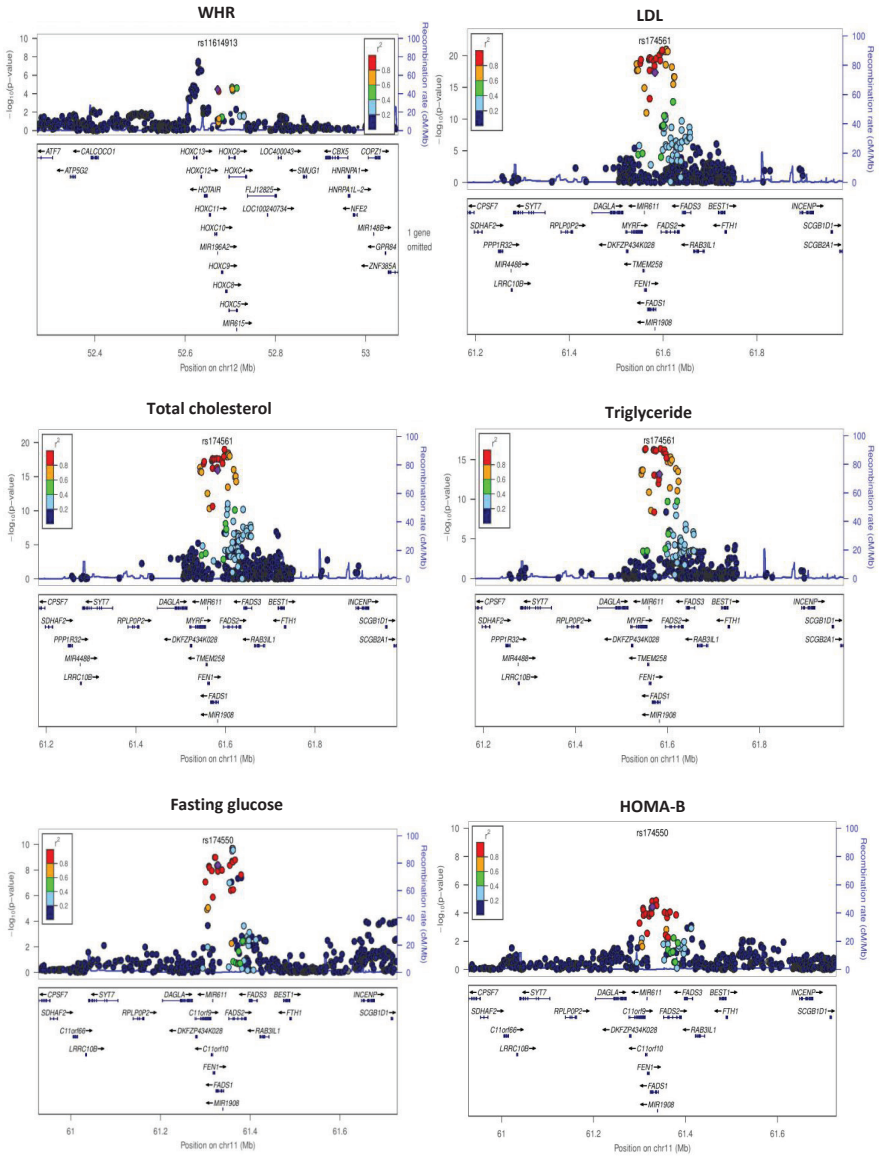
Rs11614913 is a common variant located in pre-miR-196a2 that has linked to several disorders and phenotypes. This variant contributes to the increased risk of breast cancer, lung cancer and cancers of digestive system [28, 29]. A meta-analysis of GWAS on bone mineral density (BMD) demonstrates that the risk allele of rs11614913 is inversely associated with lumbar spine and femoral neck BMD [30]. Furthermore, rs11614913 has been reported to be associated with the risk of cardiovascular disease in type 2 diabetes patients [31] and congenital heart disease [32]. However, despite using a large sample size (22,233 cases and 64,762 controls) from the CARDIOGRAM consortium, we did not find a significant association between rs11614913 and risk of CAD [19]. Our findings showed a significant association between the rs11614913 mutant allele and higher WHR. We found that there is a difference between MFE of the thermodynamic ensemble of pre-miR-196a2 with the mutant versus the wild type allele which may affect the structure and processing of the pre-miRNA. In agreement with our conjecture, it has been shown experimentally that rs11614913 affects the processing of pre-miR-196a2 and results in altered expression levels of the mature miRNA [27, 33]. Hoffman et al by delivering expression vectors containing either wild type or mutant precursors of miR-196a2 have shown that mature miRNA levels in cells transfected with pre-miR-196a2 hosting the mutant allele (T) is significantly lower than cells transfected with wild the type allele construct [27]. In addition, Hu el al have demonstrated that rs11614913 wild type allele (C) is associated with a significant increase in mature miR-196a2 expression [33]. Therefore, an altered expression level of mature miR-196a2 due to rs11614913 may serve as a functional mechanism underlying the observed association between the variant and WHR. MiRNAs regulate phenotypes through regulation of their target genes expression. We sought to identify target genes that may mediate the effect miR-196a2 on WHR and found *SFMBT1* and *HOXC8*. We showed these genes are bona fide targets of miR-196a2. Previous studies have provided evidence for co-expression of miR-196a2 and its two highlighted targets in adipose tissue which is a pre-requisite for miRNA-mRNA interaction. MiR-196a2 has been identified as a regulator in brown adipogenesis of white fat progenitor cells [22]. Accordingly, *HOXC8* is known to increase white fat cells and the risk of obesity [34]. *SFMBT1* has reported to be associated with circulating adiponectin levels [35]. In addition, it has experimentally shown that decreasing *HOXC8* expression by overexpression of miR-196a2 lead to an increased brown adipocytes [34]. Taken together, these findings suggest that depletion of miR-196a2 by rs11614913

mutant allele T elevate *HOXC8* and *SFMBT1* expression and subsequently contribute to higher WHR, potentially through an increase in white fat cells.

Rs174561 resides in the 5' end of pre-miR-1908 and is strongly associated with fatty acid levels in plasma, serum phospholipid and adipose tissue [36]. This variant has a number of proxies in *FADS1* that have been shown to be associated with lipid metabolism, glucose homeostasis and cardiometabolic outcomes [37]. However, the determination of the causal variant(s) in this locus remains a challenge. Here we suggest a potential functional mechanism for rs174561 to influence serum fatty acid levels through affecting miR-1908 expression. We found that rs174561 change MFE of the thermodynamic ensemble of pre-miR-1908 sequence, which potentially affects the processing of the pre-miRNA. MiR-1908 is highly expressed in mature human adipocytes and is highlighted as an important mediator in the development of obesity-related complications and lipid metabolism [38, 39]. Our findings may indicate that the reduced levels of miR-1908 due to rs174561 and subsequent change in target genes expression is a functional mechanism behind the identified associations. We highlighted six target genes of miR-1908 as potential mediators on these associations. Among them, *NEUROD2* has reported to be associated with HDL [17] and *SLC12A5* has linked to body mass index and type 2 diabetes [40]. This is the first report about the associations of *VPS37D,* O*TX1* and *PRRT1* with lipid traits and *EPHB3* with fasting glucose. We did not find a significant association between rs174561 and transcript expression levels of these genes in the *trans*-eQTL analysis. Lack of significant associations can be due to a limited statistical power. Furthermore, previous studies have shown that trans-regulatory effects of gene expression are highly complex and with small effects [22]. Therefore, we cannot rule out the effect of rs174561 on the target genes expression. A second mechanism may explain the correlation of rs174561 with lipid and fasting glucose concentrations through *FADS1*. This gene is known to be involved in cholesterol homeostasis [36] and we found variations in *FASD1* transcript levels through different genotypes of rs174561 in adipose tissue.

We found two common functional variants in miR-196a2 and miR-1908 to be significantly associated with WHR and serum lipid and glucose levels respectively. We suggested a number of target genes of these miRNAs as potential mediators of their effects on the associated phenotypes. These findings may improve our understanding of the miRNA-mediated regulations of fat distribution and lipid metabolism. MiR-196a2 may have potential for further translation into the clinic to combat obesity. Furthermore, we successfully applied a new approach to identify miRNAs involved in complex disorders by investigating the association between miRNA-SNPs and selected phenotypes. Given that the first phase of GWA studies are completed, this approach can be applied to a wide range of phenotypes and diseases to detect candidate functional SNPs in miRNAs and may contribute to improving the annotation of GWAS findings.

Supplement available online at:
http://onlinelibrary.wiley.com/doi/10.1002/oby.20975/full



Supplemental Figure S1. Regional plots showing the association of rs11614913 in miR-196a2 with WHR and rs174561 (or its proxy SNP rs174550) in miR-1908 with lipid traits, fasting glucose and HOMA-B.

REFERENCES

1. Isomaa B, Almgren P, Tuomi T, Forsen B, Lahti K, Nissen M, Taskinen MR, Groop L: Cardiovascular morbidity and mortality associated with the metabolic syndrome. Diabetes Care 2001, 24:683-689.
2. Kong YW, Ferland-McCollough D, Jackson TJ, Bushell M: microRNAs in cancer management. Lancet Oncol 2012, 13:e249-258.
3. Small EM, Olson EN: Pervasive roles of microRNAs in cardiovascular biology. Nature 2011, 469:336-342.
4. Richardson K, Nettleton JA, Rotllan N, Tanaka T, Smith CE, Lai CQ, Parnell LD, Lee YC, Lahti J, Lemaitre RN, et al: Gain-of-function lipoprotein lipase variant rs13702 modulates lipid traits through disruption of a microRNA-410 seed site. Am J Hum Genet 2013, 92:5-14.
5. Zampetaki A, Mayr M: MicroRNAs in vascular and metabolic disease. Circ Res 2012, 110:508-522.
6. Bartel DP: MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 2004, 116:281-297.
7. Lee RC, Feinbaum RL, Ambros V: The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 1993, 75:843-854.
8. Petersen CP, Bordeleau ME, Pelletier J, Sharp PA: Short RNAs repress translation after initiation in mammalian cells. Mol Cell 2006, 21:533-542.
9. Bartel DP: MicroRNAs: target recognition and regulatory functions. Cell 2009, 136:215-233.
10. Wu M, Jolicoeur N, Li Z, Zhang L, Fortin Y, L'Abbe D, Yu Z, Shen SH: Genetic variations of microRNAs in human cancer and their effects on the expression of miRNAs. Carcinogenesis 2008, 29:1710-1716.
11. Ryan BM, Robles AI, Harris CC: Genetic variation in microRNA networks: the implications for cancer research. Nat Rev Cancer 2010, 10:389-402.
12. Ghanbari M, de Vries PS, de Looper H, Peters MJ, Schurmann C, Yaghootkar H, Dorr M, Frayling TM, Uitterlinden AG, Hofman A, et al: A Genetic Variant in the Seed Region of Mir-4513 Shows Pleiotropic Effects on Lipid and Glucose Homeostasis, Blood Pressure and Coronary Artery Disease. Hum Mutat 2014.
13. Bao L, Zhou M, Wu L, Lu L, Goldowitz D, Williams RW, Cui Y: PolymiRTS Database: linking polymorphisms in microRNA target sites with complex traits. Nucleic Acids Res 2007, 35:D51-54.
14. Gong J, Tong Y, Zhang HM, Wang K, Hu T, Shan G, Sun J, Guo AY: Genome-wide identification of SNPs in microRNA genes and the SNP effects on microRNA target binding and biogenesis. Hum Mutat 2012, 33:254-263.
15. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Ripatti S, Chasman DI, Willer CJ, et al: Biological, clinical and population relevance of 95 loci for blood lipids. Nature 2010, 466:707-713.
16. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, Lango Allen H, Lindgren CM, Luan J, Magi R, et al: Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet 2010, 42:937-948.
17. Global Lipids Genetics C, Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, Ganna A, Chen J, Buchkovich ML, et al: Discovery and refinement of loci associated with lipid levels. Nat Genet 2013, 45:1274-1283.
18. International Consortium for Blood Pressure Genome-Wide Association S, Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, Chasman DI, Smith AV, Tobin MD, Verwoert GC, et al: Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature 2011, 478:103-109.

19. Randall JC, Winkler TW, Kutalik Z, Berndt SI, Jackson AU, Monda KL, Kilpelainen TO, Esko T, Magi R, Li S, et al: Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. PLoS Genet 2013, 9:e1003500.
20. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, Wheeler E, Glazer NL, Bouatia-Naji N, Gloyn AL, et al: New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet 2010, 42:105-116.
21. Lorenz R, Bernhart SH, Honer Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL: ViennaRNA Package 2.0. Algorithms Mol Biol 2011, 6:26.
22. Consortium CAD, Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, Thompson JR, Ingelsson E, Saleheen D, Erdmann J, et al: Large-scale association analysis identifies new risk loci for coronary artery disease. Nat Genet 2013, 45:25-33.
23. Meenhuis A, van Veelen PA, de Looper H, van Boxtel N, van den Berge IJ, Sun SM, Taskesen E, Stern P, de Ru AH, van Adrichem AJ, et al: MiR-17/20/93/106 promote hematopoietic cell expansion by targeting sequestosome 1-regulated pathways in mice. Blood 2011, 118:916-925.
24. Altuvia Y, Landgraf P, Lithwick G, Elefant N, Pfeffer S, Aravin A, Brownstein MJ, Tuschl T, Margalit H: Clustering and conservation patterns of human microRNAs. Nucleic Acids Res 2005, 33:2697-2706.
25. Hilton C, Neville MJ, Karpe F: MicroRNAs in adipose tissue: their role in adipogenesis and obesity. Int J Obes (Lond) 2013, 37:325-332.
26. Georges M, Clop A, Marcq F, Takeda H, Pirottin D, Hiard S, Tordoir X, Caiment F, Meish F, Bibe B, et al: Polymorphic microRNA-target interactions: a novel source of phenotypic variation. Cold Spring Harb Symp Quant Biol 2006, 71:343-350.
27. Hoffman AE, Zheng T, Yi C, Leaderer D, Weidhaas J, Slack F, Zhang Y, Paranjape T, Zhu Y: microRNA miR-196a-2 and breast cancer: a genetic and epigenetic association study and functional analysis. Cancer Res 2009, 69:5970-5977.
28. Hu Z, Liang J, Wang Z, Tian T, Zhou X, Chen J, Miao R, Wang Y, Wang X, Shen H: Common genetic variants in pre-microRNAs were associated with increased risk of breast cancer in Chinese women. Hum Mutat 2009, 30:79-84.
29. Tian T, Shu Y, Chen J, Hu Z, Xu L, Jin G, Liang J, Liu P, Zhou X, Miao R, et al: A functional genetic variant in microRNA-196a2 is associated with increased susceptibility of lung cancer in Chinese. Cancer Epidemiol Biomarkers Prev 2009, 18:1183-1187.
30. Estrada K, Styrkarsdottir U, Evangelou E, Hsu YH, Duncan EL, Ntzani EE, Oei L, Albagha OM, Amin N, Kemp JP, et al: Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. Nat Genet 2012, 44:491-501.
31. Buraczynska M, Zukowski P, Wacinski P, Ksiazek K, Zaluska W: Polymorphism in microRNA-196a2 contributes to the risk of cardiovascular disease in type 2 diabetes patients. J Diabetes Complications 2014, 28:617-620.
32. Xu J, Hu Z, Xu Z, Gu H, Yi L, Cao H, Chen J, Tian T, Liang J, Lin Y, et al: Functional variant in microRNA-196a2 contributes to the susceptibility of congenital heart disease in a Chinese population. Hum Mutat 2009, 30:1231-1236.
33. Hu Z, Chen J, Tian T, Zhou X, Gu H, Xu L, Zeng Y, Miao R, Jin G, Ma H, et al: Genetic variants of miRNA sequences and non-small cell lung cancer survival. J Clin Invest 2008, 118:2600-2608.
34. Mori M, Nakagami H, Rodriguez-Araujo G, Nimura K, Kaneda Y: Essential role for miR-196a in brown adipogenesis of white fat progenitor cells. PLoS Biol 2012, 10:e1001314.

35. Dastani Z, Hivert MF, Timpson N, Perry JR, Yuan X, Scott RA, Henneman P, Heid IM, Kizer JR, Lyytikainen LP, et al: Novel loci for adiponectin levels and their influence on type 2 diabetes and metabolic traits: a multi-ethnic meta-analysis of 45,891 individuals. PLoS Genet 2012, 8:e1002607.
36. Lattka E, Illig T, Koletzko B, Heinrich J: Genetic variants of the FADS1 FADS2 gene cluster as related to essential fatty acid metabolism. Curr Opin Lipidol 2010, 21:64-69.
37. Malerba G, Schaeffer L, Xumerle L, Klopp N, Trabetti E, Biscuola M, Cavallari U, Galavotti R, Martinelli N, Guarini P, et al: SNPs of the FADS gene cluster are associated with polyunsaturated fatty acids in a cohort of patients with cardiovascular disease. Lipids 2008, 43:289-299.
38. Jiang X, Yang L, Pang L, Chen L, Guo X, Ji C, Shi C, Ni Y: Expression of obesityrelated miR1908 in human adipocytes is regulated by adipokines, free fatty acids and hormones. Mol Med Rep 2014, 10:1164-1169.
39. Pencheva N, Tran H, Buss C, Huh D, Drobnjak M, Busam K, Tavazoie SF: Convergent multi-miRNA targeting of ApoE drives LRP1/LRP8-dependent melanoma metastasis and angiogenesis. Cell 2012, 151:1068-1082.
40. Lewis JP, Palmer ND, Ellington JB, Divers J, Ng MC, Lu L, Langefeld CD, Freedman BI, Bowden DW: Analysis of candidate genes on chromosome 20q12-13.1 reveals evidence for BMI mediated association of PREX1 with type 2 diabetes in European Americans. Genomics 2010, 96:211-219.

CHAPTER 2.3

Genetic variants in miRNA-binding sites affect miRNA-mediated regulation of cardiovascular genes

Mohsen Ghanbari, Oscar H. Franco, Hans W.J. de Looper,
Albert Hofman, Stefan J. Erkeland, Abbas Dehghan

Manuscript based on this chapter

Genetic variations in microRNA binding sites affect miRNA-mediated regulation of several genes associated with cardiometabolic disorders.

Circulation Cardiovascular Genetics. 2015, 8(3):473-86.

ABSTRACT

Background. Genome-wide association studies (GWAS) enabled us to discover a large number of variants at multiple genomic loci contributing to cardiovascular and metabolic disorders. Since the majority of the identified variants are thought to merely be proxies for other functional variants, the causal variants remain to be elucidated. We hypothesized that part of the functional variants involved in deregulating cardiometabolic genes are located in microRNA (miRNA) binding sites within gene 3'UTRs.

Methods and Results. Using data from the available GWAS on glycemic indices, lipid traits, anthropometric measures, blood pressure, coronary artery diseases and type 2 diabetes, we found 11,353 variants that are associated with different cardiometabolic phenotypes. Of these, 191 variants (at 129 genomic loci) were located in putative miRNA binding sites. Thirty-four out of 191 variants were found to fulfill our predefined criteria for being potentially functional in their loci. Ten variants were subsequently selected for experimental validation based on GWAS results, eQTL analyses and evidence for co-expression of their host genes and regulatory miRNAs. Luciferase reporter assays showed allele-specific regulations of mRNAs hosting the binding site variants by miRNAs. These co-transfection experiments revealed that rs174545 (*FADS1*:miR-181a-2), rs1059611 (*LPL*:miR-136), rs13702 (*LPL*:miR-410), rs1046875 (*FN3KRP*:miR-34a), rs7956 (*MKRN2*:miR-154), rs3217992 (*CDKN2B*:miR-138-2-3p) and rs11735092 (*HSD17B13*:miR-375) abrogate miRNA-mediated regulation of the genes. Conversely, two variants, rs6857 (*PVRL2*:miR-320e) and rs907091 (*IKZF3*:miR-326), were shown to enhance the interaction between miRNAs and their host genes.

Conclusion. This study provides evidence for a model in which polymorphisms in miRNA binding sites can both positively and negatively affect miRNA-mediated regulation of cardiometabolic genes.

INTRODUCTION

Cardiometabolic disorders including cardiovascular disease (CVD) and a number of metabolic abnormalities that increase CVD risk are major worldwide health problem [1]. Over the past few decades, remarkable progress has been made in identifying the genes that are responsible for cardiometabolic disorders. The recent development of novel techniques, including genome-wide association studies (GWAS) has led to the identification of a large number of single-nucleotide polymorphisms (SNPs) and specific genomic loci that contribute to the induction of different cardiometabolic traits and diseases [2-4]. A large proportion of the identified SNPs associated with cardiometabolic disorders are located in non-coding regions of the genome that are responsible for fine-tuning gene expression regulation [2, 3, 5, 6]. Many studies thus far have focused on annotating GWAS hits and functional variants that are located in the 3'-untranslated regions (3'UTR) [6, 7]. Increasing evidence suggests that the genes 3'UTRs contain regulatory elements that play important roles in gene expression [7, 8]. Genetic variations in these regions may affect gene expression through different mechanisms. This includes altering recognition sites for microRNAs (miRNA), RNA-binding proteins, and the polyadenylation machinery [8, 9]. MiRNAs play an important role in modulating the susceptibility to and onset of complex disorders, therefore, research has focused on identifying polymorphisms in miRNA binding sites [10, 11].

MiRNAs are small non-coding RNAs, consisting of 20-24 nucleotides, that regulate numerous biological processes by post-transcriptional regulation of gene expression [12]. MiRNAs can repress translation and, to a lesser extent, decrease the stability of mRNAs by interacting with the complementary sequences in the 3'UTR of target genes [12]. The seed sequence, which has been defined as nucleotides 2-7/8 of a miRNA, is essential for target recognition and regulation [12]. Genetic variation in either a miRNA's seed sequence or the complementary sequence in the 3'UTR of target gene could adversely affect target recognition, resulting in aberrant gene expression [10, 13]. Both computational and experimental studies have revealed that variants in miRNA binding sites can modulate the risk for disease induction by disrupting the original recognition site or creating an alternative binding site for miRNAs [14-17]. MiRNAs have been shown to regulate the expression of genes that are involved in induction of cardiovascular disease and regulation of various metabolic processes, such as insulin production, glucose homeostasis, lipid metabolism and obesity [18, 19]. However, thus far, only a few pathogenic miRNA-related variants have been identified that affect the regulation of genes associated with cardiometabolic phenotypes [10, 11, 13].

In the present study, we hypothesized that genetic variants in miRNA binding sites could constitute part of the functional SNPs involved in deregulating cardiometabolic genes. To test this hypothesis, we have investigated whether variants that were previ-

ously identified by GWAS of cardio-metabolic phenotypes reside in miRNA binding sites. In addition, we linked several publicly available datasets and tools (such as GWAS findings, eQTL data, miRNA and gene expression profiles) and combined them with experimental studies to identify functional variants that can alter miRNA-mediated regulation of their genes involved in cardiometabolic disorders.

MATERIALS AND METHODS

Retrieval of cardiometabolic-associated variants

We used data from the recent GWAS meta-analyses on 17 cardiometabolic traits and diseases to identify genetic variants that are associated with cardiometabolic phenotypes. To increase our statistical power and robustness, we used the largest GWAS available on these phenotypes. The GWAS summary statistics of glycemic homeostasis indices (fasting glucose, serum glucose after 2hr, fasting insulin, fasting pro-insulin, HbA1c, HOMA-B, and HOMA-IR) from the Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC) [20-24], lipid phenotypes (plasma concentrations of total cholesterol, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol and triglyceride) from the Global Lipid Genetics Consortium (GLGC) [2], anthropometric measures (body mass index and waist to hip ratio) from the Genetic Investigation of ANthropometric Traits (GIANT) consortium [25, 26], systolic and diastolic blood pressure from the Global Blood Pressure Genetics (Global BPgen) consortium [27], coronary artery disease from the Coronary ARtery DIsease Genome wide Replication and Meta-analysis plus The Coronary Artery Disease Genetics (CARDIoGRAMplusC4D) consortium [4], and type 2 diabetes from the DIAGRAM consortium [3] were analyzed. A description of the GWAS meta-analyses on these cardiometabolic phenotypes, the consortia responsible for these analyses and the number of SNPs associated with each trait or disease is provided in **Table 1**. In total 11,067 unique SNPs (in 629 genomic loci) were found to be associated with 17 cardiometabolic phenotypes at a genome-wide statistical significance level (p-value < 5×10^{-8}). We calculated the number of independent SNPs using Linkage disequilibrium based SNP pruning in PLINK tool-set with $R^2 > 0.8$ (http://pngu.mgh.harvard.edu/~purcell /plink/). On this basis, 2,369 of these SNPs were independent.

Genetic polymorphisms in miRNA binding sites

Multiple target prediction algorithms including the TargetScan (http://TargetScan.org/), miRTarBase (http://mirtarbase.mbc.nctu.edu.tw) and miRecords (http://miRecords.umn.edu/miRecords) data bases were used to identify genes that are predicted to be targeted by human miRNAs. We subsequently determined all variants were located in miRNA binding sites in the 3'UTR of these genes using the following online databases:

Table 1. Description of GWAS meta-analyses of 17 cardiovascular risk factors and disease used in this study

Phenotype	Consortium	Sample size	All GWAS associated SNPs	miR binding site associated SNPs (loci)
Glycemic indices				
Fasting glucose	MAGIC	133,010	505	10 (6)
Glucose after2h	MAGIC	42,854	4	-
Fasting insulin	MAGIC	108,557	34	-
Pro-insulin	MAGIC	10,701	407	9 (7)
HbA1c	MAGIC	46,368	188	6 (4)
HOMA-B	MAGIC	46,186	119	-
HOMA-IR	MAGIC	46,186	-	-
Type 2 diabetes	DIAGRAM	12,171 cases/56,862 controls	236	7 (4)
Lipid traits				
Total cholesterol	GLGC	187,000	4,146	88 (58)
Triglyceride	GLGC	178,000	3,242	60 (39)
HDL-cholesterol	GLGC	187,000	3,523	69 (47)
LDL-cholesterol	GLGC	173,000	3,058	70 (43)
Anthropometric traits				
Body mass index	GIANT	339,224	620	19 (18)
Waist to hip ratio	GIANT	224,459	725	12 (7)
Cardiovascular disease				
Systolic blood pressure	Global BPGen	71,225	111	2 (2)
Diastolic blood pressure	Global BPGen	71,225	88	1 (1)
Coronary artery disease	CARDIoGRM plusC4D	63,746 cases/ 130,681controls	369	13 (8)

PolymiRTS [16], Patrocles [17] and miRdSNP [28]. By adding the data from these databases together, we made a comprehensive dataset of all variants located in miRNA binding sites. Since GWAS meta-analyses on cardiometabolic phenotypes have been performed using the HapMap imputed data, we focused on those variants that were present in the list of 2.5 million HapMap SNPs, release 22 (n=23,650). No copy number variants (CNVs) was studied as the HapMap project does not include CNVs. The approach we used to determine which variants were identified as cardiometabolic-associated SNPs and were localized in miRNA binding sites is visualized in **Figure 1**.

Expression quantitative trait loci (eQTLs) survey

Functional polymorphisms in miRNA binding sites are expected to alter the expression levels of their host genes by affecting the activity of miRNAs on the target transcripts. We examined cis-eQTL meta-analysis data that was generated from 8,086 peripheral blood

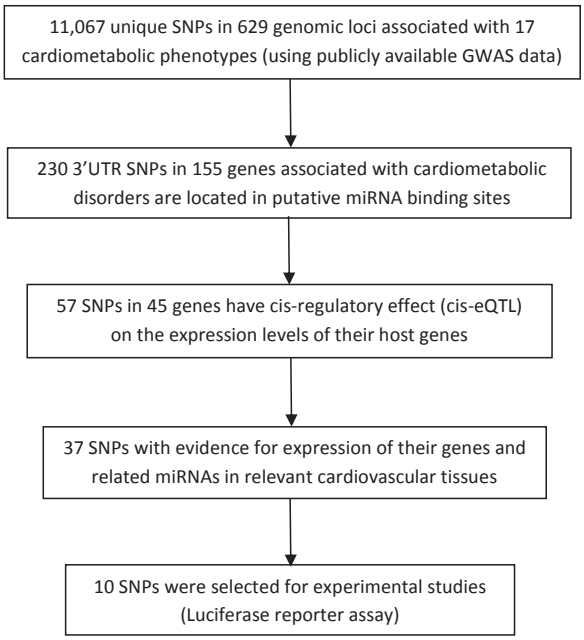


Figure 1. Identification of cardiometabolic-associated single-nucleotide polymorphisms (SNPs) that are located in miRNAbinding sites. eQTL , expression quantitative trait loci; GWAS, genome-wide association studies.

samples available through GeneNetwork (http://genenetwork.nl/bloodeqtlbrowser/) to determine the association between identified cardiometabolic SNPs in miRNA binding sites and the expression levels of their host transcripts. This particular cis-eQTL dataset was used because of its large sample size and power. In addition, two other online web browsers, including Genevar (http://www.sanger.ac.uk/resources/software/genevar) and GTEx (http://www.broadinstitute.org/gtex/), were searched for cis-eQTL data in relevant tissues such as adipose tissue, liver, left ventricle, and muscle skeletal. The Genevar platform provides information on eQTL in adipose, skin and lymphoblastoid cell lines (LCLs) collected from 856 healthy female twins of the MuTHER Study [29]. To illustrate the strength of the correlation between alleles and gene expression levels, we used the eQTL plots were generated by the Genevar web tool with a subset of the MuTHER Study (n=160), in which the expression of target genes in both adipose tissue and blood was correlated to individual genotypes based on the Spearman's rank correlation coefficient. SNPs determined to exhibit a cis-effect on the expression levels of their host transcripts (increase or decrease) were selected for further analyses.

Expression of genes and miRNAs in cardiometabolic tissues

A functional pre-requisite for miRNA-dependent gene regulation is that a miRNA and its target mRNAs to be expressed in the same tissue [9]. To determine whether the genes

hosting the identified SNPs and their related miRNAs express in relevant cardiovascular tissues, multiple web tools were used. We examined expression of the target genes across adipose tissue, liver, heart, skeletal muscle and blood using the Human BodyMap 2.0 data (http://www.ensembl.info/blog/2011/05/24/ human-bodymap-2-0-data-from-illumina/). To scan miRNA expression in cardiometabolic tissues, we used the mimiRNA (http://mimirna.centenary.org.au), PhenomiR (http://mips.helmholtz-muenchen.de/phenomir), and HMDD (http://cmbi.bjmu.edu.cn/hmdd) online databases. We subsequently used Pubmed to search through literature for those miRNAs that were not included in the above mentioned databases. Furthermore, to determine whether the interactions of the highlighted miRNAs and their target genes have previously been validated using functional experiments, multiple online databases, including miRTarBase, Tarbase (http://www.diana.pcbi.upenn.edu/tarbase) and miRwalk (http://www.umm.uni-heidelberg.de/apps/zmf/mirwalk) were searched. Additional information on miRNA sequences, host genes, miRNA families and clusters, and miRNA conservation between species was obtained from the miRBase (release 20) and TargetScan (6.2).

Selection of miRNA binding site variants for functional analysis

Specific criteria have been suggested to establish whether variants in miRNA binding sites are functional [9]. These criteria include an established association between the variant and the phenotype, the expression of the miRNA and its target gene in a relevant tissue and an allele-specific expression of the gene hosting the variant. We used the above mentioned criteria to prioritize the SNPs and experimentally examine the impact of a number of them on the miRNA and target mRNA interactions. Ten variants were subsequently selected for the validation phase based on the GWAS results (the significance of association), the eQTL analysis (effect and *p*-value in a relevant tissue) and the evidence for co-expression of their host genes with regulatory miRNAs in relevant cardiometabolic tissues. In addition, when a gene is targeted by more than one related miRNA in our collection, the experiments were performed using a miRNA that is highly conserved, as the conserved miRNAs across mammals and /or primates are considered of functional significance [30].

Luciferase reporter assay

Luciferase reporter assays were performed to first validate the selected genes as miRNA targets and then to investigate the effect of target site polymorphisms on the interaction between miRNAs and target genes. The 3'UTR sequences of genes were amplified from genomic DNA of subjects homozygous for either major or minor alleles using a forward primer containing a *XhoI* restriction site and a reverse primer containing a *EcoRI* restriction site. The amplified 3'UTR fragments were subsequently cloned into the pGL3 luciferase reporter vector downstream of the Luciferase open reading frame (Promega).

Similarly, mature miRNA sequences were amplified using a forward primer containing a *XhoI* restriction site and a reverse primer containing a *EcoRI* restriction site. The amplified miRNA sequences were then cloned into the MSCV-BC vector [31]. Sanger sequencing was performed to verify the DNA sequence of the plasmids. HEK293 cells, plated into 12-well plates (Costar), were co-transfected with a 3'UTR luciferase reporter construct and a miRNA or empty vector control. Luciferase values were normalized to the Renilla luciferase value to correct for differences in transfection efficiencies. All experiments were performed in triplicate.

Linkage Disequilibrium (LD) structure analysis of the validated SNPs

The list of cardiometabolic SNPs that were experimentally validated to have functional impact on the miRNA activity were submitted to the SNAP web tool (http://www.broadinstitute.org/mpg/snap/id) using R^2 threshold > 0.8, limit distance 500 kb, and population panel CEU to retrieve their proxy SNPs in the HapMap project. To examine the functional potential of the SNPs, we utilized the HaploReg database (http://www.broadinstitute.org/mammals/haploreg/) which predict the effect of a SNP on protein structure, gene regulation, and splicing using data from the ENCODE project.

RESULTS

Cardiometabolic-associated variants in miRNA binding sites

A total of 11,067 unique SNPs in 629 genomic loci were identified to be associated with 17 cardiometabolic phenotypes at a statistically significant level (p-value < 5×10^{-8}). **Table 1** shows a description of the cardiometabolic phenotypes, the number of variants associated with each phenotype and the consortia that generated these data. The proportion of these cardiometabolic-associated SNPs in different parts of the genes is shown in **Figure 2**.

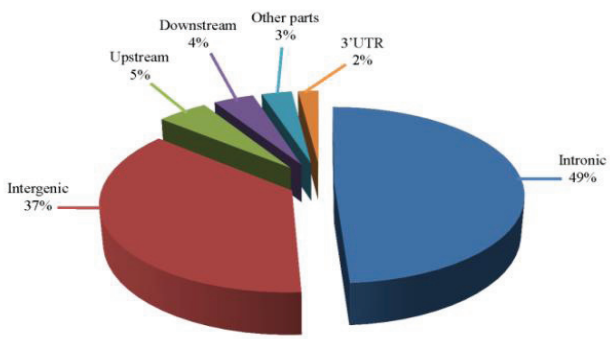


Figure 2. The proportion of variants associated with cardiometabolic phenotypes in different parts of the genes. UTR indicates untranslated region.

Searching through literature and various online databases, enabled us to identify 409,853 variants that are located in putative miRNA binding sites within the 3'UTRs of genes. Subsequent analysis of the 2.5 million HapMap SNPs showed that 23,650 of these miRNA binding site variants were present in this list, indicating that around 1% of imputed SNPs from the HapMap project (release 22) are likely to reside in miRNA binding sites. We identified 16,370 independent SNPs in the collection of 23,650 miRNA binding site SNPs that were present in the HapMap, using the LD based SNP pruning function implemented in PLINK ($R^2 > 0.8$). It is estimated that the number of independent common SNPs in the human genome are 1 million. This indicates that approximately 1.7% of imputed independent SNPs from the HapMap are predicted to be in miRNA binding sites. Here we found 230 SNPs (157 independent SNPs) out of 11,067 SNPs (2,369 independent SNPs) associated with cardiometabolic phenotypes to be located in miRNA binding sites. Therefore, the frequency of cardiometabolic GWAS findings in miRNA binding sites is ~2 times larger than other parts of the genome (p-value < 0.05 as determined by Fisher's exact test). If we consider the independent SNPs, this frequency is ~4 times higher than the expected number in other parts of the genome. The 230 identified SNPs represent 2% and 24%, respectively, of all GWAS identified SNPs and loci associated with cardiometabolic phenotypes. As shown in **Table S1**, some of these SNPs were found to be associated with more than one phenotype. We generated regional association plots of 155 loci that harbored these 230 SNPs using the locuszoom web tool [33]. Many of these variants appeared to be either the lead SNP which showed the strongest association with phenotype in the given genomic region or were one of the SNPs with the strongest associations. **Table S2** shows a list of miRNAs that, based on their binding sites, are likely to be affected by these 230 SNPs.

MiRNA binding site variants with cis-effect on the host gene expression

Cis-eQTL data was used to determine the correlation between the 230 SNPs and the expression levels of their host transcripts in tissues relevant for cardiometabolic phenotypes, such as blood and adipose tissue. From 230 variants, 57 SNPs (located in 45 genes) were significantly associated with altered expression levels of their host transcripts (**Table 2**). More information about the observed GWAS associations of these 57 SNPs with cardiometabolic phenotypes are depicted in **Table S3**. A significant association was detected between the variants located in the 3'UTRs of *FN3KRP, FADS1, CPEB4, LPL, MKRN2, GRINA, MMAB, KIAA0892* and *NT5C2* and their transcripts expression levels in multiple tissues, suggesting the functional significance of these variants on the related gene expression in cardiometabolic disorders. We could generate eQTL plots for some of these variants using the Genevar web tool to illustrate the correlations between individuals carrying different genotypes, the eQTL plots of the eight candidate SNPs were shown in **Figure 3** and plots of other SNPs are shown in **Figure S1**.

Table 2. Cardiometabolic-associated SNPs reside in miRNA binding sites with cis-effect on their host genes expression levels

SNP ID	Chr.	Position	Effect allele	Associated Phenotype	Host gene	Gene Network (Blood) *P*-value, Z-score	MuTHER Study (Adipose) *P*-value, effect	MuTHER Study (LCLs) *P*-value, effect	GTEx dataset (Different tissues) *P*-value, tissue
rs174545	11	61569306	G	FG,HDL, LDL, TC, TG	FADS1	1.53×10^{-31}, 11.69	1.5×10^{-5}, 0.214	1.8×10^{-2}, 0.041	-
rs174546	11	61569830	T	FG,HDL, LDL, TC, TG	FADS1	2.54×10^{-31}, 11.64	1.6×10^{-5}, 0.213	2.3×10^{-2}, 0.040	-
rs10786736	10	104849116	C	SBP	NT5C2	3.18×10^{-47}, 14.43	1.3×10^{-29}, 0.16	4.4×10^{-2}, 0.04	-
rs7085	15	75095483	T	SBP, DBP	CSK	1.55×10^{-131}, 24.40	6.2×10^{-1}, 0.009	3.8×10^{-11}, 0.079	-
rs1810126	6	160872151	T	CAD	SLC22A3	Na	1.9×10^{-4}, 0.023	7.7×10^{-1}, 0.002	-
rs3217992	9	22003223	T	CAD	CDKN2B	1.07×10^{-3}, 3.27	1.2×10^{-3}, 0.094	4.7×10^{-2}, 0.011	-
rs9818870	3	138122122	T	CAD	MRAS	2.35×10^{-4}, 3.68	1.5×10^{-1}, -0.035	7.4×10^{-1}, -0.003	-
rs12740374	1	109817590	T	CAD, LDL,TC, HDL	CELSR2	Na	2.1×10^{-1}, 0.008	Na	1.4×10^{-8}, Muscle
rs1046317	4	6322477	T	T2D	WFS1	1.13×10^{-5}, -5.3	2.7×10^{-2}, -0.045	3.7×10^{-1}, 0.019	-
rs1044661	17	80901020	A	HbA1c	TBCD	9.89×10^{-7}, 4.89	8.7×10^{-2}, 0.031	8.03×10^{-6}, 0.065	-
rs1046875	17	80685426	A	HbA1c	FN3KRP	2.51×10^{-190}, -29.43	3.5×10^{-14}, -0.14	6.2×10^{-3}, 0.041	1.2×10^{-7}, Adip, Muscle
rs1046896	17	80685533	T	HbA1c	FN3KRP	3.34×10^{-191}, -29.50	4.1×10^{-14}, -0.14	4.5×10^{-3}, 0.043	1.2×10^{-7}, Adip, Muscle
rs1046917	17	80685655	A	HbA1c	FN3KRP	3.73×10^{-191},-29.49	4.1×10^{-14}, -0.14	4.5×10^{-3}, 0.043	1.2×10^{-7}, Adip, Muscle
rs1057233	11	47376448	G	Pro-Ins	SPI1	9.81×10^{-198},-47.65	2.7×10^{-8}, -0.108	3.8×10^{-44}, -0.196	-
rs11988	11	47261260	A	Pro-Ins, FG	ACP2	9.36×10^{-11},6.48	1.3×10^{-1}, 0.026	1.5×10^{-2}, 0.036	-
rs9909	11	47799775	G	Pro-Ins, HDL	NUP160	1.631×10^{-5},4.31	3.1×10^{-2}, 0.039	1.9×10^{-1}, 0.027	-
rs6857	19	45392254	T	TG, LDL, HDL, TC	PVRL2	9.50×10^{-5},-3.90	6.9×10^{-1}, -0.009	5.1×10^{-1}, -0.005	-
rs4938353	11	117046197	G	TG, TC, HDL	SIDT2	9.81×10^{-198}, -52.25	9.2×10^{-3}, 0.051	6.3×10^{-1}, 0.012	1.3×10^{-6}, Blood
rs11067231	12	109993603	C	HDL, TC	MMAB	5.81×10^{-24}, 10.10	3.4×10^{-6}, 0.025	4.7×10^{-12}, 0.042	-
rs2744937	6	34557246	T	HDL, TC	C6orf106	1.86×10^{-14}, 7.66	2.96×10^{-1}, 0.028	7.5×10^{-7}, 0.128	2.5×10^{-5}, Blood
rs1059611	8	19824563	C	HDL, TG	LPL	1.27×10^{-14}, 7.71	3.0×10^{-2}, 0.102	4.5×10^{-2}, 0.016	-
rs3735964	8	19824045	A	HDL, TG	LPL	1.67×10^{-14}, 7.67	3.4×10^{-2}, 0.102	4.4×10^{-2}, 0.016	-

Table 2. (continued)

SNP ID	Chr.	Position	Effect allele	Associated Phenotype	Host gene	Gene Network (Blood) *P*-value, Z-score	MuTHER Study (Adipose) *P*-value, effect	MuTHER Study (LCLs) *P*-value, effect	GTEx dataset (Different tissues) *P*-value, tissue
rs13702	8	19868772	C	HDL,TG	LPL	5.56×10^{-5}, 4.03	5.4×10^{-2}, 0.066	1.3×10^{-1}, 0.009	-
rs1169312	12	121441461	T	TC, LDL	C12orf43	8.19×10^{-77}, 18.55	4.02×10^{-1}, 0.01	1.2×10^{-2}, 0.038	-
rs2285627	19	19292629	T	TG, TC, LDL	KIAA0892	1.5×10^{-198}, 30.2	1.03×10^{-17}, 0.09	3.27×10^{-21}, 0.12	-
rs13964	19	19292629	G	TG, TC, LDL	KIAA0892	1.5×10^{-197}, 30	1.6×10^{-15}, 0.092	7.3×10^{-20}, 0.117	-
rs2285628	19	19467996	A	TG, TC, LDL	KIAA0892	9.81×10^{-198}, -51.18	5.9×10^{-46}, -0.196	1.8×10^{-55}, -0.24	-
rs1572343	1	93387886	T	HDL	TMED5	1.6×10^{-133}, 24.6	1.9×10^{-15}, 0.154	1.3×10^{-10}, 0.092	-
rs1054623	19	8586086	A	HDL	MYO1F	1.35×10^{-3}, -3.21	6.04×10^{-2}, 0.025	5.06×10^{-2}, 0.024	-
rs1109166	16	67977382	C	HDL	LCAT	1.17×10^{-7}, 5.30	4.8×10^{-2}, 0.023	5.6×10^{-1}, 0.005	-
rs11350	16	68335392	C	HDL	SLC7A6	2.10×10^{-6}, -4.74	5.3×10^{-2}, -0.042	5.5×10^{-3}, -0.048	-
rs11700	16	67232684	G	HDL	E2F4	4.65×10^{-6}, -4.58	2.9×10^{-5}, 0.069	Na	-
rs2229714	1	26900708	A	HDL	RPS6KA1	1.03×10^{-5}, -4.41	4.1×10^{-1}, 0.021	8.4×10^{-5}, 0.097	-
rs877710	12	109993976	C	HDL	MMAB	5.73×10^{-24}, 10.10	3.4×10^{-6}, -0.025	4.7×10^{-12}, -0.042	5.9×10^{-10}, Lymphoblast
rs907091	17	37921742	T	HDL	IKZF3	3.92×10^{-5}, -4.11	7.5×10^{-1}, -0.002	9.6×10^{-1}, -0.001	-
rs8102380	19	10801185	G	LDL	ILF3	4.73×10^{-8}, 5.46	1.7×10^{-1}, 0.020	5.5×10^{-1}, 0.009	-
rs9100	8	145067467	T	LDL	GRINA	1.82×10^{-147}, -25.86	8.06×10^{-4}, 0.071	2.5×10^{-2}, 0.054	-
rs1054284	19	19357641	G	TC	GATAD2A	9.8×10^{-198}, 30.8	1.3×10^{-4}, 0.034	2.1×10^{-1}, 0.013	-
rs1052248	6	31662955	T	TC	LST1	Na	9.0×10^{-5}, -0.093	9.0×10^{-5}, -0.093	-
rs1050488	8	59496490	G	TC	NSMAF	3.83×10^{-92}, 20.36	7.3×10^{-2}, 0.022	9.6×10^{-7}, 0.052	-
rs1050504	8	59496416	T	TC	NSMAF	6.95×10^{-92}, 20.33	4.8×10^{-2}, 0.024	6.8×10^{-7}, 0.053	-
rs7194	6	32412480	G	TC	HLA-DRA	1.50×10^{-37}, 12.81	5.8×10^{-2}, 0.065	5.1×10^{-1}, 0.008	8.2×10^{-8}, Liver
rs7956	3	12624763	C	TC	MKRN2	1.16×10^{-82}, 19.26	7.9×10^{-6}, 0.075	6.5×10^{-2}, 0.026	1.6×10^{-6}, Muscle
rs1063966	19	19357641	A	TG	GATAD2A	9.8×10^{-198}, 30.8	1.1×10^{-4}, 0.034	2.2×10^{-1}, 0.013	-

Table 2. (continued)

SNP ID	Chr.	Position	Effect allele	Associated Phenotype	Host gene	Gene Network (Blood) *P*-value, Z-score	MuTHER Study (Adipose) *P*-value, effect	MuTHER Study (LCLs) *P*-value, effect	GTEx dataset (Different tissues) *P*-value, tissue
rs1051921	7	73593613	A	TG	MLXIPL	Na	8.2×10^{-16}, 0.24	2.6×10^{-1}, 0.007	-
rs11735092	4	88226231	C	TG	HSD17B13	7.02×10^{-10}, 6.17	8.9×10^{-17}, 0.196	3.1×10^{-2}, 0.011	-
rs8024	1	201845575	C	BMI	IPO9	4.63×10^{-7}, 5.04	2.4×10^{-3}, 0.03	4.7×10^{-2}, 0.028	-
rs879620	16	4015729	T	BMI	ADCY9	6.07×10^{-8}, -5.42	7.0×10^{-3}, -0.023	1.9×10^{-1}, -0.008	-
rs2531995	16	4013467	T	BMI	ADCY9	7.82×10^{-8}, -5.37	6.4×10^{-3}, -0.023	1.6×10^{-1}, -0.008	1.4×10^{-8}, Liver
rs1049376	12	26491475	T	WHR	ITPR2	2.89×10^{-6}, 4.68	3.9×10^{-4}, 0.054	1.2×10^{-5}, 0.087	-
rs12321	22	52745087	G	WHR	ZNRF3	Na	5.8×10^{-3}, -0.034	4.5×10^{-1}, 0.006	-
rs1976074	5	173957660	T	WHR	CPEB4	9.81×10^{-198}, -56.08	2.8×10^{-26},- 0.15	1.3×10^{-39}, -0.265	-
rs2179129	22	29450923	G	WHR	ZNRF3	Na	7.6×10^{-3}, 0.032	2.6×10^{-1}, 0.009	-
rs3768	12	124015292	T	WHR	ZNF664	Na	1.5×10^{-3}, 0.042	9.5×10^{-1}, 0.001	-
rs4654	12	26490155	T	WHR	ITPR2	1.93×10^{-10}, 6.37	2.01×10^{-4}, 0.053	4.70×10^{-7}, 0.094	-
rs9863	12	123936906	T	WHR	CCDC92	1.94×10^{-11}, -6.71	9.6×10^{-9}, -0.101	6.9×10^{-7}, -0.11	-
rs4823006	22	29055683	G	WHR	ZNRF3	Na	5.8×10^{-3}, 0.034	Na	2.5×10^{-6}, Liver

F.G, Fasting Glucose; Pro.In, Pro-Insulin; TC, Total cholesterol; TG, Triglyceride; HDL, High Density Lipoprotein; LDL, Low Density Lipoprotein; SBP, Systolic Blood Pressure; DBP, Diastolic Blood Pressure; T2D, Type 2 Diabetes; Coronary Artery Disease, CAD; BMI, Body Mass Index; WHR, Waist to Hip Ratio; Adip, Adipose tissue; Muscle, Muscle skeletal; Lymphoblast, Lymphoblastoid; Na, Not available.

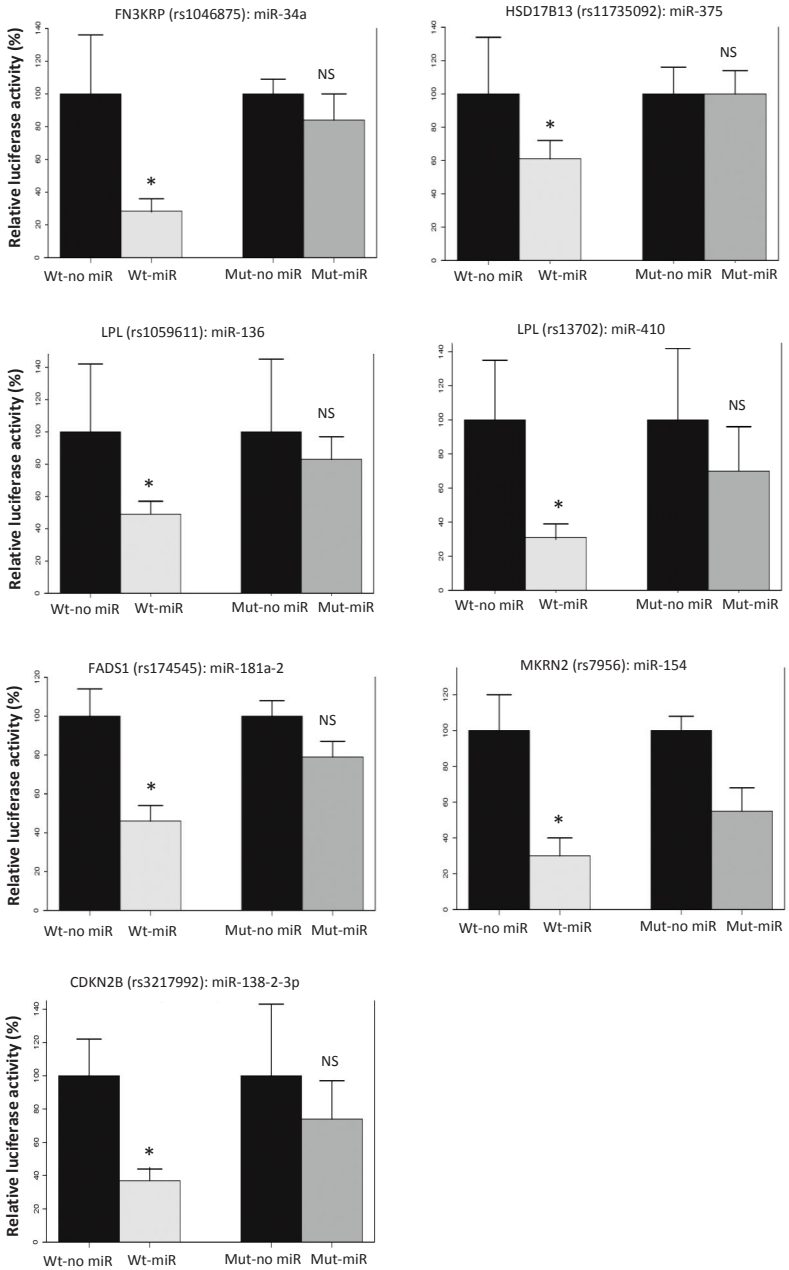


Figure 3. Experimental validation of 7 variants that were predicted to decrease the interactions between miRNAs and their target genes. These experiments revealed that miRNA-mediated inhibition of FADS1, LPL, HSD17B13, MKRN2, CDKN2B, and FN3KRP is significantly reduced or abrogated when the 3' untranslated regions harbor the minor allele (Mut). Data were depicted as fold reduction compared with empty vector (no miR), and wild-type (Wt) allele controls. Error bars represent SD of the mean. n=3 replicates for each condition. * indicates significant effect and NS means non-significant.

Expression of the identified target genes and miRNAs in relevant cardiometabolic tissues

Expression of a miRNA and its target genes in the same tissue is a pre-requisite for miRNA mediated regulation of target genes. We searched multiple online databases and literature to find evidence for expression of 45 genes that harbored the 57 identified variants with their regulatory miRNAs in cardiometabolic tissues. **Table S4** shows the genes expression across different tissues, indicating that almost all of the studied genes are expressed in cardiovascular tissues. In addition, expression of their related miRNAs in adipose tissue, liver, heart, pancreas and skeletal muscle were investigated. These analyses revealed that 29 out of the 45 genes (hosting 37 of the variants) are expressed with their predicted miRNAs in relevant cardiometabolic tissues (**Table 3**).

Functional impact of miRNA binding site variants on the miRNA-mRNA interactions

Ten of the 37 highlighted miRNA binding site SNPs were selected for functional validation based on their GWAS results, eQTL analysis and expression of their related genes and miRNAs in cardiometabolic tissues, **Table S5**. Eight of these SNPs were predicted to disrupt miRNA-mRNA interactions (rs174545 (*FADS1*and miR-181a-2), rs1059611 (*LPL* and miR-136), rs13702 (*LPL* and miR-410), rs11735092 (*HSD17B13* and miR-375), rs7956 (*MKRN2* and miR-154), rs2857101 (*MMAB* and miR-593-5p), rs3217992 (*CDKN2B* and miR-138-2-3p), and rs1046875 (*FN3KRP* and miR-34a)), while the other two SNPs were predicted to enhance the interaction between miRNAs and their target genes (rs6857 (*PVRL2* and miR-320e) and rs907091 (*IKZF3* and miR-326)). The 3'UTR sequences of the selected genes were amplified from genomic DNA of subjects homozygous for either major or minor alleles and cloned into the pGL3 luciferase reporter vector (**Table S6**). All primers are shown in **Table S7**. In addition, mature miRNA sequences were amplified and cloned into the MSCV-BC vecto.r

Co-transfection of luciferase constructs with the major alleles for the 8 variants predicted to disrupt miRNA-mRNA interactions with miRNAs, resulted in a significant decrease in luciferase activity in 7 variants (rs174545:G, rs1059611:T, rs13702:T, rs11735092:A, rs7956:T, rs3217992:G and rs1046875:T). This experiment confirmed that miRNAs can regulate the 3'UTRs harboring the wild type alleles. In addition, the co-transfection experiments revealed that the miRNA mediated inhibition of *FADS1, LPL, HSD17B13, MKRN2*, *CDKN2B* and *FN3KRP* is significantly reduced (p-value < 0.05 as determined by the Mann-Whitney test) when the 3'UTRs contain the mutant alleles (**Figure 4**). No significant differences were observed between the luciferase activities of *MMAB* constructs containing either major or minor alleles (rs877710) in presence of miR-593-5p, indicating that this particular variant does not affect the miRNA-mediated regulation of *MMAB* at mRNA level.

Table 3. Putative miRNA binding sites with wild-type and mutant alleles in the 3'UTR of genes associated with cardiometabolic traits

SNP ID	A1/A2	Gene ID	Binding sites for miRNA with wild-type allele (A1)	Binding sites for miRNAs with mutant allele (A2)	Evidence for miR-Exp in cardiometabolic tissues
rs1044661	G/A	TBCD	miR-5582-5p|miR-6866-5p	miR-466|miR-4643|miR-4789-3p	-
rs1046317	T/C	WFS1	miR-3153|miR-6733-5p|miR-6739-5p miR-6768-3p	miR-3202|miR-4271|miR-4512|miR-4725-3p miR-4747-5p|miR-5196-5p|miR-6780b-5p	-
rs1046875	A/G	FN3KRP	*miR-34a-5p*|***miR-34b,c-5p|miR-449a miR-449b,c-5p**|miR-2682-5p|miR-3978	**miR-885-3p**	miR-34a-5p (A,H,L,P,M) miR-449a (H), miR-885-3p (L)
rs1046896	C/T	FN3KRP	**miR-208a-3p|miR-208b-3p**	miR-641|miR-3617-5p	miR-208a-3p (H,L,M) miR-641 (H,M)
rs1046917	A/G	FN3KRP	miR-4435	NONE	-
rs1050488	T/C	NSMAF	NONE	**miR-196a-3p**|miR-5197-5p	miR-196a-3p (A,H,L,P,M)
rs1050504	G/A	NSMAF	**miR-154-3p**|miR-487a-3p	miR-3144-5p|miR-3191-3p|miR-3675-5p|miR-4652-5p|miR-5007-3p|miR-6810-5p|miR-6890	miR-154-3p (H,L,M)
rs1051921	C/T	MLXIPL	NONE	miR-455-3p|miR-4274	-
rs1052248	T/A	LST1	NONE	**miR-153-5p**|miR-1250-3p	miR-153-5p (A,H,L,P,M)
rs1054284	A/G	GATAD2A	**miR-215-3p|miR-223-3p**|miR-4697-3p miR-4704-3p|miR-6773-3p	NONE	miR-215-3p (A,H,L,P,M) miR-223-3p (A,H,L,P,M)
rs1054623	G/T	MYO1F	**miR-221-5p**|miR-8073	miR-1254|miR-3116	miR-221-5p (A,H,L,P,M)
rs1057233	C/T	SPI1	**miR-569(Val)|miR-655-5p**	miR-3150b-3p|miR-4784| miR-6888-5p	miR-655-5p (H)
rs1059611	T/C	LPL	**miR-136-5p**	miR-1468-5p	miR-136-5p (A,H,L,M)
rs1063966	G/A	GATAD2A	**miR-548u**|miR-3192-5p|miR-4463 miR-7161-5p	**miR-30b, c-3p**|miR-1273h-5p|miR-3689a-3p miR-3689b-3p|miR-3689c|miR-6779-5p miR-6780a-5p|miR-6788-5p|miR-887-5p	miR-30b,c-3p (A,H,L,P,M)
rs10786736	C/G	NT5C2	**miR-488-5p**|miR-5586-5p	miR-5092	-
rs11067231	G/T	MMAB	miR-1972	miR-4742-5p|miR-4761-5p|miR-6809-5p	-
rs1109166	A/G	LCAT	miR-622	miR-6849-3p	miR-622 (H,L)
rs11350	T/C	SLC7A6	NONE	miR-548e-5p(Val)	miR-548e-5p (H,L)
rs1169312	C/A	C12orf43	miR-3611|miR-4677-3p|miR-4679	miR-1264|miR-3184-3p|miR-4652-3p	-
rs11700	A/G	E2F4	miR-520f-5p|miR-760|miR-4799-5p miR-6822-3p|miR-8057	**miR-18b-3p**|miR-328-3p|miR-1291| miR-5009-3p|miR-6775-3p|miR-6851-3p	miR-520f-5p (H,L,M) miR-18b-3p (H,L,M) miR-328-3p (A,H,L,P)

Table 3. (continued)

SNP ID	A1/A2	Gene ID	Binding sites for miRNA with wild-type allele (A1)	Binding sites for miRNAs with mutant allele (A2)	Evidence for miR-Exp in cardiometabolic tissues
rs11735092	A/G	HSD17B13	**miR-375**|miR-1282|miR-5707	miR-1468-5p|miR-3655	miR-375 (A,H,L,P,M)
rs11988	C/T	ACP2	miR-548ao-3p|miR-3657|miR-6774-3p	miR-591|miR-6502-3p|miR-6834-3p	-
rs12740374	G/T	CELSR2	miR-3663-3p|miR-6865-3p|miR-6871-3p	NONE	-
rs13702	A/G	LPL	**miR-410(Val)**	NONE	miR-410 (A,H,L,M)
rs13964	C/G	KIAA0892	miR-3197|miR-4663|miR-6087|miR-6765	**miR-125a,b-5p**|miR-1291|miR-4319|miR-5008|miR-6089|miR-6775-3p|miR-7976	miR-125a-5p (H)
rs1572343	A/G	TMED5	**miR-511-5p**|miR-627-3p|miR-6830-3p	NONE	miR-511-5p (A,H,L,P,M) miR-627-3p (H,L,M)
rs174545	G/C	FADS1	**miR-181a-2-3p**|miR-4252|miR-4786-3p	**miR-124-3p**|**miR-506-3p**|miR-605-3p|miR-6770	miR-181a-2-3p (A,H,L,P,M) miR-124-3p (H,L)
rs174546	G/A	FADS1	**miR-212-5p**|miR-4312|miR-5001-3p	miR-1251-3p|miR-6728-3p	miR-212-5p (A,H,L,P,M)
rs1810126	C/T	SLC22A3	**miR-124-3p(Val)**|**miR-506-3p**|miR-4708-3p miR-5582-5p	miR-4696	miR-124-3P (H,L) miR-506 (H,M)
rs2229714	G/A	RPS6KA1	miR-663b|miR-1538|miR-4745-3p	miR-3140-5p|miR-6735-3p	-
rs2285627	T/C	KIAA0892	miR-4323|miR-5699-5p	NONE	-
rs2285628	T/A	KIAA0892	miR-3915|miR-3928-3p|miR-4476 miR-6876-5p	**miR-199a, b-5p**|miR-3190-3p|miR-3202 miR-3911|miR-4302|miR-4533|miR-6833	miR-199a-5p (H,L)
rs2744937	G/A	C6orf106	miR-6746-3p	**miR-22-3p**|miR-1207-3p|miR-4252	miR-22-3p (L)
rs3735964	C/A	LPL	miR-1277-3p|miR-5197-3p	NONE	-
rs4823006	A/G	ZNRF3	**miR-221-5p**|miR-5089-3p|miR-8073	**miR-331-3p**|miR-5195-5p	miR-221 (A,H,L,P,M) miR-331-3P (L)
rs4938353	G/A	SIDT2	NONE	**miR-183-3p**|**miR-219a-2-3p**|**miR-219b-3p** miR-4452	miR-183-3p (A,H,L,P,M) miR-219-3p (A,H,L,P,M)
rs6857	C/T	PVRL2	miR-645|miR-3929|miR-4419b|miR-4478 miR-4505|miR-5787	**miR-320e**	miR-320e (A,H,L,P,M)
rs3217992	G/A	CDKN2B	**miR-138-2-3p**|**miR-205-3p**	**miR-374b-5p (Val)**|**miR-655-3p**	miR-138-3P (L,H,M) miR-205-3p (L,P,M) miR-374b-5p (L), miR-655(H)
rs7085	T/C	CSK	miR-6504-3p|miR-7849-3p	NONE	-
rs7194	G/A	HLA-DRA	NONE	miR-6507-3p	-

Table 3. (continued)

SNP ID	A1/A2	Gene ID	Binding sites for miRNA with wild-type allele (A1)	Binding sites for miRNAs with mutant allele (A2)	Evidence for miR-Exp in cardiometabolic tissues
rs7956	T/C	MKRN2	**miR-154-5p (Val)**	**miR-218-5p**	miR-154-5p (H,L,M) miR-218-5p (A,H,L,P,M)
rs8102380	G/A	ILF3	**miR-101-3p|miR-144-3p**	NONE	miR-101-3p (A,H,L,P,M) miR-144-3p (L,H,M)
rs877710	G/C	MMAB	**miR-593-5p**	**miR-564**	miR-593-5p, miR-564 (H,L,M)
rs907091	G/A	IKZF3	miR-4497	**miR-326|miR-330-5p|miR-204-3p**|miR-3191 miR-3192-5p|miR-3649|miR-4314|miR-4646 miR-518c-5p|miR-619-5p|miR-6506-5p| miR-6764-3p|miR-6824-3p	miR-326 (A,H,L,P,M) miR-204-3p (A,H,L,P,M) miR-518c-5p (P) miR-619-5p (P)
rs9100	G/T	GRINONE	miR-548ao-3p|miR-3667-5p|miR-6859-3p	miR-548b-3p|miR-4511	-
rs9818870	C/T	MRAS	miR-6506-3p	miR-3074-3p|miR-5007-3p|miR-664a-3p	miR-664a-3p (M)
rs9909	G/C	NUP160	**let-7f-2-3p**|miR-1185-1-3p|miR-1185-2-3p miR-5683	miR-3976|miR-5186	let-7f-2-3p (A,H,L,P,M)
rs1049376	A/G	ITPR2	miR-4503|miR-6792-5p|miR-7856-5p	miR-4263	-
rs12321	G/C	ZNRF3	**let-7f-2-3p**|miR-1185-1-3p|miR-1185-2-3p	miR-3123|miR-3925-5p|miR-3976	let-7f-2-3p (A,H,L,P,M)
rs1976074	T/A	CPEB4	NONE	**miR-500a-5p|miR-501-5p**	-
rs2179129	A/G	ZNRF3	**miR-411-5p**	miR-4436a|miR-4645-3p|miR-4700-3p miR-5000-3p|miR-936	-
rs2531995	G/A	ADCY9	miR-3925-5p	**miR-346|miR-632|miR-654-3p**|miR-3123 miR-4288|	miR-346 (M, L)
rs3768	C/T	ZNF664	miR-2355-5p	miR-6793-3p	-
rs4654	G/A	ITPR2	miR-1283|miR-3162-3p|miR-4698	**miR-454-5p**|miR-3163|miR-3662|miR-6883-3p	miR-454-5p (L)
rs8024	C/A	IPO9	miR-6753-3p|miR-7107-3p	**miR-148b-5p**|miR-3168|miR-4652-3p|miR-5584 miR-6874-3p	miR-148-5p (A,H,L,P,M)
rs879620	G/A	ADCY9	miR-4318|miR-4693-5p|miR-892a	**miR-126-5p|miR-656-3p**|miR-4795-3p|miR-5701	miR-126-5p (A,H,L,P,M)
rs9863	A/G	CCDC92	miR-3611|miR-3650	NONE	-

A, Adipose; H, Heart; L, Liver; P, Pancreas; M, Skeletal muscle; Val, the miR-mRNA interaction is experimentally validated; Conserved miRNAs are shown in bold.

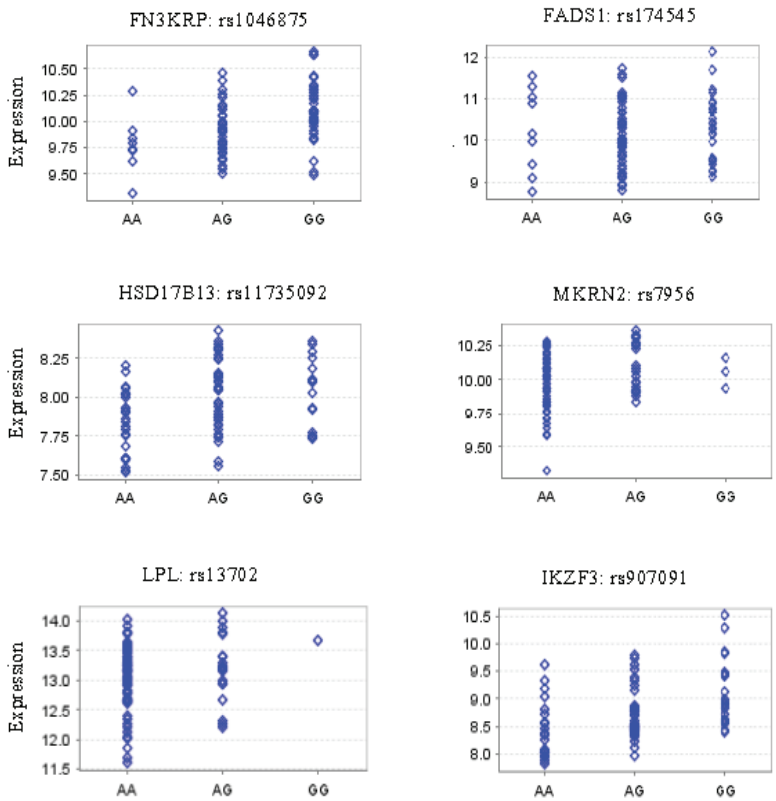


Figure 4. Plots showing the association between six of the miRNA binding site variants and expression levels of their host genes generated using the Genevar data set.

Similar experiments were performed to identify the role of other two variants in the miRNA-mRNA interactions. A dose-dependent effect was observed of miR-320e and miR-326 on the expression levels of *PVRL2* and *IKZF3*, respectively, in presence of the minor alleles (**Figure S2**). The presence of a minor allele significantly increased the interaction between miR-320e and the 3'UTR of *PVRL2* as was visualized by a decreased Luciferase activity (from 90% to 50%). Similarly, while co-transfection of the major allele construct of IKZF3 (rs907091) with miR-326, resulted in a non-significant reduction in luciferase activity, co-transfection of the minor allele further reduced luciferase activity (from 95% to 60%). These findings indicate that the mutant alleles in *PVRL2* and *IKZF3* enable miR-320e and miR-326, respectively, to more efficiently regulate the expression of their host transcripts (**Figure 5**).

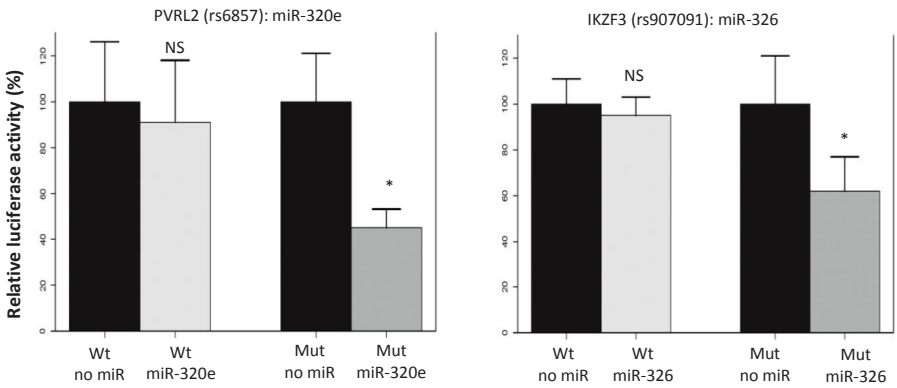


Figure 5. Experimental validation of two variants that were predicted to increase the interactions between miRNAs and their targetgenes. These variants significantly enhance miRNA-mediated regulation of their host genes (PVRL2 co-transfected with miR-320e, and IKZF3 co-transfected with miR-326). Data were depicted as fold reduction compared with empty vector (no miR), wild-type (Wt) allele controls. Error bars represent SD of the mean. n=3 replicates for each condition. * indicates significant effect and NS means non-significant.

Evaluation of the LD structure of the loci containing the experimentally validated SNPs

Proxy SNPs ($R^2 > 0.8$ and MAF > 0.10) in strong LD with the nine selected variants were extracted and their predicted effects on protein structure and gene regulation, using the HaploReg database v3 were investigated. Detailed information of these proxies and their predicted functions was provided in **Table S8**. This data further showed that the nine studied variants have the potential to be functional variants in their loci associated with cardiometabolic phenotypes.

DISCUSSION

In this study, we performed a genome-wide scan to identify variants that may associate with cardiometabolic phenotypes through affecting miRNA-mediated regulation of cardiometabolic genes. We annotated 230 SNPs have been identified by GWAS for different cardiometabolic traits and diseases to miRNA binding sites. We determined that 37 of these SNPs (representing 29 genes) fulfill our predefined criteria for being potentially functional in their genomic loci. Using luciferase reporter assay for ten prioritized variants, we showed that rs174545 (*FADS1*), rs1059611 and rs13702 (*LPL*), rs1046875 (*FN3KRP*), rs7956 (*MKRN2*), rs3217992 (*CDKN2B*) and rs11735092 (*HSD17B13*) decrease miRNA-dependent regulation of their corresponding transcripts. Conversely, rs6857 (*PVRL2*) and rs907091 (*IKZF3*) were shown to enhance the activity of miRNAs on their

host transcripts. Our results suggest that allele-specific miRNA regulation of transcripts because of variants in miRNA binding sites could serves as a functional mechanism underlying GWAS findings.

Although GWAS have led to the discovery of thousands of variants that appear to contribute to cardiometabolic disorders, most of the identified variants are located in non-coding regions of the genome and their functionality remains to be investigated. To determine the most likely functional candidate SNPs in non-coding regions novel approaches are required. A potential mechanism for GWAS identified variants is to influence gene expression by affecting the interactions between miRNAs and target mRNAs. As the seed sequence in miRNAs have to be perfectly complementary to the miRNA binding site for target recognition and interaction, genetic variants in miRNA seed regions or miRNA binding sites can alter disease susceptibility [8, 10, 13, 15]. From the time that an association between a mutation in the binding site of miR-189 and the pathogenesis of Tourette's syndrome has been demonstrated [14], a considerable level of variation in miRNA target sites has been reported. To date, catalogues and online databases for this variability have been developed which provide a valuable resource for epidemiologic studies [16, 17, 28]. However, these databases are mainly based on bioinformatics approaches and include only a limited number of functional annotations.

In this study, we linked several publicly available datasets and tools and combined them with experimental studies to identify the likely functional variants located in the 3'UTR of cardiometabolic genes. We identified 230 cardiometabolic risk SNPs that were predicted to either abolish the existing miRNA binding sites or create new putative binding sites. This number is higher than the expected by chance and may suggest an enrichment of GWAS findings in miRNA binding sites. In addition, since miRNAs are highly conserved, and negative selection is stronger on their putative binding sites than on other conserved sequence motifs in the 3'UTR of genes [34], the vast majority of miRNA binding site variants are rare and absent in the GWAS data imputed from the HapMap project. Therefore, a higher enrichment level of cardiometabolic-associated variants in miRNA binding sites can be expected if dense genotyping methods (GWAS from 1000 genome imputed data) or sequencing will implement in association studies.

Specific criteria have previously been suggested to establish whether variants in miRNA binding sites are functional, including an established association between the variant and the phenotype, the expression of the host gene and regulatory miRNA in a relevant tissue and an allele-specific expression of the host gene [9]. We determined 37 from the 230 SNPs that were found to be associated with cardiometabolic phenotypes and located in miRNA binding sites fulfill the above-mentioned criteria for being functional. These 37 SNPs are likely functional candidate variants in their genomic loci and propose for experimental validation studies. These data may also indicate that approximately 16% of SNPs that are located in miRNA binding sites could be functional.

Ten SNPs were selected for functional validation. Our results, first demonstrate the interaction between studied miRNAs and the 3'UTRs of *FADS1* (miR-181a-2), *LPL* (miR-136 and miR-410), *HSD17B13* (miR-375), *MKRN2* (miR-154), *CDKN2B* (miR-138-2-3p) and *FN3KRP* (miR-34a), resulting in reduced gene expression. We subsequently showed that the miRNA-dependent regulation of these genes is significantly decreased when mutant alleles are present in their 3'UTRs. Among all, the rs1173092 mutant allele completely abrogates the miR-375 binding site and results in the normal expression of *HSD17B13*. Other variants partially disrupt their related miRNA binding sites. This variability in disrupting the miRNA binding sites may be fully explained by compensation of the miRNA binding to its target by match sequences outside the "seed" that is called "compensatory 3'-end pairing [35]. In addition, the position of mutated nucleotide in a miRNA target site is important. Sometimes one SNP could completely disrupt the target site and strongly affect the binding of miRNA to the target mRNA, and sometimes just leads to the activity of miRNA on the target transcript with lower affinity.

Both *FADS1* and its specified miRNA, miR-181a-2-3p, have been shown to be involved in regulation of lipid metabolism [36, 37]. In this study, we showed that the minor allele of rs174545 in the 3'UTR of *FADS1* disrupts the binding site of miR-181a-2-3p. In addition, eQTL analysis revealed that carriers of this particular mutant allele display an elevated *FADS1* expression. Thus, an allele-specific miR-181a-2-3p mediated regulation of *FADS1* expression may partly explain the observed GWAS association with lipid traits. *LPL*, one of the other genes shown to be regulated by miRNAs, exhibits a strong association with the levels of serum triglyceride and cholesterol and is implicated in the pathogenesis of cardiovascular disease [2, 38]. Its regulatory miRNAs, miR-410 and miR-136-5p, are highly expressed in adipocytes and are known regulators of adipogenesis [10, 39]. In this study, we determined that the rs13702 and rs1059611 mutant alleles can modulate respectively the interaction of miR-410 and miR-136-5p with the *LPL* 3'UTR. Similarly, Richardson et al have demonstrated that the interaction between miR-410 and *LPL* is disrupted by variant rs13702 [10]. Although the importance of rs13702 in disruption of miRNA mediated regulation of *LPL* expression has previously been shown, our results revealed that miRNA-mediated regulation of *LPL* can also be abrogated by variant rs1059611. Dysregulation of *LPL* by these miRNA binding site SNPs may be one of the functional explanation for the identified GWAS association between *LPL* and the level of serum cholesterol.

Our results demonstrated that, similar to *FADS1* and *LPL*, mutant alleles significantly reduce the interactions between *HSD17B13* (rs11735092) and *MKRN2* (rs7956) and their regulatory miRNAs, miR-375 and miR-154-5p. Aberrant regulation of these genes by miR-375 and miR-154-5p could be a reason behind the associations between *HSD17B13* and triglycerides and between *MKRN2* and total-cholesterol levels. In addition, while *HSD17B13* has been shown to be expressed in adipose tissue, miR-375 is known to be

involved in regulation of glucose and lipid metabolism [40]. The expression patterns of both *MKRN2* and miR-154-5p in adipocytes were also in accordance with our expectations [41].

CDKN2B has been shown to be associated with coronary artery disease (CAD), myocardial infarction and type 2 diabetes [2, 42]. GWA and experimental studies have revealed that the rs3217992 minor allele is protective against incident CAD [4, 43]. MiR-138-2-3p has been reported to be required for normal cardiac development and function [44]. Furthermore, Horswell et al have previously been exhibited that this miRNA has an effect on both *CDKNB2* mRNA and protein levels [45]. In this study, we showed that the rs3217992 minor allele disrupts miR-138-2-3p mediated regulation of *CDKN2B*, thereby increasing the transcript levels of the gene. These findings may indicate that allelic-specific regulation of *CDKN2B* by miR-138-2-3p as a potential explanation for the association between *CDKN2B* (rs3217992) and the susceptibility to CAD. Finally, we revealed that although miR-34a-5p can interact with the wild type 3'UTR of *FN3KRP,* this interaction was significantly reduced by the mutant allele (rs1046875). MiR-34a-5p is highly expressed in liver and plays important regulatory role in pancreatic development and insulin secretion [46]. The disruption of the interaction between miR-34a and *FN3KRP* by rs1046857 may provide a reason, at least in part, for the association of *FN3KRP* with HbA1c [23].

Although a large number of variants in 3'UTRs have been demonstrated to disrupt the existing miRNA binding sites, only a few studies have provided experimental evidence showing variants that improve the original recognition sites or create novel binding sites [47]. In this study, we investigated the functionality of two variants that were predicted to enhance the binding of miR-320e and miR-326 to *PVRL2* and *IKZF3*, respectively. Previous GWAS on lipid traits have revealed associations between *PVRL2* and lipid traits and between *IKZF3* and HDL-cholesterol. Furthermore, both miR-320e and miR-326 are highly expressed in adipose tissue and pancreas [40, 48, 49]. Luciferase experiments revealed that the mutant alleles rs6857 and rs907091 enhanced the binding of miR-320e to *PVRL2* and miR-326 to *IKZF3*, respectively. It has previously been shown that target gene regulation requires a specific concentration of miRNAs [50]. This explains why, in our experiments, a dose-dependent effect could be observed. These results demonstrate an allele-specific regulation of *PVRL2* and *IKZF3* by miR-320e and miR-326 that may be the underlying cause of their GWAS associations with lipid traits.

A limitation of our study needs to be addressed. We tried to provide functional evidence for the miRNA-target gene interactions and the impact of miRNA binding site SNPs on these interactions. It should be noted that miRNA activities on the transcripts of interest are largely dependent on the cell type and expression levels of both miRNA and their competing targets. Therefore, it is optimal to do quantitative measure of co-expression for the identified genes and related miRNAs in the target cardiovascular tissues (such as liver, skeletal muscle or adipose tissue). In addition, the chain of evidences would be

more complete if we examine the influences of SNPs on post-transcriptional regulation of the miRNA targets. Unfortunately, we did not have a model system available to test expression levels of the protein products of the identified miRNA target genes in the right cellular context. However, to gain more insights in the function of the SNPs, we used the well-accepted *in vitro* experiments (Luciferase reporter assays) in combination with eQTL data, and miRNA and gene expression profiles that provide strong support for the functionality of these 3'UTR polymorphisms.

In conclusion, we have demonstrated that 3'UTR SNPs in miRNA binding sites can affect miRNA-mediated regulation of genes involved in cardiometabolic disorders. These findings may improve our understanding of the roles of miRNAs in the pathophysiology of cardiovascular disease and metabolic syndromes. Our findings may also be of clinical importance since they indicate that specific miRNAs can modify gene expression profiles and affect cardiometabolic phenotypes. This study shows that a multidisciplinary approach, including bioinformatics, epidemiology, and molecular biology can be utilized to identify the most likely functional candidate variants in specific genomic regions. This approach could be applied to a wide range of phenotypes and may contribute to improving the annotation of GWAS findings.

Supplement available online at:
http://circgenetics.ahajournals.org/content/early/2015/03/26/CIRCGENETICS.114.000968/suppl/DC1

Supplementary Table S5. Thirty-seven miRNA binding site SNPs fulfill criteria for being potentially functional in their loci

SNP ID	GWAS	eQTL	miR Exp	Asso. Phenotype	Gene	miRNA	Functional
rs174545	✓✓	✓✓	✓✓	FG,HDL,LDL,TC,TG	FADS1	miR-181a-3p	Selected
rs174546	✓✓	✓✓	✓✓	FG,HDL,LDL,TC,TG	FADS1	miR-212-5p	* High LD
rs1046875	✓✓	✓✓	✓✓	HbA1c	FN3KRP	miR-34a-5p	Selected
rs1046896	✓✓	✓✓	✓✓	HbA1c	FN3KRP	miR-208a-3p	*High LD
rs11735092	✓✓	✓✓	✓✓	TG	HSD17B13	miR-375	Selected
rs3217992	✓✓	✓✓	✓✓	CAD	CDKN2B	miR-138-3P	Selected
rs877710	✓✓	✓✓	✓✓	HDL	MMAB	miR-593-5p	Selected
rs1059611	✓✓	✓✓	✓✓	HDL,TG	LPL	miR-136-5p	Selected
rs907091	✓✓	✓	✓✓	HDL	IKZF3	miR-326	Selected
rs6857	✓✓	✓	✓✓	TG, LDL, HDL,TC	PVRL2	miR-320e	Selected
rs7956	✓	✓✓	✓✓	TC	MKRN2	miR-154-5p	Selected
rs13702	✓	✓✓	✓✓	HDL,TG	LPL	miR-410	Selected
rs11350	✓✓	✓✓	✓✓	HDL	SLC7A6	miR-548e-5p	Validated
rs11700	✓✓	✓✓	✓	HDL	E2F4	miR-520f-5p	**
rs1063966	✓	✓✓	✓✓	TG	GATAD2A	miR-30b-3p	**
rs1572343	✓	✓✓	✓✓	HDL	TMED5	miR-511-5p	**
rs4938353	✓	✓✓	✓✓	TG, TC, HDL	SIDT2	miR-183-3p	**
rs1050488	✓	✓✓	✓✓	TC	NSMAF	miR-196a-3p	**
rs1050504	✓	✓✓	✓✓	TC	NSMAF	miR-154-3p	**
rs1054284	✓	✓✓	✓✓	TC	GATAD2A	miR-215-3p	**
rs1810126	✓✓	✓	✓	CAD	SLC22A3	miR-124-3P	**
rs8102380	✓	✓	✓✓	LDL	ILF3	miR-101-3p	**
rs1054623	✓	✓	✓✓	HDL	MYO1F	miR-221-5p	**
rs9909	✓	✓	✓✓	Pro-Ins, HDL	NUP160	let-7f-2-3p	**
rs1052248	✓	✓	✓✓	TC	LST1	miR-153-5p	**
rs1057233	✓	✓✓	✓	Pro-Ins	SPI1	miR-655-5p	**
rs13964	✓	✓✓	✓	TG, TC, LDL	KIAA0892	miR-125a-5p	**
rs2285628	✓	✓✓	✓	TG, TC, LDL	KIAA0892	miR-199a-5p	**
rs2744937	✓	✓✓	✓	HDL,TC	C6orf106	miR-22-3p	**
rs9818870	✓	✓	✓	CAD	MRAS	miR-664a-3p	**
rs1109166	✓	✓	✓	HDL	LCAT	miR-622	**
rs4823006	✓✓	✓	✓✓	WHR	ZNRF3	miR-221	**†
rs879620	✓✓	✓	✓✓	BMI	ADCY9	miR-126-5p	**†

Supplementary Table S5. (continued)

SNP ID	GWAS	eQTL	miR Exp	Asso. Phenotype	Gene	miRNA	Functional
rs8024	✓✓	✓	✓✓	BMI	IPO9	miR-148-5p	**†
rs2531995	✓✓	✓	✓	BMI	ADCY9	miR-346	**†
rs12321	✓	✓	✓✓	WHR	ZNRF3	let-7f-2-3p	**†
rs4654	✓	✓	✓	WHR	ITPR2	miR-454-5p	**†

Shown are 37 binding site SNPs that fulfill our criteria for being likely functional in their corresponding loci. * These SNPs are in high LD with twoother SNPs selected for functional experiments; ** Candidate for future studies; †New GWAS data; Validated, the miR-mRNA interaction is validated by previous studies. Ten out of 37 SNPs were selected for experimental study based on the following criteria:
GWAS results (✓✓: top SNP or one of the top 10% SNPs with the strongest association with the phenotype on the given genomic loci, ✓: SNP is associated with the phenotype).
eQTL analysis (✓✓: a very significant association of SNP with the host gene expression in the right direction and in different relevant tissues, ✓: an association of SNP with the host gene expression).
miRNA expression (✓✓: evidence for the expression of the regulatory miRNA in multiple relevant cardio-metabolic tissues, ✓: evidence for miRNA expression in a relevant tissue.

Supplemental Table S6. The paired miRNA-target gene sequences candidate for functional experiments.

miRNA and target gene corresponding sequences		Binding site variant
miR-34a	3′ uguuggucgauucuGUGACGGu 5′	**rs1046875**
	|||||||	mirSVR score: -0.1433
FN3KRP (Major A allele)	5′ cauagguaccuuucC**A**CUGCCu 3′	PhastCons score: 0.4468
FN3KRP (Minor G allele)	5′ cauagguaccuuucC**G**CUGCCu 3′	
miR-181a-2	3′ ccauGUCA-GUUGCCAGUCACCa 5′	**rs174545**
	::|| || : ||||||	mirSVR score:-0.0681
FADS1 (Major G allele)	5′ ucacUGGUCCAUUUCUCAGUG**G**c 3′	PhastCons score:0.5359
FADS1 (Minor C allele)	5′ ucacUGGUCCAUUUCUCAGUG**C**c 3′	
miR-136	3′ agguaguaguuuUGUUUACCUCa 5′	**rs1059611**
	:: ||||||||	mirSVR score: -0.6354
LPL (Major T allele)	5′ augugguggauguGUAAA**U**GGAGc 3′	PhastCons score: 0.6850
LPL (Minor A allele)	5′ augugguggauguGUAAA**A**GGAGc 3′	
miR-138-2-3P	3′ uugggaccacaGCACUUUAUCg 5′	**rs3217992**
	||| | ||| |||||||	mirSVR score:-0.6073
CDKN2B (Major G allele)	5′ CCACCUGGUUGUAUU**G**AAAUAGu 3′	PhastCons score: 0.5099
CDKN2B (Minor T allele	5′ CCACCUGGUUGUAUU**U**AAAUAGu 3′	
miR-154	3′ gcUUCCGUUGUGCCUAUUGGAu 5′	**rs7956**
	|| | |:| :|||||	mirSVR score:-0.2392
MKRN2 (Major T allele)	5′ ugAAAGAUAUA-AAG**U**AACCUa 3′	PhastCons score: 0.5550
MKRN2 (Minor C allele)	5′ ugAAAGAUAUA-AAG**C**AACCUa 3′	
miR-593-5p	3′ ACUCGUUACGGA----CCGACCACGGa 5′	**rs877710**
	||: : ||||| | : |||||	mirSVR score:NA
MMAB (Major G allele)	5′ UGGCUUAUGCCUCUGAGAUC**G**GUGCCa 3′	PhastCons score: NA
MMAB (Minor C allele)	5′ UGGCUUAUGCCUCUGAGAUC**C**GUGCCa 3′	
miR-410	3′ uguccggUAGACACAAUAUAa 5′	**rs13702**
	|||||||||	mirSVR score:-1.1589
LPL (Major T allele)	5′ uccgaaaAACUUUGU**U**AUAUa 3′	PhastCons score: 0.6217
LPL (Minor C allele)	5′ uccgaaaAACUUUGU**C**AUAUa 3′	
miR-375	3′ agugCGCUCGGCUUGCUUGUUu 5′	**rs11735092**
	| ||:| ||||||||	mirSVR score: -0.7667
HSD17B13 (major A allele)	5′ agcaGC-AGUCAAACG**A**ACAAg 3′	PhastCons score: 0.5849
HSD17B13 (Minor G allele)	5′ agcaGC-AGUCAAACG**G**ACAAg 3′	
miR-326	3′ CCUCCAGUUCCCGGGUCUCC 5′	**rs907091**
	||| || |||||||	mirSVR score: NA
IKZF3 (Minor A allele)	5′ GGAAUGAGUGGUCCC**A**GAGA 3′	PhastCons score: NA
IKZF3 (Major G allele)	5′ GGAAUGAGUGGUCCC**G**GAGA 3′	
miR-320e	3′ GGAAGAGUUGGGUCGAAA 5′	**rs6857**
	|| |::|||||	mirSVR score: -0.085
PVRL2 (Minor T allele)	5′ CUGUUGAUCCCUUAGC**U**UUc 3′	PhastCons score: 0.5628
PVRL2 (Major G allele)	5′ CCTTGATGGCCCCAGC**C**TTG 3′	

REFERENCES

1. Isomaa B, Almgren P, Tuomi T, Forsen B, Lahti K, Nissen M, Taskinen MR, Groop L: Cardiovascular morbidity and mortality associated with the metabolic syndrome. Diabetes Care 2001, 24:683-689.
2. Global Lipids Genetics C, Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, Ganna A, Chen J, Buchkovich ML, et al: Discovery and refinement of loci associated with lipid levels. Nat Genet 2013.
3. Scott RA, Lagou V, Welch RP, Wheeler E, Montasser ME, Luan J, Magi R, Strawbridge RJ, Rehnberg E, Gustafsson S, et al: Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. Nat Genet 2012, 44:991-1005.
4. Consortium CAD, Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, Thompson JR, Ingelsson E, Saleheen D, Erdmann J, et al: Large-scale association analysis identifies new risk loci for coronary artery disease. Nat Genet 2013, 45:25-33.
5. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA: Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A 2009, 106:9362-9367.
6. Kumar V, Westra HJ, Karjalainen J, Zhernakova DV, Esko T, Hrdlickova B, Almeida R, Zhernakova A, Reinmaa E, Vosa U, et al: Human disease-associated genetic variation impacts large intergenic non-coding RNA expression. PLoS Genet 2013, 9:e1003201.
7. Hesketh J: 3'-Untranslated regions are important in mRNA localization and translation: lessons from selenium and metallothionein. Biochem Soc Trans 2004, 32:990-993.
8. Skeeles LE, Fleming JL, Mahler KL, Toland AE: The impact of 3'UTR variants on differential expression of candidate cancer susceptibility genes. PLoS One 2013, 8:e58609.
9. Ryan BM, Robles AI, Harris CC: Genetic variation in microRNA networks: the implications for cancer research. Nat Rev Cancer 2010, 10:389-402.
10. Richardson K, Nettleton JA, Rotllan N, Tanaka T, Smith CE, Lai CQ, Parnell LD, Lee YC, Lahti J, Lemaitre RN, et al: Gain-of-function lipoprotein lipase variant rs13702 modulates lipid traits through disruption of a microRNA-410 seed site. Am J Hum Genet 2013, 92:5-14.
11. Miller CL, Haas U, Diaz R, Leeper NJ, Kundu RK, Patlolla B, Assimes TL, Kaiser FJ, Perisic L, Hedin U, et al: Coronary heart disease-associated variation in TCF21 disrupts a miR-224 binding site and miRNA-mediated regulation. PLoS Genet 2014, 10:e1004263.
12. Bartel DP: MicroRNAs: target recognition and regulatory functions. Cell 2009, 136:215-233.
13. Ghanbari M, de Vries PS, de Looper H, Peters MJ, Schurmann C, Yaghootkar H, Dorr M, Frayling TM, Uitterlinden AG, Hofman A, et al: A genetic variant in the seed region of miR-4513 shows pleiotropic effects on lipid and glucose homeostasis, blood pressure, and coronary artery disease. Hum Mutat 2014, 35:1524-1531.
14. Abelson JF, Kwan KY, O'Roak BJ, Baek DY, Stillman AA, Morgan TM, Mathews CA, Pauls DL, Rasin MR, Gunel M, et al: Sequence variants in SLITRK1 are associated with Tourette's syndrome. Science 2005, 310:317-320.
15. Georges M, Clop A, Marcq F, Takeda H, Pirottin D, Hiard S, Tordoir X, Caiment F, Meish F, Bibe B, et al: Polymorphic microRNA-target interactions: a novel source of phenotypic variation. Cold Spring Harb Symp Quant Biol 2006, 71:343-350.
16. Bao L, Zhou M, Wu L, Lu L, Goldowitz D, Williams RW, Cui Y: PolymiRTS Database: linking polymorphisms in microRNA target sites with complex traits. Nucleic Acids Res 2007, 35:D51-54.
17. Hiard S, Charlier C, Coppieters W, Georges M, Baurain D: Patrocles: a database of polymorphic miRNA-mediated gene regulation in vertebrates. Nucleic Acids Res 2010, 38:D640-651.

18. Gauthier BR, Wollheim CB: MicroRNAs: 'ribo-regulators' of glucose homeostasis. Nat Med 2006, 12:36-38.
19. Zampetaki A, Mayr M: MicroRNAs in vascular and metabolic disease. Circ Res 2012, 110:508-522.
20. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, Wheeler E, Glazer NL, Bouatia-Naji N, Gloyn AL, et al: New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet 2010, 42:105-116.
21. Manning AK, Hivert MF, Scott RA, Grimsby JL, Bouatia-Naji N, Chen H, Rybin D, Liu CT, Bielak LF, Prokopenko I, et al: A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. Nat Genet 2012, 44:659-669.
22. Saxena R, Hivert MF, Langenberg C, Tanaka T, Pankow JS, Vollenweider P, Lyssenko V, Bouatia-Naji N, Dupuis J, Jackson AU, et al: Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. Nat Genet 2010, 42:142-148.
23. Soranzo N, Sanna S, Wheeler E, Gieger C, Radke D, Dupuis J, Bouatia-Naji N, Langenberg C, Prokopenko I, Stolerman E, et al: Common variants at 10 genomic loci influence hemoglobin A(1)(C) levels via glycemic and nonglycemic pathways. Diabetes 2010, 59:3229-3239.
24. Strawbridge RJ, Dupuis J, Prokopenko I, Barker A, Ahlqvist E, Rybin D, Petrie JR, Travers ME, Bouatia-Naji N, Dimas AS, et al: Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes. Diabetes 2011, 60:2624-2634.
25. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, Powell C, Vedantam S, Buchkovich ML, Yang J, et al: Genetic studies of body mass index yield new insights for obesity biology. Nature 2015, 518:197-206.
26. Shungin D, Winkler TW, Croteau-Chonka DC, Ferreira T, Locke AE, Magi R, Strawbridge RJ, Pers TH, Fischer K, Justice AE, et al: New genetic loci link adipose and insulin biology to body fat distribution. Nature 2015, 518:187-196.
27. International Consortium for Blood Pressure Genome-Wide Association S, Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, Chasman DI, Smith AV, Tobin MD, Verwoert GC, et al: Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature 2011, 478:103-109.
28. Gong J, Tong Y, Zhang HM, Wang K, Hu T, Shan G, Sun J, Guo AY: Genome-wide identification of SNPs in microRNA genes and the SNP effects on microRNA target binding and biogenesis. Hum Mutat 2012, 33:254-263.
29. Grundberg E, Small KS, Hedman AK, Nica AC, Buil A, Keildson S, Bell JT, Yang TP, Meduri E, Barrett A, et al: Mapping cis- and trans-regulatory effects across multiple tissues in twins. Nat Genet 2012, 44:1084-1089.
30. Friedman RC, Farh KK, Burge CB, Bartel DP: Most mammalian mRNAs are conserved targets of microRNAs. Genome Res 2009, 19:92-105.
31. Meenhuis A, van Veelen PA, de Looper H, van Boxtel N, van den Berge IJ, Sun SM, Taskesen E, Stern P, de Ru AH, van Adrichem AJ, et al: MiR-17/20/93/106 promote hematopoietic cell expansion by targeting sequestosome 1-regulated pathways in mice. Blood 2011, 118:916-925.
32. International HapMap C: A haplotype map of the human genome. Nature 2005, 437:1299-1320.
33. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ: LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 2010, 26:2336-2337.
34. Chen K, Rajewsky N: Deep conservation of microRNA-target relationships and 3'UTR motifs in vertebrates, flies, and nematodes. Cold Spring Harb Symp Quant Biol 2006, 71:149-156.

35. Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP: MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol Cell 2007, 27:91-105.
36. Wu C, Gong Y, Yuan J, Zhang W, Zhao G, Li H, Sun A, KaiHu, Zou Y, Ge J: microRNA-181a represses ox-LDL-stimulated inflammatory response in dendritic cell by targeting c-Fos. J Lipid Res 2012, 53:2355-2363.
37. Williams A, Henao-Mejia J, Harman CC, Flavell RA: miR-181 and Metabolic Regulation in the Immune System. Cold Spring Harb Symp Quant Biol 2013.
38. Rip J, Nierman MC, Wareham NJ, Luben R, Bingham SA, Day NE, van Miert JN, Hutten BA, Kastelein JJ, Kuivenhoven JA, et al: Serum lipoprotein lipase concentration and risk for future coronary artery disease: the EPIC-Norfolk prospective population study. Arterioscler Thromb Vasc Biol 2006, 26:637-642.
39. Zhang Z, Kang Y, Zhang Z, Zhang H, Duan X, Liu J, Li X, Liao W: Expression of microRNAs during chondrogenesis of human adipose-derived stem cells. Osteoarthritis Cartilage 2012, 20:1638-1646.
40. Poy MN, Spranger M, Stoffel M: microRNAs and the regulation of glucose and lipid metabolism. Diabetes Obes Metab 2007, 9 Suppl 2:67-73.
41. Zaragosi LE, Wdziekonski B, Brigand KL, Villageois P, Mari B, Waldmann R, Dani C, Barbry P: Small RNA sequencing reveals miR-642a-3p as a novel adipocyte-specific microRNA and miR-30 as a key regulator of human adipogenesis. Genome Biol 2011, 12:R64.
42. Harismendy O, Notani D, Song X, Rahim NG, Tanasa B, Heintzman N, Ren B, Fu XD, Topol EJ, Rosenfeld MG, Frazer KA: 9p21 DNA variants associated with coronary artery disease impair interferon-gamma signalling response. Nature 2011, 470:264-268.
43. Kral BG, Mathias RA, Suktitipat B, Ruczinski I, Vaidya D, Yanek LR, Quyyumi AA, Patel RS, Zafari AM, Vaccarino V, et al: A common variant in the CDKN2B gene on chromosome 9p21 protects against coronary artery disease in Americans of African ancestry. J Hum Genet 2011, 56:224-229.
44. Morton SU, Scherz PJ, Cordes KR, Ivey KN, Stainier DY, Srivastava D: microRNA-138 modulates cardiac patterning during embryonic development. Proc Natl Acad Sci U S A 2008, 105:17830-17835.
45. Horswell SD, Fryer LG, Hutchison CE, Zindrou D, Speedy HE, Town MM, Duncan EJ, Sivapackianathan R, Patel HN, Jones EL, et al: CDKN2B expression in adipose tissue of familial combined hyperlipidemia patients. J Lipid Res 2013, 54:3491-3505.
46. Chakraborty C, George Priya Doss C, Bandyopadhyay S: miRNAs in insulin resistance and diabetes-associated pancreatic cancer: the 'minute and miracle' molecule moving as a monitor in the 'genomic galaxy'. Curr Drug Targets 2013, 14:1110-1117.
47. Caussy C, Charriere S, Marcais C, Di Filippo M, Sassolas A, Delay M, Euthine V, Jalabert A, Lefai E, Rome S, Moulin P: An APOA5 3' UTR variant associated with plasma triglycerides triggers APOA5 downregulation by creating a functional miR-485-5p binding site. Am J Hum Genet 2014, 94:129-134.
48. Ling HY, Ou HS, Feng SD, Zhang XY, Tuo QH, Chen LX, Zhu BY, Gao ZP, Tang CK, Yin WD, et al: CHANGES IN microRNA (miR) profile and effects of miR-320 in insulin-resistant 3T3-L1 adipocytes. Clin Exp Pharmacol Physiol 2009, 36:e32-39.
49. Sebastiani G, Grieco FA, Spagnuolo I, Galleri L, Cataldo D, Dotta F: Increased expression of microRNA miR-326 in type 1 diabetic patients with ongoing islet autoimmunity. Diabetes Metab Res Rev 2011, 27:862-866.
50. Shu J, Xia Z, Li L, Liang ET, Slipek N, Shen D, Foo J, Subramanian S, Steer CJ: Dose-dependent differential mRNA target selection and regulation by let-7a-7f and miR-17-92 cluster microRNAs. RNA Biol 2012, 9:1275-1287.

CHAPTER 2.4

Genetic variants in long non-coding RNAs associate with cardiometabolic disorders

Mohsen Ghanbari, Marjolein J. Peters, Paul S. de Vries, Yu-Chi Lee, Cindy G. Boer, Jeroen G.J. van Rooij, Vinod Kumar, BIOS consortium, Andre G. Uitterlinden, Albert Hofman, Cisca Wijmenga,
Jose M. Ordovas, Caren E. Smith, Joyce B.J. van Meurs,
Stefan J. Erkeland, Oscar H. Franco, Abbas Dehghan

Manuscript based on this chapter

Integrative genome-wide analysis highlights the role of long non-coding RNAs in cardiometabolic disorders.

2017, Submitted.

ABSTRACT

Background. Genome-wide association studies (GWAS) have identified many susceptibility loci for cardiovascular risk factors and disease. Most of the associated variants reside in non-coding genomic regions including long non-coding RNAs (lncRNAs) that play critical regulatory roles in diverse biological processes. We aimed to investigate the extent to which genetic associations with cardiometabolic traits in non-coding regions may act through lncRNAs.

Methods and results. We leveraged data from the available GWAS on lipid and obesity-related traits, blood pressure, type 2 diabetes and coronary artery disease and identified 179 associated SNPs in 102 lncRNAs (p-value < 2.3×10^{-7}). Of these, 55 SNPs, either the sentinel SNP (n=17) or in strong linkage-disequilibrium with the sentinel SNP (n=38), were selected for further investigations. Our in-silico predictions and functional annotations for these SNPs as well as expression and DNA methylation analysis for their lncRNAs demonstrated that several lncRNA-SNPs fulfill predefined criteria for being functional in their loci. Further, we highlighted a number of loci that the current annotations in the literature have overlooked the role of lncRNAs in the association. For instance, we provide several pieces of evidence showing LOC157273 to be involved in regulating serum lipid levels. In particular, rs4841132 in the second exon and cg17371580 in the promoter of LOC157273 associate with lipids, the lncRNA is expressed in liver and regulates expression of its nearby-coding gene, *PPP1R3B*.

Conclusion. Our results show several cardiometabolic-associated SNPs likely to be functional and act through lncRNAs. Further, we highlight a number of loci for which the role of lncRNA may have been overlooked in the literature. The identified lncRNAs could serve as targets for future studies.

INTRODUCTION

Recent data from high-throughput sequencing platforms have established that much of the human genome has been transcribed [1, 2]. These data also display that protein-coding genes make only a very small proportion (~2%) of the genome, highlighting a lack of understanding in the possible contribution of non-coding elements to disease biology. Although the non-coding regions of the genome do not encode proteins, they are often transcribed into non-coding RNAs (ncRNAs) [3]. The ncRNAs can be roughly categorized, based on their transcript size, into small ncRNAs (such as microRNAs (miRNAs)) and long ncRNAs (lncRNAs) [4, 5]. The lncRNAs are defined as non-protein coding transcripts over 200 nucleotides in length [6, 7]. Similar to messenger RNAs (mRNAs), lncRNAs are transcribed by RNA polymerase II, 5'-capped and polyadenylated [8, 9]. Moreover, many of them contain multiple exons and are subjected to alternative splicing [3, 10]. Over the past few years, it has become increasingly evident that lncRNAs act as important players in the epigenetic, post-transcriptional, and translational coordination of gene expression in developmental processes and human diseases [11-14].The expression and function of a number of lncRNAs have been shown to be dysregulated in cardiovascular disease, such as *MHRT* in heart failure [15]. Genome-wide association studies (GWAS) have also reported relationship between lncRNA variants and cardiovascular disease, including variants in *CDKN2B-AS* (*ANRIL*) that are associated with the risk of coronary artery disease and variants in *MIAT* that confer susceptibility to myocardial infarction [16-18]. Nevertheless, despite some well-characterized lncRNAs, such as *ANRIL*, little is known about the general features of most lncRNAs and their possible molecular mechanisms in cardiometabolic disorders.

Large-scale GWAS have enabled the discovery of genetic variants at multiple loci associated with cardiovascular risk factors and disease [6, 19-21]. The vast majority of the associated variants map to non-coding regions and their biological relevance to the disease remain poorly understood [22]. The functional impact of non-coding variants are particularly difficult to study since the functional consequences of them should be investigated beyond the context of protein-coding genes [23, 24]. Recently, many studies have focused on functional annotation of GWAS hits that are localized in ncRNAs [23-27]. We and others have shown previously that variants in miRNAs can explain some of the GWAS associations with cardiometabolic disorders [28-31]. However, to date, the association of lncRNA variants with cardiometabolic disorders has not been investigated systematically. In the present study, we aimed to identify associations with cardiovascular disease and its risk factors in non-coding regions that may act through lncRNAs. To this end, we examined the association of genetic variants in lncRNAs with cardiometabolic disorders using publicly available data from the recent GWAS. We subsequently performed various *in silico* analyses and functional annotations (such as annotation to

regulatory features, eQTL and miRNA-binding sites) for the associated variants as well as expression and DNA methylation analysis for their lncRNAs to determine lncRNAs that might play a role in cardiometabolic disorders.

MATERIALS AND METHODS

Identification of genetic variants in lncRNAs

We extracted genetic variants that are located in human lncRNA transcripts using the lncRNASNP database [31]. This is a comprehensive database that includes 495,729 SNPs in 32,108 lncRNA transcripts of 17,436 lncRNA genes. This database is based on two online resources: dbSNP (build 138) (http://www.ncbi.nlm.nih.gov/SNP/) and LNCipedia (http://www.lncipedia.org/). We included only SNPs with minor allele frequencies (MAF) > 0.01 in our analysis. We focused on the 21,915 SNPs (in 8,829 lncRNAs) that are present in HapMap (release 22) [32]. A flow chart of our approach to detect variants in lncRNAs that are associated with cardiovascular traits is visualized in **Figure 1**.

Genome-wide association studies on cardiometabolic disorders

Data from the publicly available GWAS meta-analyses of 10 cardiometabolic disorders were used for association analysis. The cardiometabolic disorders in our analysis included four lipid traits (plasma concentrations of total cholesterol, low-density lipoprotein cholesterol (LDL), high-density lipoprotein cholesterol (HDL), and triglycerides) from the Global Lipids Genetics Consortium (GLGC) [19], two anthropometric measures (body mass index (BMI) and waist to hip ratio (WHR)) from the Genetic Investigation of ANthropometric Traits (GIANT) consortium [33, 34], systolic and diastolic blood pressure from the Global BPgen consortium [21], coronary artery disease (CAD) from the CARDIoGRAMplusC4D consortium [6], and type 2 diabetes (T2D) from the DIAGRAM consortium [20]. Descriptions of the GWAS meta-analyses, the consortia responsible for these analyses, the sample sizes and the number of variants associated with each trait are provided in **Table S1**. The number of statistical tests was calculated by multiplying the number of examined SNPs by the number of traits (21,915×10). The Bonferroni correction method was then used to establish the significance threshold (p-value < 2.3×10^{-7}). Manhattan and regional plots showing the association of lncRNA variants and flanking variants in the corresponding loci with cardiometabolic disorders were generated by R and the LocusZoom web tool [35]. The conditional and joint association analysis was performed using the GCTA tool [36] to explore whether the lncRNA variants and their proxies in strong LD drive independent signals, the possibility of multiple causal variants at the associated loci (see supplementary methods for details).

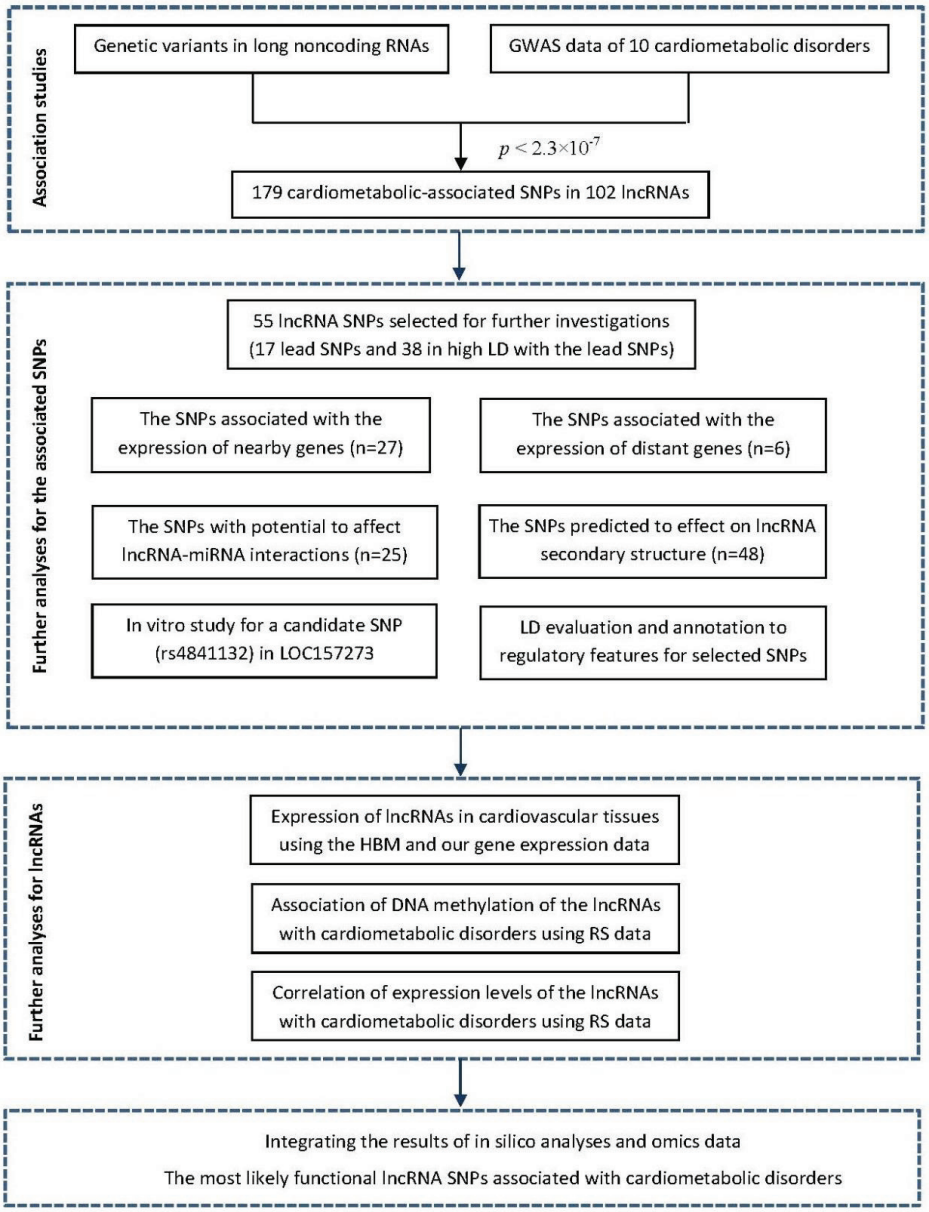


Figure 1. Identification and functional analysis of lncRNA variants associated with cardiovascular traits

Association of lncRNA variants with gene expression

The correlation between lncRNA variants and expression levels of hosting lncRNAs, nearby and distant coding genes were examined in blood and adipose tissue. The expression quantitative trait loci analysis (eQTL) was done in adipose tissue, since a high percentage

of the identified variants were found to be associated with lipid traits, BMI and WHR. To this end, data from samples of 856 healthy female twins included in the MuTHER Study was used [37]. In addition, associations of the identified variants with gene expression levels in whole blood were examined in 5,311 individuals, data was accessed through the Gene Network database (http://genenetwork.nl/blood eqtlbrowser/) [38]. For cis-eQTL analysis (< 1MB), the significance threshold was set by dividing the significance level of 0.05 by the number of studied SNPs. For trans-eQTL analysis (≥ 1MB), given the large number of tests (the number of lncRNA SNPs multiplied by 23,644 probes for all transcripts) and the relatively small effect sizes of trans-eQTLs, a false discovery rate (FDR) of 1% was used [37].

Annotating the regulatory features of lncRNA variants

The identified lncRNA variants were annotated to regulatory features, including promoter and enhancer regulatory motifs, DNase foot printing sites and conserved sequences using the HaploReg v4.0 (http://www.broadinstitute.org/mammals/haploreg/ haploreg.php). Further, for each of the variants, the LD region ($R^2 > 0.8$) was determined using the 1000G Phase 1 population. For each set of variants in strong LD with a given lncRNA variant, we investigated whether the variant was located in a potential regulatory region using the Roadmap consortium reference epigenomes dataset (see supplementary methods for details) [39, 40]. The reference epigenomes used were chromatin state models based on imputed data of 12 histone marks [40], identifying separate epigenetic chromatin states for each of the lncRNA variants. Heat maps were constructed by calculating the percentage of variants in LD with the GWAS found variant located in DNA regulatory regions as defined by the chromatin states. In addition, to determine the functionality of the GWAS hits, we investigated whether known protein-coding variants were in high LD with the associated lncRNA variants. We examined the genes surrounding the variants using the publicly available database, Online Mendelian Inheritance in Man (OMIM) [41], to see if any of the surrounding genes are known to be involved in Mendelian cardiovascular disorders.

Potential impact of lncRNA variants on the interaction between lncRNAs and miRNAs

To predict whether the identified lncRNA variants are located in putative miRNA-binding sites and could potentially affect the binding of miRNAs to lncRNAs, the lncRNASNP database was used [32]. Since a functional prerequisite for miRNA-dependent gene regulation is that the miRNA and its target gene are co-expressed in the same tissue, multiple web tools were used to investigate whether lncRNAs and related miRNAs in our list were expressed in relevant tissues (see supplementary methods for details). Additionally, we performed Luciferase reporter assays for one of the identified lncRNA

variant to examine the interaction between the lncRNA and related miRNA and to determine the effect of the variant on the lncRNA-miRNA interaction. To do this, primers were designed to amplify the lncRNA sequence (wild type and mutated), the forward and reverse primers included the restriction enzyme sites XbaI and ApaI, respectively. All primers are shown in **Table S2**. The lncRNA sequences containing the binding site of miRNA, were amplified and cloned into the pGL3 Luciferase reporter vector (Promega) downstream of the Luciferase open reading frame [42]. The inserts of all constructs were confirmed by Sanger sequencing. The HepG2 cells (n=10,000) were plated into 96-well plates and co-transfected with 1 μg of pGL3 containing the lncRNA sequence with either the major or the minor allele, miRNA mimic (mirVana TM Mimics) and a plasmid expressing the Renilla Luciferase, which served as transfection control, using LipofectamineR RNAiMAX (Invitrogen). Luciferase activity was determined using the Dual-Glo Luciferase Assay System according to manufacturer's protocol (Promega). Renilla activity was used for normalization of the data. All experiments were performed in triplicate.

Prediction of the effect of lncRNA variants on the lncRNA secondary structure

For variants that were found to be significantly associated with cardiometabolic traits, we used both the lncRNASNP database and the RNAsnp web tool to predict the impact of these variants on the secondary structure of host lncRNAs [31, 42]. The lncRNA transcript sequences were extracted from human reference genome (GRCh37/hg19) according to the lncRNA transcript BED file as reference transcripts. We changed the corresponding allele in the given reference-transcript to the alternative allele as alternative-transcript. The RNAfold program (http://rna.tbi.univie.ac.at/) was subsequently used to calculate the minimal free energy (MFE, ΔG) and illustrate the lncRNA secondary structure. Energy change of RNA structures (ΔΔG) was quantified as the difference between the level of MFE of the thermodynamic ensemble of alternative versus reference transcript (ΔΔG= ΔG alternative transcript - ΔG reference transcript).

Expression of lncRNAs in cardiometabolic tissues

The lncRNA transcripts, on average, express at lower levels and show a higher degree of tissue specificity in comparison to protein-coding transcripts [43, 44]. Hence, it is important to examine the expression of lncRNAs in a relevant tissue. At present, the expression levels of lncRNAs are only available for a limited number of tissues. The Human Body Map catalog is one of the most complete set of expression data for lncRNAs that includes the RNA-seq across 22 human tissues and cell lines [45]. In addition, we used gene expression data from 83 subcutaneous adipose tissue (SAT), 77 visceral adipose tissue (VAT), 74 liver samples and 62 muscle skeletal samples from our cohort of 85 unrelated obese Dutch individuals [44, 46]. This analysis enabled us to detect expression levels of lncRNAs with probe available in the Illumina Human HT12v3 microarray.

Association of DNA methylation and expression of lncRNAs with cardiometabolic disorders

We extracted the cytosine-phosphate-guanine (CpG) sites within (+/-1 kb) of the identified lncRNAs and examined the association of DNA methylation levels at these CpG sites with cardiometabolic disorders using data from the Rotterdam Study (RS). The design of the Rotterdam Study has previously been described elsewhere [47]. Briefly, the RS is a prospective, population-based cohort study that comprises three cohorts. The initial cohort (RS-I) started out in 1990 with 7,983 participants aged 55 years and over from the neighborhood Ommoord in Rotterdam. In 2000-2001, a second cohort was established (RS-II) with 3,011 inhabitants. The third cohort (RS-III) started in 2006 with 3,932 inhabitants aged 45 years and over. In the current study, we used the methylation data from a set of 767 participants from the third visit of the second cohort (RSII-3) and the second visit of the third cohort (RSIII-2) as the discovery panel (See Supplementary Methods for details). Moreover, we sought replication in a set of 731 participants from RS-III-1 with DNA methylation data available [38, 48]. None of the participants included in the replication study were included in the discovery cohort. Methylation levels were measured in whole blood using the Illumina Methylation 450 array. Associations between cardiometabolic disorders and DNA methylation beta-values were examined using linear mixed-effect models. All models were adjusted for sex, age, smoking, white blood cell proportions, array number, and position on array. The baseline characteristics of participants in the study are shown in **Table S3**. In addition, we examined the association between expression levels of the lncRNAs (with probe available in the Illumina Human HT12v3 microarray) and cardiometabolic disorders using gene expression data from a random subset of 881 individuals in RSIII-1. A linear regression model was used to examine the association between the traits and the lncRNA expression levels, adjusting for age, sex, BMI, RNA Quality Score, plate ID, and cell counts.

RESULTS

lncRNA variants associated with cardiometabolic disorders

A flowchart of our approach to identify lncRNA variants associated with cardiometabolic disorders is visualized in **Figure 1**. We examined the associations of 21,915 SNPs (with MAF > 0.01 and available in the GWAS data) in 8,829 lncRNAs with lipid traits, BMI, WHR, systolic and diastolic blood pressure, T2D and CAD. Manhattan plots showing the associations of these lncRNA SNPs and cardiometabolic disorders are presented in **Figure S1**. We found significant associations between 179 unique SNPs in 102 lncRNAs and the studied traits (*p*-value threshold of 2.3×10^{-7}) (**Table S4**). Out of the 179 associated SNPs, 90 SNPs are located in long intergenic ncRNAs (lincRNAs), 53 in intronic lncRNAs and

36 in other lncRNA subtypes, including antisense transcripts. Ninety-nine of the 102 lncRNAs hosting the associated SNPs are located in the loci that have been previously reported in the original GWAS [6, 19-21, 33, 34]. Here, we report three additional loci for cardiometabolic disorders (*p*-value between 5×10^{-8} and 2.3×10^{-7}). These include lnc-XRCC3-1 (rs709400, C/T, chr14:103683138) associated with BMI, lnc-CACNA1E-2 (rs7537211, A/G, chr1:182089421) associated with HDL, and lnc-MAP3K4-1 (rs9355860, A/G, chr6:160969755) associated with LDL (**Table S5**). Our analysis showed that from the 179 associated SNPs, 17 are lead SNPs showing the strongest association with the traits (**Figure S2**) and 38 are in strong LD ($R^2 > 0.8$) with the lead SNPs in the related loci. We used GCTA and explored the rest of the SNPs (n=119) for the possibility of a secondary signal driven by the lncRNA SNPs independent of the lead SNP. Using this approach, we did not find any of the 119 SNPs to drive an independent signal. We thus focused on the 55 variants (17 lead SNPs and 38 proxy-lead SNPs) hosting by 55 different lncRNAs for further investigations (**Table S6**).

Association of lncRNA variants with gene expression

The cis-eQTL data were used to determine the correlation between 55 selected SNPs and expression levels of their nearby genes (<1 MB). The significance threshold for this analysis was set at *p*-value < 9.1×10^{-4} (0.05/55). Twenty-seven SNPs were found to be significantly associated with the expression levels of their nearby genes either in blood (26 SNPs) (**Table S7**) or adipose tissue (17 SNPs) (**Table S8**). Out of the 17 SNPs with cis-eQTL in adipose tissue, 16 SNPs showed the eQTL also in blood, while rs709400 in lnc-XRCC3-1 associated with BMI had the eQTL effect only in adipose tissue. Some of the eQTL SNPs were found to be associated with the expression levels of multiple nearby genes (sometimes in both tissues) including rs3099844 in linc-HLA-B-1, rs3177647 in lnc-TMEM116-2 (also known as linc-BRAP-2), rs9925964 in lnc-BCKDK-1, rs9934328 in lnc-ARGP-2, rs2301826 in lnc-RCCD1-1, rs2898290 in lnc-BLK-1, and rs6760828 in lnc-SNX17-1.

We subsequently checked the correlation between 55 selected SNPs and 23,644 probes for all transcripts (trans-eQTL analysis) in adipose tissue and whole blood. We used a false discovery rate (FDR) of 1%, which corresponded to p-value < 2.0×10^{-6} in adipose tissue and p-value < 1.5×10^{-7} in blood. Four SNPs were found to be significantly associated with the expression levels of distant genes in the adipose tissue, including rs17427875 (in lnc-EVX1-5) with *C19orf12*, rs2070959 (in lnc-HJURP-2) with *PRDM1*, rs456598 (in lnc-SLC22A2-1) with *BPIL1* and *UCP3*, and rs7537211 (in lnc-CACNA1E-2) with *PHF21A* and *ZNF746* (**Table S9**). Further, two SNPs were significantly associated with the expression levels of distant genes in blood, including rs243018 (in lnc-BCL11A-3) with *ESPN* and rs3099844 (in lnc-ATP6V1G2) with *TMEM154, BTN3A2* and *HIST1H2BD* (**Table S9**).

Functional annotations and in silico predictions for the selected lncRNA variants

We used HaploReg v4.0 to explore alterations in regulatory motifs, including promoter and enhancer sequences, associated with the 55 selected SNPs. This analysis showed that most of the SNPs were not in strong LD with any non-synonymous variants in the related loci (**Table S10**). For each set of variants in high LD with a given lncRNA SNP, we further examined if the variants are located in a potential regulatory region using the Roadmap consortium reference epigenome data set across relevant tissues (**Table S10**). Our analysis indicated that while most of the SNPs are not preferentially located in regulatory regions in cardiometabolic-related tissues, some of the (potential) promoter or enhancer regions, in which the SNPs are located, are enriched for the specific cell types (**Figure 2**). For example, the regulatory region in which the lipid-associated SNP rs4841132 in lnc-RP11-10A14.4 (LOC157273) is located, was enriched in liver, which is a relevant tissue for lipid metabolism. Likewise, the regulatory region in which rs10965215 in lnc-MTAP-1 (*CDKN2B-AS1*), associated with coronary artery disease, is located, was enriched in fetal heart.

LncRNAs have been shown to indirectly regulate the expression levels of coding genes either by acting as miRNA decoys [32, 49] or as competitors for miRNAs by binding to their target genes [50]. Our in silico analysis showed that 25 out of the 55 selected SNPs are localized in sequences that match to the seed sequence of miRNAs (**Table S11**). It has been shown that lncRNAs can only function as molecular sponges for miRNAs, if the localization of the lncRNA is appropriate and the lncRNA expression level is sufficiently high [51, 52]. We thus tested expression levels of the lncRNAs (with probe available in the Illumina microarray) harboring SNPs in putative miRNA-binding sites in cardiovascular tissues using gene expression data from our cohort of 85 Dutch individuals [45, 46]. We found that five of the lncRNAs are abundantly expressed in adipose tissue, liver and skeletal muscle (**Table S12**). Using the RNA-seq data from the Human Body Map catalog, we found evidence for expression of another four lncRNAs in cardiovascular relevant tissues (**Figure S3**). Further, we used multiple miRNA expression databases to identify evidence for expression of the interacting miRNAs in cardiovascular tissues. The lncRNAs and miRNAs that are expressed in relevant tissues and their putative interactions might be affected by the lncRNA SNPs are highlighted in **Table S11**.

Previous studies have shown correlations between the presence of variants in lncRNAs that affect the secondary structure and expression of the lncRNAs, and the development of specific diseases [53-55]. We used both the lncRNASNP database and the RNAsnp web tool to predict the effects of the 55 selected SNPs on the secondary structures of the lncRNAs. Differences in minimum free energy (ΔMFE > 0.1) of the thermodynamic ensemble of mutant versus wild-type lncRNA sequences were observed for 48 out of the 55 SNPs (**Table S13**), suggesting them as structure-disruptive variants. Since the

main functions of lncRNAs include interaction with DNA, RNA and protein, the secondary structure is important for the normal lncRNA functions. It can be hypothesized that disruption of the secondary structure by these SNPs might change the processing or function of the lncRNAs [53].

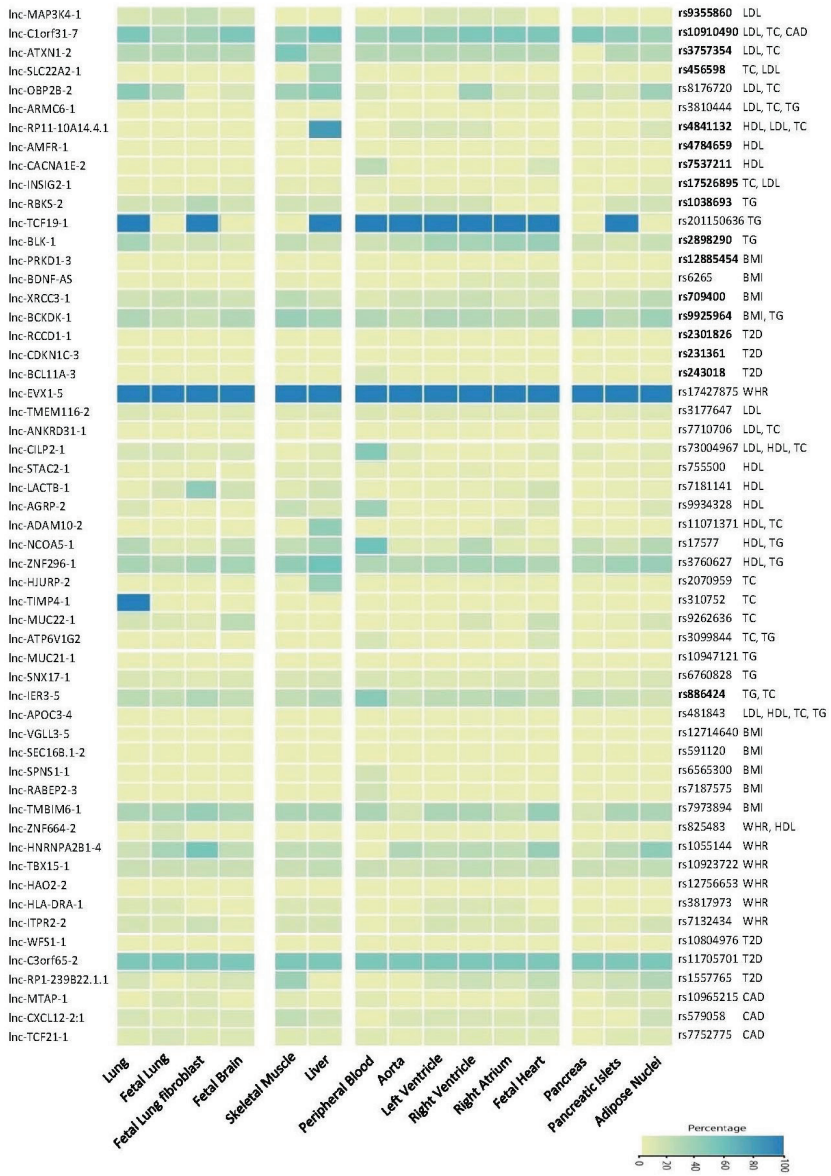


Figure 2. The heat map was constructed by calculating the percentage of 55 lncRNA SNPs and their proxies in strong LD (R^2> 0.8) located in DNA regulatory regions as defined by the chromatin states using the Roadmap consortium reference epigenome data set across different tissues.

Table 1. The association of top 10 CpG methylation sites annotated to lncRNAs with cardiovascular traits

lncRNA ID	Asso. trait in GWAS	All CpGs	CpG stie	Chr:position	Trait	Discovery (n=767)		Replication (n=731)		Meta-analysis *p*-value
						Regression coefficient	*p*-value	Regression coefficient	*p*-value	
lnc-RP11-10A14.4.1 (LOC157273)	HDL, TC, LDL	3	cg17371580	Chr8:9182401	HDL	0.0091	6.98×10^{-5}	0.0802	2.04×10^{-2}	**3.65×10^{-5}**
					LDL	-0.0028	6.43×10^{-3}	-0.0694	4.15×10^{-2}	5.20×10^{-3}
lnc-NCOA5-1 (RP11-465L10.10)	HDL, TG	21	cg07191594	Chr20:44650220	HDL	-0.0119	3.49×10^{-4}	-0.0424	4.69×10^{-1}	2.83×10^{-4}
			cg17920479	Chr20:44650246	HDL	-0.0103	9.33×10^{-4}	-0.0199	735×10^{-1}	8.40×10^{-4}
					TG	0.0036	2.76×10^{-3}	0.0413	4.95×10^{-1}	2.55×10^{-3}
lnc-CDKN1C-3 (KCNQ1OT1)	T2D	67	cg13444966	Chr11:2705357	HDL	0.0090	3.13×10^{-5}	0.0058	8.37×10^{-1}	2.70×10^{-5}
			cg05740879	Chr11:2721866	T2D	-0.0073	6.92×10^{-3}	-0.0037	1.77×10^{-1}	3.98×10^{-3}
lnc-MTAP-1 (CDKN2B-AS1)	CAD	10	cg08390209	Chr9:22005563	HDL	-0.0130	6.88×10^{-5}	-0.0687	2.69×10^{-1}	4.92×10^{-5}
					CHD	-0.0061	3.53×10^{-1}	-0.0227	1.11×10^{-1}	1.38×10^{-2}
lnc-SPNS1-1 (RP11-264B17.3)	BMI	18	cg00474657	Chr16:28989308	T2D	-0.0090	9.46×10^{-3}	-0.0078	2.31×10^{-1}	5.61×10^{-4}
					BMI	-0.0004	8.30×10^{-2}	-0.0004	5.07×10^{-1}	8.90×10^{-3}
lnc-ANKRD31 (RP11-229C3.2)	LDL, TC	8	cg11926485	Chr5:74346260	HDL	-0.0070	5.68×10^{-5}	-0.0378	1.71×10^{-1}	3.61×10^{-5}
					TC	-0.0016	2.96×10^{-2}	-	-	2.96×10^{-2}
lnc-IER3-5 (LINC00243)	TG, TC	11	cg22028727	Chr6:30796243	TG	0.0038	4.61×10^{-3}	0.0245	6.94×10^{-1}	4.35×10^{-3}
lnc-XRCC3-1 (RP11-894P9.1)	BMI	1	cg15322516	Chr14:104145810	HDL	-0.0096	7.91×10^{-3}	-0.0532	5.26×10^{-1}	7.16×10^{-3}
					BMI	0.0006	7.82×10^{-2}	0.0006	1.93×10^{-1}	2.83×10^{-2}

Values are regression coefficients based on linear mixed models and reflect differences in methylation beta values per increase in unit related to each traits. Models are adjusted for age, gender, current smoking, leukocyte proportions, array number, and position on array. We adjusted the p-value for the number of CpGs in each of the studied lncRNAs. In addition, the level of significance is p-value < 9.65×10^{-5} which is based on 0.05 / 518 CpGs in all studied lncRNAs.

Association of DNA methylation and expression of lncRNAs with cardiometabolic traits

We assessed the association between 518 CpG methylation sites annotated to 38 of the identified lncRNAs and cardiometabolic traits using data from the Rotterdam Study. **Table 2** shows the top 10 CpGs associated with the studied traits which are in line with the GWAS findings. After applying the Bonferroni correction, the significance cut-off sets at 9.65×10-5 (0.05 / 518), only one CpG remained significant. This includes cg17371580 annotated to lnc-RP11-10A14.4.1 (LOC157273) that was associated with HDL-cholesterol (meta-analysis p-value= 3.65×10^{-5}, Effect= 0.0094). The associations of all CpG sites nominally associated with cardiometabolic traits are shown in **Table S14**. Next, we assessed the association between expression level of 13 of the lncRNAs, with probe available in the Illumina Human HT12v4 microarray, and cardiometabolic traits using the Rotterdam Study data. After applying the Bonferroni correction, the level of significance sets at 3.85×10^{-3} (0.05 / 13), we found no significant association (**Table S15**). The expression levels of two lncRNAs were nominally associated with their traits, including lnc-IER3-5 (ILMN_1657996) associated with total cholesterol (p-value= 4.65×10^{-3} and effect= -4.09) and lnc-MTAP-1 (*CDKN2B-AS*) (ILMN_2376723) associated with CAD (p-value= 1.24×10^{-2} and effect= 0.42).

Prioritizing the potential functional lncRNA variants associated with cardiometabolic disorders

We integrated the results of our analyses for the 55 selected lncRNA SNPs and prioritized the variants that are more likely to affect cardiometabolic disorders through lncRNAs (**Table S16**). A summary overview of our findings for the top 20 lncRNA SNPs associated with cardiometabolic disorders is provided in Table 2. The SNP with the highest credibility for functionality in this list is rs4841132 located in LOC157273 (**Figure 3**). One of the prediction for rs4841132 was to affect the binding of LOC157273 to miR-1287. We performed Luciferase reporter assay to test the regulatory interaction between these two types of ncRNAs, and to investigate the impact of SNP on the lncRNA-miRNA interaction. Co-transfection of LOC157273 luciferase constructs with miR-1287 mimic in Hep2G cells showed no significant difference between the luciferase activities of the constructs containing the major and the minor allele at the SNP site rs4841132 (**Figure S4**). This experiment suggests that the effect of rs4841132 on the binding of LOC157273 and miR-1287 is minor and not detectable with this experimental setup. The negative result can also be explained by compensation of the miR-1287 binding to LOC157273 by match sequences outside the miRNA seed sequence (**Figure S4**) [56].

Table 2. An overview of the results of our analyses for thetop 20 lncRNA SNPs associated with cardiovascular traits

SNP ID	lncRNA ID	Associated trait(s)	Proxy SNPs		eQTL analysis		Regulatory features		miRNA binding site		ΔMFE	Exp.	Rotterdam Study data	
			All	Non-syn	Cis	Trans	Promoter	Enhancer	Disruption	Creation			Expression	Methylation
rs4841132	lnc-RP11-10A14.4	Lipids	7	0	1	-	Yes	Yes	miR-1287	No	0.41	Liver	-	Lipids
rs3757354	lnc-ATXN1-2	LDL, TC	3	0	2	-	Yes	Yes	No	miR-370	1.20	Low	n/a	LDL
rs709400	lnc-XRCC3-1	BMI	20	1	2	-	No	No	-	-	1.30	n/a	n/a	BMI, CMD
rs886424	lnc-IER3-5	TG,TC	50	0	4	-	Yes	Yes	No	miR-452	0.20	n/a	TC	TG, CMD
rs231361	lnc-CDKN1C-3	T2D	0	0	-	-	Yes	Yes	-	-	0.0	Low	-	T2D, CMD
rs9925964	lnc-BCKDK-1	TG, BMI	20	1	6	-	Yes	Yes	-	-	0.11	n/a	n/a	WHR
rs12885454	lnc-PRKD1-3	BMI	20	0	-	-	Yes	Yes	No	miR-514a	2.50	Low	n/a	CMD
rs9355860	lnc-MAP3K4-1	LDL	20	0	-	-	Yes	Yes	-	-	1.65	Low	CMD	CMD
rs4784659	lnc-AMFR-1	HDL	3	0	2	-	No	No	miR-18a-5p	No	2.10	n/a	n/a	n/a
rs7537211	lnc-CACNA1E-2	HDL	8	0	-	2	No	No	miR-152-5p	miR-760	0.30	Low	n/a	n/a
rs10910490	lnc-C1orf31-7	LDL, TC	13	0	-	-	Yes	Yes	No	miR-544a	0.80	Liver	n/a	CMD
rs17526895	lnc-INSIG2-1	LDL, TC	100	1	-	-	Yes	Yes	miR-580-5p	No	0.40	Low	n/a	n/a
rs456598	lnc-SLC22A2-1	LDL, TC	2	0	-	2	No	No	miR-432-3p	No	0.0	Liver	n/a	n/a
rs1038693	lnc-RBKS-2	TG	7	0	-	-	Yes	Yes	miR-4672	No	0.0	liver	n/a	n/a
rs2898290	lnc-BLK-1	TG	20	0	4	-	Yes	Yes	miR-141-3p	No	0.10	Low	n/a	-
rs2301826	lnc-RCCD1-1	T2D	1	0	2	-	No	No	-	-	0.80	n/a	-	CMD
rs243018	lnc-BCL11A-3	T2D	10	0	-	1	No	No	-	-	1.70	Low	n/a	CMD
rs1055144*	lnc-HNRNPA2B1	WHR	20	0	-	-	Yes	Yes	-	-	3.64	Low	-	n/a
rs10965215*	lnc-MTAP-1	CAD	40	0	-	-	No	No	-	-	0.40	Low	CHD	CHD, CMD
rs17427875*	lnc-EVX1-5	WHR	10	0	1	1	No	No	-	-	0.84	Adipose	-	WHR

LDL, Low Density Lipoprotein; HDL, High Density Lipoprotein; TC, Total cholesterol; TG, Triglyceride; BMI, Body Mass Index; WHR, Waist to Hip Ratio; T2D, Type 2 Diabetes; n/a, Not available; Non-syn, Non-synonymous proxy; ΔMFE, Minimum free energy change; eQTL, expression quantitative trait loci; Cis and Trans, association of SNP with the expression level of nearby or distant genes (number of genes); The traits mentioned for expression and DNA methylation analysis using the RS data show nominally association (CMD, association with cardiometabolic disorders but not with the GWAS associated trait). Exp, gene expression data from the Dutch cohort and RNA-seq data from the Human Body Map catalog (Low, expressed at low level).

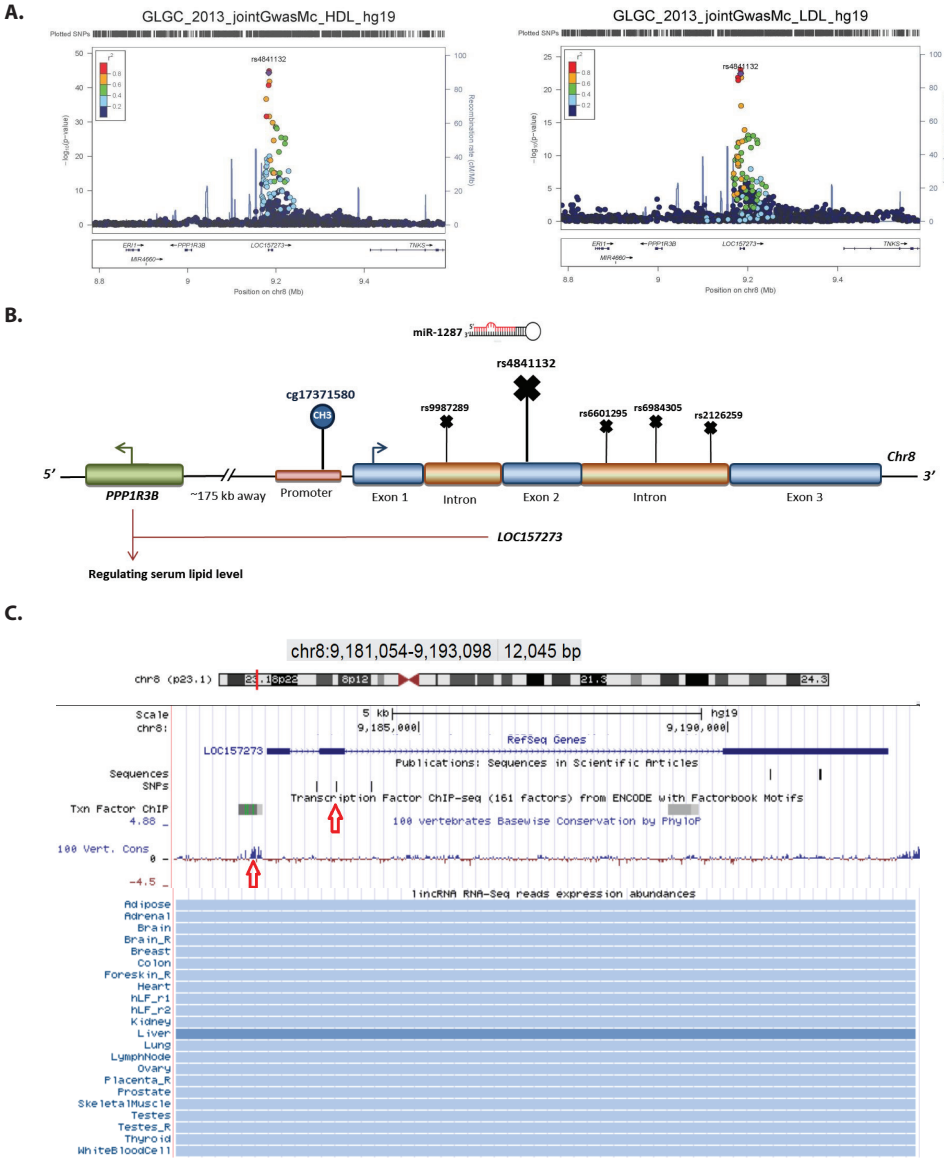


Figure 3. Functional evidence supporting the role of LOC157273 in regulating cholesterol and lipid metabolism.

A) The regional association plots show the association of variants in LOC157273 with HDL and LDL-cholesterol. **B)** A schematic showing the lncRNA structure, the nearest protein coding gene PPP1R3B, the positions of SNPs and CpG site associated with lipid traits. **C)** A figure generated by the UCSC Genome Browser on 8p23.1 locus hosting LOC157273. The position of SNP rs4841132 within the second exon of LOC157273 and CpG site cg17371580 within the conserved promoter region of lncRNA are depicted by red arrows. The figure also illustrates expression of LOC157273 across different tissues using the Human Body Map catalog, indicating that the lncRNA is exclusively expressed in liver.

DISCUSSION

In a systematic investigation of lncRNAs associated with cardiometabolic disorders, we identified 179 associated SNPs in 102 lncRNAs. We selected 55 of the SNPs, either the lead SNP or in strong LD with the lead SNP in the corresponding locus, and performed various in silico analyses for the SNPs as well as expression and DNA methylation studies for the lncRNAs in relation to cardiometabolic disorders. Our results demonstrated a number of lncRNA SNPs that fulfill predefined criteria for being potentially functional in their loci, emphasizing the importance of lncRNAs in the genetic susceptibility of cardiovascular disease. Further, we highlighted some of the loci harboring these SNPs for which genetic associations have either been incorrectly annotated to the closest protein-coding gene in the literature or the mediatory role of lncRNAs have been overlooked.

Our results demonstrated 179 lncRNA SNPs that are significantly associated with cardiometabolic disorders. Many of these lncRNA SNPs were localized adjacent to protein-coding genes. Due to this close proximity to the coding regions, it was difficult to ascertain whether the observed GWAS association is driven by the lncRNA SNP or its proxy coding variants in the corresponding locus. We therefore limited the SNP set to those that are either the sentinel SNP or one of the top SNPs in high LD with the sentinel SNP in the associated loci. This filtration was applied since in case of a single functional variant, the functional SNP is expected to show the strongest association with the trait. In case of multiple functional variants in one locus, we should have either identified multiple independent signals or find the functional variants within the top SNPs [36]. This filtration enriched the set of associations for those that are more likely to be mediated by the lncRNAs. Notably, the majority of 55 selected SNPs that remained after this filtration had no non-synonymous proxy variants in their vicinity.

In order to examine whether lncRNA SNPs could functionally affect the lncRNAs, several criteria have been suggested [27, 32, 51, 57]. An established association between the variant and a trait, the correlation of variant with expression of the lncRNA or protein-coding genes, the localization of variant in regulatory regions and the potential of variant for structural perturbations within the lncRNA are a number of such criteria. These criteria mainly assess the functionality of SNPs and extend them to the host lncRNA. As a second step, we investigated whether the identified lncRNAs are related to the trait of interest by performing expression and DNA methylation analyses. This integrative approach enabled us to highlight a number of genetic associations that are likely to affect cardiometabolic disorders through lncRNAs. Our results further posited that at least some of these associations may have been incorrectly annotated in the literature, where the nearest protein-coding gene to the SNP is reported as the trait-associated gene, while the lncRNA directly encompassing the SNP is ignored or the mediatory role of lncRNA is overlooked. For example, a number of GWAS have reported genetic as-

sociations for lipid levels, fasting insulin, fasting glucose and obesity in a chromosome 8 locus (8p23.1) [19, 58-61]. The association was annotated to *PPP1R3B* in the literature. In our study, we found a CpG site within the conserved promoter region of LOC157273 to be significantly associated with serum HDL-cholesterol. The lead SNPs in this locus are rs4841132 and rs9987289, which are in perfect LD (R2=1). Rs4841132 is located in the second exon of LOC157273, the nearest protein-coding gene is PPP1R3B (~175 kb away) and its variants are not in LD with the SNP. Our fine-mapping further showed that rs4841132 is mapped to promoter histone marks in liver, a relevant tissue for lipid metabolism. The second SNP, rs9987289, resides in the first intron of the lncRNA and is a cis-eQTL for *PPP1R3B* in human liver [58, 62]. In addition, Lipovich et al have shown more recently that LOC157273 knockdown increased the *PPP1R3B* mRNA levels in hepatocytes, indicating that the lncRNA represses *PPP1R3B* expression [63, 64]. *PPP1R3B* is a relevant gene for the associated traits since it encodes protein phosphatase 1 regulatory subunit 3B, which inhibits glycogen breakdown by regulating the interaction of phosphorylated protein 1 (PP1) with glycogen metabolism enzymes [65]. LOC157273 is shown to be exclusively expressed in human hepatocytes [44]. Altogether, these data suggest that LOC157273 to be the key regulator that links the identified variants with *PPP1R3B* activity and with lipid profile.

The second SNP is located in the well-known *KCNQ1* locus on chromosome 11 (11p15.5) associated with type 2 diabetes [66]. The top hit in this locus, rs231361, overlaps both the lnc-CDKN1C-3 (also known as *KCNQ1OT1*) transcript and intron 11 of *KCNQ1*. The *KCNQ1* gene has been reported in several studies to be associated with cardiovascular events and type 2 diabetes [67-73]. This gene encodes a voltage-gated potassium channel and plays an important role in repolarization of the cardiac action potential [74, 75]. Korostowski et al. have demonstrated that the expression level of *KCNQ1* gene is up-regulated in mice expressing a truncated *KCNQ1OT1*, indicating that the lncRNA represses *KCNQ1* expression [76]. Moreover, two independent studies have revealed that *KCNQ1OT1* negatively regulates the expression of *CDKN1C*, a known regulator of beta-cell development [77, 78]. Further, the SNP rs231361 has been shown to influence the expression of nearby genes including *CDKN1C* [79, 80]. These data suggest that rs231361-induced aberrant expression of *KCNQ1OT1* may be sufficient to cause disease through deregulation of the expression levels of *KCNQ1* and *CDKN1C*, two coding genes that are most likely to mediate type 2 diabetes susceptibility at the *KCNQ1* locus.

The third SNP is located in in chromosome 6 harboring IER3-5 (also known as LINC00243) associated with triglyceride and total cholesterol. In our study, we found that the expression level of LINC00243 is associated with total cholesterol. In addition, we observed that methylation of a CpG site annotated to this lncRNA is associated with triglyceride. The lead variant in this locus is rs886424 resides in the second exon of LINC00243, ~70 kb away from *IER3* that was thought to be the disease candidate

gene. Through the eQTL analysis, we observed a correlation between rs886424 and the expression level of LINC00243 and its nearby coding genes including *IER3, VARS2, FLOT1, MICB*, *HLA-C, HLA-H* and *HLA-G*. Performing a pathway analysis (using IPA) with these genes, we found that they are enriched in the immunological disease pathway. In particular, *IER3* has been reported to inhibit the production of the pro-inflammatory cytokines [81]. Since lncRNAs have been suggested to modulate the same biological processes as regulated by their nearby genes [82-84], LINC00243 can be hypothesized to play a regulatory role in immune-function-related pathways that need to be confirmed in further experimental studies.

While the above-mentioned examples endorsed the annotation in the literature, however, suggests a lncRNAs to mediate the observed association, another part of our findings indicated potential incorrect annotations to close by coding genes. For instance, rs17427875 was found to be associated with WHR and was annotated to *HOXA11*, which is a nearby coding gene and ~95 kb downstream of the variant. Nevertheless, this sentinel SNP resides within the first exon of *HOXA11* antisense transcript (*HOXA11-AS*). This lncRNA is located in the 5' region of the *HOXA* gene cluster, is highly conserved amongst several species [85], and is abundantly expressed in adipose tissue [44]. Previous studies have suggested that antisense transcript lncRNAs can directly or indirectly regulate the expression of the sense genes, however, experimental studies for *HOXA11-AS* have shown no effect on *HOXA11* expression or function [86, 87]. Instead, Sun et al have reported that knockdown of *HOXA11-AS* upregulates *KLF2*, which is a well-known anti-adipogeneic gene, through functioning as a scaffold to control the gene expression. In addition, our cis-eQTL data showed a correlation between rs17427875 and expression level of *HOXA5*, a nearby gene with an established association with the obesity-related phenotypes [88-90]. Through trans-eQTL analysis in adipose tissue, we further observed a correlation between rs17427875 and the expression level of *C19orf12* (also known as *NBIA4*), which is itself shown to be up-regulated during adipocyte differentiation [91]. Together, these data suggest a role for *HOXA11-AS* in fat distribution that may take place through regulating the expression of the aforementioned coding genes, and not *HOXA11*.

Likewise, rs243018, is a lead SNP associated with type 2 diabetes in an evolutionary conserved locus on chromosome 2 (2p16.1). The variant is localized in the last exon of lnc-BCL11A-3, which overlaps with miR-4432 host gene. However, the genetic association is annotated to the nearest coding gene, *BCL11A*, in the literature and the miRNA host gene that directly encompass the variant is overlooked [92, 93]. *BCL11A* gene is upstream (~90 kb away) of rs243018 and its variants are not in LD with the variant. The whole blood tran-eQTL data showed the association of rs243018 with the expression of *ESPN*, a gene which is thought to play a role in the regulation of the insulin signaling pathway [94].

In this study we also suggest three new susceptibility loci for cardiometabolic disorders that need to be replicated in future association studies. These include lnc-XRCC3-1 for BMI, lnc-MAP3K4-1 for LDL and lnc-CACNA1E-2 for HDL-cholesterol. Among these, lnc-XRCC3-1 located in a chromosome 14 locus (14q32) is expressed in adipose tissue [44]. We observed that the sentinel SNP rs709400 in this locus is enriched in promoter histone marks in adipose tissue and has the potential to change the lncRNA secondary structure. Moreover, we found a correlation between rs709400 and expression levels of two protein-coding genes, PPP1R13B and KLC1, in adipose tissue. Our results also showed that the only CpG site (cg15322516) annotated to lnc-XRCC3-1 is nominally associated with BMI. These data propose a potential role for lnc-XRCC3-1 in obesity-related pathways that warrant further investigation.

This study has several strengths as well as limitations that should be considered with the interpretation of the reported results. The main strength of this study is applying an integration approach by integrating omics data (genomics and epigenomics) and various in silico studies. The availability of summary statistics from the largest GWAS meta-analyses of cardiometabolic disorders enabled us to detect multiple lncRNAs associated with the studied traits. In addition, the availability of genome-wide DNA methylation and expression data in the Rotterdam Study, a population-based setting, allowed us to further investigate the potential causal mechanisms in the identified regions. Our study, however, has some limitations that need to be addressed. Although we had a reasonable sample size for methylation and expression analysis, weaker associations could have been overlooked due to lack of power. Another limitation is the use of whole blood samples to determine DNA methylation and expression levels of the lncRNAs, which is not the most relevant tissue. DNA methylation is cell type specific. LncRNA transcripts, on average, express at lower levels and show a higher degree of tissue specificity in comparison to protein-coding transcripts. Further, many of the identified lncRNAs had either no probe available in the Illumina expression array or the probes did not pass quality control, we thereby could not investigate the association between expression levels of these lncRNAs and cardiometabolic traits. Although this could have resulted overlooking some of the lncRNAs, the found associations are valid. In an optimal setting one should examine the associations using expression arrays covering all lncRNAs or next generation sequencing in target tissues such as heart, liver and adipose tissue. However, such infrastructure is not yet available in large epidemiologic studies with validated clinical data.

CONCLUSION

The present study provides a systematic investigation of genetic variants in lncRNAs associated with cardiometabolic disorders. We identified a number of genetic associations that are likely to affect cardiometabolic traits and diseases through lncRNAs. Further, we highlighted a number of loci associated with cardiometabolic disorders for which the role of lncRNA may have been overlooked in the literature. The identified lncRNAs could serve as candidates for future studies to determine their functions at different stages of cardiac development and metabolic processes as well as their potential as diagnostic biomarkers and therapeutic targets for cardiometabolic disorders.

Table S1. Description of GWAS meta-analyses of 10 cardiometabolic disorders

Cardiometabolic traits	Consortium	Sample size	All SNPs in GWAS	Associated lncRNA SNPs
Total cholesterol	GLGC	187,000	4,196	104 (36)
Triglyceride	GLGC	178,000	3,249	70 (26)
HDL-cholesterol	GLGC	187,000	3,524	48 (25)
LDL-cholesterol	GLGC	173,000	3,078	84 (26)
Body mass index	GIANT	339,229	620	16 (12)
Waist to hip ratio	GIANT	224,459	725	11 (8)
Coronary artery disease	CARDIoGRMplusC4D	63,746 cases/ 130,681controls	369	7 (4)
Type 2 diabetes	DIAGRAM	26,488 cases/ 83,964 controls	236	12 (7)
Systolic blood pressure	Global BPGen	71,225	111	0 (0)
Diastolic blood pressure	Global BPGen	71,225	88	1 (1)

SNPs, single-nucleotide polymorphisms; DIAGRAM indicates DIAbetes Genetics Replication And Meta-analysis; GIANT, Genetic Investigation of Anthropometric Traits; GLGC, Global Lipid Genetics Consortium; MAGIC, Meta-Analyses of Glucose and Insulin-related traits Consortium. Numbers inside the parentheses are associated lncRNA loci.

Supplemental Table S5. The lead variants at three novel lncRNA loci associated with cardiovascular disorders

SNP ID	A1/A2	Chr.	lncRNA ID	Alternative lncRNA ID	Trait	GWAS effect	GWAS *p*-value
rs7537211	G/A	1	lnc-CACNA1E-2	linc-RGSL1-2	HDL	0.02	7.26×10^{-8}
rs9355860	A/G	6	lnc-MAP3K4-1	RP3-428L16.1	LDL	0.02	7.84×10^{-8}
rs709400	A/G	14	lnc-XRCC3-1	RP11-894P9.1	BMI	0.02	1.38×10^{-7}

A1, Allele 1; A2, Allele 2; Chr., Chromosome; HDL, High Density Lipoprotein; LDL, Low Density Lipoprotein; BMI, Body Mass Index.

Supplemental Table S6. Fifty-five selected lncRNA SNPs associated with cardiovascular disorders

SNP ID	lncRNA ID	Associated trait	SNP status in the locus
rs12885454	lnc-PRKD1-3	BMI	Top
rs709400	lnc-XRCC3-1	BMI	Top
rs9925964	lnc-BCKDK-1	BMI, TG	Top
rs4784659	lnc-AMFR-1	HDL	Top
rs7537211	lnc-CACNA1E-2	HDL	Top
rs9355860	lnc-MAP3K4-1	LDL	Top
rs4841132	lnc-RP11-10A14.4.1-4	HDL, LDL, TC	Top
rs3757354	lnc-ATXN1-2	LDL, TC	Top
rs456598	lnc-SLC22A2-1	TC, LDL	Top
rs10910490	lnc-C1orf31-7	LDL, TC, CAD	Top
rs17526895	lnc-INSIG2-1	TC, LDL	Top
rs2898290	lnc-BLK-1	TG	Top
rs1038693	lnc-RBKS-2	TG	Top
rs886424	lnc-IER3-5	TG, TC	Top
rs243018	lnc-BCL11A-3	T2D	Top
rs231361	lnc-CDKN1C-3	T2D	Top
rs2301826	lnc-RCCD1-1	T2D	Top
rs6265	lnc-BDNF-AS	BMI	Proxy top
rs6565300	lnc-SPNS1-1	BMI	Proxy top
rs591120	lnc-SEC16B.1-2	BMI	Proxy top
rs7187575	lnc-RABEP2-3	BMI	Proxy top
rs12714640	lnc-VGLL3-5	BMI	Proxy top
rs7973894	lnc-TMBIM6-1	BMI	Proxy top
rs10965215	lnc-MTAP-1	CAD	Proxy top
rs7752775	lnc-TCF21-1	CAD	Proxy top
rs579058	lnc-CXCL12-2:1	CAD	Proxy top
rs9934328	lnc-AGRP-2	HDL	Proxy top
rs7181141	lnc-LACTB-1	HDL	Proxy top
rs755500	lnc-STAC2-1	HDL	Proxy top
rs11071371	lnc-ADAM10-2	HDL, TC	Proxy top
rs3760627	lnc-ZNF296-1	HDL, TG	Proxy top
rs17577	lnc-NCOA5-1	HDL, TG	Proxy top
rs8176720	lnc-OBP2B-2	LDL, TC	Proxy top
rs3810444	lnc-ARMC6-1	LDL, TC, TG	Proxy top
rs3177647	lnc-TMEM116-2	LDL	Proxy top
rs73004967	lnc-CILP2-1	LDL, HDL, TC	Proxy top
rs7710706	lnc-ANKRD31-1	LDL, TC	Proxy top
rs11705701	lnc-C3orf65-2	T2D	Proxy top
rs1557765	lnc-RP1-239B22.1.1-1	T2D	Proxy top

Supplemental Table S6. (continued)

SNP ID	lncRNA ID	Associated trait	SNP status in the locus
rs10804976	lnc-WFS1-1	T2D	Proxy top
rs2070959	lnc-HJURP-2	TC	Proxy top
rs9262636	lnc-MUC22-1	TC	Proxy top
rs310752	lnc-TIMP4-1	TC	Proxy top
rs3099844	lnc-ATP6V1G2-DDX39B-1	TC, TG	Proxy top
rs201150636	lnc-TCF19-1	TG	Proxy top
rs6760828	lnc-SNX17-1	TG	Proxy top
rs10947121	lnc-MUC21-1	TG	Proxy top
rs481843	lnc-APOC3-4	TG, TC, HDL, LDL	Proxy top
rs17427875	lnc-EVX1-5	WHR	Proxy top
rs3817973	lnc-HLA-DRA-1	WHR	Proxy top
rs1055144	lnc-HNRNPA2B1-4	WHR	Proxy top
rs10923722	lnc-TBX15-1	WHR	Proxy top
rs12756653	lnc-HAO2-2	WHR	Proxy top
rs7132434	lnc-ITPR2-2	WHR	Proxy top
rs825483	lnc-ZNF664-2	WHR, HDL	Proxy top

In this table, SNPs are ranked in two groups based on the significance of association with cardiometabolic traits in the related loci. Top, SNP with the strongest association in the locus; Proxy top, one of the top SNPs in high LD ($R^2 > 0.8$) with the top SNP in the locus.

Supplemental Table S16. An overview of the results of our analyses for 55 selected lncRNA SNPs associated with cardiometabolic disorders

SNP ID	lncRNA ID	Associated trait(s)	Proxy SNPs		eQTL		Regulatory features		miRNA binding site		ΔMFE	Exp.	Rotterdam Study data	
			All	Non-syn	Cis	Trans	Promoter	Enhancer	Disruption	Creation			Expression	Methylation
rs12885454	lnc-PRKD1-3	BMI	20	0	-	-	Yes	Yes	No	miR-514a	2.50	Low	n/a	CMD
rs709400	lnc-XRCC3-1	BMI	20	1	2	-	No	No	-	-	1.30	n/a	n/a	BMI, CMD
rs4841132	lnc-RP11-10A14.4	HDL, LDL,TC	7	0	-	-	Yes	Yes	miR-1287	No	0.41	Liver	-	HDL, LDL
rs4784659	lnc-AMFR-1	HDL	3	0	2	-	No	No	miR-18a-5p	No	2.10	n/a	n/a	n/a
rs7537211	lnc-CACNA1E-2	HDL	8	0	-	2	No	No	miR-152-5p	miR-760	0.30	Low	n/a	n/a
rs9355860	lnc-MAP3K4-1	LDL	20	0	-	-	Yes	Yes	-	-	1.65	Low	T2D, BP	CMD
rs10910490	lnc-C1orf31-7	LDL, TC	13	0	-	-	Yes	Yes	No	miR-544a	0.80	Liver	n/a	CMD
rs17526895	lnc-INSIG2-1	LDL, TC	100	1	-	-	Yes	Yes	miR-580-5p	No	0.40	Low	n/a	n/a
rs3757354	lnc-ATXN1-2	LDL, TC	3	0	2	-	Yes	Yes	No	miR-370	1.20	Low	n/a	LDL
rs456598	lnc-SLC22A2-1	LDL, TC	2	0	-	2	No	No	miR-432-3p	No	0.0	Liver	n/a	n/a
rs9925964	lnc-BCKDK-1	TG, BMI	20	1	6	-	Yes	Yes	-	-	0.11	n/a	n/a	WHR
rs1038693	lnc-RBKS-2	TG	7	0	-	-	Yes	Yes	miR-4672	No	0.0	n/a	n/a	n/a
rs2898290	lnc-BLK-1	TG	20	0	4	-	Yes	Yes	miR-141-3p	No	0.10	Low	n/a	-
rs2301826	lnc-RCCD1-1	T2D	1	0	2	-	No	No	-	-	0.80	n/a	-	CMD
rs231361	lnc-CDKN1C-3	T2D	0	0	-	-	Yes	Yes	-	-	0.0	Low	-	T2D, CMD
rs243018	lnc-BCL11A-3	T2D	10	0	-	1	No	No	-	-	1.70	Low	n/a	CMD
rs886424	lnc-IER3-5	TG,TC	50	0	4	-	Yes	Yes	No	miR-452	0.20	n/a	n/a	TG, CMD
rs1055144	lnc-HNRNPA2B1	WHR	20	0	-	-	Yes	Yes	-	-	3.64	Low	-	n/a
rs10965215	lnc-MTAP-1	CAD	40	0	-	-	No	Yes	-	-	0.40	Low	TG, CHD	CHD, HDL
rs3099844	lnc-ATP6V1G2	TC, TG	8	0	4	3	Yes	Yes	-	-	1.42	Blood	HDL, CHD	CMD
rs17427875	lnc-EVX1-5	WHR	0	0	1	1	No	No	-	-	0.84	Adipose	-	WHR, CMD
rs6265	lnc-METTL15-4	BMI	20	1	-	-	No	No	-	-	2.50	n/a	n/a	CMD
rs8176720	lnc-OBP2B-2	LDL, TC	10	0	2	-	Yes	No	miR-455-5p	No	2.71	Multiple	n/a	CMD

Supplemental Table S16. (continued)

SNP ID	lncRNA ID	Associated trait(s)	Proxy SNPs		eQTL		Regulatory features		miRNA binding site		ΔMFE	Exp.	Rotterdam Study data	
			All	Non-syn	Cis	Trans	Promoter	Enhancer	Disruption	Creation			Expression	Methylation
rs3810444	lnc-ARMC6-1	LDL, TC, TG	2	0	-	-	No	No	-	-	1.70	Low	n/a	-
rs201150636	lnc-TCF19-1	TG	0	0	-	-	No	Yes	-	-	4.20	Adipose	n/a	CMD
rs3177647	lnc-TMEM116-2	LDL	100	1	5	-	No	Yes	miR-553	No	1.10	Multiple	n/a	CMD
rs7710706	lnc-ANKRD31-1	LDL ,TC	0	0	-	-	Yes	Yes	No	miR-135a	No	n/a	n/a	CMD
rs73004967	lnc-CILP2-1	LDL, TC HDL	18	0	-	-	Yes	No	miR-766-3p	-	1.70	n/a	n/a	n/a
rs481843	lnc-APOC3-4	All lipids	30	0	-	-	No	No	No	miR-1206	No	n/a	n/a	n/a
rs7181141	lnc-LACTB-1	HDL	25	0	2	-	No	No	-	-	0.70	n/a	n/a	CMD
rs755500	lnc-STAC2-1	HDL	50	0	3	-	No	No	-	-	2.04	n/a	n/a	n/a
rs11071371	lnc-ADAM10-2	HDL,TC	11	0	-	-	No	Yes	-	-	2.02	Liver	n/a	n/a
rs17577	lnc-NCOA5-1	HDL, TG	15	0	-	-	Yes	No	miR-7-5p	miR-125a	2.90	n/a	n/a	TG, CMD
rs3760627	lnc-ZNF296-1	HDL, TG	50	0	-	-	Yes	Yes	-	-	No	-	n/a	-
rs310752	lnc-TIMP4-1	TC	0	0	2	-	No	No	-	-	0.10	n/a	n/a	n/a
rs10947121	lnc-MUC21-1	TG	11	0	1	-	No	No	miR-627-5p	No	No	n/a	n/a	-
rs2070959	lnc-HJURP-2	TC	15	1	-	1	Yes	No	miR-181a	miR-586	0.60	n/a	n/a	-
rs9262636	lnc-MUC22-1	TC	20	0	2	-	No	No	miR-513b	No	1.20	n/a	n/a	TC, CMD
rs6760828	lnc-SNX17-1	TG	100	0	4	-	Yes	No	-	-	2.17	n/a	n/a	TG, CMD
rs9934328	lnc-AGRP-2	HDL	10	0	6	-	Yes	No	-	-	0.60	n/a	n/a	CMD
rs7973894	lnc-TMBIM6-1	BMI	10	0	-	-	No	Yes	-	-	0.70	n/a	n/a	BMI, CMD
rs12714640	lnc-VGLL3-5	BMI	50	0	-	-	Yes	Yes	-	-	1.40	n/a	n/a	n/a
rs7187575	lnc-RABEP2-3	BMI	7	0	2	-	Yes	Yes	-	-	0.30	n/a	n/a	WHR, CMD
rs591120	lnc-SEC16B.1-2	BMI	1	0	-	-	No	Yes	miR-601	miR-1301	3.46	n/a	n/a	BMI, CMD
rs6565300	lnc-SPNS1-1	BMI	7	0	2	-	Yes	No	miR-144-3p	No	1.90	n/a	n/a	BMI
rs825483	lnc-ZNF664-2	WHR, HDL	20	0	1	-	No	Yes	-	-	0.20	n/a	n/a	n/a

Supplemental Table S16. (continued)

SNP ID	lncRNA ID	Associated trait(s)	Proxy SNPs		eQTL		Regulatory features		miRNA binding site		ΔMFE	Exp.	Rotterdam Study data	
			All	Non-syn	Cis	Trans	Promoter	Enhancer	Disruption	Creation			Expression	Methylation
rs3817973	lnc-HLA-DRA-1	WHR	12	1	2	-	No	Yes	-	-	2.50	Low	n/a	n/a
rs12756653	lnc-HAO2-2	WHR	3	0	1	-	No	Yes	miR-500a	No	0.20	n/a	n/a	CMD
rs7132434	lnc-ITPR2-2	WHR	25	0	-	-	No	Yes	miR-452-3p	No	3.30	n/a	n/a	CMD
rs10923722	lnc-TBX15-1	WHR	20	0	1	-	No	Yes	-	-	2.50	n/a	n/a	BMI, CMD
rs1557765	lnc-RP1-239B22	T2D	18	3	1	-	No	Yes	-	-	0.40	Multiple	n/a	n/a
rs11705701	lnc-C3orf65-2	T2D	1	0	1	-	No	No	-	-	1.50	Low	n/a	CMD
rs10804976	lnc-WFS1-1	T2D	15	0	1	-	No	Yes	-	-	1.46	n/a	n/a	n/a
rs7752775	lnc-TCF21-1	CAD	20	0	-	-	No	Yes	-	-	2.96	n/a	n/a	CHD, CMD
rs579058	lnc-CXCL12-2	CAD	100	0	-	-	Yes	Yes	No	miR-1178	1.70	n/a	n/a	n/a

This table provide a results overview of our functional analyses for the 55 selected lncRNA SNPs associated with cardiometabolic traits. LDL, Low Density Lipoprotein; HDL, High Density Lipoprotein; TC, Total cholesterol; TG, Triglyceride; BMI, Body Mass Index; WHR, Waist to Hip Ratio; T2D, Type 2 Diabetes; n/a, Not available; Non-syn, Non-synonymous proxy SNPs; ΔMFE, Minimum free energy change difference between wild type and mutant lncRNA; eQTL, expression quantitative trait loci; Cis, the effect of SNP on the expression level of nearby gene; Trans, the effect of SNP on the expression level of distant gene. The traits mentioned in the expression and DNA methylation analyses from the Rotterdam Study data show nominally significant association. Exp, gene expression using data from the Dutch cohort and RNA-seq from the Human Body Map 2.0.

REFERENCES

1 Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C. et al. (2005) The transcriptional landscape of the mammalian genome. Science, 309, 1559-1563.

2 Consortium, E.P. (2012) An integrated encyclopedia of DNA elements in the human genome. Nature, 489, 57-74.

3 Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P. et al. (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature, 458, 223-227.

4 Pennisi, E. (2012) Genomics. ENCODE project writes eulogy for junk DNA. Science, 337, 1159, 1161.

5 Consortium, E.P., Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T. et al. (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature, 447, 799-816.

6 Consortium, C.A.D., Deloukas, P., Kanoni, S., Willenborg, C., Farrall, M., Assimes, T.L., Thompson, J.R., Ingelsson, E., Saleheen, D., Erdmann, J. et al. (2013) Large-scale association analysis identifies new risk loci for coronary artery disease. Nat Genet, 45, 25-33.

7 Kapranov, P., Cheng, J., Dike, S., Nix, D.A., Duttagupta, R., Willingham, A.T., Stadler, P.F., Hertel, J., Hackermuller, J., Hofacker, I.L. et al. (2007) RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science, 316, 1484-1488.

8 Ponting, C.P., Oliver, P.L. and Reik, W. (2009) Evolution and functions of long noncoding RNAs. Cell, 136, 629-641.

9 Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A. and Rinn, J.L. (2011) Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev, 25, 1915-1927.

10 Clark, B.S. and Blackshaw, S. (2014) Long non-coding RNA-dependent transcriptional regulation in neuronal development and disease. Front Genet, 5, 164.

11 Esteller, M. (2011) Non-coding RNAs in human disease. Nat Rev Genet, 12, 861-874.

12 Huarte, M. (2015) The emerging role of lncRNAs in cancer. Nat Med, 21, 1253-1261.

13 Chung, S., Nakagawa, H., Uemura, M., Piao, L., Ashikawa, K., Hosono, N., Takata, R., Akamatsu, S., Kawaguchi, T., Morizono, T. et al. (2011) Association of a novel long non-coding RNA in 8q24 with prostate cancer susceptibility. Cancer Sci, 102, 245-252.

14 Wapinski, O. and Chang, H.Y. (2011) Long noncoding RNAs and human disease. Trends Cell Biol, 21, 354-361.

15 Han, P., Li, W., Lin, C.H., Yang, J., Shang, C., Nurnberg, S.T., Jin, K.K., Xu, W., Lin, C.Y., Lin, C.J. et al. (2014) A long noncoding RNA protects the heart from pathological hypertrophy. Nature, 514, 102-106.

16 Yan, B., Yao, J., Liu, J.Y., Li, X.M., Wang, X.Q., Li, Y.J., Tao, Z.F., Song, Y.C., Chen, Q. and Jiang, Q. (2015) lncRNA-MIAT Regulates Microvascular Dysfunction by Functioning as a Competing Endogenous RNA. Circ Res, 116, 1143-1156.

17 Broadbent, H.M., Peden, J.F., Lorkowski, S., Goel, A., Ongen, H., Green, F., Clarke, R., Collins, R., Franzosi, M.G., Tognoni, G. et al. (2008) Susceptibility to coronary artery disease and diabetes is encoded by distinct, tightly linked SNPs in the ANRIL locus on chromosome 9p. Hum Mol Genet, 17, 806-814.

18 Pasmant, E., Sabbagh, A., Vidaud, M. and Bieche, I. (2011) ANRIL, a long, noncoding RNA, is an unexpected major hotspot in GWAS. FASEB J, 25, 444-448.
19 Global Lipids Genetics, C., Willer, C.J., Schmidt, E.M., Sengupta, S., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M.L. et al. (2013) Discovery and refinement of loci associated with lipid levels. Nat Genet, in press.
20 Replication, D.I.G., Meta-analysis, C., Asian Genetic Epidemiology Network Type 2 Diabetes, C., South Asian Type 2 Diabetes, C., Mexican American Type 2 Diabetes, C., Type 2 Diabetes Genetic Exploration by Nex-generation sequencing in muylti-Ethnic Samples, C., Mahajan, A., Go, M.J., Zhang, W., Below, J.E. et al. (2014) Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. Nat Genet, 46, 234-244.
21 International Consortium for Blood Pressure Genome-Wide Association, S., Ehret, G.B., Munroe, P.B., Rice, K.M., Bochud, M., Johnson, A.D., Chasman, D.I., Smith, A.V., Tobin, M.D., Verwoert, G.C. et al. (2011) Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature, 478, 103-109.
22 Hindorff, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S. and Manolio, T.A. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A, 106, 9362-9367.
23 Zhang, F. and Lupski, J.R. (2015) Non-coding genetic variants in human disease. Hum Mol Genet, 24, R102-110.
24 Ritchie, G.R., Dunham, I., Zeggini, E. and Flicek, P. (2014) Functional annotation of noncoding sequence variants. Nat Methods, 11, 294-296.
25 Ward, L.D. and Kellis, M. (2012) Interpreting noncoding genetic variation in complex traits and human disease. Nat Biotechnol, 30, 1095-1106.
26 Ionita-Laza, I., McCallum, K., Xu, B. and Buxbaum, J.D. (2016) A spectral approach integrating functional genomic annotations for coding and noncoding variants. Nat Genet, 48, 214-220.
27 Huan, T., Rong, J., Tanriverdi, K., Meng, Q., Bhattacharya, A., McManus, D.D., Joehanes, R., Assimes, T.L., McPherson, R., Samani, N.J. et al. (2015) Dissecting the roles of microRNAs in coronary heart disease via integrative genomic analyses. Arterioscler Thromb Vasc Biol, 35, 1011-1021.
28 Ghanbari, M., Sedaghat, S., de Looper, H.W., Hofman, A., Erkeland, S.J., Franco, O.H. and Dehghan, A. (2015) The association of common polymorphisms in miR-196a2 with waist to hip ratio and miR-1908 with serum lipid and glucose. Obesity (Silver Spring), 23, 495-503.
29 Ghanbari, M., de Vries, P.S., de Looper, H., Peters, M.J., Schurmann, C., Yaghootkar, H., Dorr, M., Frayling, T.M., Uitterlinden, A.G., Hofman, A. et al. (2014) A genetic variant in the seed region of miR-4513 shows pleiotropic effects on lipid and glucose homeostasis, blood pressure, and coronary artery disease. Hum Mutat, 35, 1524-1531.
30 Li, Q., Chen, L., Chen, D., Wu, X. and Chen, M. (2015) Influence of microRNA-related polymorphisms on clinical outcomes in coronary artery disease. Am J Transl Res, 7, 393-400.
31 Locke, A.E., Kahali, B., Berndt, S.I., Justice, A.E., Pers, T.H., Day, F.R., Powell, C., Vedantam, S., Buchkovich, M.L., Yang, J. et al. (2015) Genetic studies of body mass index yield new insights for obesity biology. Nature, 518, 197-206.
32 Shungin, D., Winkler, T.W., Croteau-Chonka, D.C., Ferreira, T., Locke, A.E., Magi, R., Strawbridge, R.J., Pers, T.H., Fischer, K., Justice, A.E. et al. (2015) New genetic loci link adipose and insulin biology to body fat distribution. Nature, 518, 187-196.
33 Gong, J., Liu, W., Zhang, J., Miao, X. and Guo, A.Y. (2014) lncRNASNP: a database of SNPs in lncRNAs and their potential functions in human and mouse. Nucleic Acids Res, in press.

34 Li, J.H., Liu, S., Zhou, H., Qu, L.H. and Yang, J.H. (2014) starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. Nucleic Acids Res, 42, D92-97.

35 Wang, Y., Xu, Z., Jiang, J., Xu, C., Kang, J., Xiao, L., Wu, M., Xiong, J., Guo, X. and Liu, H. (2013) Endogenous miRNA sponge lincRNA-RoR regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-renewal. Dev Cell, 25, 69-80.

36 Ryan, B.M., Robles, A.I. and Harris, C.C. (2010) Genetic variation in microRNA networks: the implications for cancer research. Nat Rev Cancer, 10, 389-402.

37 Uchida, S. and Dimmeler, S. (2015) Long noncoding RNAs in cardiovascular diseases. Circ Res, 116, 737-750.

38 Fu, J., Wolfs, M.G., Deelen, P., Westra, H.J., Fehrmann, R.S., Te Meerman, G.J., Buurman, W.A., Rensen, S.S., Groen, H.J., Weersma, R.K. et al. (2012) Unraveling the regulatory mechanisms underlying tissue-dependent genetic variation of gene expression. PLoS Genet, 8, e1002431.

39 Kumar, V., Westra, H.J., Karjalainen, J., Zhernakova, D.V., Esko, T., Hrdlickova, B., Almeida, R., Zhernakova, A., Reinmaa, E., Vosa, U. et al. (2013) Human disease-associated genetic variation impacts large intergenic non-coding RNA expression. PLoS Genet, 9, e1003201.

40 Mirza, A.H., Kaur, S., Brorsson, C.A. and Pociot, F. (2014) Effects of GWAS-associated genetic variants on lncRNAs within IBD and T1D candidate loci. PLoS One, 9, e105723.

41 Guttman, M. and Rinn, J.L. (2012) Modular regulatory principles of large non-coding RNAs. Nature, 482, 339-346.

42 Wang, K.C. and Chang, H.Y. (2011) Molecular mechanisms of long noncoding RNAs. Mol Cell, 43, 904-914.

43 Grimson, A., Farh, K.K., Johnston, W.K., Garrett-Engele, P., Lim, L.P. and Bartel, D.P. (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol Cell, 27, 91-105.

44 Yang, J., Lee, S.H., Goddard, M.E. and Visscher, P.M. (2011) GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet, 88, 76-82.

45 Manning, A.K., Hivert, M.F., Scott, R.A., Grimsby, J.L., Bouatia-Naji, N., Chen, H., Rybin, D., Liu, C.T., Bielak, L.F., Prokopenko, I. et al. (2012) A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. Nat Genet, 44, 659-669.

46 Hayes, M.G., Urbanek, M., Hivert, M.F., Armstrong, L.L., Morrison, J., Guo, C., Lowe, L.P., Scheftner, D.A., Pluzhnikov, A., Levine, D.M. et al. (2013) Identification of HKDC1 and BACE2 as genes influencing glycemic traits during pregnancy through genome-wide association studies. Diabetes, 62, 3282-3291.

47 Inouye, M., Ripatti, S., Kettunen, J., Lyytikainen, L.P., Oksala, N., Laurila, P.P., Kangas, A.J., Soininen, P., Savolainen, M.J., Viikari, J. et al. (2012) Novel Loci for metabolic networks and multi-tissue expression studies reveal genes for atherosclerosis. PLoS Genet, 8, e1002907.

48 Liu, C.T., Raghavan, S., Maruthur, N., Kabagambe, E.K., Hong, J., Ng, M.C., Hivert, M.F., Lu, Y., An, P., Bentley, A.R. et al. (2016) Trans-ethnic Meta-analysis and Functional Annotation Illuminates the Genetic Architecture of Fasting Glucose and Insulin. Am J Hum Genet, 99, 56-75.

49 Teslovich, T.M., Musunuru, K., Smith, A.V., Edmondson, A.C., Stylianou, I.M., Koseki, M., Pirruccello, J.P., Ripatti, S., Chasman, D.I., Willer, C.J. et al. (2010) Biological, clinical and population relevance of 95 loci for blood lipids. Nature, 466, 707-713.

50 Lipovich L, Brown B, Raghavan S, Wessel J, Meigs J. (2015) Long non-coding RNA genes are direct disease causal candidates in human metabolic disorders: insights from the GWAS-transcriptome interface. RNA Meeting 2015.

51 Manning AK, Lipovich L, Goustin A, Leong A, Morrison AC, Brody J, Liu C, Brown JB, Chen Y, Rotter I, Meigs JB, CHARGE Consortium. (2016) The fasting glucose associated PPP1R3B locus maps to LOC157273, a long non-coding RNA which represses PPP1R3B expression in primary human hepatocytes. SAC Meeting 2016.

52 Depetris, R.S., Hu, J., Gimpelevich, I., Holt, L.J., Daly, R.J. and Hubbard, S.R. (2005) Structural basis for inhibition of the insulin receptor by the adaptor protein Grb14. Mol Cell, 20, 325-333.

53 Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J. and Pachter, L. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol, 28, 511-515.

54 Travers, M.E., Mackay, D.J., Dekker Nitert, M., Morris, A.P., Lindgren, C.M., Berry, A., Johnson, P.R., Hanley, N., Groop, L.C., McCarthy, M.I. et al. (2013) Insights into the molecular mechanism for type 2 diabetes susceptibility at the KCNQ1 locus from temporal changes in imprinting status in human islets. Diabetes, 62, 987-992.

55 Duchatelet, S., Crotti, L., Peat, R.A., Denjoy, I., Itoh, H., Berthet, M., Ohno, S., Fressart, V., Monti, M.C., Crocamo, C. et al. (2013) Identification of a KCNQ1 polymorphism acting as a protective modifier against arrhythmic risk in long-QT syndrome. Circ Cardiovasc Genet, 6, 354-361.

56 Olszak-Waskiewicz, M., Dziuk, M., Kubik, L., Kaczanowski, R. and Kucharczyk, K. (2008) Novel KCNQ1 mutations in patients after myocardial infarction. Cardiol J, 15, 252-260.

57 Pfeufer, A., Jalilzadeh, S., Perz, S., Mueller, J.C., Hinterseer, M., Illig, T., Akyol, M., Huth, C., Schopfer-Wendels, A., Kuch, B. et al. (2005) Common variants in myocardial ion channel genes modify the QT interval in the general population: results from the KORA study. Circ Res, 96, 693-701.

58 Shamgar, L., Ma, L., Schmitt, N., Haitin, Y., Peretz, A., Wiener, R., Hirsch, J., Pongs, O. and Attali, B. (2006) Calmodulin is essential for cardiac IKS channel gating and assembly: impaired function in long-QT mutations. Circ Res, 98, 1055-1063.

59 Unoki, H., Takahashi, A., Kawaguchi, T., Hara, K., Horikoshi, M., Andersen, G., Ng, D.P., Holmkvist, J., Borch-Johnsen, K., Jorgensen, T. et al. (2008) SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations. Nat Genet, 40, 1098-1102.

60 Yasuda, K., Miyake, K., Horikawa, Y., Hara, K., Osawa, H., Furuta, H., Hirota, Y., Mori, H., Jonsson, A., Sato, Y. et al. (2008) Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. Nat Genet, 40, 1092-1097.

61 Vausort, M., Wagner, D.R. and Devaux, Y. (2014) Long noncoding RNAs in patients with acute myocardial infarction. Circ Res, 115, 668-677.

62 Inanobe, A., Tsuzuki, C. and Kurachi, Y. (2015) An Epithelial Ca2+-Sensor Protein is an Alternative to Calmodulin to Compose Functional KCNQ1 Channels. Cell Physiol Biochem, 36, 1847-1861.

63 Sanguinetti, M.C. and Tristani-Firouzi, M. (2006) hERG potassium channels and cardiac arrhythmia. Nature, 440, 463-469.

64 Korostowski, L., Sedlak, N. and Engel, N. (2012) The Kcnq1ot1 long non-coding RNA affects chromatin conformation and expression of Kcnq1, but does not regulate its imprinting in the developing heart. PLoS Genet, 8, e1002956.

65 Arima, T., Kamikihara, T., Hayashida, T., Kato, K., Inoue, T., Shirayoshi, Y., Oshimura, M., Soejima, H., Mukai, T. and Wake, N. (2005) ZAC, LIT1 (KCNQ1OT1) and p57KIP2 (CDKN1C) are in an imprinted gene network that may play a role in Beckwith-Wiedemann syndrome. Nucleic Acids Res, 33, 2650-2660.

66 Asahara, S., Etoh, H., Inoue, H., Teruyama, K., Shibutani, Y., Ihara, Y., Kawada, Y., Bartolome, A., Hashimoto, N., Matsuda, T. et al. (2015) Paternal allelic mutation at the Kcnq1 locus reduces

pancreatic beta-cell mass by epigenetic modification of Cdkn1c. Proc Natl Acad Sci U S A, 112, 8332-8337.

67 Kassem, S.A., Ariel, I., Thornton, P.S., Hussain, K., Smith, V., Lindley, K.J., Aynsley-Green, A. and Glaser, B. (2001) p57(KIP2) expression in normal islet cells and in hyperinsulinism of infancy. Diabetes, 50, 2763-2769.

68 Voight, B.F., Scott, L.J., Steinthorsdottir, V., Morris, A.P., Dina, C., Welch, R.P., Zeggini, E., Huth, C., Aulchenko, Y.S., Thorleifsson, G. et al. (2010) Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet, 42, 579-589.

69 Lai, W.S., Parker, J.S., Grissom, S.F., Stumpo, D.J. and Blackshear, P.J. (2006) Novel mRNA targets for tristetraprolin (TTP) identified by global analysis of stabilized transcripts in TTP-deficient fibroblasts. Mol Cell Biol, 26, 9196-9208.

70 Yang, K.C., Yamada, K.A., Patel, A.Y., Topkara, V.K., George, I., Cheema, F.H., Ewald, G.A., Mann, D.L. and Nerbonne, J.M. (2014) Deep RNA sequencing reveals dynamic regulation of myocardial noncoding RNAs in failing human heart and remodeling with mechanical circulatory support. Circulation, 129, 1009-1021.

71 Vance, K.W., Sansom, S.N., Lee, S., Chalei, V., Kong, L., Cooper, S.E., Oliver, P.L. and Ponting, C.P. (2014) The long non-coding RNA Paupar regulates the expression of both local and distal genes. EMBO J, 33, 296-311.

72 Luo, H., Sun, S., Li, P., Bu, D., Cao, H. and Zhao, Y. (2013) Comprehensive characterization of 10,571 mouse large intergenic noncoding RNAs from whole transcriptome sequencing. PLoS One, 8, e70835.

73 Potter, S.S. and Branford, W.W. (1998) Evolutionary conservation and tissue-specific processing of Hoxa 11 antisense transcripts. Mamm Genome, 9, 799-806.

74 Chau, Y.M., Pando, S. and Taylor, H.S. (2002) HOXA11 silencing and endogenous HOXA11 antisense ribonucleic acid in the uterine endometrium. J Clin Endocrinol Metab, 87, 2674-2680.

75 Richards, E.J., Permuth-Wey, J., Li, Y., Chen, Y.A., Coppola, D., Reid, B.M., Lin, H.Y., Teer, J.K., Berchuck, A., Birrer, M.J. et al. (2015) A functional variant in HOXA11-AS, a novel long non-coding RNA, inhibits the oncogenic phenotype of epithelial ovarian cancer. Oncotarget, 6, 34745-34757.

76 Parrillo, L., Costa, V., Raciti, G.A., Longo, M., Spinelli, R., Esposito, R., Nigro, C., Vastolo, V., Desiderio, A., Zatterale, F. et al. (2016) Hoxa5 undergoes dynamic DNA methylation and transcriptional repression in the adipose tissue of mice exposed to high-fat diet. Int J Obes (Lond), 40, 929-937.

77 Singh, S., Rajput, Y.S., Barui, A.K., Sharma, R. and Datta, T.K. (2016) Fat accumulation in differentiated brown adipocytes is linked with expression of Hox genes. Gene Expr Patterns, 20, 99-105.

78 Karpe, F. and Pinnick, K.E. (2015) Biology of upper-body and lower-body adipose tissue--link to whole-body phenotypes. Nat Rev Endocrinol, 11, 90-100.

79 Hartig, M.B., Iuso, A., Haack, T., Kmiec, T., Jurkiewicz, E., Heim, K., Roeber, S., Tarabin, V., Dusi, S., Krajewska-Walasek, M. et al. (2011) Absence of an orphan mitochondrial protein, c19orf12, causes a distinct clinical subtype of neurodegeneration with brain iron accumulation. Am J Hum Genet, 89, 543-550.

80 Langenberg, C., Sharp, S.J., Franks, P.W., Scott, R.A., Deloukas, P., Forouhi, N.G., Froguel, P., Groop, L.C., Hansen, T., Palla, L. et al. (2014) Gene-lifestyle interaction and type 2 diabetes: the EPIC interact case-cohort study. PLoS Med, 11, e1001647.

81 Morris, A.P., Voight, B.F., Teslovich, T.M., Ferreira, T., Segre, A.V., Steinthorsdottir, V., Strawbridge, R.J., Khan, H., Grallert, H., Mahajan, A. et al. (2012) Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet, 44, 981-990.

82 Angunsri, N., Taura, A., Nakagawa, T., Hayashi, Y., Kitajiri, S., Omi, E., Ishikawa, K. and Ito, J. (2011) Insulin-like growth factor 1 protects vestibular hair cells from aminoglycosides. Neuroreport, 22, 38-43.

83 Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R. and Willer, C.J. (2010) LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics, 26, 2336-2337.

84 Grundberg, E., Small, K.S., Hedman, A.K., Nica, A.C., Buil, A., Keildson, S., Bell, J.T., Yang, T.P., Meduri, E., Barrett, A. et al. (2012) Mapping cis- and trans-regulatory effects across multiple tissues in twins. Nat Genet, 44, 1084-1089.

85 Westra, H.J., Peters, M.J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., Christiansen, M.W., Fairfax, B.P., Schramm, K., Powell, J.E. et al. (2013) Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet, in press.

86 Bernstein, B.E., Stamatoyannopoulos, J.A., Costello, J.F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M.A., Beaudet, A.L., Ecker, J.R. et al. (2010) The NIH Roadmap Epigenomics Mapping Consortium. Nat Biotechnol, 28, 1045-1048.

87 Roadmap Epigenomics, C., Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J. et al. (2015) Integrative analysis of 111 reference human epigenomes. Nature, 518, 317-330.

88 McKusick, V.A. (2007) Mendelian Inheritance in Man and its online version, OMIM. Am J Hum Genet, 80, 588-604.

89 Meenhuis, A., van Veelen, P.A., de Looper, H., van Boxtel, N., van den Berge, I.J., Sun, S.M., Taskesen, E., Stern, P., de Ru, A.H., van Adrichem, A.J. et al. (2011) MiR-17/20/93/106 promote hematopoietic cell expansion by targeting sequestosome 1-regulated pathways in mice. Blood, 118, 916-925.

90 Sabarinathan, R., Tafer, H., Seemann, S.E., Hofacker, I.L., Stadler, P.F. and Gorodkin, J. (2013) RNAsnp: efficient detection of local RNA secondary structure changes induced by SNPs. Hum Mutat, 34, 546-556.

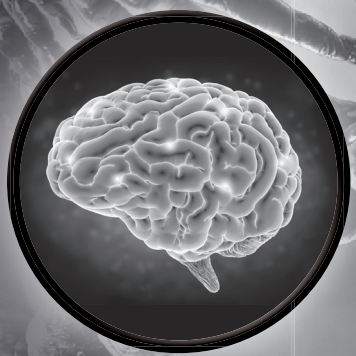
91 Kornienko, A.E., Dotter, C.P., Guenzl, P.M., Gisslinger, H., Gisslinger, B., Cleary, C., Kralovics, R., Pauler, F.M. and Barlow, D.P. (2016) Long non-coding RNAs display higher natural expression variation than protein-coding genes in healthy humans. Genome Biol, 17, 14.

92 Hofman, A., Brusselle, G.G., Darwish Murad, S., van Duijn, C.M., Franco, O.H., Goedegebure, A., Ikram, M.A., Klaver, C.C., Nijsten, T.E., Peeters, R.P. et al. (2015) The Rotterdam Study: 2016 objectives and design update. Eur J Epidemiol, 30, 661-708.

93 Steenaard, R.V., Ligthart, S., Stolk, L., Peters, M.J., van Meurs, J.B., Uitterlinden, A.G., Hofman, A., Franco, O.H. and Dehghan, A. (2015) Tobacco smoking is associated with methylation of genes related to coronary artery disease. Clin Epigenetics, 7, 54.

Chapter 3

Contribution of microRNAs to neurodegenerative diseases



CHAPTER 3.1

Genome-wide identification of miRNA-related variants associated with Alzheimer's disease

Mohsen Ghanbari, M. Arfan Ikram, Hans de Looper, Stefan J. Erkeland, Oscar H. Franco, Abbas Dehghan

Manuscript based on this chapter

Genome-wide identification of microRNA-related variants associated with risk of Alzheimer's disease

Scientific Reports. 2016 Jun 22;6:28387.

ABSTRACT

Background. MicroRNAs (miRNAs) serve as key post-transcriptional regulators of gene expression. Genetic variations in miRNAs and miRNA-binding sites can affect miRNA function and contribute to disease risk.

Methods and results. We investigated the extent to which variants within miRNA-related sequences could constitute a part of the functional variants involved in Alzheimer's disease (AD), using the largest available genome-wide association study. First, among 237 variants in miRNAs, we found rs2291418 in the precursor miR-1229 sequence to be significantly associated with AD (p-value= 6.8×10^{-5}, OR= 1.2). Our *in silico* analysis and *in vitro* miRNA expression experiments demonstrated that the variant's minor allele enhances the production of miR-1229-3p. Next, we found miR-1229-3p target genes that are associated with AD and might be downstream targets of the miRNA in relation to AD. Among them, we demonstrated that miR-1229-3p directly controls the expression of its top AD-associated target gene (*SORL1*) using luciferase reporter assays. Additionally, we showed that miR-1229-3p and *SORL1* are both expressed in the brain. Second, among 42,855 variants in miRNA-binding sites, we identified 10 variants (in the 3'UTR of 9 genes) that are significantly associated with AD, including rs6857 that increases miR-320e-mediated regulation of *PVRL2*.

Conclusion. Our results demonstrate that variants in miRNA-related sequences are associated with AD and may affect miRNA-dependent regulation of AD genes.

INTRODUCTION

Alzheimer's disease (AD) is the most common form of age-related neurodegenerative diseases worldwide manifested by the progressive loss of memory and cognitive decline [1]. A number of cellular and molecular mechanisms lead to AD occurrence [2-4]. Over the past few decades, enormous efforts have been made to discover risk factors that play a role in development of AD and new biomarkers that may help in early diagnosis of AD [5, 6]. Recently, microRNAs (miRNAs), a class of small non-coding RNAs with 20–24 nucleotides (nt) long, have gained widespread attention as important modulators of different biological processes. They have been shown to be involved in biological and pathological processes in the brain [7-9]. In addition, a number of dysregulated miRNAs have been reported to be associated with AD and are suggested as potential diagnostic biomarker for AD [10-12]. MiRNAs repress translation or decrease the stability of target messenger RNAs (mRNAs), by binding of the nucleotides 2-7/8 from their 5′ end (known as the seed region) to complementary sequences in the 3′ UTR regions of target mRNAs [13]. MiRNAs are predicted to regulate the translation of more than half of all protein coding genes, indicating the involvement of miRNAs in most cellular processes [14].

Genetic variants in miRNA-encoding sequences can affect miRNA biogenesis and function [15, 16]. Furthermore, variants that are located in the seed-matching regions of target genes may interfere with the interaction between miRNA and its target genes, resulting in an altered expression level of the target transcript [17, 18]. More recently, we and others have been able to show a number of variants in genomic sequences encoding miRNAs or in the 3′UTRs of miRNA target genes that contribute to phenotypic variations and disease risk [18-22]. However, the association of such variants with the risk of AD has not yet been systematically investigated.

In the present study, we hypothesized that miRNA-related variants could constitute a part of the functional genomic variants influencing the risk of AD. To test this hypothesis, we investigated the association of all genetic variants located in miRNA genes as well as miRNA-binding sites in the 3′ UTR of coding genes with risk of AD using data from the largest available GWAS on late-onset AD [23]. We subsequently integrated our results with publicly available databases (e.g. miRNA and gene expression profiles) and performed experimental studies to provide evidence for the potential function of identified variants in relation to AD.

MATERIALS AND METHODS

Identification of genetic variants in miRNAs and miRNA-binding sites

We made a data set of genetic variants that are located in miRNA-related sequences using two online databases: miRNASNP and PolymiRTS [24, 25]. The mature miRNAs of 20–24 nt in length are produced by processing of the precursor miRNAs (pre-miRNA) (60-80 nt). We thus investigated all variants that are located in human precursor and mature miRNA sequences. We retrieved a total of 2,420 variants in all miRNA-encoding sequences. We excluded variants with minor allele frequency (MAF) < 0.01. Of these, we included 237 single-nucleotide polymorphisms (SNPs) in 206 different miRNAs that were available in the recent GWAS of AD [23]. In addition, we retrieved almost 401,000 variants in the 3'UTR of human miRNA target genes that are predicted to affect the match to the seed region of miRNAs. We included 42,855 SNPs with MAF> 0.01 and available in the AD GWAS [23].

Genome-wide association study on AD

To examine the association of variants in miRNA-related sequences with AD, we used the summary statistics data (stage 1) from the thus far largest GWAS on late-onset AD, including data from 17,008 AD cases and 37,154 controls on 7,055,881 SNPs, was reported by IGAP consortium. A description of the consortium data sets and participants are described elsewhere [23]. We adjusted *p*-value using the Bonferroni correction for the number of tests and significant threshold was set at 2.1×10^{-4} (0.05 /237) for SNPs in miRNAs and 1.2×10^{-6} (0.05 / 42,855) for SNPs in miRNA-binding sites.

Prediction of the variant effect on the miRNA structure

When a miRNA-variant was associated with AD, we used the Vienna RNAfold algorithm (ViennaRNA package 2.0) to predict the effect of that variant on the secondary structure of the pre-miRNA (hairpin structure) [26]. The differences in minimum free energy (MFE) of the thermodynamic ensemble of pre-miRNA sequences containing the mutant versus the wild type alleles may affect the pre-miRNA that is processed to form mature miRNA.

Impact of variant on the miRNA expression

We cloned the pre-miRNA sequence containing either wild type or mutant allele behind the gene encoding green fluorescent protein (GFP) in the expression plasmid MSCV-BC (Murine Stem Cell Virus-Bar Coded) [27], resulting in GFP-miRNA fusion transcripts. The inserts of all constructs were validated by Sanger sequencing. HEK293 cell transfection, total RNA isolation and quantitative PCRs were performed as previously described [27]. All primers are shown in **Table S1**. The experiment was performed in triplicate.

Association of miR-1229-3p target genes with AD

We compiled a list of all predicted target genes of candidate miRNA using two online databases, TargetScan v7.0 [28] and miRanda [29]. The miRNA target genes that are listed in both databases were used for our analyses. We then examined which of the miRNA putative targets are expressed in the human brain using the Human Body Map 2.0 data. Subsequently, we used the AD-GWAS data in a candidate gene approach to identify those target genes that are likely to be involved in AD. We retrieved the summary statistics for the association of all genetic variants in the miRNA target genes with AD from the GWAS data. The significance threshold was set using the Bonferroni correction based on the number of studied SNPs. Additionally, we explored whether target genes of the identified miRNA may play a role in neurologic-related pathways. To do this, we used Ingenuity Pathway Analysis (IPA) (http://www.ingenuity.com/products/ipa/), a knowledge database generated from peer-reviewed scientific publications that enables the discovery of highly represented biological mechanisms, pathways or functions most relevant to the genes of interest from large, quantitative datasets. We uploaded all miRNA target genes and performed a core IPA analysis with default settings. We mapped the miRNA target genes to biological functions or canonical pathways to see whether they are enriched in neurologic-related networks.

Luciferase reporter assay

Primers were designed to amplify the 3'UTR sequence of target gene included in the restriction enzyme sites *XbaI* for the forward primer and *ApaI* for the reverse. The gene 3'UTR sequences (wild-type and mutated), containing the putative binding site of miRNA, were amplified and cloned into the pGL3 luciferase reporter vector downstream of the Luciferase open reading frame (Promega). The primers are shown in **Table S2**. Similarly, mature miRNA sequence was amplified using a forward primer containing a *XhoI* restriction site and a reverse primer containing a *EcoRI* restriction site. The amplified miRNA sequence was then cloned into the pMSCV-BC vector as previously described [27]. The inserts of all constructs were validated by Sanger sequencing. HEK293 cells were plated into 96-well plates and co-transfected with MSCV-miRNA, pGL3 containing the different 3'UTR and a plasmid expressing the Renilla transfection control. Luciferase activity was determined with the Dual-Glo Luciferase Assay System according to manufacturer's protocol (Promega). Renilla luciferase activity was used for normalization. All experiments were performed in triplicate.

Expression of the identified miRNAs and target genes in the human brain

We examined the expression of miR-1229-3p in the human brain. The brain tissues (3 gray matter and 3 white matter) were obtained from the Netherlands Brain Bank (Amsterdam, The Netherlands). All samples were free of neurological disease. For isolation

of total RNA, five cryopreserved sections of 40µm were homogenized in 250µl Trizol reagent (Invitrogen, Carlsbad, CA, USA). Total RNA was isolated from 6 brain samples and. the concentration and purity of RNA samples were determined with a NanoDrop ND-1000 spectrophotometer (NanoDrop, Wilmington, DE). The expression level of miR-1229-3p was determined with TaqMan MicroRNA Assay according to manufacturer's protocol (Applied Biosystems, Foster City, CA, USA). RNU6B was used as an endogenous control. All experiments were performed in triplicate. The Human Body Map 2.0 data was used to examine the expression of miRNA target genes across different tissues. To scan the expression of other miRNAs in the brain, we used HMED (http://bioinfo.life.hust.edu.cn/smallRNA/index.php) and miRmine (http://guanlab.ccmb.med.umich.edu/mirmine/help.html) databases. Moreover, we used human miRNA expression data from four independent studies showing miRNAs that are differentially expressed (both up- and down-regulated) in different brain tissues and also AD patients [30-33].

Functional characteristics of the identified miRNA-binding site variants

The list of identified variants in miRNA-binding sites associated with AD were submitted to the SNAP web tool (http://www.broadinstitute.org/mpg/snap/id) using R^2 threshold > 0.8, limit distance 500 kb, and population panel CEU to retrieve their proxy SNPs in the 1000 Genomes project. We used the HaploReg (v4) web tool to predicts the effect of these SNPs on protein structure, gene regulation, and splicing. This analysis helps us to know whether there are other variants in high LD with the identified SNPs that may drive the association with AD. We scanned the expression quantitative trait loci (cis-eQTL) data using GTEx (http://www.broadinstitute.org/gtex/) and HaploReg (4.1) that show the correlation of the variants with the expression of nearby genes in several tissues [34, 35]. Other information, including miRNA sequences, host gene, miRNAs family and cluster amongst others, and miRNA conservation in different species was obtained from miRBase (release 20) [36] and TargetScan (v7.0) databases.

RESULTS

A polymorphism in miR-1229 was associated with AD

A flow chart of our approach to detect genetic variants in miRNAs and in miRNA-binding sites associated with AD is shown in **Figure 1**. Out of 237 variants located in 206 miRNAs, we found rs2291418:A>G (Chr5:179225324), a low-frequency variant (MAF= 0.02) in the pre-miR-1229 sequence, to be significantly associated with the risk of AD (*p*-value= 6.8×10^{-5} and OR= 1.2). **Figure S1** shows the association of rs2291418 and other variants in the related locus with AD. The associations of all miRNA-variants that are nominally (*p*-value < 0.05) associated with AD are shown in **Table S3**.

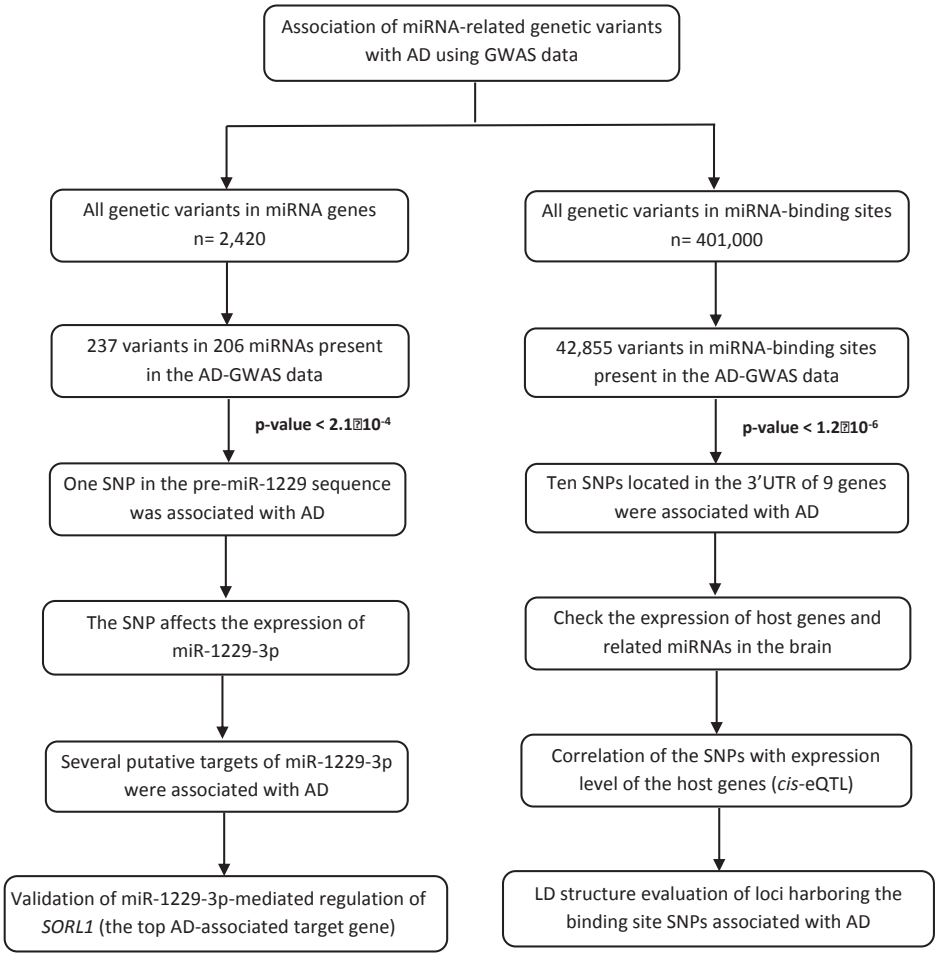


Figure 1 Our approach to detect variants in miRNAs and miRNA-binging sites that are associated with AD.

Impact of rs2291418 on the miR-1229 structure and expression

We generated the pre-miR-1229 hairpin structures containing the rs2291418 major allele (G) and minor allele (A) using the Vienna RNAfold algorithm [26]. We observed a -4.2 kcal/mol difference in the minimum free energy of the thermodynamic ensemble of the predicted structure of pre-miR-1229 mutant compared to wild type (**Figure 2**). This analysis indicated that the variant's mutant allele enhances the stability of the pre-miR-1229 hairpin, suggesting an improvement of the pre-miRNA processing into the mature miR-1229.

To experimentally show the impact of rs2291418 on the pre-miR-1229 processing, we examined the expression levels of mature miR-1229 from the wild type and mutant

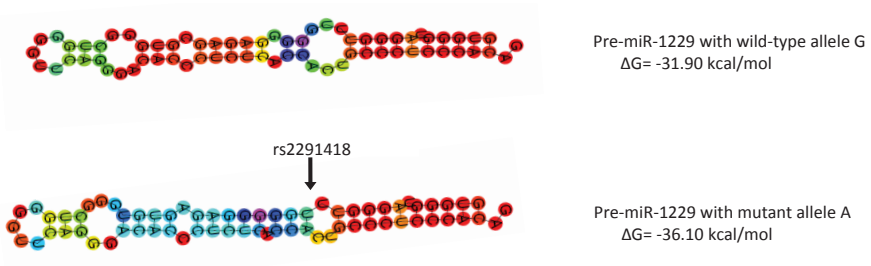


Figure 2. Schematic view of the RNAfold predicted secondary structures of the pre-miR-1229 with the wild-type and the mutant allele. The pre-miRNA structure with lower minimum free energy of the thermodynamic ensemble (ΔG) is expected to thermodynamically be more stable.

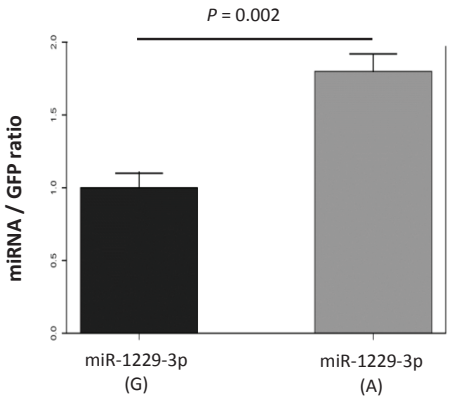


Figure 3. The effect of rs2291418 on the expression level of miR-1229-3p. This figure illustrates a significant increase in the expression level of mature miR-1229-3p from the minor allele A compared to the major allele G relative to GFP.

pre-miRNA construct. To this end, HEK293 cells were transfected with pMCSV vectors expressing transcripts with EGFP and wild type or mutant pre-miR-1229. Both the pre-miR-1229 wild type and pre-miR-1229 mutant transfected cells had equal levels of the EGFP-miRNA transcripts. However, the mature miR-1299-3p levels were significantly increased by 70% in cells transfected with pre-miR-1299 containing the mutant allele compared to the wild-type allele (*p*-value= 0.002) (**Figure 3**), indicating that the variant improves the processing or stability of miR-1229.

miR-1229-3p target genes associated with AD

We assessed which target genes of miR-1229-3p are implicated in pathways in nervous system or neurological disorders and may have a role in developing AD. To this end, we compiled a list of all predicted target genes of miR-1299-3p using the two most commonly used miRNA target prediction algorithms (TargetScan and miRanda) (n=960). Both miRNA and its target genes should be expressed in the same tissue for any biological function to be exerted. Thus, we checked the expression of miR-1229-3p in the brain. Our data showed that miR-1229-3p is abundantly expressed (average Ct-value: 28.8) in both white and gray matter of the human brain (**Table S4**). In addition, miRNA expression

databases (HMED and miRmine) show that miR-1229-3p is expressed at higher levels in the brain compared to other tissues, indicating an active regulatory role of miR-1229-3p on its target genes in the brain. Using the Human Body Map 2.0 data, we subsequently found that 750 miR-1229-3p target genes are expressed in the brain,.

We then used the AD-GWAS data in a candidate gene approach to identify those target genes that are likely to be involved in developing AD [23]. We assessed the association of genetic variants in the 750 brain-expressed target genes of miR-1229-3p with AD. **Table S5** shows top ten target genes of miR-1229-3p with the most significant association with AD. In addition, our pathway analysis using IPA demonstrated several target genes of miR-1229-3p to be involved in the network of Nervous System Development and Neurological disease (**Table S6**).

miR-1229-3p-mediated regulation of *SORL1*

We selected *SORL1* which is the most significant AD-association target gene and examined whether miR-1229-3p controls its expression level *in vitro*. We generated expression vector containing the pre-miR-1229-3p sequence and co-transfected the construct with Luciferase reporters containing the 3' UTR (wild-type or mutant miR-1229-3p binding site) of *SORL1*. We found that over-expression of miR-1229-3p significantly decreases the Luciferase activity of the reporter containing the wild type *SORL1* fragment, compared to the mutated *SORL1* reporter (*p*-value= 0.003) (**Figure 4**). In addition, this experiment showed a dose-dependent regulation of *SORL1* by miR-1229-3p (**Figure S2**). As both the miRNA and *SORL1* are expressed in thebrain, our data suggest for miR-1229-3p-mediated regulation of *SORL1* in the human brain.

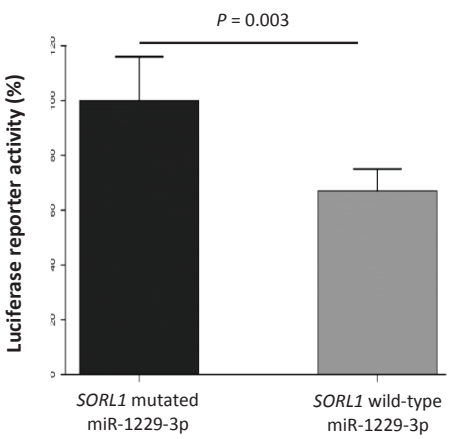


Figure 4. miR-1229-3p-mediated regulation of *SORL1*. This figure illustrates a significant decrease of the mean relative luciferase activity of the wild-type 3'UTR of *SORL1* reporter, compared to the mutated 3'UTR of *SORL1* reporter in the presence of miR-1229-3p.

Ten miRNA binding site variants were associated with AD

We examined the association of 42,855 variants in putative miRNA-binding sites with AD that are illustrated in a Manhattan plot (**Figure 5**). Of these, ten variants located in the 3'UTR of 9 different genes passed the Bonferroni corrected significance threshold (p-value < 1.2×10^{-6}). These variants would potentially interfere with miRNA-mediated regulation of their host genes by disrupting, creating or modifying a number of miRNA-binding sites that are depicted in **Table S7**. Out of 9 genes hosting the identified variants, the association of 7 genes with AD have been reported in the original GWAS (p-value < 5×10^{-8}) [23]. In addition, we found two new susceptibility genes to be potentially associated with AD, including *DMWD* (p-value= 8.48×10^{-7}) and *HBEGF* (p-value= 3.96×10^{-7}). Regional plots showing association of the ten miRNA-binding site SNPs and their close by variants with AD are shown in **Figure S3**.

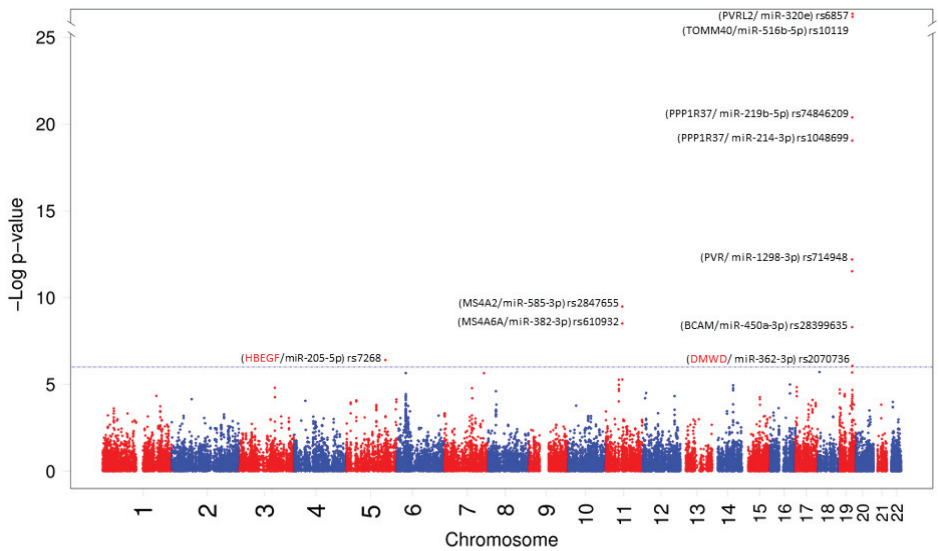


Figure 5. The Manhattan plot shows the association of 42,855 SNPs in miRNA-binding sites with AD. Dashed line indicates the significant study threshold which was set at 1.2×10^{-6}. Genes hosting the variants significantly associated with AD are highlighted. The two newly identified genes for AD are depicted in red.

Functional annotation of the miRNA-binding site variants associated with AD

We searched the literature and several web tools to identify evidence supporting the functional role of ten miRNA-binding site SNPs associated with AD. **Table S8** shows a list of all web tools and databases that we used for our analyses. First, we examined the cis-expression quantitative trait loci (eQTLs) for 9 genes hosting the identified SNPs. We found that rs6857, rs2070736, rs2847655, rs10119 and rs610932 are correlated with differential expression levels of their host genes in blood (**Table 1**), showing that these SNPs alter expression levels of their host genes. In addition, the cis-regulatory effects of

Table 1. A summary of evidence supporting the potential functional role for the 10 identified miRNA-binding site variants associated with Alzheimer's disease

		rs6857	rs2847655	rs610932	rs74846209	rs1048699	rs10119	rs714948	rs28399635	rs2070736	rs7268
GWAS	*p*-value	2.5×10^{-575}	3.3×10^{-10}	3.1×10^{-9}	4.1×10^{-21}	9.0×10^{-20}	1.2×10^{-342}	6.3×10^{-13}	5.0×10^{-9}	8.5×10^{-7}	4.0×10^{-7}
	OR	3.2	1.1	0.9	2.3	1.2	1.3	1.1	0.9	1.1	1.1
Gene	Host gene	PVRL2	MS4A2	MS4A6A	PPP1R37	PPP1R37	TOMM40	PVR	BCAM	DMWD	HBEGF
	Brain Exp.	10.0	NA	6.2	11.0	11.0	9.7	5.7	3.0	11.5	9.2
miRBS	miRNA ID	miR-320e	miR-585-3p	miR-382-3p	miR-219b-5p	miR-214-3p	miR-516b-5p	miR-1298-3p	miR-450a-3p	miR-362-3p	miR-205-5p
	score change	-0.19	-0.48	-0.10	-0.29	-0.08	-0.06	-0.08	-0.16	-0.15	-0.12
eQTL	Blood	9.5×10^{-5}(-)	6.5×10^{-22}(-)	2.1×10^{-35}(+)	-	-	+	-	-	2.4×10^{-10} (+)	-
	Brain	-	-	+	-	-	-	+	-	-	-
Proxy	nr. proxies	0	94	72	18	18	0	1	0	18	8
	Non-synon	0	0	1	1	1	0	0	0	0	0

miRBS, miRNA binding site; OR, odd ratio; Exp, Expression; Non-synon, non-synonymous SNP; NA; Not available. miRNAs in this table are highly conserved miRNAs related to each SNPs. The score change is the Context score change using PolymiRTS database (v3.0).

rs610932 and rs714948 on the expression of *MS4A6A* and *PVR* in cerebellum and temporal cortex have been reported [37]. Using the HaploReg web tool (v4), we found that 7 of the identified SNPs have no non-synonymous proxy SNPs in their loci, might suggesting the identified SNPs in miRNA binding sites to drive the observed associations (**Table S9**). The Human Body Map 2.0 data show that 8 out of the 9 host genes are expressed in the human brain. In addition, miRNA expression databases show that miR-214-3p, miR-212-5p, miR-204-3p, miR-362-3p, miR-450a and miR-320 are expressed in the brain. A summary of our findings regarding on potential functionality of the ten identified variants in miRNA-binding sites are shown in **Table 1** and **Table S9**.

DISCUSSION

Recent studies have shown the critical role of miRNAs in neuronal development and function [19, 38-40]. In addition, a number of miRNAs have been reported to be implicated in the development of AD [7-12]. However, these studies are mainly focused on differentially expressed miRNAs using expression arrays in a small number of samples. A few studies have linked miRNAs with neuro-degenerative disorders using genetic association in a candidate gene approach [41, 42]. In this study, we performed a genome-wide investigation to identify genetic variants in miRNAs and miRNA-binding sites that are associated with AD using data from the thus far largest GWAS on AD [23].

We found that rs2291418 in the miR-1229 precursor, a miRNA which is abundantly expressed in the brain, is associated with AD. The pre-miR-1299 is precursor of two mature miRNAs (miR-1229-3p and miR-1229-5p), however, the 3p miRNA has been shown to be the predominant product [43]. *In silico* analysis indicated that the mutant allele improves the stability of the hairpin structure of pre-miR-1229. Our miRNA expression experiments further showed that the mutant allele enhances the stability of mature miR-1229-3p. It has been shown previously that variants located at the stem region of miRNA precursors could alter Drosha-mediated processing [22, 44, 45]. Although many studies have shown that genetic variants within pre-miRNA sequences reduce the mature miRNA levels [22, 44, 45], a few studies have reported such variants to enhance miRNA processing and expression [45].

We revealed miR-1229-3p target genes that are associated with AD and might be downstream targets of the miRNA in relation to AD. An optimal approach would have been to check the co-expression of all potential target genes of miR-1229-3p in brain tissues and examine their correlation with the miRNA expression to distill the potential mediators. However, collecting samples carrying the rare mutant allele (MAF=0.03) of miR-1229-3p is challenging and was not possible for the investigators. Alternatively, we used GWAS data to identify the target genes that are more likely to be involved in

the development of AD. By this approach, we could highlight a number of miR-1229-3p target genes that are known for neurological-related pathways and diseases, such as *SORL1, MCFD2, COL25A1* and *BMP2* [47-49]. This approach, however, may overlook certain target genes that mediate the miRNA function on AD. Next, we experimentally verified that miR-1229-3p interacts with the predicted target site in the 3'UTR of *SORL1,* the most significant AD-associated miR-1229-3p target, which is abundantly expressed in the human brain. These data indicate that miR-1229-3p directly controls the expression of *SORL1* in the brain. Several studies have reported decreased *SORL1* expression to be mechanistically involved in causing AD [49]. It has been shown that *SORL1* directs trafficking of the amyloid precursor protein (APP) into recycling pathways and that when *SORL1* is reduced, APP is sorted into amyloid beta peptide in AD [47]. Higher expression levels of miR-1229-3p in rs2291418 mutant allele carriers may reduce *SORL1* level and subsequently increase AD risk.

We identified ten miRNA-binding site variants associated with AD. Among them, rs6857 in *PVRL2* has the most significant association with AD [50, 51]. The rs6857 mutant allele is predicted to create a target site for miR-320e in the 3'UTR of *PVRL2*. Previously, we have demonstrated that expression of the mutant *PVRL2* is regulated by miR-320e, resulting in decreased levels of the *PVRL2* transcript [18]. The eQTL data also show a correlation between the rs6857 mutant allele and lower expression levels of *PVRL2*. Both *PVRL2* and miR-320e are also expressed in the brain [33]. These data may indicate that rs6857 increases AD risk through down-regulation of *PVRL2* expression by miR-320e.

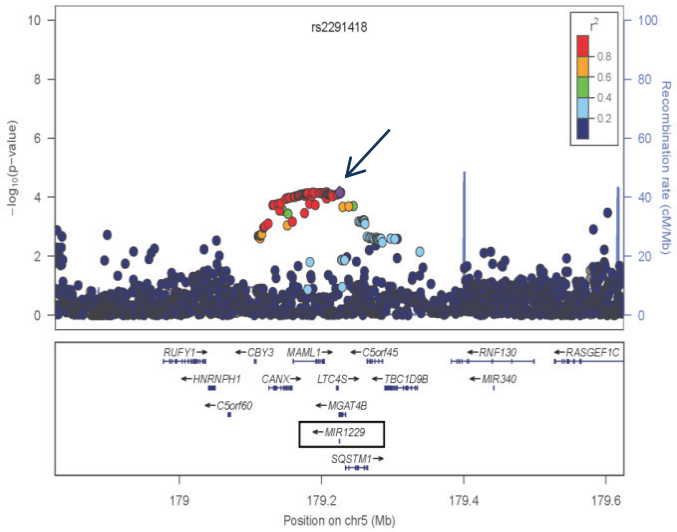
We found that two SNPs in miRNA-binding sites of *MS4A6A* and *MS4A2* are associated with AD susceptibility. The first SNP is rs610932 in the 3'UTR of *MS4A6A* predicted to disrupt a binding site of miR-382-3p. *MS4A6A* is a well-known gene for susceptibility to AD [37, 52, 53], and miR-382-3p is expressed in the brain. We observed that carriers of the minor allele containing rs610932 SNP, have a higher expression level of *MS4A6A* transcript in the blood eQTL data. The variant has been also shown to be associated with an altered expression level of *MS4A6A* in cerebellum and temporal cortex [37]. Therefore, an allele-specific regulation of *MS4A6A* by miR-328-3p can serve as an explanation for the observed association with AD. We also identified two SNPs in the 3'UTR of *PPP1R37* which potentially disrupt binding sites of miR-214-3p and miR204-3p. *PPP1R37* gene has previously been reported to be associated with AD and is expressed in the brain [23, 54]. Both miRNAs have been shown to be implicated in neurodegenerative disorders including Parkinson's and Huntington's disease [55]. A disruption of miR-214-3p and miR-204-3p mediated regulations of *PPP1R37* may be considered as a reason for the association of rs74846209 and rs1048699 with AD. Future experimental studies warrant to determine the function of the variants and miRNAs in AD.

We report *DMWD* and *HBEGF* as two new susceptibility genes for AD that need to be replicated in future association studies. Rs2070736 in the 3'UTR of *DMWD* creates

potential binding sites for miR-362-3p and miR-329-3p. These miRNAs and *DMWD* are expressed in the brain [32, 56] and the eQTL analysis showed that the mutant allele is associated with lower expression level of the *DMWD* transcript. These data suggest that rs2070736 may change the expression level of *DMWD* by affecting miRNA-mediated regulation. *DMWD* plays a role in developing mental symptoms in severe cases of myotonic dystrophy [57]. Interestingly, Alzheimer's neurofibrillary changes in brain are reported to be present in myotonic dystrophy [58]. Together, the link between myotonic dystrophy and AD is notable and may indicate a potential role for *DMWD* in AD as well. Rs7268 in the 3'UTR of *HBEGF* is expected to create a binding site for miR-205-5p, a brain-expressed miRNA. It has been reported that the membrane bound for *HBEGF* is constitutively expressed and present on the blood-brain barrier. *HBEGF* expression is also shown to be abundantly up-regulated in cerebral blood vessels in inflammatory situations such as AD, Parkinson's disease, stroke, epilepsy and encephalitis [59, 60]. Therefore, the aberrant miR-205-5p regulation of *HBEGF* could be a potential mechanism underlying the identified association.

Collectively, in a genome-wide investigation we identified a variant in pre-miR-1229 and ten variants in miRNA-binding sites associated with AD. Our *in silico* and *in vitro* analyses showed a miR-1229-mediated regulation of *SORL1*, which is a known gene involved in AD. Further experimental studies on the identified variants as well as miRNA profiling in AD patients are needed to determine the role of highlighted miRNAs and target genes in the pathogenesis of AD. Our study further suggest altered miRNA-dependent regulation of genes due to variants in 3'UTRs may explain part of the associations identified by GWAS for AD.

Supplement available online at:
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4916596/



Supplemental Figure S1. Regional plot showing the association of miR-1229 variant with Alzheimer's disease generated by Locuszoom web tool.

Supplemental Table S5. Putative miR-1229-3p target genes with the most significant association with AD

Gene	SNP ID	P-value	Gene name	Expression in the brain
**SORL1*	rs11218343	4.98E-11	Sortilin-Related Receptor, L(DLR Class) A Repeats Containing	26.94
PFDN1	rs6580473	8.73E-06	Prefoldin Subunit 1	54.75
**MCFD2*	rs6715234	8.82E-06	Neural Stem Cell-Derived Neuronal Survival Protein	34.90
RAB31	rs1015228	1.20E-05	Ras-Related Protein Rab-22B	61.99
LRRC32	rs1893306	3.35E-05	Leucine Rich Repeat Containing 32.	10.29
**COL25A1*	rs11736110	3.63E-05	Alzheimer Disease Amyloid-Associated Protein	1.12
ZNF594	rs8081019	3.72E-05	Zinc Finger Protein 594	1.02
JARID2	rs764650	5.90E-05	Jumonji, AT Rich Interactive Domain 2	4.10
**BMP2*	rs6054767	5.92E-05	Bone Morphogenetic Protein 2	2.56
RFX3	rs10972598	6.59E-05	Regulatory Factor X 3	4.60

The *p*-value for each target gene is *p*-value of the SNP in the gene with the most significant association in GWAS of AD (Lambert, et al., 2013). The expression values are as fragments per kb of exon per million reads (FPKMs), which is a measure of gene expression normalized to size of the gene and RNA-seq library size in the Human Body Map 2.0 data. *Genes that have been shown to be involved in neurological-related pathways (GeneCards database).

REFERENCES

1. Brookmeyer R, Johnson E, Ziegler-Graham K, Arrighi HM: Forecasting the global burden of Alzheimer's disease. Alzheimers Dement 2007, 3:186-191.
2. Schmechel DE, Saunders AM, Strittmatter WJ, Crain BJ, Hulette CM, Joo SH, Pericak-Vance MA, Goldgaber D, Roses AD: Increased amyloid beta-peptide deposition in cerebral cortex as a consequence of apolipoprotein E genotype in late-onset Alzheimer disease. Proc Natl Acad Sci U S A 1993, 90:9649-9653.
3. Frankfort SV, Tulner LR, van Campen JP, Verbeek MM, Jansen RW, Beijnen JH: Amyloid beta protein and tau in cerebrospinal fluid and plasma as biomarkers for dementia: a review of recent literature. Curr Clin Pharmacol 2008, 3:123-131.
4. Crews L, Masliah E: Molecular mechanisms of neurodegeneration in Alzheimer's disease. Hum Mol Genet 2010, 19:R12-20.
5. Hampel H, Frank R, Broich K, Teipel SJ, Katz RG, Hardy J, Herholz K, Bokde AL, Jessen F, Hoessler YC, et al: Biomarkers for Alzheimer's disease: academic, industry and regulatory perspectives. Nat Rev Drug Discov 2010, 9:560-574.
6. Kivipelto M, Helkala EL, Laakso MP, Hanninen T, Hallikainen M, Alhainen K, Soininen H, Tuomilehto J, Nissinen A: Midlife vascular risk factors and Alzheimer's disease in later life: longitudinal, population based study. BMJ 2001, 322:1447-1451.
7. Geekiyanage H, Chan C: MicroRNA-137/181c regulates serine palmitoyltransferase and in turn amyloid beta, novel targets in sporadic Alzheimer's disease. J Neurosci 2011, 31:14820-14830.
8. Long JM, Lahiri DK: MicroRNA-101 downregulates Alzheimer's amyloid-beta precursor protein levels in human cell cultures and is differentially expressed. Biochem Biophys Res Commun 2011, 404:889-895.
9. Patel N, Hoang D, Miller N, Ansaloni S, Huang Q, Rogers JT, Lee JC, Saunders AJ: MicroRNAs can regulate human APP levels. Mol Neurodegener 2008, 3:10.
10. Leidinger P, Backes C, Deutscher S, Schmitt K, Mueller SC, Frese K, Haas J, Ruprecht K, Paul F, Stahler C, et al: A blood based 12-miRNA signature of Alzheimer disease patients. Genome Biol 2013, 14:R78.
11. Tan L, Yu JT, Tan L: Causes and Consequences of MicroRNA Dysregulation in Neurodegenerative Diseases. Mol Neurobiol 2015, 51:1249-1262.
12. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, Powell C, Vedantam S, Buchkovich ML, Yang J, et al: Genetic studies of body mass index yield new insights for obesity biology. Nature 2015, 518:197-206.
13. Bartel DP: MicroRNAs: target recognition and regulatory functions. Cell 2009, 136:215-233.
14. 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.
15. Ghanbari M, de Vries PS, de Looper H, Peters MJ, Schurmann C, Yaghootkar H, Dorr M, Frayling TM, Uitterlinden AG, Hofman A, et al: A genetic variant in the seed region of miR-4513 shows pleiotropic effects on lipid and glucose homeostasis, blood pressure, and coronary artery disease. Hum Mutat 2014, 35:1524-1531.
16. Dorn GW, 2nd, Matkovich SJ, Eschenbacher WH, Zhang Y: A human 3' miR-499 mutation alters cardiac mRNA targeting and function. Circ Res 2012, 110:958-967.
17. Georges M, Clop A, Marcq F, Takeda H, Pirottin D, Hiard S, Tordoir X, Caiment F, Meish F, Bibe B, et al: Polymorphic microRNA-target interactions: a novel source of phenotypic variation. Cold Spring Harb Symp Quant Biol 2006, 71:343-350.

18. Ghanbari M, Franco OH, de Looper H, Hofman A, Erkeland S, Dehghan A: Genetic Variations in miRNA Binding Sites Affect miRNA-Mediated Regulation of Several Genes Associated with Cardiometabolic Phenotypes. Circ Cardiovasc Genet 2015.
19. Richardson K, Nettleton JA, Rotllan N, Tanaka T, Smith CE, Lai CQ, Parnell LD, Lee YC, Lahti J, Lemaitre RN, et al: Gain-of-function lipoprotein lipase variant rs13702 modulates lipid traits through disruption of a microRNA-410 seed site. Am J Hum Genet 2013, 92:5-14.
20. Ghanbari M, Sedaghat S, de Looper HW, Hofman A, Erkeland SJ, Franco OH, Dehghan A: The association of common polymorphisms in miR-196a2 with waist to hip ratio and miR-1908 with serum lipid and glucose. Obesity (Silver Spring) 2015, 23:495-503.
21. Landi D, Gemignani F, Naccarati A, Pardini B, Vodicka P, Vodickova L, Novotny J, Forsti A, Hemminki K, Canzian F, Landi S: Polymorphisms within micro-RNA-binding sites and risk of sporadic colorectal cancer. Carcinogenesis 2008, 29:579-584.
22. Ghanbari M, Darweesh SK, de Looper HW, van Luijn MM, Hofman A, Ikram MA, Franco OH, Erkeland SJ, Dehghan A: Genetic Variants in MicroRNAs and their Binding Sites are Associated with the Risk of Parkinson Disease. Hum Mutat 2015.
23. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, DeStafano AL, Bis JC, Beecham GW, Grenier-Boley B, et al: Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet 2013, 45:1452-1458.
24. Bao L, Zhou M, Wu L, Lu L, Goldowitz D, Williams RW, Cui Y: PolymiRTS Database: linking polymorphisms in microRNA target sites with complex traits. Nucleic Acids Res 2007, 35:D51-54.
25. Gong J, Tong Y, Zhang HM, Wang K, Hu T, Shan G, Sun J, Guo AY: Genome-wide identification of SNPs in microRNA genes and the SNP effects on microRNA target binding and biogenesis. Hum Mutat 2012, 33:254-263.
26. Lorenz R, Bernhart SH, Honer Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL: ViennaRNA Package 2.0. Algorithms Mol Biol 2011, 6:26.
27. Meenhuis A, van Veelen PA, de Looper H, van Boxtel N, van den Berge IJ, Sun SM, Taskesen E, Stern P, de Ru AH, van Adrichem AJ, et al: MiR-17/20/93/106 promote hematopoietic cell expansion by targeting sequestosome 1-regulated pathways in mice. Blood 2011, 118:916-925.
28. Agarwal V, Bell GW, Nam JW, Bartel DP: Predicting effective microRNA target sites in mammalian mRNAs. Elife 2015, 4.
29. Betel D, Koppal A, Agius P, Sander C, Leslie C: Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. Genome Biol 2010, 11:R90.
30. Cogswell JP, Ward J, Taylor IA, Waters M, Shi Y, Cannon B, Kelnar K, Kemppainen J, Brown D, Chen C, et al: Identification of miRNA changes in Alzheimer's disease brain and CSF yields putative biomarkers and insights into disease pathways. J Alzheimers Dis 2008, 14:27-41.
31. Moreau MP, Bruse SE, Jornsten R, Liu Y, Brzustowicz LM: Chronological changes in microRNA expression in the developing human brain. PLoS One 2013, 8:e60480.
32. Liu DZ, Tian Y, Ander BP, Xu H, Stamova BS, Zhan X, Turner RJ, Jickling G, Sharp FR: Brain and blood microRNA expression profiling of ischemic stroke, intracerebral hemorrhage, and kainate seizures. J Cereb Blood Flow Metab 2010, 30:92-101.
33. Jeyaseelan K, Lim KY, Armugam A: MicroRNA expression in the blood and brain of rats subjected to transient focal ischemia by middle cerebral artery occlusion. Stroke 2008, 39:959-966.
34. Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, Christiansen MW, Fairfax BP, Schramm K, Powell JE, et al: Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet 2013.

35. Grundberg E, Small KS, Hedman AK, Nica AC, Buil A, Keildson S, Bell JT, Yang TP, Meduri E, Barrett A, et al: Mapping cis- and trans-regulatory effects across multiple tissues in twins. Nat Genet 2012, 44:1084-1089.
36. Kozomara A, Griffiths-Jones S: miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res 2014, 42:D68-73.
37. Hollingworth P, Harold D, Sims R, Gerrish A, Lambert JC, Carrasquillo MM, Abraham R, Hamshere ML, Pahwa JS, Moskvina V, et al: Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. Nat Genet 2011, 43:429-435.
38. Dweep H, Sticht C, Pandey P, Gretz N: miRWalk--database: prediction of possible miRNA binding sites by "walking" the genes of three genomes. J Biomed Inform 2011, 44:839-847.
39. Kumar M, Nath S, Prasad HK, Sharma GD, Li Y: MicroRNAs: a new ray of hope for diabetes mellitus. Protein Cell 2012, 3:726-738.
40. Chen L, Song J, Cui J, Hou J, Zheng X, Li C, Liu L: microRNAs regulate adipocyte differentiation. Cell Biol Int 2013.
41. Qi L, Hu Y, Zhan Y, Wang J, Wang BB, Xia HF, Ma X: A SNP site in pri-miR-124 changes mature miR-124 expression but no contribution to Alzheimer's disease in a Mongolian population. Neurosci Lett 2012, 515:1-6.
42. Saba R, Medina SJ, Booth SA: A functional SNP catalog of overlapping miRNA-binding sites in genes implicated in prion disease and other neurodegenerative disorders. Hum Mutat 2014, 35:1233-1248.
43. Berezikov E, Chung WJ, Willis J, Cuppen E, Lai EC: Mammalian mirtron genes. Mol Cell 2007, 28:328-336.
44. Cammaerts S, Strazisar M, De Rijk P, Del Favero J: Genetic variants in microRNA genes: impact on microRNA expression, function, and disease. Front Genet 2015, 6:186.
45. Sun G, Yan J, Noltner 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.
46. Ryan BM, Robles AI, Harris CC: Genetic variation in microRNA networks: the implications for cancer research. Nat Rev Cancer 2010, 10:389-402.
47. Rogaeva E, Meng Y, Lee JH, Gu Y, Kawarai T, Zou F, Katayama T, Baldwin CT, Cheng R, Hasegawa H, et al: The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. Nat Genet 2007, 39:168-177.
48. Tong Y, Xu Y, Scearce-Levie K, Ptacek LJ, Fu YH: COL25A1 triggers and promotes Alzheimer's disease-like pathology in vivo. Neurogenetics 2010, 11:41-52.
49. Vardarajan BN, Zhang Y, Lee JH, Cheng R, Bohm C, Ghani M, Reitz C, Reyes-Dumeyer D, Shen Y, Rogaeva E, et al: Coding mutations in SORL1 and Alzheimer disease. Ann Neurol 2015, 77:215-227.
50. Nicodemus KK, Stenger JE, Schmechel DE, Welsh-Bohmer KA, Saunders AM, Roses AD, Gilbert JR, Vance JM, Haines JL, Pericak-Vance MA, Martin ER: Comprehensive association analysis of APOE regulatory region polymorphisms in Alzheimer disease. Neurogenetics 2004, 5:201-208.
51. Consortium EP, Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, et al: Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 2007, 447:799-816.
52. Naj AC, Jun G, Beecham GW, Wang LS, Vardarajan BN, Buros J, Gallins PJ, Buxbaum JD, Jarvik GP, Crane PK, et al: Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. Nat Genet 2011, 43:436-441.

53. Deng YL, Liu LH, Wang Y, Tang HD, Ren RJ, Xu W, Ma JF, Wang LL, Zhuang JP, Wang G, Chen SD: The prevalence of CD33 and MS4A6A variant in Chinese Han population with Alzheimer's disease. Hum Genet 2012, 131:1245-1249.
54. Miyashita A, Koike A, Jun G, Wang LS, Takahashi S, Matsubara E, Kawarabayashi T, Shoji M, Tomita N, Arai H, et al: SORL1 is genetically associated with late-onset Alzheimer's disease in Japanese, Koreans and Caucasians. PLoS One 2013, 8:e58618.
55. Heman-Ackah SM, Hallegger M, Rao MS, Wood MJ: RISC in PD: the impact of microRNAs in Parkinson's disease cellular and molecular pathogenesis. Front Mol Neurosci 2013, 6:40.
56. Alwazzan M, Newman E, Hamshere MG, Brook JD: Myotonic dystrophy is associated with a reduced level of RNA from the DMWD allele adjacent to the expanded repeat. Hum Mol Genet 1999, 8:1491-1497.
57. Westerlaken JH, Van der Zee CE, Peters W, Wieringa B: The DMWD protein from the myotonic dystrophy (DM1) gene region is developmentally regulated and is present most prominently in synapse-dense brain areas. Brain Res 2003, 971:116-127.
58. Yoshimura N, Otake M, Igarashi K, Matsunaga M, Takebe K, Kudo H: Topography of Alzheimer's neurofibrillary change distribution in myotonic dystrophy. Clin Neuropathol 1990, 9:234-239.
59. Rip J, Schenk GJ, de Boer AG: Differential receptor-mediated drug targeting to the diseased brain. Expert Opin Drug Deliv 2009, 6:227-237.
60. Schenk GJ, Haasnoot PC, Centlivre M, Legrand N, Rip J, de Boer AG, Berkhout B: Efficient CRM197-mediated drug targeting to monocytes. J Control Release 2012, 158:139-147.

CHAPTER 3.2

Genetic variants in miRNAs and miRNA-binding sites are associated with Parkinson's disease

Mohsen Ghanbari, Sirwan K. Darweesh, Hans de Looper,
Marvin van Luijn, M. Arfan Ikram, Oscar H. Franco, Stefan J. Erkeland,
Abbas Dehghan

Manuscript based on this chapter

Genetic variants in microRNAs and their binding sites are associated with the risk of Parkinson disease.

Human Mutation. 2016, 37(3):292-300.

ABSTRACT

Background. MicroRNAs (miRNAs) are small non-coding RNAs that serve as key regulators of gene expression. MiRNAs have been shown to be involved in a wide-range of biological processes including neurodegenerative diseases. Genetic variants in miRNAs and in miRNA-binding sites on their target genes could affect miRNA function and contribute to disease risk.

Methods and results. We investigated the association of miRNA-related genetic variants with Parkinson's disease (PD) using data from the largest available GWAS on PD. Out of 243 variants in miRNAs, we identified rs897984:T>C in miR-4519 (p-value= 1.3×10^{-5}, OR= 0.93) and rs11651671:A>G in miR-548at-5p (p-value= 1.1×10^{-6}, OR= 1.09) to be associated with PD. We showed that the variant's mutant alleles change the secondary structure and decrease expression level of their related miRNAs. Subsequently, we highlighted target genes that might mediate the effects of miR-4519 and miR-548at-5p on PD. Among them, we experimentally showed that *NSF* is a direct target of miR-4519. Furthermore, among 48,844 variants in miRNA-binding sites, we found 32 variants (in 13 genes) that are associated with PD. Four of the host genes, *CTSB, STX1B, IGSF9B* and *HSD3B7*, had not previously been reported to be associated with PD. We provide evidence supporting the potential impact of the identified miRNA binding site-variants on miRNA-mediated regulation of their host genes.

Conclusion. Our findings support the idea that miRNAs play a role in PD and highlight a number of variants in miRNA-related sequences that may affect miRNA-mediated regulation of PD genes.

INTRODUCTION

Neurodegenerative diseases collectively represent one of the major worldwide causes of morbidity and healthcare costs to society [1]. Parkinson disease (PD) is the second most common neurodegenerative disorder [2], and its prevalence and burden at the population level are projected to grow dramatically as the size of elderly populations increases [3]. Clinically, PD is characterized by a combination of motor symptoms, known as parkinsonism, and a range of non-motor symptoms, such as cognitive decline and autonomic dysfunction, that contribute to a devastating loss of quality of life [4, 5]. PD is a complex disease, and genetic factors have a substantial impact on the phenotypic variation of the disease [6]. Over the past few decades, enormous efforts have been done to discover genetic factors that play a role in development of PD. In recent years, the large-scale genome-wide association studies (GWAS) have enabled the discovery of hundreds genetic variants that are associated with PD [7, 8]. Nevertheless, most of the identified variants are mapped to non-coding regions of the genome, and their causal mechanisms remain to be investigated.

MicroRNAs (miRNA) are a class of small non-coding RNAs, that post-transcriptionally regulate gene expression [9, 10]. MiRNAs have been shown to be involved in a wide range of biological processes and human diseases including neurodegenerative disorders [11-13]. In addition, a number of dysregulated miRNAs have been reported to be associated with PD in patients and animal models [14-16]. The regulatory functions of miRNAs are accomplished through binding of the nucleotides 2-8 from their 5′-end (the seed region) to the complementary sequences at the target mRNAs, resulting in repression of translation or a decreased stability of target mRNAs [9, 10]. Genetic variants that fall in miRNA-related sequences may affect miRNA function and due to aberrant expression of the miRNA target genes, the variants could modify susceptibility to disease [17]. Recently, we and others have been able to show a number of polymorphisms in miRNAs or their target genes that may contribute to phenotypic variations [18-22]. Further, no systematic investigation of the impact of such variants on the risk of PD has been published to date.

In this study, we examined the association of variants in miRNAs as well as miRNA binding sites with PD using data from the largest available GWAS. We integrated GWAS findings with various *in silico* information and performed experimental studies to provide evidence for functionality of the miRNA-related variants associated with PD.

MATERIALS AND METHODS

Identification of genetic variants in miRNAs and miRNA-binding sites

We retrieved genetic variants that are localized in miRNAs and in miRNA binding sites within the 3'UTR of their target genes by reviewing the literature and using two online databases: miRNASNP (v2.0) [23] and PolymiRTS (v3.0) [24]. We retained a total of 2,420 variants in all human precursor (60-80 nt) and mature (20-24 nt) miRNA sequences. We excluded variants with minor allele frequency (MAF) < 0.01. Of the remaining variants, we included 243 single-nucleotide polymorphisms (SNPs) in 214 miRNAs that were present in the recent GWAS of PD [8]. Furthermore, we retained around 401,000 miRNA binding site-variants that were predicted to affect the match to the seed region of miRNAs. Of these, 48,845 SNPs with MAF > 0.01 and present in the GWAS of PD were included (**Figure 1**).

Genome-wide association study on PD

We examined the association of retrieved variants in miRNAs and miRNA-binding sites with PD using summary statistics data from the largest GWAS on PD across 13,708 PD cases (39% female) and 95,282 controls (46% female) [8]. The GWAS data were imputed to 1000 Genomes project reference panel, providing data for 7,893,274 variants. All participating studies in the PD-GWAS had provided informed consent for participation in genetics studies and were approved by their local ethical committees. More details about the consortium and participants are described elsewhere [8]. We used the Bonferroni correction, to adjust *p*-value for the number of tests, and significant threshold was set at 2.06×10^{-4} (0.05 / 243) for variants in miRNAs and 1.02×10^{-6} (0.05 / 48,844) for variants in miRNA binding sites.

Analyzing the effect of miRNA variants on the miRNA structure and expression

For miRNA variants that were associated with PD, we used the Vienna RNAfold algorithm (package 2.0) to predict the variant effect on the hairpin structure of miRNA [25]. Difference in minimum free energy (MFE) of the thermodynamic ensemble of precursor miRNA (pre-miRNA) sequence containing the mutant versus the wild type allele may indicate an altered miRNA processing. Furthermore, to experimentally examine the variant's effect on the expression level of mature miRNA, we cloned the pre-miRNA sequences containing either the wild type or mutant allele behind the gene encoding green fluorescent protein (GFP) in the expression plasmid MSCV-BC, resulting in GFP-miRNA fusion transcripts [26]. The inserts of all constructs were validated by Sanger sequencing. HEK293 cell transfection, total RNA isolation and quantitative PCRs were performed as previously described [26]. The primers are shown in **Table S1**. The experiment was performed in triplicate.

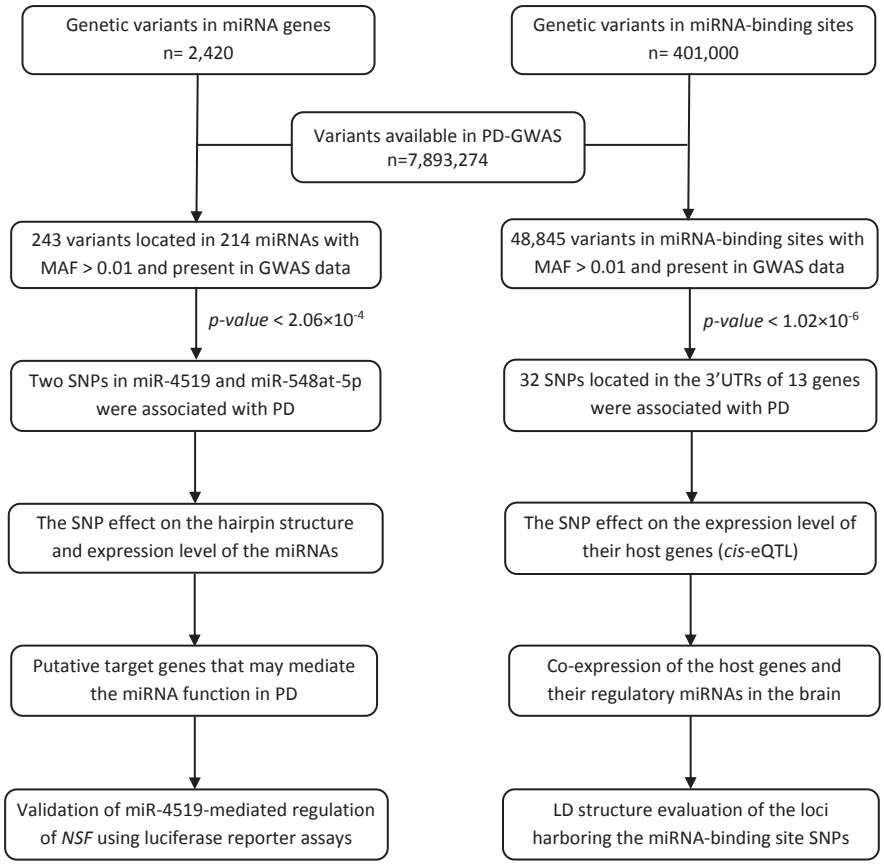


Figure 1. This flow chart describes our approach to identify genetic variants in miRNAs and miRNA-binding sites that are associated with Parkinson disease. MAF, minor allele frequency; LD, linkage disequilibrium.

quantitative-PCR of miRNAs

We examined whether miRNAs hosting the variants associated with PD are expressed in the human brain. The brain samples were obtained from the Netherlands Brain Bank (Amsterdam, the Netherlands). All samples were free of neurological disease. For isolation of total RNA, five cryopreserved sections of 40 µm were homogenized in 250 µl Trizol reagent (Invitrogen, Carlsbad, CA, USA). Total RNA was isolated from 6 brain samples (3 white matter and 3 gray matter). The concentration and purity of RNA samples were determined with a NanoDrop ND-1000 spectrophotometer (NanoDrop, Wilmington, DE). The expression levels of miRNAs were determined with TaqMan MicroRNA Assays according to manufacturer's protocols (Applied Biosystems, Foster City, CA, USA). RNU6B was used as an endogenous control. All experiments were performed in triplicate.

Association of miRNA target genes with PD

For miRNAs that were associated with PD, we further examined which of the target genes may mediate the effect of miRNAs on PD. To do this, we extracted all predicted target genes of the identified miRNAs using two online databases, TargetScan v7.0 (http://www.targetscan.org) [27] and miRDB (http://mirdb.org/miRDB) [28]. The miRNA target genes that are listed in both databases were used for our analyses. Next, we used the PD-GWAS data in a candidate gene approach to identify those target genes that are likely to be involved in PD. We retrieved the summary statistics for the association of all genetic variants in the target genes with PD. The significance threshold for this analysis was set using the Bonferroni correction based on the number of studied variants. Additionally, we performed Ingenuity Pathway Analysis (IPA) to explore the pathways in which target genes of the identified miRNAs may play a role (http://www.ingenuity.com/products/ipa/). A list of all putative target genes of each miRNA were uploaded and a core IPA analysis were performed using the default settings. We mapped the miRNA target genes to biological functions or canonical pathways to determine whether they are enriched in neurological networks. The *p*-values are calculated using the right-tailed Fisher Exact Test and a p-value less than 0.05 indicates a statistically significant, non-random association.

Luciferase reporter assay

The pre-miRNAs were amplified and cloned in pMSCV-BC as previously described [26]. The 3'UTR sequence of the candidate target gene (either wild-type or mutated), containing the putative binding site of the miRNA, were cloned downstream of the Luciferase gene in the pGL3 vectors as previously described [26]. The primers are shown in **Table S1** and **Table S2**. The inserts of all constructs were validated by Sanger sequencing. COS cells were plated into 96-well plates and co-transfected with pMSCV-miRNA, pGL3 containing the different 3'UTRs and a plasmid expressing the Renilla transfection control. Luciferase activity was determined with the Dual-Glo Luciferase Assay System according to manufacturer's protocol (Promega, Madison, WI). Renilla luciferase activity was used for normalization. All experiments were performed five times.

Expression quantitative trait loci (eQTLs)

We scanned cis-eQTL data to examine the correlation between miRNA-binding site variants and the related transcript expression levels. We used two online web browsers: Genenetwork (http://genenetwork.nl/bloodeqtlbrowser/) and GTEx V4 (http://www.broadinstitute.org/gtex/). The GTEx platform provides information on eQTL in different tissues including brain. In addition, we used the eQTL data in whole blood from the Gene network because of a very large sample size (n= 5,311). The designs of these studies have been described in detail elsewhere [29, 30].

Expression of the identified target genes and related miRNAs in the brain

To search for expression of the identified target genes and their regulatory miRNAs in relevant tissues, we employed several web tools. The Illumina's Human Body Map 2.0 data (http://www.ensembl.info/blog/2011/05/24/human-bodymap-2-0-data-from-illumina/) were used to examine the expression of the genes hosting miRNA binding site-variants across different tissues. This database provides RNASeq data of 16 human tissue types, including brain. To scan the expression of related miRNAs in the brain, we used the Human MiRNA Expression Database (HMED) [31], mimiRNA [32] and PhenomiR [33] databases. We further searched the literature via PubMed using the search terms of miRNA ID and expression for those not implicated in the listed databases.

Analyzing functional characteristics of the identified miRNA binding site-variants

We evaluated LD blocks of the identified miRNA binding site variants to examine whether there are other SNPs in high LD in the related loci that may drive the observed GWAS associations. To this end, the list of PD-associated variants in miRNA binding sites were submitted to the SNAP web tool (http://www.broadinstitute.org/mpg/snap/id) using R^2 threshold > 0.8, limit distance 500 kb, and population panel CEU to retrieve their proxy SNPs in the 1000 Genomes project. We then utilized the HaploReg web tool v3 (http://www.broadinstitute.org/mammals/haploreg/haploreg_v3.php) to predict the effect of SNPs on protein structure, gene regulation, and splicing. Other information, including miRNA sequences, miRNA host genes and miRNA conservation in different species was obtained from TargetScan v7.0 and miRBase (release 20) [34] databases.

RESULTS

Two miRNA variants were associated with PD

We studied the association of 243 SNPs (with MAF > 0.01) located in 214 miRNAs with PD. Of these, rs11651671:A>G (Chr17:42494785) in miR-548at-5p (p-value= 1.06×10^{-6}, OR= 1.09) and rs897984:T>C (Chr16:30875322) in miR-4519 (p-value= 1.34×10^{-5}, OR= 0.93) were significantly associated with PD. **Table S3** shows miRNA variants that are associated with PD with a p-value < 0.05. The forest association plots showing meta-analysis of the association of the two identified miRNA variants with PD are shown in **Figure S1**.

The effect of variants on miR-4519 and miR-548at-5p structure and expression

We generated the hairpin structures of miR-4519 and miR-548at-5p containing the wild type and the mutant allele using the Vienna RNAfold algorithm [25]. We noted a 6.1 kcal/mol difference in the minimum free energy of the predicted thermodynamic ensemble

of the mutant versus the wild type miR-548at-5p structure (**Figure S2**), which may affect the processing of pre-miRNA. The predicted change in minimum free energy of the thermodynamic ensemble of the hairpin structure of miR-4519 containing the mutant versus the wild type was 0.3 kcal/mol (**Figure S3**). We then examined the expression levels of these miRNAs from two instances: the wild type pre-miRNA and the mutant pre-miRNA. We cloned the pre-miRNAs sequences (containing either wild-type or mutant allele) behind the GFP in the expression plasmid. Transient transfection of the miRNAs in HEK293 cells showed significantly reduced levels of the mature miRNAs from the mutant constructs relative to GFP compared with the wild-type constructs, where the rs897984 mutant allele reduced the expression level of miR-4519 by 90% (*p*-value= 0.003) and the rs11651671 mutant allele decreased the expression level of miR-548at-5p by 30% (*p*-value= 0.049) (**Figure 2**).

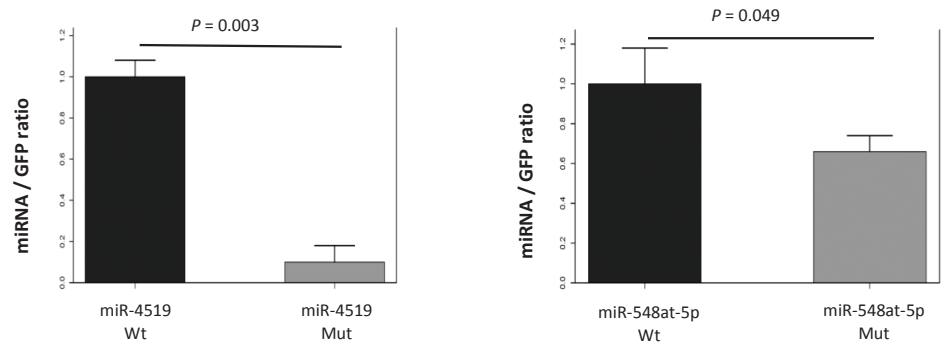


Figure 2. This figure shows the impact of rs897984 in miR-4519 and rs11651671 in miR-548at-5p on the expression levels of the mature miRNAs. As shown, the levels of miR-4519 and miR-548at-5p are significantly reduced from the mutant constructs (minor allele) relative to GFP compared to the wild-type constructs.

Association of miR-4519 and miR-548at-5p target genes with PD

MiRNAs act through regulation of their target gene expression. We further assessed whether 342 putative target genes of miR-4519 and 676 putative target genes of miR-548at-5p are implicated in neurological pathways using IPA. This analysis indicated that several target genes of miR-4519 are directly or indirectly linked with Nervous System Development and Function Networks, **Table S4**. We then examined the association of genetic variants in all putative target genes of miR-4519 and miR-548at-5p with PD using GWAS data. We studied 76,457 SNPs in 342 target genes of miR-4519, with the significant threshold of 6.5×10^{-7}, and found four target genes, *NSF, TMEM163, CCNT2* and *SH3GL2,* to be significantly associated with PD (**Table 1)**. For miR-548at-5p, we assessed 153,018 SNPs in 676 target genes, with the significance threshold of 3.3×10^{-7}, and identi-

fied *GCH1, MMRN1, CCNT2* and *DCUN1D1* to be associated with PD (**Table 1**). The Human Body Map data showed that the highlighted target genes of miR-4519 and miR-548at-5p are expressed in the brain (**Table S5**). Next, we asked whether the identified miRNAs are expressed in the brain. Our data showed that miR-4519 is expressed at detectable levels (average Ct-value: 32.5) in both white and gray matter of the human brain (**Table S6**). Subsequently, we examined whether miR-4519 control the expression level of its top identified target gene, *NSF, in vitro*. We generated expression vector containing the pre-miR-4519 sequence and co-transfected the construct with Luciferase reporters containing the wild-type and mutant 3' UTR of *NSF*. We found that expression of miR-4519 significantly decreases the Luciferase activity of wild type *NSF* reporter, compared to the mutated *NSF* reporter, *p*-value= 2.0×10^{-4} (**Figure 3**). These data indicate that *NSF* is a direct target of miR-4519.

Table 1. Association of two variants in miR-4519 and miR-548at-5p with Parkinson disease

miRNA ID	SNP ID (A1/A2)	Sample size in GWAS	OR GWAS	*p*-value GWAS	PD-associated target genes (*p*-value)
miR-4519	rs897984 (G/A)	108,990	0.93	1.3×10^{-5}	*NSF* (3.9×10^{-29}) *TMEM163* (2.7×10^{-13}) *CCNT2* (8.6×10^{-9}) *SH3GL2* (1.3×10^{-7})
miR-548at-5p	rs11651671 (A/G)	108,990	1.09	1.1×10^{-6}	*MMRN1* (4.4×10^{-15}) *GCH1* (2.0×10^{-10}) *CCNT2* (8.6×10^{-9}) *DCUN1D1* (8.7×10^{-8})

A1, Allele 1; A2, Allele 2; OR, Odd ratio; GWAS, Genome-wide association study of PD (Nalls et al., 2014). The *p*-value for the highlighted target genes is *p*-value of the SNP in the gene with the most significant association with PD.

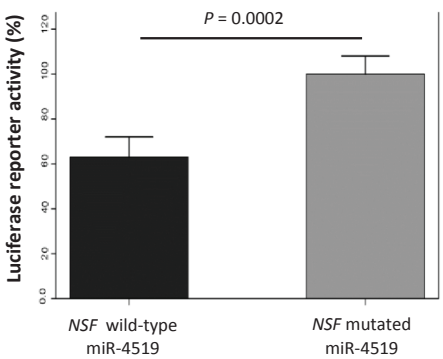


Figure 3. miR-4519-mediated regulation of *NSF*. This figure illustrates a significant difference between the mean relative luciferase activity of the wild-type *NSF* 3'UTR reporter compared and the mutated *NSF* 3'UTR reporter in the presence of miR-4519.

Multiple miRNA binding site-variants were associated with PD

We examined the associations of 48,844 miRNA binding site-variants with PD, (**Figure 4**). Of these, 32 SNPs located in the 3'UTR of 13 genes were significantly associated with PD (p-value < 1.02×10^{-6}) (**Table 2**). These SNPs are predicted to affect miRNA-mediated regulation of their host genes by disrupting, creating or modifying a number of miRNA binding sites that are depicted in **Table S7** and **Table S8**. Out of 13 genes hosting the 32 SNPs, the association of 9 genes with PD had already been reported by GWAS [8]. Four others have not previously reported for PD include *HSD3B7* (p-value= 5.2×10^{-8}), *IGSF9B* (p-value= 2.6×10^{-7}), *CSTB* (p-value= 4.8×10^{-7}) and *STX1B* (p-value= 1.2×10^{-7}).

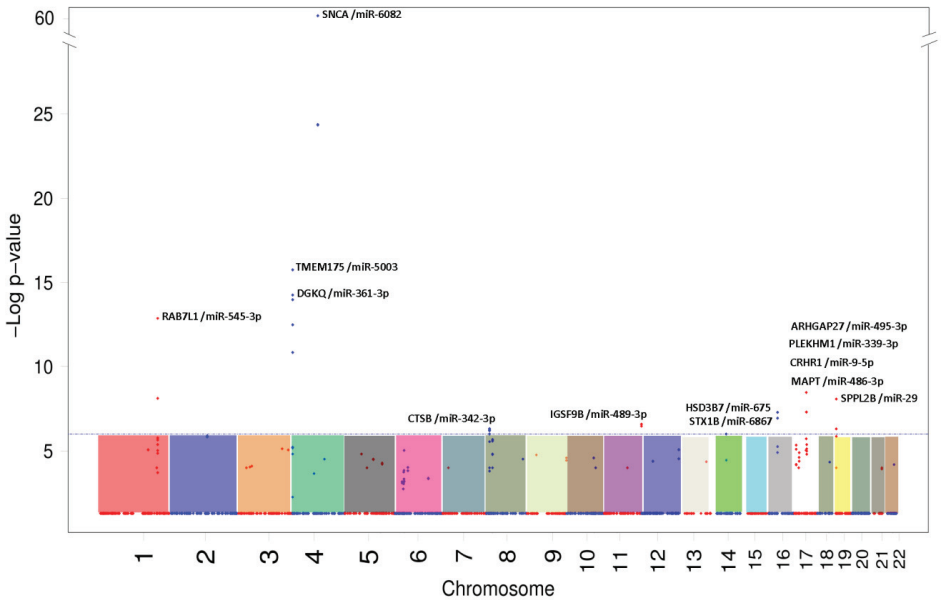


Figure 4. The Manhattan plot shows the association of miRNA-binding site variants (with a p-value < 0.05) with Parkinson's disease. Dashed line indicates the significant study threshold (1.02×10^{-6}).

Supporting evidence for potential functionality of the identified miRNA-binding site variants

We searched for cis-eQTL of 13 genes hosting the 32 identified miRNA-binding site variants and found 6 SNPs that are correlated with expression levels of their host genes. This includes rs10420958, which was associated with *SPPL2B* expression in the brain, and five SNPs (including rs708723 and rs2645425) that were associated with expression levels of *CTSB* and *RAB7L1* in blood. Using the HaploReg web tool v3, we searched for potential functionality of the 32 SNPs and their proxies in high LD on gene regulation. This analysis showed that some of the miRNA binding site SNPs have no proxy or only week proxies in the related loci, such as rs8327, rs356165, rs708723 and rs1128402 (**Table 3**).

Table 2. miRNA binding site variants associated with Parkinson disease

SNP ID	Chr: position	A1/A2	MAF	Gene	P-value GWAS	OR GWAS
rs4763	chr17:43471489	A/G	0.14	ARHGAP27	5.0×10^{-8}	<1
rs8327	chr17:43472507	G/A	0.24	ARHGAP27	3.5×10^{-9}	1.21
rs16940681	chr17:43912159	C/G	0.17	CRHR1	5.0×10^{-8}	<1
rs4482334	chr17:43912830	T/C	0.22	CRHR1	5.0×10^{-8}	>1
rs4640231	chr17:43912786	C/G	0.22	CRHR1	5.0×10^{-8}	<1
rs878886	chr17:43912490	C/G	0.22	CRHR1	5.0×10^{-8}	>1
rs878888	chr17:43912635	A/G	0.22	CRHR1	5.0×10^{-8}	>1
rs2740592	chr8:11701253	A/G	0.21	CTSB	4.8×10^{-7}	0.91
rs2645425	chr8:11701278	A/G	0.21	CTSB	5.2×10^{-7}	0.91
rs4839	chr8:11701933	C/T	0.21	CTSB	6.0×10^{-7}	0.91
rs3947	chr8:11702375	A/G	0.21	CTSB	6.4×10^{-7}	0.91
rs4583705	chr4:973036	T/C	0.14	DGKQ	1.5×10^{-11}	1.24
rs3733349	chr4:954311	T/C	0.43	DGKQ	$3.4x^{10-13}$	0.89
rs3733345	chr4:954247	T/G	0.45	DGKQ	1.0×10^{-14}	0.88
rs4690326	chr4:953698	C/A	0.45	DGKQ	5.4×10^{-15}	0.88
rs2305880	chr16:30999462	T/C	0.25	HSD3B7	5.2×10^{-8}	0.91
rs3802922	chr11:133786945	C/A	0.19	IGSF9B	2.6×10^{-7}	1.11
rs3802921	chr11:133786993	C/T	0.20	IGSF9B	2.7×10^{-7}	1.11
rs3802920	chr11:133787001	T/G	0.19	IGSF9B	3.4×10^{-7}	1.11
rs1052594	chr17:44102689	C/G	0.20	MAPT	5.0×10^{-8}	<1
rs11012	chr17:43513441	T/C	0.15	PLEKHM1	5.0×10^{-8}	<1
rs62064654	chr17:43513896	T/C	0.15	PLEKHM1	5.0×10^{-8}	<1
rs62064655	chr17:43514954	A/G	0.42	PLEKHM1	5.0×10^{-8}	<1
rs11557080	chr1:205737739	G/A	0.23	RAB7L1	7.6×10^{-9}	0.87
rs708723	chr1:205739266	C/T	0.44	RAB7L1	1.4×10^{-13}	0.89
rs356165	chr4:90646886	A/G	0.48	SNCA	2.9×10^{-62}	0.76
rs3857053	chr4:90645674	T/C	0.11	SNCA	4.3×10^{-25}	1.34
rs1045722	chr4:90645671	A/T	0.11	SNCA	4.6×10^{-25}	1.34
rs10420958	chr19:2353368	A/G	0.48	SPPL2B	4.9×10^{-7}	0.92
rs1128402	chr19:2353150	A/C	0.28	SPPL2B	8.4×10^{-9}	1.12
rs8060857	chr16:31002720	G/A	0.24	STX1B	1.2×10^{-7}	0.91
rs748483	chr4:952409	G/A	0.14	TMEM175	1.8×10^{-16}	1.22

Chr, chromosome; A1, ancestor allele; A2, alternative allele; MAF, Minor Allele Frequency in 1000G CEU; OR, Odd ratio; GWAS, Genome-wide association study on PD.

Table 3. Functional characteristics of the identified miRNA binding site-variants in the 3'UTR of 13 genes associated with Parkinson disease

		rs8327	rs708723	rs356165	rs16940681	rs2645425	rs4690326	rs2305880	rs3802921	rs1052594	rs62064654	rs1128402	rs8060857	rs748483
GWAS	*p*-value	3.5×10^{-9}	1.4×10^{-13}	2.9×10^{-62}	5.0×10^{-8}	5.2×10^{-7}	5.4×10^{-15}	5.2×10^{-8}	2.7×10^{-7}	5.0×10^{-8}	5.0×10^{-8}	4.9×10^{-7}	1.2×10^{-7}	1.8×10^{-16}
	OR	1.21	0.89	0.76	<1	0.91	0.88	0.91	1.1	>1	<1	0.92	0.91	1.22
Gene	Host gene	*ARHGAP27*	*RAB7L1*	*SNCA*	*CRHR1*	*CTSB*	*DGKQ*	*HSD3B7*	*IGSF9B*	*MAPT*	*PLEKHM1*	*SPPL2B*	*STX1B*	*TMEM175*
	Brain Exp	1.4	7.5	187	2.0	135	1.4	1.1	0.3	54	6.4	4.7	44	5.0
Proxy	All SNPs	0	3	2	1150	35	4	56	5	1067	133	4	55	13
	Nonsynon	0	0	0	13	0	0	1	0	12	1	0	1	0
eQTL	Blood	-	2.3×10^{-42}	-	-	7.3×10^{-172}	-	-	-	-	-	-	-	-
	Brain	-	-	-	-	-	-	-	-	-	-	1.1×10^{-9}	-	-
miR-binding site	Dis (A1)	miR-495-3p	NC	miR-6082	miR-9-5p	None	NC	NC	miR-489-3p	NC	miR-339-3p	None	miR-6867	None
	ΔScore	-0.02	-	-0.12	NA	-	-	-	0.02	-	-0.19	-	0.14	-
	Cre (A2)	None	miR-545-3p	None	NC	miR-342-3p	miR-361-3p	miR-675-5p	miR-514-3p	miR-486-3p	miR-657	miR-29a-5p	None	miR-5003
	ΔScore	-	-0.08	-	-	NA	-0.23	-0.48	0.1	-0.03	-0.09	0.03	-	-0.15

Shown are supporting evidences for 13 miRNA binding site-variants found to be associated with PD. For genes hosting more than one variant in miRNA binding sites, variant with the higher probability to be being functional has reported in the table. In addition, for variants that are predicted to affect more than one miRNA, the highly conserved miRNA in each situation has mentioned (full information are provided in the supplementary data). miR, miRNA; OR, Odd ratio; Exp, Expression; Nonsynon, Non-Synonymous SNPs (using the HaploReg web tool v3); ΔScore, Context score change (using the PolymiRTS database v3); Dis, Disruption binding site; Cre, Creation binding site; A1, Allele 1; A2, Allele 2; None, No binding site for miRNAs; NC, Binding site for a non-conserved miRNA; NA, Not available. The value for eQTL is *p*-value; the value for brain expression is fragments per kb of exon per million reads (FPKMs) (Human Body Map 2.0 data).

Furthermore, through scanning the expression of 13 host genes using the Human Body Map 2.0 data, we found that *SNCA, CTSB, MAPT* and *STX1B* are abundantly expressed in the brain (**Table S9**). We found evidence for expression of several of the regulatory miRNAs in the brain, in particular miR-342-3p, miR-29a-5p, miR-9-5p, using the miRNA expression databases (**Table S7**). A summary of the evidence that we found suggesting the potential functionality of the identified miRNA binding site-SNPs in their related loci are depicted in **Table 3**.

DISCUSSION

In this study, we investigated the association of variants located in miRNAs and miRNA-binding sites with PD using genetic data. We found two common miRNA variants, rs897984 in miR-4519 and rs11651671 in miR-548at-5p, that are associated with PD. We showed that the variant's mutant alleles have the potential to affect the hairpin secondary structures of pre-miRNAs and decrease the expression levels of mature miR-4519 and miR-548at-5p. We highlighted target genes that might mediate the effects of miR-4519 and miR-548at-5p in relation to PD. We experimentally showed that *NSF* is a direct target of miR-4519. Furthermore, we identified 32 miRNA binding site-variants (within 13 genes) that are associated with PD. Four of the host genes, *CTSB, STX1B, IGSF9B* and *HSD3B7*, had not previously been reported for PD. Finally, we provide evidence supporting some of the identified miRNA binding site-variants to affect miRNA-mediated gene regulation in PD.

An increasing number of studies have shown the critical role of miRNAs in neurodegenerative disorders including PD [35-37]. Most of these studies are mainly focused on differentially expressed miRNAs and genes detected by expression arrays in a small sample size. In addition, some studies have linked a number of miRNAs with neurodegenerative disorders using the candidate gene approach. For example, Saba et al. have recently reported a catalog of SNPs overlapping miRNA-binding sites in a subset of genes that are implicated in neurological diseases [38]. Here, we systematically investigated the association of all miRNA-related genetic variants with PD using population level data in a large scale.

The GWAS data shows that carriers of the rs897984 mutant allele in miR-4519 have a decreased risk of PD. Our functional experiments showed that the mutant allele dramatically decreases the expression level of mature miR-4519. SNP rs897984 occurs 1 nt after the 5′ end of pre-miR-4519 and overlaps the site of Drosha cleavage. The maturation of miRNAs is a complex and highly regulated process, which is characterized by two-step sequential processing by RNase III enzymes, Drosha and Dicer [39, 40]. It has been shown that polymorphisms residing within (±)1 nt of the Drosha or Dicer cleavage sites

may affect miRNA biogenesis [41, 42]. Therefore, the observed effect of rs897984 on the miR-4519 level can be explained with the SNP impact on the processing of miR-4519 by Drosha enzyme. We showed that miR-4519 is expressed in both gray and white matter of the human brain. Interestingly, data from the GTEx database showed that miR-4519 is expressed in higher levels in substantia nigra, the primary area of the brain that is affected by PD [43]. In agreement with our findings, two recent studies have demonstrated this miRNA to be expressed in the human brain tissues [44, 45].

We suggested that the effect of miR-4519 on PD might be mediated by four target genes (*NSF*, *TMEM163, CCNT2* and *SH3GL2*) that were found to be potentially associated with PD. Among them, *NSF* showed the most significant association with PD (in the GWAS data) and is the most abundantly expressed gene in the brain. We further experimentally showed that *NSF* is a direct target of miR-4519. *NSF* is known as a crucial factor in intracellular membrane-fusion events, such as the fusion of synaptic vesicles with the presynaptic membrane during neurotransmission [46]. This gene has also been shown to be associated with different neurodegenerative disorders including PD [47, 48]. Furthermore, previous studies have shown that expression of *NSF* is decreased in PD substantia nigra [49-51]. These data may indicate that the rs897984 mutant allele influences the risk of PD through reducing the level of miR-4519 and increasing the expression of *NSF* gene.

We found that the rs11651671 mutant allele in miR-548at-5p is associated with an increased risk of PD. *In silico* analysis indicated that the SNP rs11651671 may affect the hairpin structure of precursor miR-548at-5p. In agreement with this conjecture, our *in vitro* experiments demonstrated that the mutant allele significantly reduces the level of miR-548at-5p. In our expression analysis, miR-548at-5p was no detectable in the gray and white matter of the human brain, however, we were not able to examine its expression level in more relevant tissue such as substantia nigra.

Interestingly, *ATP6V0A1*, which hosts miR-548at-5p in a sense direction, is abundantly expressed in the brain tissues (using the Human Body Map data). Previous studies have suggested that the expression of an miRNA host gene can be used as a proxy for miRNA expression when both the host gene and the miRNA are in the same direction [52]. Future studies are thereby needed to examine the miR-548at-5p expression in the more relevant tissues. Moreover, our results showed four target genes of miR-548at-5p, including *GCH1, MMRN1, CCNT2* and *DCUN1D1*, that are associated with PD and are likely to be potential mediators of the miRNA effects on PD. Among them, *GCH1*, also called Dopa-Responsive Dystonia1, has been shown to be associated with movement disorders including early-onset PD [53-55]. These results suggest that decreased levels of miR-548at-5p in carriers of the rs11651671 minor allele may influence PD risk through altering the expression of these target genes.

We identified 32 miRNA binding site-SNPs that are associated with PD. Among 13 genes hosting the variants, *SNCA* (alpha-synuclein) plays an important role for normal brain function and is a major risk factor for PD [56]. Genetic variants in *SNCA* have been shown to be associated with the common sporadic form of PD [57-59]. However, some of the associated variants are mapped downstream of the *SNCA* gene and a direct functional effect is thus unlikely. SNP rs356165 is located in the 3'UTR of *SNCA* and is one of the top associated SNPs with PD in the related locus. We found that the rs356165 mutant allele is predicted to disrupt a binding site of miR-6508, presumably resulting in an elevated level of *SNCA* expression. An allele-specific regulation of *SNCA* transcript by miR-6508 might serve as a functional explanation behind the association of rs356165 with PD. Future studies are needed to determine the effect of rs356165 on *SNCA* expression and the function of miR-6508 in PD patients. On the other hand, our results showed that some of the identified miRNA-binding site SNPs improve the original recognition sites or create new binding sites for miRNAs. For example, rs1128402 is predicted to create a binding site for miR-29a-5p in the 3'UTR of *SPPL2B*. This variant is one of the top SNPs associated with PD in the related locus and has no non-synonymous proxy. In addition, several independent studies have shown miR-29a-5p to be involved in PD and other neurodegenerative diseases [60-63]. Down-regulation of *SPPL2B* by miR-29a-5p may be a functional reason for the association of rs1128402 with PD.

We suggested four new genes, *CTSB, STX1B, IGSF9B* and *HSD3B7*, that are potentially associated with PD. Of these, *CTSB* (Cathepsin B1) encodes a protein that is known as amyloid precursor protein secretase and is involved in the proteolytic processing of amyloid precursor protein (APP). Incomplete proteolytic processing of APP has been suggested to be a causative factor in Alzheimer disease [64, 65]. We found three miRNA-binding site SNPs in the 3'UTR of *CTSB* that are associated with PD. Of these, rs2645425 is predicted to create a potential binding site for miR-342-3p. Using blood eQTL analysis, we showed that rs2645425 mutant allele carriers have lower expression levels of *CTSB*. The expression data further demonstrated that the *CTSB* gene is abundantly expressed in the brain. Notably, miR-342-3p has been shown to be upregulated in the brain of patients with prion disease [62, 66] and has been suggested as a biomarker for Alzheimer's disease [67]. The observed association of rs2645425 with PD be through increasing the miR-342-3p-dependent regulation of *CTSB*. Among the other three genes, *STX1B* (syntaxin-1B) has been found to be directly implicated in the process of calcium-dependent synaptic transmission in the rat brain and has been suggested to play an important role in the excitatory pathway of synaptic transmission [68]. *IGSF9B* is known as Immunoglobulin Superfamily Member 9B and has reported as a novel, brain-specific, homophilic adhesion molecule that is strongly expressed in GABAergic interneurons [69]. Finally, *HSD3B7* has been shown to be associated with congenital bile acid synthesis defect and

liver diseases [70]. This gene has also been linked with the etiology of Alzheimer disease through deactivation pathway of LXR ligands in the brain [71].

Collectively, we identified two miRNA variants and multiple miRNA binding site variants that are associated with PD and could potentially affect miRNA-mediated regulation of a number of genes involved in PD. These results may contribute to increase our understanding of the role of miRNAs in the pathogenesis of PD. Our findings may also be of clinical importance as they suggest a number of miRNAs that modify gene expression profiles and affect PD risk. Experimental assays of the identified variants and miRNA profiling in PD patients will be the next step towards determining the functionality of the variants, related target genes, and miRNAs in PD.

Supplement available online at:
http://onlinelibrary.wiley.com/doi/10.1002/humu.22943/suppinfo

REFERENCES

1. Cowan WM, Kandel ER: Prospects for neurology and psychiatry. JAMA 2001, 285:594-600.
2. de Lau LM, Breteler MM: Epidemiology of Parkinson's disease. Lancet Neurol 2006, 5:525-535.
3. Dorsey ER, Constantinescu R, Thompson JP, Biglan KM, Holloway RG, Kieburtz K, Marshall FJ, Ravina BM, Schifitto G, Siderowf A, Tanner CM: Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030. Neurology 2007, 68:384-386.
4. Lohle M, Storch A, Reichmann H: Beyond tremor and rigidity: non-motor features of Parkinson's disease. J Neural Transm 2009, 116:1483-1492.
5. Lansbury PT, Lashuel HA: A century-old debate on protein aggregation and neurodegeneration enters the clinic. Nature 2006, 443:774-779.
6. Shulman JM, De Jager PL, Feany MB: Parkinson's disease: genetics and pathogenesis. Annu Rev Pathol 2011, 6:193-222.
7. Lill CM, Roehr JT, McQueen MB, Kavvoura FK, Bagade S, Schjeide BM, Schjeide LM, Meissner E, Zauft U, Allen NC, et al: Comprehensive research synopsis and systematic meta-analyses in Parkinson's disease genetics: The PDGene database. PLoS Genet 2012, 8:e1002548.
8. Nalls MA, Pankratz N, Lill CM, Do CB, Hernandez DG, Saad M, DeStefano AL, Kara E, Bras J, Sharma M, et al: Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. Nat Genet 2014, 46:989-993.
9. Bartel DP: MicroRNAs: target recognition and regulatory functions. Cell 2009, 136:215-233.
10. Ambros V: The functions of animal microRNAs. Nature 2004, 431:350-355.
11. Tan L, Yu JT, Tan L: Causes and Consequences of MicroRNA Dysregulation in Neurodegenerative Diseases. Mol Neurobiol 2015, 51:1249-1262.
12. Muller M, Jakel L, Bruinsma IB, Claassen JA, Kuiperij HB, Verbeek MM: MicroRNA-29a Is a Candidate Biomarker for Alzheimer's Disease in Cell-Free Cerebrospinal Fluid. Mol Neurobiol 2015.
13. Issler O, Chen A: Determining the role of microRNAs in psychiatric disorders. Nat Rev Neurosci 2015, 16:201-212.
14. Harraz MM, Dawson TM, Dawson VL: MicroRNAs in Parkinson's disease. J Chem Neuroanat 2011, 42:127-130.
15. Khoo SK, Petillo D, Kang UJ, Resau JH, Berryhill B, Linder J, Forsgren L, Neuman LA, Tan AC: Plasma-based circulating MicroRNA biomarkers for Parkinson's disease. J Parkinsons Dis 2012, 2:321-331.
16. Kim J, Inoue K, Ishii J, Vanti WB, Voronov SV, Murchison E, Hannon G, Abeliovich A: A MicroRNA feedback circuit in midbrain dopamine neurons. Science 2007, 317:1220-1224.
17. Ryan BM, Robles AI, Harris CC: Genetic variation in microRNA networks: the implications for cancer research. Nat Rev Cancer 2010, 10:389-402.
18. Richardson K, Nettleton JA, Rotllan N, Tanaka T, Smith CE, Lai CQ, Parnell LD, Lee YC, Lahti J, Lemaitre RN, et al: Gain-of-function lipoprotein lipase variant rs13702 modulates lipid traits through disruption of a microRNA-410 seed site. Am J Hum Genet 2013, 92:5-14.
19. Ghanbari M, de Vries PS, de Looper H, Peters MJ, Schurmann C, Yaghootkar H, Dorr M, Frayling TM, Uitterlinden AG, Hofman A, et al: A genetic variant in the seed region of miR-4513 shows pleiotropic effects on lipid and glucose homeostasis, blood pressure, and coronary artery disease. Hum Mutat 2014, 35:1524-1531.
20. Ghanbari M, Franco OH, de Looper H, Hofman A, Erkeland S, Dehghan A: Genetic Variations in miRNA Binding Sites Affect miRNA-Mediated Regulation of Several Genes Associated with Cardiometabolic Phenotypes. Circ Cardiovasc Genet 2015.

21. Gong J, Tong Y, Zhang HM, Wang K, Hu T, Shan G, Sun J, Guo AY: Genome-wide identification of SNPs in microRNA genes and the SNP effects on microRNA target binding and biogenesis. Hum Mutat 2012, 33:254-263.
22. Hughes AE, Bradley DT, Campbell M, Lechner J, Dash DP, Simpson DA, Willoughby CE: Mutation altering the miR-184 seed region causes familial keratoconus with cataract. Am J Hum Genet 2011, 89:628-633.
23. Gong J, Liu C, Liu W, Wu Y, Ma Z, Chen H, Guo AY: An update of miRNASNP database for better SNP selection by GWAS data, miRNA expression and online tools. Database (Oxford) 2015, 2015:bav029.
24. Bao L, Zhou M, Wu L, Lu L, Goldowitz D, Williams RW, Cui Y: PolymiRTS Database: linking polymorphisms in microRNA target sites with complex traits. Nucleic Acids Res 2007, 35:D51-54.
25. Lorenz R, Bernhart SH, Honer Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL: ViennaRNA Package 2.0. Algorithms Mol Biol 2011, 6:26.
26. Meenhuis A, van Veelen PA, de Looper H, van Boxtel N, van den Berge IJ, Sun SM, Taskesen E, Stern P, de Ru AH, van Adrichem AJ, et al: MiR-17/20/93/106 promote hematopoietic cell expansion by targeting sequestosome 1-regulated pathways in mice. Blood 2011, 118:916-925.
27. Friedman RC, Farh KK, Burge CB, Bartel DP: Most mammalian mRNAs are conserved targets of microRNAs. Genome Res 2009, 19:92-105.
28. Wong N, Wang X: miRDB: an online resource for microRNA target prediction and functional annotations. Nucleic Acids Res 2015, 43:D146-152.
29. Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, Christiansen MW, Fairfax BP, Schramm K, Powell JE, et al: Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet 2013.
30. Grundberg E, Small KS, Hedman AK, Nica AC, Buil A, Keildson S, Bell JT, Yang TP, Meduri E, Barrett A, et al: Mapping cis- and trans-regulatory effects across multiple tissues in twins. Nat Genet 2012, 44:1084-1089.
31. Gong J, Wu Y, Zhang X, Liao Y, Sibanda VL, Liu W, Guo AY: Comprehensive analysis of human small RNA sequencing data provides insights into expression profiles and miRNA editing. RNA Biol 2014, 11:1375-1385.
32. Ritchie W, Flamant S, Rasko JE: mimiRNA: a microRNA expression profiler and classification resource designed to identify functional correlations between microRNAs and their targets. Bioinformatics 2010, 26:223-227.
33. Ruepp A, Kowarsch A, Schmidl D, Buggenthin F, Brauner B, Dunger I, Fobo G, Frishman G, Montrone C, Theis FJ: PhenomiR: a knowledgebase for microRNA expression in diseases and biological processes. Genome Biol 2010, 11:R6.
34. Kozomara A, Griffiths-Jones S: miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res 2014, 42:D68-73.
35. Hebert SS, De Strooper B: Molecular biology. miRNAs in neurodegeneration. Science 2007, 317:1179-1180.
36. Martins M, Rosa A, Guedes LC, Fonseca BV, Gotovac K, Violante S, Mestre T, Coelho M, Rosa MM, Martin ER, et al: Convergence of miRNA expression profiling, alpha-synuclein interacton and GWAS in Parkinson's disease. PLoS One 2011, 6:e25443.
37. Serafin A, Foco L, Zanigni S, Blankenburg H, Picard A, Zanon A, Giannini G, Pichler I, Facheris MF, Cortelli P, et al: Overexpression of blood microRNAs 103a, 30b, and 29a in L-dopa-treated patients with PD. Neurology 2015, 84:645-653.

38. Saba R, Medina SJ, Booth SA: A functional SNP catalog of overlapping miRNA-binding sites in genes implicated in prion disease and other neurodegenerative disorders. Hum Mutat 2014, 35:1233-1248.
39. Winter J, Jung S, Keller S, Gregory RI, Diederichs S: Many roads to maturity: microRNA biogenesis pathways and their regulation. Nat Cell Biol 2009, 11:228-234.
40. Denli AM, Tops BB, Plasterk RH, Ketting RF, Hannon GJ: Processing of primary microRNAs by the Microprocessor complex. Nature 2004, 432:231-235.
41. Sun G, Yan J, Noltner 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.
42. Obsteter J, Dovc P, Kunej T: Genetic variability of microRNA regulome in human. Mol Genet Genomic Med 2015, 3:30-39.
43. Croisier E, Moran LB, Dexter DT, Pearce RK, Graeber MB: Microglial inflammation in the parkinsonian substantia nigra: relationship to alpha-synuclein deposition. J Neuroinflammation 2005, 2:14.
44. Hong X, Sin WC, Harris AL, Naus CC: Gap junctions modulate glioma invasion by direct transfer of microRNA. Oncotarget 2015, 6:15566-15577.
45. Stamova B, Ander BP, Barger N, Sharp FR, Schumann CM: Specific Regional and Age-Related SmallNoncoding RNA Expression Patterns Within Superior Temporal Gyrus of Typical Human Brains Are Less Distinct in Autism Brains. J Child Neurol 2015.
46. Haas A: NSF--fusion and beyond. Trends Cell Biol 1998, 8:471-473.
47. Liu X, Cheng R, Verbitsky M, Kisselev S, Browne A, Mejia-Sanatana H, Louis ED, Cote LJ, Andrews H, Waters C, et al: Genome-wide association study identifies candidate genes for Parkinson's disease in an Ashkenazi Jewish population. BMC Med Genet 2011, 12:104.
48. Lencz T, Lambert C, DeRosse P, Burdick KE, Morgan TV, Kane JM, Kucherlapati R, Malhotra AK: Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. Proc Natl Acad Sci U S A 2007, 104:19942-19947.
49. Capurro A, Bodea LG, Schaefer P, Luthi-Carter R, Perreau VM: Computational deconvolution of genome wide expression data from Parkinson's and Huntington's disease brain tissues using population-specific expression analysis. Front Neurosci 2014, 8:441.
50. Elstner M, Morris CM, Heim K, Bender A, Mehta D, Jaros E, Klopstock T, Meitinger T, Turnbull DM, Prokisch H: Expression analysis of dopaminergic neurons in Parkinson's disease and aging links transcriptional dysregulation of energy metabolism to cell death. Acta Neuropathol 2011, 122:75-86.
51. Simunovic F, Yi M, Wang Y, Stephens R, Sonntag KC: Evidence for gender-specific transcriptional profiles of nigral dopamine neurons in Parkinson disease. PLoS One 2010, 5:e8856.
52. Radfar MH, Wong W, Morris Q: Computational prediction of intronic microRNA targets using host gene expression reveals novel regulatory mechanisms. PLoS One 2011, 6:e19312.
53. Cobb SA, Wider C, Ross OA, Mata IF, Adler CH, Rajput A, Rajput AH, Wu RM, Hauser R, Josephs KA, et al: GCH1 in early-onset Parkinson's disease. Mov Disord 2009, 24:2070-2075.
54. Zirn B, Steinberger D, Troidl C, Brockmann K, von der Hagen M, Feiner C, Henke L, Muller U: Frequency of GCH1 deletions in Dopa-responsive dystonia. J Neurol Neurosurg Psychiatry 2008, 79:183-186.
55. Momma K, Funayama M, Li Y, Ichinose H, Motoyoshi K, Hattori N, Mizuno Y, Kamakura K: A new mutation in the GCH1 gene presents as early-onset Parkinsonism. Parkinsonism Relat Disord 2009, 15:160-161.

56. Chandra S, Gallardo G, Fernandez-Chacon R, Schluter OM, Sudhof TC: Alpha-synuclein cooperates with CSPalpha in preventing neurodegeneration. Cell 2005, 123:383-396.
57. Mizuta I, Satake W, Nakabayashi Y, Ito C, Suzuki S, Momose Y, Nagai Y, Oka A, Inoko H, Fukae J, et al: Multiple candidate gene analysis identifies alpha-synuclein as a susceptibility gene for sporadic Parkinson's disease. Hum Mol Genet 2006, 15:1151-1158.
58. Mueller JC, Fuchs J, Hofer A, Zimprich A, Lichtner P, Illig T, Berg D, Wullner U, Meitinger T, Gasser T: Multiple regions of alpha-synuclein are associated with Parkinson's disease. Ann Neurol 2005, 57:535-541.
59. Tan EK, Skipper LM: Pathogenic mutations in Parkinson disease. Hum Mutat 2007, 28:641-653.
60. Chandrasekaran S, Bonchev D: A network view on Parkinson's disease. Comput Struct Biotechnol J 2013, 7:e201304004.
61. Margis R, Margis R, Rieder CR: Identification of blood microRNAs associated to Parkinsonis disease. J Biotechnol 2011, 152:96-101.
62. Bellingham SA, Coleman BM, Hill AF: Small RNA deep sequencing reveals a distinct miRNA signature released in exosomes from prion-infected neuronal cells. Nucleic Acids Res 2012, 40:10937-10949.
63. Villa C, Ridolfi E, Fenoglio C, Ghezzi L, Vimercati R, Clerici F, Marcone A, Gallone S, Serpente M, Cantoni C, et al: Expression of the transcription factor Sp1 and its regulatory hsa-miR-29b in peripheral blood mononuclear cells from patients with Alzheimer's disease. J Alzheimers Dis 2013, 35:487-494.
64. Schechter I, Ziv E: Cathepsins S, B and L with aminopeptidases display beta-secretase activity associated with the pathogenesis of Alzheimer's disease. Biol Chem 2011, 392:555-569.
65. Sundelof J, Sundstrom J, Hansson O, Eriksdotter-Jonhagen M, Giedraitis V, Larsson A, Degerman-Gunnarsson M, Ingelsson M, Minthon L, Blennow K, et al: Higher cathepsin B levels in plasma in Alzheimer's disease compared to healthy controls. J Alzheimers Dis 2010, 22:1223-1230.
66. Montag J, Hitt R, Opitz L, Schulz-Schaeffer WJ, Hunsmann G, Motzkus D: Upregulation of miRNA hsa-miR-342-3p in experimental and idiopathic prion disease. Mol Neurodegener 2009, 4:36.
67. Tan L, Yu JT, Tan MS, Liu QY, Wang HF, Zhang W, Jiang T, Tan L: Genome-wide serum microRNA expression profiling identifies serum biomarkers for Alzheimer's disease. J Alzheimers Dis 2014, 40:1017-1027.
68. Smirnova T, Stinnakre J, Mallet J: Characterization of a presynaptic glutamate receptor. Science 1993, 262:430-433.
69. Woo J, Kwon SK, Nam J, Choi S, Takahashi H, Krueger D, Park J, Lee Y, Bae JY, Lee D, et al: The adhesion protein IgSF9b is coupled to neuroligin 2 via S-SCAM to promote inhibitory synapse development. J Cell Biol 2013, 201:929-944.
70. Fischler B, Bodin K, Stjernman H, Olin M, Hansson M, Sjovall J, Bjorkhem I: Cholestatic liver disease in adults may be due to an inherited defect in bile acid biosynthesis. J Intern Med 2007, 262:254-262.
71. Ogundare M, Theofilopoulos S, Lockhart A, Hall LJ, Arenas E, Sjovall J, Brenton AG, Wang Y, Griffiths WJ: Cerebrospinal fluid steroidomics: are bioactive bile acids present in brain? J Biol Chem 2010, 285:4666-4679.

CHAPTER 3.3

A genome-wide scan for long non-coding RNAs associated with Alzheimer's disease

Manuscript based on this chapter

A genome-wide analysis of long non-coding RNAs reveals chromosome17q22 hosting BZRAP1-AS1/miR-142 as a new susceptibility locus for late-onset Alzheimer's disease

2017, In prepration.

ABSTRACT

Background. The majority of genetic variants identified by genome-wide association studies (GWAS) for Alzheimer's disease (AD) map to non-coding regions of the genome. These include long non-coding RNAs (lncRNAs) that although do not encode proteins, they have shown as a potentially new and crucial layer of biological regulation. In this study, we aimed to identify lncRNAs associated with AD using GWAS data.

Methods and results. We leveraged data from the largest available GWAS on late-onset AD (in > 50,000 individuals) and examined the association of genetic variants in lncRNAs with AD. We identified significant association between five lncRNAs and AD (p-value < 9.02×10^{-7}). Among these, the association of 17q22 locus with AD has not been reported previously. The lead variant in this locus is rs2632516 and overlaps the promoter region of miR-142 host gene and the intronic region of *BZRAP1-AS1*. *In silico* analysis suggested that the variant is more likely to work through affecting the promoter activity of miR-142. Our expression analysis showed that both miR-142-3p and miR-142-5p are expressed in the brain. Further, miR-142 expression has been previously shown to be upregulated in the hippocampus of AD patients. We highlighted the target genes of miR-142-3p and miR-142-5p that are expressed in the brain and are associated with AD. Among them, we found evidence suggesting that *SIRT1*, a known AD gene with neuroprotective roles, as a potential target that may mediate the downstream effect of miR-142 in relation to AD.

Conclusion. Our results show 17q22 as a new susceptibility locus for AD and suggest that up-regulation of miR-142 located in this locus may contribute to increased risk of AD.

INTRODUCTION

Alzheimer's disease (AD) is the most common age-related neurodegenerative disease worldwide manifested by the progressive loss of memory and cognitive decline [1]. AD is a complex disease which is influenced by both environmental and genetic factors [2-4]. Enormous efforts have been made over the past decades to discover risk factors that play a role in development of AD and new biomarkers that help in early diagnosis of the disease [5, 6]. The cause of early-onset AD is known with mutations in one of three genes, *APP, PSEN1* and *PSEN2* [7]. The late-onset AD (LOAD) is the common form of disease and is genetically heterogeneous with a heritability between 60% and 80% [8]. In addition to the apolipoprotein E (*APOE*) polymorphisms that could explain almost 25% of the LOAD heritability, genome-wide association studies (GWAS) have enabled to identify more than 20 common genetic loci associated with LOAD [9, 10]. Most of the associated variants map to non-coding genomic regions and their biological relevance to the disease remain poorly understood.

Recent developments in RNA-sequencing technologies have revealed the complexity of our genome, displaying that protein-coding RNAs make only ~2% of the transcriptome [11]. This highlights a lack of understanding in the possible contribution of non-coding RNAs (ncRNAs) to disease biology. Today, understanding the significance of ncRNAs is one of the most important challenges facing biology. The ncRNAs can be roughly categorized, based on their transcript size, into small ncRNAs (such as microRNAs (miRNAs)) and long ncRNAs (lncRNAs) [12]. MiRNAs, with approximately 21-23 nucleotides in length, are the best-characterized ncRNAs. Previous *in vitro* and *in vivo* studies have shown the crucial roles of miRNAs in various biological processes including developmental and pathological processes in the brain [13-15]. Accordingly, we and others have shown previously that genetic variants in miRNAs affect miRNA-mediate regulation of genes involved in AD [16, 17]. LncRNAs comprise a large and diverse class of transcribed RNA molecules with over 200 nucleotides that do not encode proteins [18]. They may be classified into different subtypes (including antisense, intergenic, overlapping, intronic, and bidirectional) according to the position and direction of transcription in relation to other genes [19, 20]. It has become increasingly evident that lncRNAs are important regulators of gene expression through their divers biological functions including epigenetic, post-transcriptional and translational coordination of gene expression [21, 22]. LncRNAs (with nearly 30,000 different transcripts) account for the major part of the non-coding transcriptome, however, a few lncRNAs have been characterized in details to date [21-23]. In the present study, we performed a genome-wide scan for lncRNAs associated with LOAD using data from the thus far largest GWAS [9]. Further, we performed *in silico* and *in vitro* experiments to provide evidence for functionality of the associated variants in relation to AD.

MATERIALS AND METHODS

Genome-wide association study on LOAD

The summary statistics data (Stage 1 and 2) from the recent GWAS meta-analysis on LOAD, reported by IGAP consortium [9], were used to examine the association between genetic variants in lncRNAs and AD. In Stage 1 (including 7,055,881 SNPs) the combined sample included 17,008 AD cases and 37,154 controls and in Stage 2 loci with *p*-values (combined over all SNPs at the locus) less than 10^{-4} were examined for replication for 8,572 AD cases and 11,312 controls of European ancestry. A description of the consortium data sets and participants are described elsewhere [9].

Genetic variants in lncRNAs

We extracted genetic variants that are located in human lncRNA transcripts using lncRNASNP database (http://bioinfo.life.hust.edu.cn/lncRNASNP/), a comprehensive database including 495,729 SNPs in 32,108 lncRNA transcripts of 17,436 lncRNA genes [24]. We excluded SNPs with minor allele frequencies (MAF) < 0.01 in our analysis. Of the remaining SNPs, we analyzed 55,418 SNPs in 12,578 lncRNAs that were present in the AD GWAS data [9]. The Bonferroni correction was used to adjust p-value for the number of tests and the significance threshold was set at 9.0×10^{-7} (0.05 / 55418).

Functional mapping and annotation of variants associated with AD

We examined whether SNPs in lncRNAs associated with AD are annotated to regulatory features, including promoter and enhancer regulatory motifs, DNase footprinting sites and conserved sequences using the HaploReg v4.1 (http://www.broadinstitute.org/mammals/haploreg/haploreg. php). This databases was also used to check eQTL of the associated SNPs. In addition, for each of the associated variants, the LD region ($R^2 > 0.8$) was determined using the 1000G Phase 1 population. For each set of variants in strong LD with a given associated lncRNA SNP, we investigated whether the SNP was located in a potential regulatory region using the Roadmap consortium reference epigenomes dataset [25]. The reference epigenomes used were chromatin state models based on imputed data of 12 histone marks, identifying separate epigenetic chromatin states for each of the SNPs. In addition, to determine the functionality of the GWAS hits, we investigated whether known protein-coding variants were in high LD with the associated lncRNA SNPs. Regional plots showing the association of lncRNA SNPs and flanking variants in the corresponding loci with AD were generated by the LocusZoom web tool (http://locuszoom.org/).

Gene expression analysis in human brain samples

The brain samples (3 gray matter and 3 white matter) were obtained from the Netherlands Brain Bank (Amsterdam, The Netherlands) and were a kind gift of Dr. Marvin van Luijn. All samples were free of neurological disease. Total RNA from the brain cryopreserved sections was isolated using Trizol LS reagent (Invitrogen, Carlsbad, CA, USA) according to manufacturer's protocols. The concentration of total RNA was determined using a NanoDrop ND-1000 spectrophotometer (NanoDrop, Wilmington, DE). TaqMan q-PCR Assays were used according to the manufacturer's protocols (Applied Biosystems, Foster City, CA, USA) to determine expression levels of mature miR-142 and *BZRAP1-AS1*. RNU6B was used as an internal control for miRNA expression analysis. All experiments were performed in triplicate and repeated three times.

Association of miR-142 target genes with AD

The miR-142 precursor encodes two mature miRNAs (miR-142-3p and miR-142-5p). A list of putative target genes of miR-142 were retrieved from five widely used online miRNA target prediction databases, miRtarget2 [26] (http://mirdb.org/miRDB), picTar [27] (http://pictar.mdc-berlin.de), PITA [28] (https://genie.weizmann.ac.il/pubs/mir07/index.html), TargetScan [29] (http://www.targetscan.org) and miRanda [30] (http://www.microrna.org/microrna). To reduce the number of false targets of miR-142 among these prediction tools, we included the miRNA target genes overlapped in at least three databases. Further, we added experimentally validated target genes of miR-142-3p and -5p (miRNA-target interactions with strong evidence) reported in miRTarBase (http://mirtarbase.mbc.nctu.edu.tw/) [31] to our list. The RNA-seq data from the Human Body Map 2.0 were used to check which of the target genes are expressed in the brain. We filtered the target genes that were expressed in the brain (FPKM $\geq$ 1). To enrich our list for the target genes that are more likely to mediate the effect of miR-142 in AD, we included those target genes that were association with AD. To this end, we examined the association of genetic variants in the selected target genes with AD using the GWAS data in a candidate gene approach. The significance threshold was set using the Bonferroni correction and based on the number of tested variants the selected target genes. In addition, we performed pathway analysis using DAVID gene ontology (GO) program (https://david.ncifcrf.gov) and used AlzGene database (www.alzgene.org) to detect the target genes linked to neurological-related pathways.

Knockout miR-142 mice

In mice, miR-142 gene (ENSMUSG00000065420) is located on chromosome 11 and in the vicinity of an exon belonging to a lincRNA (ENSMUSG00000084796), also known as miR-142 host gene. The miR-142 locus also contains *Bzrap1* (ENSMUSG00000034156), a coding gene 3.5 kb down-stream of the miR-142 gene. The Knockout (KO) miR-142

mouse model has been generated recently by Shrestha et al [32], which is a mouse model with complete deletion of miR-142 and the deletion of the second exon of lncRNA (ENSMUSG00000084796) as well. The brain sections of three KO miR-142 mice and three controls (C57BL/6N) were used to test whether the selected target genes of miR-142-3p and miR-142-5p are up-regulated in the brain of KO mice.

RESULTS

lncRNA variants associated with AD

Out of 55,418 studied variants (in 12,578 lncRNAs), six SNPs in five lncRNAs passed the study significance threshold (p-value= 9.02×10^{-7}) (**Table 1**). Regional association plots showing the association of the 6 lncRNA SNPs and other variants in the corresponding loci with AD are presented in **Figure S1**. Five of these associated SNPs are localized adjacent to protein-coding genes and had proxy variants in the coding regions with smaller *p*-values (more significant association with AD). Due to this close proximity to the coding regions and since in case of a single functional variant, the functional SNP is expected to show the strongest association, it is less likely that the identified lncRNA SNPs drive the association in these loci. Therefore, we limited our investigations to 17q22 locus that the lncRNA SNP rs2632516 is the lead variant (p-value= 8.8×10^{-7}) (**Figure 1**). This SNP overlaps lincRNA miR-142 host gene (also known as miR-142hg) and the intronic part of antisense lncRNA *BZRAP1-AS1*, in the opposite strand.

Table 1. Characteristics of five lncRNAs associated with Alzheimer's disease

SNP ID	lncRNA ID	Chr.	Position	A1/A2	MAF	Beta	SE	*p*-value
rs636355	lnc-CCDC83-1:1	11	85724661	T/C	0.44	-0.119	0.014	6.1×10^{-18}
rs11039225	lnc-SPI1-1:1	11	47430599	A/G	0.37	0.072	0.013	3.9×10^{-8}
rs9273489	lnc-HLA-DQA1-5:1	6	32628189	G/A	0.16	-0.132	0.024	6.9×10^{-8}
rs8028610	lnc-TRPM7-1:1	15	51025494	T/C	0.30	-0.071	0.014	6.5×10^{-7}
rs2632516	lnc-BZRAP1-1:1	17	56409089	C/G	0.46	-0.064	0.013	8.8×10^{-7}

Chr. Chromosome; A1, Reference allele; A2, effect allele; MAF, Minor allele frequency; Beta, effect estimate; SE, Standard error. The significant study threshold is 9.02×10^{-7}.

Functional annotation of rs2632516

We performed *in silico* analysis to predict the impact of rs2632516 on the two overlapping ncRNAs. The variant is located within the intron of *BZRAP1-AS1*. We checked the possibility that the variant affects *BZRAP1-AS1* function through altering the lncRNA secondary structure. To this end, we generated the lncRNA hairpin structures with different sizes containing either the rs2632516 major allele (G) or the minor (A) allele using the Vienna

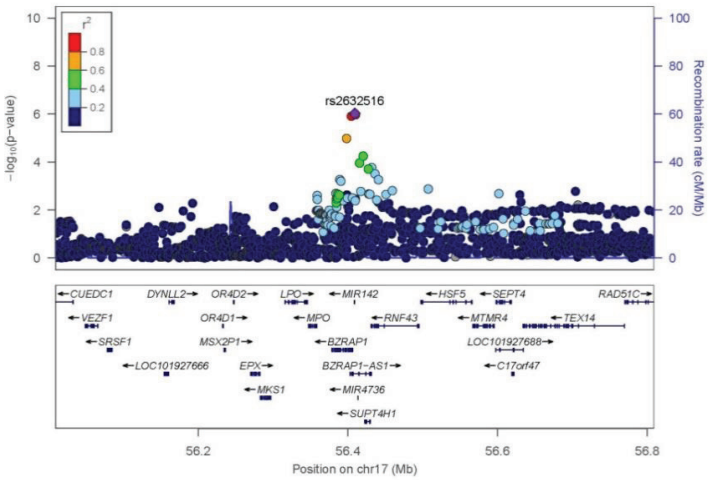


Figure 1. Regional association plot showing the association of rs2632516 and other variants at 17q22 locus (hosting *BZRAP1-AS1* /miR-142) with AD.

RNAfold algorithm [25]. We observed only -0.5 kcal/mol difference in the minimum free energy of the thermodynamic ensemble of the lncRNA predicted structure containing different alleles (**Figure S2**). Further, our analysis demonstrated that rs2632516 is located in the promotor region of miR-142 (410 nt up-stream of the precursor miR-142 sequence). Our expression analysis showed that both mature miR-142 isoforms (miR-142-3p and -5p) and *BZRAP-AS1* are expressed in the brain (both white and gray matter). It has previously been suggested that antisense lncRNAs can regulate expression levels of their nearby genes [25, 26]. However, the expression analysis of the KO miR-142 mice has been demonstrated that while the expression of miR-142 and *BZRAP1-AS1* are decreased, the expression of *BZRAP1* is not altered, suggesting no correlation between *BZRAP1-AS1* and *BZRAP1* expression. Cis-eQTL data form brain and other tissues also showed no correlation between rs2632516 and expression levels of *BZRAP1*. These data suggest that rs2632516 is more likely to work through altering miR-142hg function.

Target genes of miR-142 associated with AD

To identify target genes that may mediate the downstream effect of miR-142 in regard to AD, we compiled a list of all putative target genes of miR-142-3p and miR-142-5p. Since miRNAs and targets should be expressed in the same target tissue for any biological function to be exerted, we focused our analysis on 204 and 582 target genes of miR-142-3p and -5p that are expressed in the brain, respectively. We further filtered the brain-expressed target genes to those that were associated with AD (using the AD GWAS data) or were linked to neuronal pathways (using pathway analysis by DAVID web tool). These

filters highlighted a number of target genes that are likely to play the mediatory roles for miR-142 in AD (**Figure 2**). Among these target genes, the regulatory interactions of *SIRT1, TGFBR1, ROCK2, RAC1* with miR-142-3p, and *SIRT1* and *TP53INP1* with miR-142-5p have been previously confirmed *in vitro*.

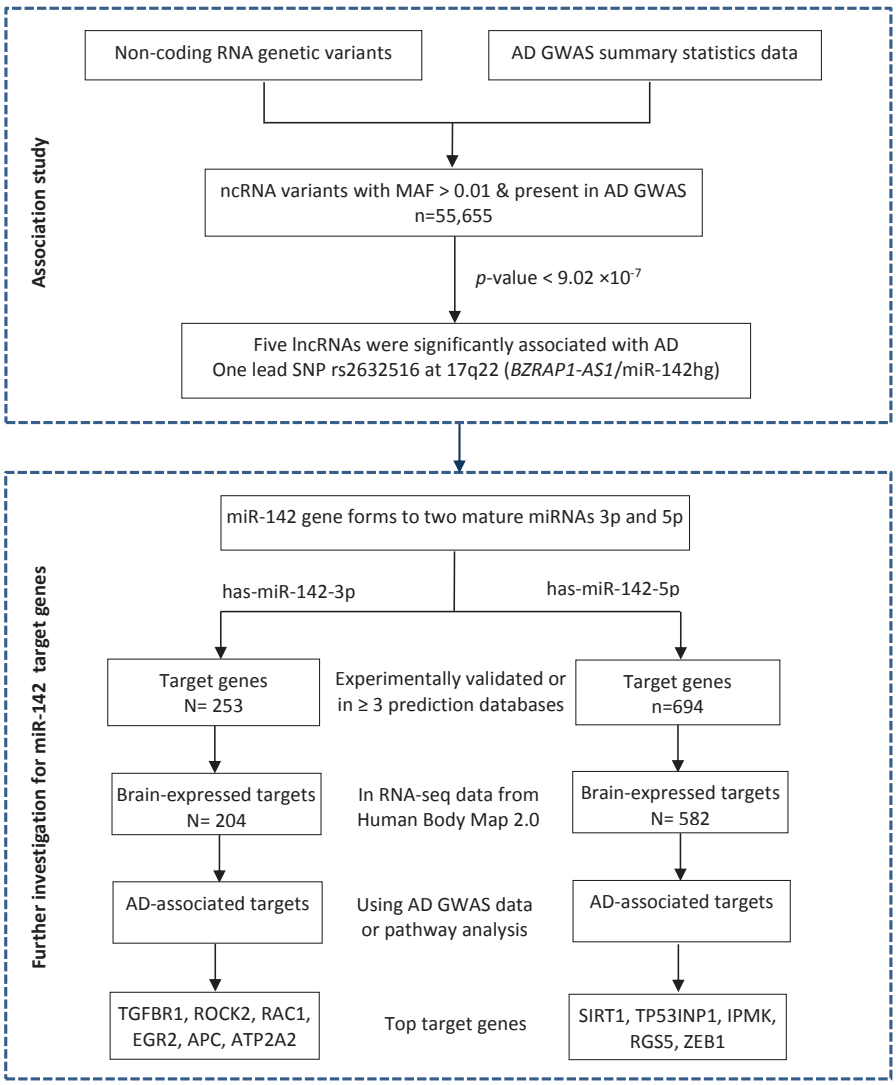


Figure 2. Identification and functional analysis of lncRNA variants association with AD.

DISCUSSION

It has become increasingly evident that lncRNAs are important regulators of gene expression, and they are thought to have a wide range of functions in cellular and developmental processes [19-23]. However, despite some well-characterized lncRNAs such as *HOTAIR*, little is known about the general features of most lncRNAs and their possible molecular mechanisms in human diseases. Recent studies have indicated the critical role of lncRNAs in neuronal development and function [35-38]. A few lncRNAs have been also shown to be dysregulated in the brain of AD patients. For example, antisense lncRNA *BACEI-AS* has been revealed to be implicated in AD through regulating the expression of *BACE1* [39], an essential gene for the production of the toxic amyloid-β (Aβ) [40]. Another dysregulated lncRNA in AD is *BC200*, which is expressed almost exclusively in neuronal cells and its expression increased in parallel with the progression of AD [41, 42]. These studies, however, are mainly focused on differentially expressed lncRNA in a small number of samples. A few studies have also linked lncRNAs to neurodegenerative disorders using genetic data in the candidate gene approach [43, 44]. Here, we performed a genome-wide scan for lncRNAs associated with AD using the largest available GWAS which provided us a proper statistical power to detect lncRNAs potentially involved in AD [9]. We found 17q22 as a new susceptibility locus for AD. The lead variant in this locus, rs2632516 overlaps *BZRAP1-AS1* and miR-142. Our *in silico* analysis proposed that the SNP is more likely to act through altering the promoter activity of miR-142, the minor allele C attenuates AD risk via reducing miR-142 expression. It has been shown previously that variants in the promotor region of miRNA genes could alter the processing and expression of miRNAs [45, 46].

The miR-142 hairpin is precursor of two functional mature strands, miR-142-3p (guide strand) and miR-142-5p (passenger strand). The miRNA is highly conserved amongst many species (**Figure S5**), suggesting that miR-142 plays an essential biological role. Accordingly, miR-142 is emerging as a major regulator of the various biological processes recently, including hematopoietic system, lung development and cardiac hypertrophy [47]. Convergent evidence from multiple investigations also indicate that miR-142 is implicated in the brain and neurodegenerative disorders [48, 49]. In this context, miR-142 has been shown to be expressed in human brain tissues including hippocampus, the prime target of neurodegenerative diseases such as AD [48]. Zovoilis el al have shown 448 miRNAs that are detectable in the mouse hippocampus by RNA-seq analysis [50], miR-142 is one of the abundantly expressed hippocampal miRNAs with more than 5000 counts. Lau et al have also recently reported 31 deregulated miRNAs in AD by miRNA expression profiling in the hippocampus of 41 LOAD patients and 23 age-matched controls [51]. Notably, they have shown that miR-142-3p is one of the 15 significantly up-regulated miRNAs in the LOAD group (with 1.7 fold change). In addition, they have

performed deep-sequencing of miRNAs in the prefrontal cortex of a second independent cohort of 12 LOAD patients and revealed that both miR-142-3p and miR-142-5p are significantly up-regulated (2 and 3.2 fold change respectively) [51]. These data together suggest that higher levels of miR-142 in the brain may increase risk of AD.

Although the alteration of miR-142 expression has been demonstrated in previous studies to be linked with AD, very little is known about the effect of chronic miR-142 upregulation in the brain and neuronal cells. The deregulation of miR-142 might be particularly relevant to AD at several levels. For instance, miR-142 has been shown to be related to neuro-inflammatory changes occurring during multiple sclerosis (MS) [49]. The innate system of immunity and inflammatory signaling are critical for brain homeostasis, repair, and neuroprotection [52, 53]. Increasing evidence suggests that AD pathogenesis is not restricted to the neuronal compartment, but includes strong interactions with immunological mechanisms for example by release of inflammatory mediators in the brain [53]. MiR-142 has been demonstrated to be up-regulated in response to inflammation in the brain [48]. Chaudhuri et al have used microarray data of prefrontal cortex samples from the same cohort of LOAD patients with miRNA data [54] and found down-regulation of several target genes of miR-142. These include *SIRT1* that our analysis also showed it as a putative target for both miR-142-3p and -5p that may mediate the effect of miR-142 in relation to AD. *SIRT1* is a NAD-dependent protein deacetylase that its neuroprotective roles have been well documented in vitro and in vivo [55]. The brain-specific knockout of *SIRT1* has been shown to impair memory and synaptic plasticity [56]. In addition, overexpression and activation of *SIRT1* have been shown to reduce beta amyloid production and toxicity in cellular and animal model of AD [57, 58]. This gene deacetylases microtubule binding protein Tau and promotes its clearance, thus preventing formation of neurotoxic tangles [59]. Chaudhuri et al have also experimentally shown that *SIRT1* is a bona fide target of miR-142-5p [48]. They have demonstrated that overexpression of miR-142 led to a 3.7-fold decrease in *SIRT1* protein levels and a 33-fold decrease in its mRNA levels. Chaudhuri et al have further shown that miR-142 is expressed in hippocampal neurons and confirmed miR-142-mediated repression of SIRT1 in primary human neurons [48]. Therefore, locating at the intersection of anti-inflammatory, antiaging and neuroprotective pathways, makes SIRT1 an ideal molecule for mediating the downstream effects of miR-142 in the pathogenesis of AD. However, it is certainly possible that other miR-142 target genes may be acting in concert with SIRT1 to alter the AD risk. For instance, Chaudhuri et al, in an independent study, have shown that miR-142 indirectly reduces *MAOA* protein level through regulating *SIRT1*, which is known as a key transcriptional up-regulator of *MAOA* [60]. MiR-142 downregulates the expression of *SIRT1*, and *SIRT1* can act on the *MAOA* promotor to upregulate its expression [60]. *MAOA* is a neurotransmitter-metabolizing enzymes and delaminates serotonin, melanin, epinephrine and norepinephrine [60]. Up-regulation of

miR-142 may thereby contribute to change in dopaminergic neurotransmission in AD by lowering *MAOA* expression and activity.

Collectively, in this study we performed a genome-wise scan for lncRNAs associated with AD and found 17q22 (harboring *BZRAP1-AS1* and miR-142) as a new susceptibility locus for AD. We provide evidence suggesting that up-regulation of miR-142 may contribute to increased risk of AD. We subsequently highlighted a number of target genes that may mediate the effect of miR-142 in relation of AD, including *SIRT1* which is a known gene with neuroprotective roles. Further experimental studies are needed to determine the functional impact of the identified variant on miR-142 expression level as well as the potential role of miR-142 in the pathogenesis of AD.

REFERENCES

1. Brookmeyer R, Johnson E, Ziegler-Graham K, Arrighi HM: Forecasting the global burden of Alzheimer's disease. Alzheimers Dement 2007, 3:186-191.
2. Schmechel DE, Saunders AM, Strittmatter WJ, Crain BJ, Hulette CM, Joo SH, Pericak-Vance MA, Goldgaber D, Roses AD: Increased amyloid beta-peptide deposition in cerebral cortex as a consequence of apolipoprotein E genotype in late-onset Alzheimer disease. Proc Natl Acad Sci U S A 1993, 90:9649-9653.
3. Frankfort SV, Tulner LR, van Campen JP, Verbeek MM, Jansen RW, Beijnen JH: Amyloid beta protein and tau in cerebrospinal fluid and plasma as biomarkers for dementia: a review of recent literature. Curr Clin Pharmacol 2008, 3:123-131.
4. Crews L, Masliah E: Molecular mechanisms of neurodegeneration in Alzheimer's disease. Hum Mol Genet 2010, 19:R12-20.
5. Hampel H, Frank R, Broich K, Teipel SJ, Katz RG, Hardy J, Herholz K, Bokde AL, Jessen F, Hoessler YC, et al: Biomarkers for Alzheimer's disease: academic, industry and regulatory perspectives. Nat Rev Drug Discov 2010, 9:560-574.
6. Kivipelto M, Helkala EL, Laakso MP, Hanninen T, Hallikainen M, Alhainen K, Soininen H, Tuomilehto J, Nissinen A: Midlife vascular risk factors and Alzheimer's disease in later life: longitudinal, population based study. BMJ 2001, 322:1447-1451.
7. Ertekin-Taner N: Genetics of Alzheimer's disease: a centennial review. Neurol Clin 2007, 25:611-667, v.
8. Gatz M, Reynolds CA, Fratiglioni L, Johansson B, Mortimer JA, Berg S, Fiske A, Pedersen NL: Role of genes and environments for explaining Alzheimer disease. Arch Gen Psychiatry 2006, 63:168-174.
9. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, DeStafano AL, Bis JC, Beecham GW, Grenier-Boley B, et al: Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet 2013, 45:1452-1458.
10. Calero M, Gomez-Ramos A, Calero O, Soriano E, Avila J, Medina M: Additional mechanisms conferring genetic susceptibility to Alzheimer's disease. Front Cell Neurosci 2015, 9:138.
11. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, et al: Landscape of transcription in human cells. Nature 2012, 489:101-108.
12. Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermuller J, Hofacker IL, et al: RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science 2007, 316:1484-1488.
13. Geekiyanage H, Chan C: MicroRNA-137/181c regulates serine palmitoyltransferase and in turn amyloid beta, novel targets in sporadic Alzheimer's disease. J Neurosci 2011, 31:14820-14830.
14. Long JM, Lahiri DK: MicroRNA-101 downregulates Alzheimer's amyloid-beta precursor protein levels in human cell cultures and is differentially expressed. Biochem Biophys Res Commun 2011, 404:889-895.
15. Patel N, Hoang D, Miller N, Ansaloni S, Huang Q, Rogers JT, Lee JC, Saunders AJ: MicroRNAs can regulate human APP levels. Mol Neurodegener 2008, 3:10.
16. Ghanbari M, Ikram MA, de Looper HW, Hofman A, Erkeland SJ, Franco OH, Dehghan A: Genome-wide identification of microRNA-related variants associated with risk of Alzheimer's disease. Sci Rep 2016, 6:28387.
17. Dorn GW, 2nd, Matkovich SJ, Eschenbacher WH, Zhang Y: A human 3' miR-499 mutation alters cardiac mRNA targeting and function. Circ Res 2012, 110:958-967.

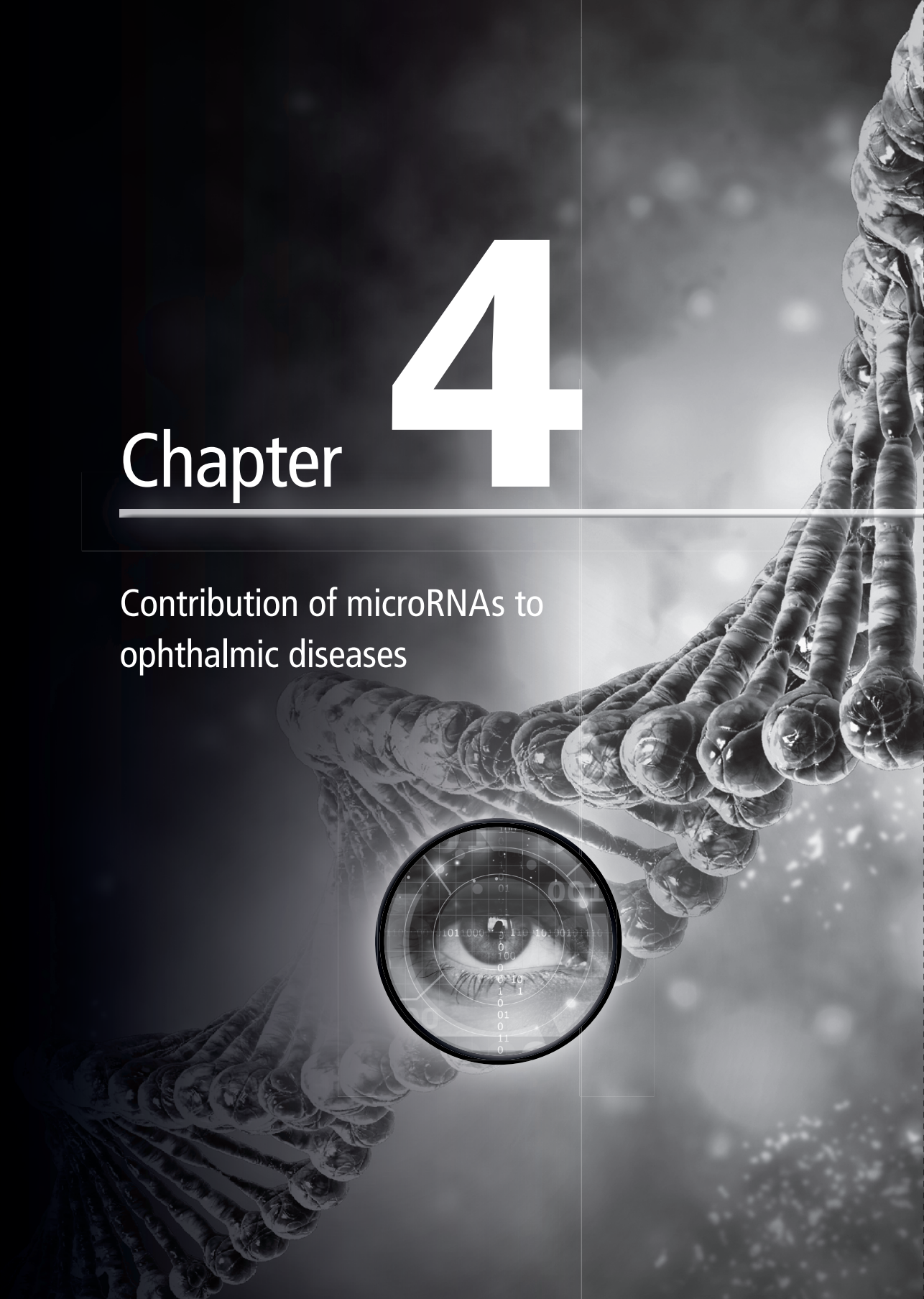
18. Rinn JL, Chang HY: Genome regulation by long noncoding RNAs. Annu Rev Biochem 2012, 81:145-166.
19. Mattick JS, Rinn JL: Discovery and annotation of long noncoding RNAs. Nat Struct Mol Biol 2015, 22:5-7.
20. Peschansky VJ, Wahlestedt C: Non-coding RNAs as direct and indirect modulators of epigenetic regulation. Epigenetics 2014, 9:3-12.
21. Wapinski O, Chang HY: Long noncoding RNAs and human disease. Trends Cell Biol 2011, 21:354-361.
22. Lalevee S, Feil R: Long noncoding RNAs in human disease: emerging mechanisms and therapeutic strategies. Epigenomics 2015, 7:877-879.
23. Bhan A, Mandal SS: Long noncoding RNAs: emerging stars in gene regulation, epigenetics and human disease. ChemMedChem 2014, 9:1932-1956.
24. Gong J, Wu Y, Zhang X, Liao Y, Sibanda VL, Liu W, Guo AY: Comprehensive analysis of human small RNA sequencing data provides insights into expression profiles and miRNA editing. RNA Biol 2014, 11:1375-1385.
25. Romanoski CE, Glass CK, Stunnenberg HG, Wilson L, Almouzni G: Epigenomics: Roadmap for regulation. Nature 2015, 518:314-316.
26. Wang X, El Naqa IM: Prediction of both conserved and nonconserved microRNA targets in animals. Bioinformatics 2008, 24:325-332.
27. Chen K, Rajewsky N: Natural selection on human microRNA binding sites inferred from SNP data. Nat Genet 2006, 38:1452-1456.
28. Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E: The role of site accessibility in microRNA target recognition. Nat Genet 2007, 39:1278-1284.
29. Agarwal V, Bell GW, Nam JW, Bartel DP: Predicting effective microRNA target sites in mammalian mRNAs. Elife 2015, 4.
30. Betel D, Koppal A, Agius P, Sander C, Leslie C: Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. Genome Biol 2010, 11:R90.
31. Chou CH, Chang NW, Shrestha S, Hsu SD, Lin YL, Lee WH, Yang CD, Hong HC, Wei TY, Tu SJ, et al: miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. Nucleic Acids Res 2016, 44:D239-247.
32. Shrestha A, Carraro G, El Agha E, Mukhametshina R, Chao CM, Rizvanov A, Barreto G, Bellusci S: Generation and Validation of miR-142 Knock Out Mice. PLoS One 2015, 10:e0136913.
33. Luo H, Sun S, Li P, Bu D, Cao H, Zhao Y: Comprehensive characterization of 10,571 mouse large intergenic noncoding RNAs from whole transcriptome sequencing. PLoS One 2013, 8:e70835.
34. Vance KW, Ponting CP: Transcriptional regulatory functions of nuclear long noncoding RNAs. Trends Genet 2014, 30:348-355.
35. Dweep H, Sticht C, Pandey P, Gretz N: miRWalk--database: prediction of possible miRNA binding sites by "walking" the genes of three genomes. J Biomed Inform 2011, 44:839-847.
36. Kumar M, Nath S, Prasad HK, Sharma GD, Li Y: MicroRNAs: a new ray of hope for diabetes mellitus. Protein Cell 2012, 3:726-738.
37. Chen L, Song J, Cui J, Hou J, Zheng X, Li C, Liu L: microRNAs regulate adipocyte differentiation. Cell Biol Int 2013.
38. Richardson K, Nettleton JA, Rotllan N, Tanaka T, Smith CE, Lai CQ, Parnell LD, Lee YC, Lahti J, Lemaitre RN, et al: Gain-of-function lipoprotein lipase variant rs13702 modulates lipid traits through disruption of a microRNA-410 seed site. Am J Hum Genet 2013, 92:5-14.

39. Modarresi F, Faghihi MA, Patel NS, Sahagan BG, Wahlestedt C, Lopez-Toledano MA: Knockdown of BACE1-AS Nonprotein-Coding Transcript Modulates Beta-Amyloid-Related Hippocampal Neurogenesis. Int J Alzheimers Dis 2011, 2011:929042.
40. Mulder SD, van der Flier WM, Verheijen JH, Mulder C, Scheltens P, Blankenstein MA, Hack CE, Veerhuis R: BACE1 activity in cerebrospinal fluid and its relation to markers of AD pathology. J Alzheimers Dis 2010, 20:253-260.
41. Mus E, Hof PR, Tiedge H: Dendritic BC200 RNA in aging and in Alzheimer's disease. Proc Natl Acad Sci U S A 2007, 104:10679-10684.
42. Iacoangeli A, Bianchi R, Tiedge H: Regulatory RNAs in brain function and disorders. Brain Res 2010, 1338:36-47.
43. Qi L, Hu Y, Zhan Y, Wang J, Wang BB, Xia HF, Ma X: A SNP site in pri-miR-124 changes mature miR-124 expression but no contribution to Alzheimer's disease in a Mongolian population. Neurosci Lett 2012, 515:1-6.
44. Saba R, Medina SJ, Booth SA: A functional SNP catalog of overlapping miRNA-binding sites in genes implicated in prion disease and other neurodegenerative disorders. Hum Mutat 2014, 35:1233-1248.
45. Cammaerts S, Strazisar M, De Rijk P, Del Favero J: Genetic variants in microRNA genes: impact on microRNA expression, function, and disease. Front Genet 2015, 6:186.
46. Luo X, Yang W, Ye DQ, Cui H, Zhang Y, Hirankarn N, Qian X, Tang Y, Lau YL, de Vries N, et al: A functional variant in microRNA-146a promoter modulates its expression and confers disease risk for systemic lupus erythematosus. PLoS Genet 2011, 7:e1002128.
47. Shrestha A, Mukhametshina RT, Taghizadeh S, Vasquez-Pacheco E, Cabrera-Fuentes H, Rizvanov A, Mari B, Carraro G, Bellusci S: MicroRNA-142 is a multifaceted regulator in organogenesis, homeostasis, and disease. Dev Dyn 2016.
48. Chaudhuri AD, Yelamanchili SV, Marcondes MC, Fox HS: Up-regulation of microRNA-142 in simian immunodeficiency virus encephalitis leads to repression of sirtuin1. FASEB J 2013, 27:3720-3729.
49. Mandolesi G, De Vito F, Musella A, Gentile A, Bullitta S, Fresegna D, Sepman H, Di Sanza C, Haji N, Mori F, et al: miR-142-3p Is a Key Regulator of IL-1beta-Dependent Synaptopathy in Neuroinflammation. J Neurosci 2017, 37:546-561.
50. Zovoilis A, Agbemenyah HY, Agis-Balboa RC, Stilling RM, Edbauer D, Rao P, Farinelli L, Delalle I, Schmitt A, Falkai P, et al: microRNA-34c is a novel target to treat dementias. EMBO J 2011, 30:4299-4308.
51. Lau P, Bossers K, Janky R, Salta E, Frigerio CS, Barbash S, Rothman R, Sierksma AS, Thathiah A, Greenberg D, et al: Alteration of the microRNA network during the progression of Alzheimer's disease. EMBO Mol Med 2013, 5:1613-1634.
52. Nguyen MD, Julien JP, Rivest S: Innate immunity: the missing link in neuroprotection and neurodegeneration? Nat Rev Neurosci 2002, 3:216-227.
53. Heneka MT, Carson MJ, El Khoury J, Landreth GE, Brosseron F, Feinstein DL, Jacobs AH, Wyss-Coray T, Vitorica J, Ransohoff RM, et al: Neuroinflammation in Alzheimer's disease. Lancet Neurol 2015, 14:388-405.
54. Bossers K, Wirz KT, Meerhoff GF, Essing AH, van Dongen JW, Houba P, Kruse CG, Verhaagen J, Swaab DF: Concerted changes in transcripts in the prefrontal cortex precede neuropathology in Alzheimer's disease. Brain 2010, 133:3699-3723.
55. Zhang F, Wang S, Gan L, Vosler PS, Gao Y, Zigmond MJ, Chen J: Protective effects and mechanisms of sirtuins in the nervous system. Prog Neurobiol 2011, 95:373-395.

56. Gao J, Wang WY, Mao YW, Graff J, Guan JS, Pan L, Mak G, Kim D, Su SC, Tsai LH: A novel pathway regulates memory and plasticity via SIRT1 and miR-134. Nature 2010, 466:1105-1109.
57. Chen J, Zhou Y, Mueller-Steiner S, Chen LF, Kwon H, Yi S, Mucke L, Gan L: SIRT1 protects against microglia-dependent amyloid-beta toxicity through inhibiting NF-kappaB signaling. J Biol Chem 2005, 280:40364-40374.
58. Donmez G, Wang D, Cohen DE, Guarente L: SIRT1 suppresses beta-amyloid production by activating the alpha-secretase gene ADAM10. Cell 2010, 142:320-332.
59. Min SW, Cho SH, Zhou Y, Schroeder S, Haroutunian V, Seeley WW, Huang EJ, Shen Y, Masliah E, Mukherjee C, et al: Acetylation of tau inhibits its degradation and contributes to tauopathy. Neuron 2010, 67:953-966.
60. Chaudhuri AD, Yelamanchili SV, Fox HS: MicroRNA-142 reduces monoamine oxidase A expression and activity in neuronal cells by downregulating SIRT1. PLoS One 2013, 8:e79579.

Chapter 4

Contribution of microRNAs to ophthalmic diseases



4.1 Genetic variants in miRNAs and their binding sites associate with age-related macular degeneration (AMD)

4.2 Association of miRNA-related genetic variants with primary open-angle glaucoma

CHAPTER 4.1

Genetic variants in miRNAs and their binding sites associate with age-related macular degeneration (AMD)

Mohsen Ghanbari, Stefan J. Erkeland, Lei Xu, Johanna M. Colijn, Albert Hofman, Andre G. Uitterlinden, Oscar H. Franco, Abbas Dehghan, Caroline C.W. Klaver, Magda A. Meester Smoor

Manuscript based on this chapter

Genetic variants in microRNAs and their binding sites within gene 3'UTRs associate with susceptibility to age-related macular degeneration (AMD)

Human Mutation. 2017; (00):1-12.

ABSTRACT

Background. Age-related macular degeneration (AMD), the leading cause of blindness in the elderly, is a complex disease that results from multiple genetic and environmental factors. MicroRNAs (miRNAs) are small non-coding RNAs that post-transcriptionally regulate target mRNAs and are frequently implicated in human disease.

Methods and Results. Through a genome-wide scan, we investigated the association of genetic variants in miRNAs and miRNA-binding sites within gene 3'UTRs with AMD using data from the largest AMD genome-wide association study. First, we identified three variants in miRNAs significantly associated with AMD. These include rs2168518 in the miR-4513 seed sequence, rs41292412 in pre-miR-122/miR-3591 and rs4351242 in the terminal-loop of pre-miR-3135b. We demonstrated that these variants reduce expression levels of the mature miRNAs in vitro and pointed the target genes that may mediate downstream effects of these miRNAs in AMD. Second, we identified 54 variants (in 31 genes) in miRNA-binding sites associated with AMD. Based on stringent prioritization criteria, we highlighted the variants that are more likely to have an impact on the miRNA-target interactions. Further, we selected rs4151672 within the *CFB* 3'UTR and experimentally showed that, while miR-210-5p downregulates expression of CFB, the variant decreases miR-210-5p-mediated repression of *CFB*.

Conclusion. Our findings support the notion that miRNAs may play a role in the pathogenesis of AMD.

INTRODUCTION

Age-related macular degeneration (AMD) is the primary cause of irreversible vision loss in the elderly, estimated to affect around 200 million people by 2020 [1]. The disease is characterized by loss of function of the retinal pigment epithelium (RPE) and subsequent loss of photoreceptors in the macula [2]. The early stages of AMD are characterized by yellow-white deposits called drusen, while advanced stages of the disease are characterized by either geographic atrophy of the RPE (dry AMD) or choroidal neovascularization (wet AMD) [2]. Susceptibility to AMD is influenced by multiple genetic variants as well as environmental and lifestyle factors such as smoking [3]. Recent genome-wide association studies (GWAS) have identified a large number of variants and genomic loci associated with AMD [4, 5]. Most of the associated variants, however, are located in regions of the genome that do not encode proteins [5]. The non-coding regions are responsible for fine-tuning regulation of gene expression and genetic variants in these regions can be linked to human diseases [6]. These include variants located in microRNAs (miRNAs) and their binding sites within the 3'-untranslated region (3'UTR) of coding genes [7, 8].

MiRNAs are defined as small non-coding RNAs, consisting of 19-22 nucleotides, that post-transcriptionally regulate gene expression [9]. The biogenesis of miRNAs is a multistep coordinated process and miRNA transcripts need to fulfill structural and sequential prerequisites in order to result in expression of the correct mature miRNAs [10]. In the nucleus, miRNA genes are initially transcribed as long primary transcripts (~1000 nt) containing an extended stem-loop structure called a pri-miRNA. These transcripts fold into a hairpin shaped structure and are processed by an RNase-III enzyme, Drosha, into a smaller (~70 nt) hairpin precursor miRNA (pre-miRNA). After being transported to the cytoplasm, a second RNase-III enzyme, Dicer, cleaves the loop structure of the pre-miRNA, resulting in a mature miRNA (~20 nt). The mature miRNA is subsequently incorporated into the RISC complex, and the active complex interacts with the 3'UTR of a target mRNA, resulting in mRNA degradation or translational repression [11, 12]. Sequence variability in miRNAs or in their binding sites could affect miRNA function, resulting in aberrant gene expression. Previous computational and experimental studies have shown that such variants may have a significant impact on disease risk [8, 13-15].

There are strong indications that miRNAs play a role in the pathogenesis of AMD. For example, miR-23 and miR-27 are involved in angiogenesis of choroidal neovascularization [16], and miR-184 promotes RPE degeneration in vitro [17]. In addition, a number of miRNAs have been reported as potential diagnostic biomarkers or therapeutic targets for AMD [18, 19]. Nevertheless, these studies are mainly focused on differentially expressed miRNAs detected by expression arrays in a small sample size, which makes it difficult to extrapolate their results to the general population. Moreover, no systematic investigation of the impact of miRNA-related genetic variation on the risk of AMD has been

published to date. In the present study, we performed a genome-wide scan for miRNAs associated with AMD using data from the largest available GWAS. Further, we performed functional validation studies through various *in silico* and *in vitro* experiments to provide evidence for the potential function of the identified miRNAs in relation to AMD.

MATERIALS AND METHODS

Genome-wide association studies on AMD

To study the association of genetic variants in miRNA-related sequences with AMD, we used the publicity available summary statistics of the largest GWAS on AMD, including 1000 Genome imputed data from 16,144 advanced AMD cases and 17,832 controls on approximately 12 million variants [5]. All participating studies in the AMD GWAS had provided informed consent for participation in genetics studies and were approved by their local ethical committees. More details about the AMD Gene Consortium and participants have been described elsewhere [4, 5].

Identification of genetic variants in miRNAs and miRNA-binding sites

A total number of 2,420 variants located in precursor miRNA sequences were retrieved from miRNASNP (v2) [20]. We excluded variants with minor allele frequency (MAF) <0.01. Of the remaining variants, we analyzed 413 single-nucleotide polymorphisms (SNPs) in 331 miRNAs that were available in the GWAS data for association with AMD. Likewise, we extracted almost 401,000 variants (including SNPs and INDELs) in predicted miRNA-binding sites within gene 3'UTRs using PolymiRTS (v3) [21]. Of these, 82,051 SNPs were included for our analysis (with MAF > 0.01 and present in the AMD GWAS data). The Bonferroni correction was used to adjust the significance threshold for the number of variants in each category. Manhattan and regional plots showing the association of miRNA-related variants and flanking variants in the corresponding loci with AMD were generated by R and LocusZoom web tool, Version 1.1 [22].

The impact of variants in miRNA genes on the secondary structure and expression of miRNAs

The Vienna RNAfold algorithm (package 2.0) was used to predict the effect of miRNA-variants on the hairpin stem-loop structure of the pri-miRNAs [23]. This program calculates the changes in minimum free energy (MFE) of the thermodynamic ensemble of the hairpin structure of miRNAs (wild type and mutant). To examine the functional impact of the variants on the expression levels of mature miRNAs, pre-miRNA sequence containing either major or minor allele were cloned behind the gene encoding green fluorescent protein (GFP) in the expression plasmid MSCV-BC, resulting in GFP-miRNA

fusion transcripts as described previously [24]. For cloning purposes, the restriction enzyme sites *XhoI* and *EcoRI* were inserted in forward and reverse primers, respectively. The primers are listed in **Table S1**. Inserts of all constructs were validated by Sanger sequencing. HEK293 cells transfection, total RNA isolation and quantitative PCRs of EGFP and miRNAs were performed as described previously [13, 24, 25]. Expression levels of miRNAs containing the major and minor allele were calculated relative to EGFP. Experiments were performed in triplicate.

Expression of miRNAs in retinal pigment epithelium (RPE) cells and human brain tissues

The retinal pigment epithelium cells (ARPE) were obtained from the Department of Ophthalmology, Erasmus MC (Rotterdam, the Netherlands). The human brain samples (3 gray matter and 3 white matter) were obtained from the Netherlands Brain Bank (Amsterdam, the Netherlands) and were a kind gift of Dr. Marvin van Luijn. RNA from RPE cells and brain cryopreserved sections was isolated using Trizol LS reagent (Invitrogen, Carlsbad, CA, USA) according to manufacturer's protocols. TaqMan MicroRNA Assays were performed according to the manufacturer's protocols (Applied Biosystems, Foster City, CA, USA) to determine the miRNA expression levels. RNU6B was used as an internal control. All experiments were performed in triplicate and repeated three times.

Association of target genes of the identified miRNAs with AMD

To identify target genes that may mediate the downstream effects of miRNAs associated with AMD, we extracted the miRNA putative target genes from TargetScan v7.0 (Total context++ score > 0.1) (http://www.TargetScan.org/) [26]. In addition, we extracted the miRNA validated target genes, based on crosslinking and immunoprecipitation sequencing data (CLIP-seq), from miRTarBase v6.0 (http://mirtarbase.mbc.nctu.edu.tw/index.php) [27]. Next, we examined the association of genetic variants in these miRNA target genes with AMD using the GWAS data [5]. The Bonferroni correction in this analysis was set based on the number of variants in the target genes of each miRNA.

Functional prioritization of miRNA-binding site variants

To determine whether variants in miRNA-binding sites are functional, specific criteria have been suggested (e.g. a strong association between the variant and the phenotype, the expression of the host gene and regulatory miRNA in a relevant tissue, an allele-specific expression of the host gene) (**Table S2**) [8, 28]. The miRNA-binding site variants associated with AMD were prioritized based on these criteria (see supplementary data for details). To this end, all proxies in high LD (R2 threshold > 0.8, limit distance 100 kb, and population panel CEU) with the binding site variants were extracted and their effects on protein structure, gene regulation and splicing were analyzed using HaploReg

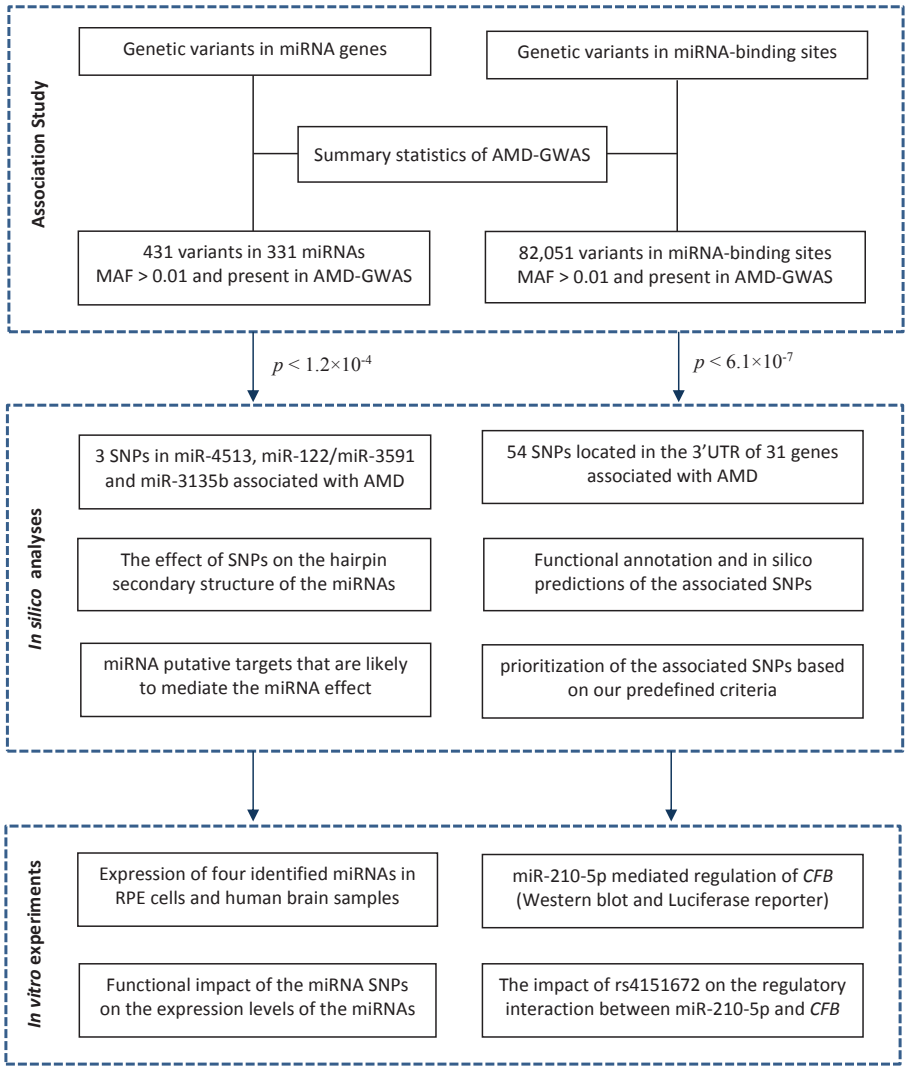


Figure 1 Schematic workflow of our analyses to identify miRNA-related variants associated with AMD. The figure shows our in silico and experimental studies to identify genetic variants in miRNAs as well as miRNA-binding sites within gene 3'UTRs that are associated with AMD. GWAS, Genome-wide association study; SNP, Single-nucleotide polymorphism; MAF, Minor allele frequency.

v4 (http://www.broadinstitute.org/mammals/haploreg/haploreg.php). Further, HaploReg was used to study the haplotypes and the functional potential of variants in the corresponding loci. The correlation between the binding site variants and expression levels of their host transcripts, were determined by GTEx V6 (http://www.gtexportal.org/home/), which provide information on the expression quantitative trait loci (cis-eQTL) in different tissues, and GeneNetwork (http://www.genenetwork.nl/bloodeqtlbrowser/).

The Ocular Tissue Database (https://genome.uiowa.edu/otdb/) and the Illumina's Human Body Map 2.0 (http://www.ensembl.info/blog/2011/05/24/human-bodymap-2-0-data-from-illumina/) were used to examine the expression of genes in eye and brain tissues. Expression of miRNAs in relevant tissues was checked in several databases were screened including miRetina (http://miretina.tigem.it/index.php), miRmine (http://guanlab.ccmb.med.umich. edu/mirmine/help.html), HMDD [29] and PhenomiR [30]. Additional information, such as conservation of miRNA in different species, was obtained from miRBase (release 21) [31]. A list of all web tools and databases that were used for our in silico analysis is shown in **Table S3**.

Luciferase reporter assay

Luciferase reporter assays were performed to assess the interactions between candidate miRNAs and the 3'UTR of their target genes. In addition, these assays were performed to determine the impact of miRNA-binding site variants on the expression levels of the target genes. Primers were designed to amplify 3'UTR sequences of miRNA target genes. Forward and reverse primers contained restriction enzyme sites XbaI and ApaI, respectively. The 3'UTR sequences, containing the putative binding site of a miRNA with either major or minor allele, were amplified and cloned into the pGL3 Luciferase reporter vector (Promega) downstream of the Luciferase open reading frame [24]. All primers are shown in **Table S4**. Inserts of all constructs were confirmed by Sanger sequencing. HEK293 cells were cultured in DMEM containing 10% fetal bovine serum (FBS) and 100 units/ml penicillin/streptomycin in a 37° C incubator with 5% CO2. Cells (n=10,000) were plated into 96-well plates and co-transfected with 1 µg of pGL3 3'UTR constructs, miRNA mimic (mirVanaTM Mimics), and a plasmid expressing the Renilla Luciferase, which served as transfection control, using LipofectamineR RNAiMAX (Invitrogen). Empty 3'UTR vector was used as control, to define the basal expression level of luciferase. The optimal miRNA concentration for regulation of the putative target was determined by testing increasing amounts of miRNA mimic (0, 7.5, 15, 30, 45, and 60 nmol/l). Luciferase activity was determined using the Dual-Glo Luciferase Assay System (Promega) according to manufacturer's protocol. Renilla activity was used for normalization of the data. All the experiments were performed triplicate and repeated in at least three times.

Western blot analysis

Western blot was performed to examine the influence of miR-210-5p on the endogenous CFB protein levels. HEK293 and A549 cells (n=250,000) were cultured in DMEM containing 10% fetal bovine serum (FBS) and 100 units/ml penicillin/streptomycin in a 37°C incubator with 5% CO_2. The next day, miRNA-mimic (50 pmol) was transfected using Lipofectamine 2000 (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. After 48 hour (h), whole cell extracts were obtained by lysis in SDS loading buffer

and were heated at 95 °C for 5 min. Proteins were subjected to a 12.5% sodium dodecyl sulphate-polyacrylamide gel (SDS-PAGE) and separated at 120V for 100 min. Proteins were electrophoretically transferred onto a polyvinylidene difluoride (PVDF) membrane (pore size: 0.45µm, Invitrogen) for 1.5h with an electric current of 250 mA. Subsequently, the membrane was blocked with blocking buffer (LI-COR Biosciences, Lincoln, USA) in 1×PBS containing 0.1% Tween-20. Membranes were incubated with primary antibodies overnight at 4°C. Rabbit anti-CFB (#HPA001817; Atlas Antibodies, Stockholm, Sweden) (1:150) antibody was diluted in 5% (mass/vol) BSA in 1×PBS containing 0.1% Tween-20 and β-actin (1:500, mouse antibody). The membrane was washed 3 times followed by incubation for 1 h with goat anti-rabbit and goat anti-mouse IRDye®-conjugated secondary antibodies (Li-COR Biosciences, Lincoln, USA) (1:5000) at room temperature. β-actin levels served as standardization of sample loading. The membrane was scanned and bands were quantified by Odyssey Infrared Imaging System (Li-COR Biosciences, Lincoln, USA). The experiments were performed triplicate and repeated at least three times.

RESULTS

Three variants in miRNAs were associated with AMD

We tested the association of 413 variants (MAF > 0.01 and present in the AMD GWAS data) in miRNA-encoding sequences with AMD (**Figure 1**). These include 42 variants in miRNA seed regions, 80 variants in mature miRNAs and 291 variants in pre-miRNA sequences. Three variants passed the Bonferroni corrected significance threshold of *p*-value < 1.2×10^{-4} (0.05/413). The three variants could potentially change the biogenesis or function of four distinct miRNAs, i.e. one of the variants overlaps two pre-miRNAs located on opposite strands. The variants include rs2168518: G>A (Chr15:75081078) in the miR-4513 seed sequence (*p*-value= 3.3×10^{-6}), rs41292412: C>T (Chr18:56118358) in the stem regions of pre-miR-122 and pre-miR-3591 (*p*-value= 2.5×10^{-5}), and rs4351242: C>T (Chr6:32749945) in the loop sequence of pre-miR-3135b (*p*-value= 4.1×10^{-4}) (**Table 1**). Regional association plots showing the association between the variants in the corresponding loci and AMD are presented in **Figure S1**.

The miRNA-variants affect expression levels of the mature miRNAs

The secondary structure of a pri-miRNA is critical for its biogenesis [34]. A variant's effect on the miRNA biogenesis can be predicted by performing structural analysis of the miRNA hairpin [25]. We examined the changes in MFE of the thermodynamic ensemble of the hairpin structure of the identified miRNAs for both wild type and mutant alleles using the Vienna RNAfold algorithm [25]. **Figure 2** shows the predicted hairpin structure

Table 1. Characteristics of the three variants in miRNAs associated with AMD

SNP ID	MAF	A1/A2	miRNA ID	SNP position in miRNA	*p*-value in GWAS	Validated target genes associated with AMD
rs2168518	0.39	A/G	miR-4513	Seed seq.	3.3×10^{-6}	*FAM126B* (4.5×10^{-6})
rs41292412	0.01	T/C	miR-122	Stem region	2.5×10^{-5}	*C3* (3.8×10^{-69}) *HLA-DQA1* (1.1×10^{-13})
			miR-3591	Mature seq.	2.5×10^{-5}	*SERPINH1* (1.01×10^{-6})
rs4351242	0.06	T/C	miR-3135b	Terminal loop	4.1×10^{-4}	*C3* (3.8×10^{-69})

MAF, minor allele frequency; Chr. Chromosome; A1, major allele; A2, minor allele. The AMD-associated target genes are the genes with most significantly association with AMD using the GWAS data. * This SNP is present only in the HapMap GWAS.

of the host miRNAs and the position of the three identified variants. The difference in MFE between mutant and wild type variants was 0.70 kcal/mol for pre-miR-4513 and 2.5 kcal/mol for pre-miR-122/miR-3591, suggesting that these variants affect the structure of the pre-miRNA hairpins and thereby would affect miRNA maturation. In contrast, the change in MFE between pre-miR-3135b mutant and wild type was small (0.01 kcal/mol). To experimentally show the impact of the variants on pre-miRNA processing, we determined mature miRNA expression levels for both wild type and variant miRNAs. We have previously demonstrated that the rs2168518 mutant allele decreases the expression of mature miR-4513 (40%, *p*-value= 0.005) (**Figure 2**) [35]. For other three miRNAs, we cloned the miRNA precursor sequences (wild type and mutant) behind GFP in the pMSCV-BC vector. HEK293 cells were transfected, resulting in expression of GFP-miRNA transcripts containing the wild type or mutant pre-miRNAs. Mature miRNA levels were significantly decreased in cells transfected with pre-miRNAs containing mutant alleles; (miR-3135b [70%, *p*-value= 0.0055], miR-122-5p [60%, *p*-value= 0.0021] and miR-3591-5p [95%, *p*-value= 0.0003]) (**Figure 2**).

Potential downstream target genes of the identified miRNAs associated with AMD

We aimed to identify the target genes that may mediate the downstream effects of the four identified miRNAs in relation to AMD. Since we expect these target genes to be associated with AMD as well, we used the GWAS data in a candidate gene approach and analyzed their associations with AMD. For each of the four miRNAs, the total number of putative target genes and those that are associated with AMD are shown in **Table S5**. The AMD-associated target genes of these miRNAs included *TIMP3* (a common target of miR-3135b, miR-4513 and miR-3591), *C3* (a common target of miR-122-5p and miR-3135b), *ARMS2* and *CFH* (two targets of miR-3135b). Of these, the interactions between *C3* and miR-3135b and miR-122-5p have been confirmed by CLIP-seq (**Table 1**) [27].

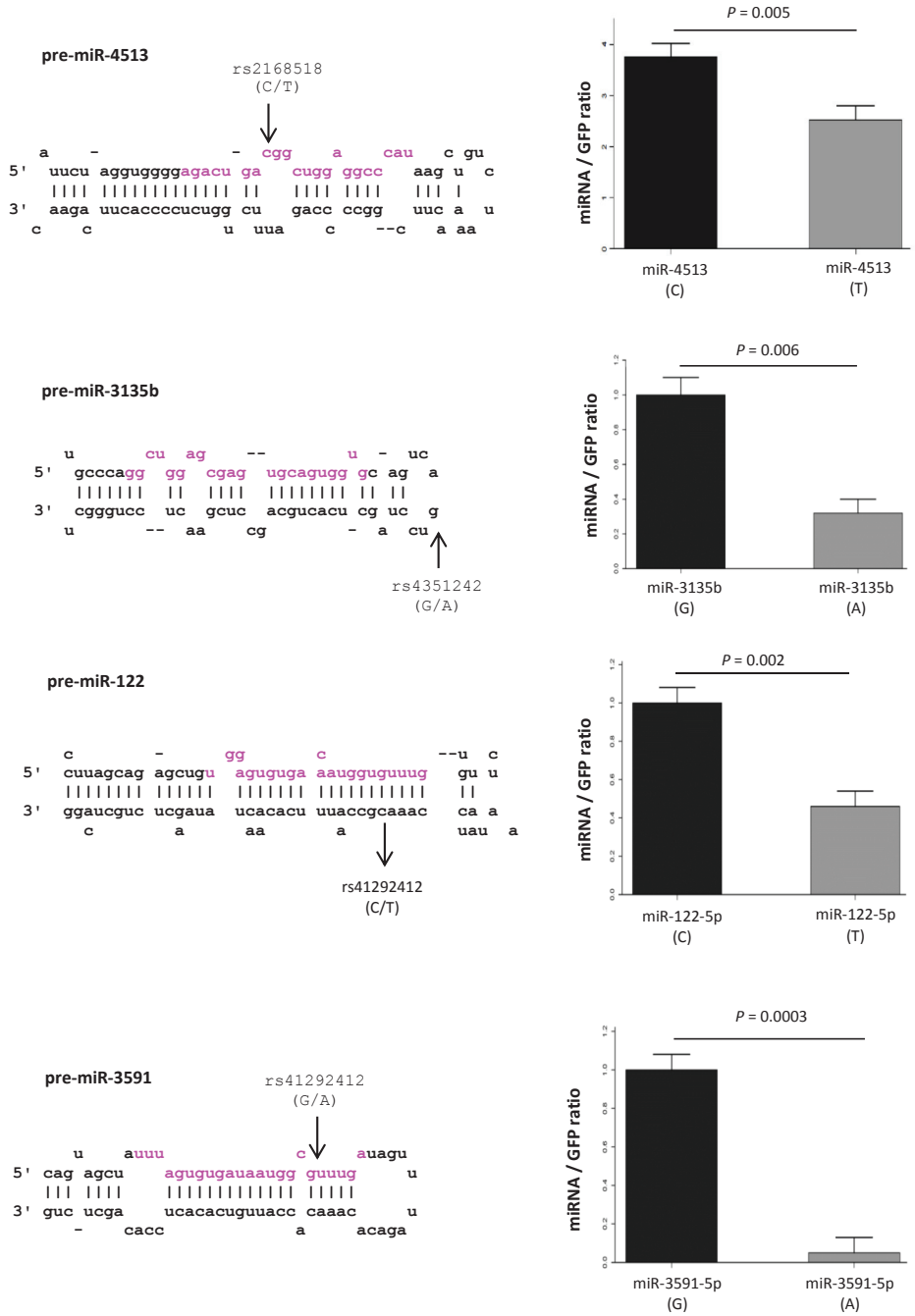


Figure 2. Predicted hairpin structure of four miRNAs hosting the AMD-associated variants and the variant effect on the miRNA expression levels. The mature miRNA sequences are shown in red and the positions of variants associated with AMD are depicted by arrows.

Since a miRNA and its target mRNA should be expressed in the same tissue in order for the miRNA to regulate the expression of its target, we examined whether the identified miRNAs and their highlighted target genes are expressed in the eye and brain tissue. Expression of miR-3135b was detected in RPE cells (average Ct-value: 33) and human brain samples (in both the white and gray matter) (average Ct-value: 30), whereas expression of the other three miRNAs was not detectable (**Table S6**). In addition, gene expression data from the Ocular Tissue Database and Human Body Map 2.0 revealed that the majority of the AMD-associated target genes of miR-3135b are also expressed in the eye and brain tissue (**Table S7**).

Fifty-four miRNA-binding site variants were associated with AMD

In the second part of our study, we examined the associations of 82,051 3'UTR variants located in miRNA-binding sites with AMD (**Figure S2**). Fifty-four variants (in the 3'UTR of 31 genes) passed the significance threshold of p-value < 6.1×10^{-7} (0.05 / 82,051). These variants are predicted to change the regulatory interaction between their host genes and miRNAs by disrupting, creating or modifying a number of miRNA-binding sites that are listed in **Table S8**. Out of the 54 identified variants, 53 are located in the loci that have previously been reported to be associated with AMD in the GWAS of the AMDGene Consortium [4, 5]. We further suggest *CDKN2B* at 9p21 locus as a new susceptibility gene for AMD. The lead variant in this locus, rs3217992 within the *CDKN2B* 3'UTR, is predicted to disrupt the binding site of miR-138-2-3p. Our LD evaluation for the 54 identified variants showed that some of them had proxy variants in strong LD ($R^2 > 0.8$) in the coding regions of related genes (**Table S9**). We performed various functional annotations and *in silico* predictions for the 54 binding site variants to prioritize those that are more likely to alter AMD risk through affecting miRNA-target interactions (**Table S2**). Through *cis*-eQTL analysis, we identified that 18 of the variants are correlated with the expression levels of their host gene in whole blood and/or other tissues (**Table S9**). Gene expression data from the Ocular Tissue Database showed that 29 out of 31 host genes are expressed in the eye (**Table S10**). Further, we found evidence for expression of several miRNAs in our collection in the eye-related tissues using the miRNA expression databases (**Tables S11**). We integrated the results of our analyses for these variants and prioritized them based on their potential functionality score (**Table S11**). **Table 2** shows the top ten variants with the highest credibility for functionality to affect miRNA-binding sites. Regional plots showing the association of these top ten miRNA-binding site variants with AMD are displayed in **Figure S3**.

Table 2 Functional characteristics of the top ten miRNA-binding site variants associated with AMD

	rs6857	rs3217992	rs334348	rs4151672	rs760070	rs705866	rs7674356	rs201075702	rs12724106	rs10119
GWAS data										
p-value	2.4×10^{-33}	4.2×10^{-7}	2.9×10^{-7}	8.9×10^{-41}	9.5×10^{-91}	10.0×10^{-9}	2.6×10^{-7}	1.7×10^{-26}	1.6×10^{-111}	1.3×10^{-7}
A2 (effect)	T (-)	A (-)	A (-)	T (-)	A (-)	C (+)	A (-)	A (-)	C (+)	A (-)
Gene Exp.										
Host gene	*PVRL2*	CDKN2B	*TGFBR1*	*CFB*	*NELFE*	*PILRB*	*CASP6*	*DENND1B*	*ZBTB41*	*TOMM40*
Retina	43.5	35.8	36	12.8	36.5	92.5	13.3	32.4	75.1	82.5
Brain	10.0	0.7	11.8	5.0	19.9	5.6	1.82	1.8	3.7	9.7
Proxy SNPs										
All SNPs	0	5	20	20	15	60	20	0	0	0
Non-syn	0	0	0	3	2	0	0	0	0	0
Cis-eQTL										
host gene	Blood	Blood	_	Multiple	Multiple	Multiple	Multiple	_	_	Blood
miRNA-BS										
A1 (Dis)	miR-645	miR-138	miR-628	miR-210	miR-548an	miR-374	miR-548at	No	miR-551b	No
Δscore	-0.14	- 0.19	NA	-0.27	-0.1	-0.06	0.09	-	0.03	-
A2 (Cre)	miR-320e	miR-374c	miR-32	miR-188	miR-552	No	miR-628	miR-153	No	miR-516b
Δscore	-0.19	-0.15	0.09	-0.18	-0.05	-	0.13	0.001	-	-0.06

Shown are ten 3'UTR SNPs in different genes that are one of the top variants in their loci associated with AMD and are predicted to affect miRNA-mediated gene regulation. Haplotype information and cis-eQTL from HaploReg (v4) database and miRNA-binding site effect from PolymiRTS (v3). For SNPs that are predicted to affect more than one miRNA-binding site, only the highly conserved miRNA has mentioned in the table. The value for retina expression is PLIER normalized level of expression (based on the Affymetrix human Exon 1.0 ST arrays) from the OTDB. Brain expression levels are indicated as fragments per kb of exon per million reads (FPKMs). A2, mutant allele; Exp., Expression; Non-syn, Non-Synonymous proxy variant; Multiple, Correlation of the SNP with host gene expression in different tissues; miR-BS, miRNA-binding site; Δscore, Context score change; Dis, Disruption of a miRNA-binding site; Cre, Creation of a miRNA-binding site.

The functional impact of miRNA-binding site variants on the miRNA-target interactions

Three of the prioritized binding site variants, including rs6857:C>T (*PVRL2*:miR-320e), rs3217992:C>T (*CDKN2B*:miR-138-3p) and rs334348:A>G (*TGFBR1*:miR-628-5p), have been confirmed previously to affect miRNA-mediated regulation of their genes *in vitro* [14, 31]. We further selected two of the variants (rs4151672:C>T and rs760070:A>G) that were expected to affect the miRNA-binding sites through two converse scenarios (disruption vs. creation). Rs4151672 in the *CFB* 3'UTR was expected to decrease the interaction between miR-210-5p and *CFB*. The regulatory connection between miR-210-5p and *CFB* was first substantiated by assessing the endogenous protein levels of CFB in two different cell lines (HEK293 and A549) using Western blot analysis (**Figure 3**). Luciferase reporter assays further verified that the relative luciferase activity of the *CFB* 3'UTR construct containing the rs4151672 major allele C is decreased in the presence of miR-210-5p, indicating that *CFB* is a direct target of miR-210-5p (**Figure 3** and **Figure S4**). Next, we showed that the miR-210-5p-mediated repression of the *CFB* reporter is reduced in the presence of minor allele T (p-value < 0.05) (**Figure 3**). Together, our results suggest that rs4151672 diminishes miR-210-5p-mediated regulation of *CFB*, which may consequently result in increased *CFB* expression in AMD patients.

Rs760070 in the 3'UTR of *NELFE* (also known as RDBP) was predicted to improve the interaction between miR-552-5p and *NELFE*. Luciferase reporter assays showed that miR-552-5p downregulates the expression of the *NELFE* 3'UTR construct containing the major allele G by 45%, although the seed sequence dose not pair perfectly (**Figure S5**). The minor allele A of rs760070 makes a perfect match between the miR-552-p seed and the 3'UTR of NELFE. Luciferase reporter assays further showed that the miR-552-5p mediated repression of *NELFE* reporter construct containing the minor allele A is slightly (~10%), but non-significantly, increased (**Figure S5**). This may indicate that the effect of rs760070 on the binding of miR-552-5p to *NELFE* is minor and not detectable with this experimental setup. The non-significant difference can also be explained by compensation of miR-552-5p binding to the *NELFE* 3'UTR by match sequences outside the miRNA seed, so called compensatory 3' pairing [32].

DISCUSSION

In this study, we performed a genome-wide scan for miRNAs associated with AMD using GWAS data. First, we identified three variants in miR-4513, miR-122/miR-3591 and miR-3135b significantly associated with AMD. We demonstrated that the variants reduce expression levels of the mature miRNAs. We then highlighted the target genes of these miRNAs that are associated with AMD and may mediate the downstream effects

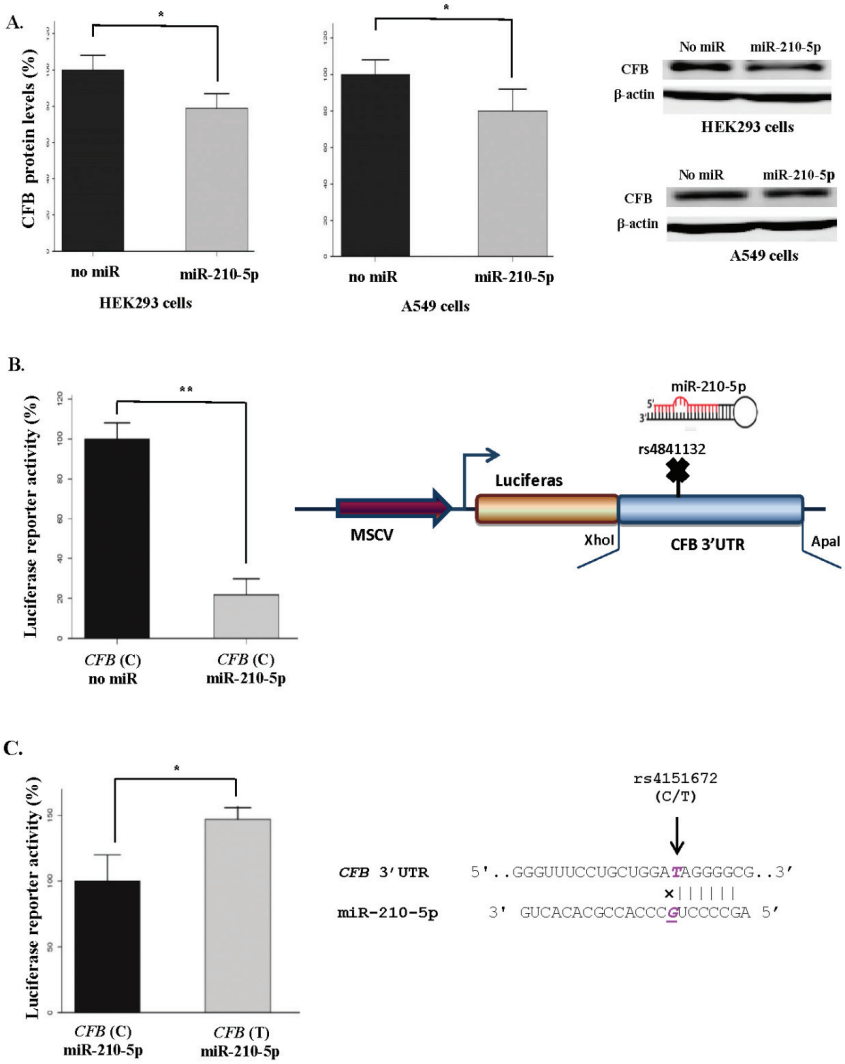


Figure 3. miR-210-5p-mediated regulation of *CFB* and the effect of rs4151672 on the miRNA-mRNA interaction. A) Downregulation of CFB protein levels by miR-210-5p.Western blot analysis was performed to determine the effect of miR-210-5p on endogenous protein levels of CFB in HEK293 and A549 cell lines. β-actin was used as a loading control. This experiment shows that miR-210-5p mimic downregulates CFB protein levels around 20% in both cell lines. B) Luciferase reporter assays indicating miR-210-5p-mediated repression of CFB. HEK293 cells were co-transfected with CFB 3'UTR luciferase reporter vector containing the major allele C and (±) miR-210-5p mimic. This experiment shows that the relative luciferase activity of the CFB 3'UTR construct is decreased in the presence of miR-210-5p. C) Luciferase reporter activity of the CFB 3'UTR construct containing the rs4151672 minor allele T is increased. This experiment suggests that the variant diminishes the regulatory interaction between miR-210-5p and CFB, and consequently results in increased CFB expression levels. Error bars, standard deviation (SD); * p-value < 0.05 and ** p-value < 0.005 compared with the control group.

of the miRNAs. Second, we found 54 variants in miRNA-binding sites (within the 3'UTR of 31 genes) that are associated with AMD. Based on stringent prioritization criteria, we highlighted ten variants that are more likely to have a functional impact on the miRNA-target recognitions. Three of these variants, including rs3217992 (*CDKN2B*:miR-138-3p), rs6857 (*PVRL2*:miR-320e) and rs334348 (*TGFBR1*:miR-628), have been previously verified to affect miRNA-mediated gene regulation in vitro. Additionally, we selected rs4151672 within the 3'UTR of *CFB* and demonstrated that, while miR-210-5p downregulates *CFB* expression, the variant diminishes miR-210-5p-mediated regulation, resulting in increased level of *CFB*.

Previous studies have suggested two main strategies to study if and which miRNAs are involved in the pathogenesis of human diseases: miRNA profiling and genetic approach [35]. In the first approach, affected tissue of patients is subjected to miRNA expression profiling and compared to expression profiles of the same tissue of healthy individuals [36]. In the genetic approach, the starting point is a linkage or an association analysis of miRNA-related variants between patients and controls. We and others have proven that the genetic approach is efficient and successful for identification of new miRNAs for complex diseases [8, 13, 25, 28, 37-39]. The main advantage of the genetic approach compared to miRNA profiling is that when a miRNA variant is associated with disease risk, it supports the idea that the miRNA has a primary effect in the disease mechanisms [35]. Nevertheless, differential expression of miRNAs is not providing evidence for a causal role of the miRNAs in relation to the disease. Expression profiling studies are subject to confounding bias or reverse causation. For instance, differential expression of miRNAs can be a secondary effect of disturbed pathways due to the disease or other characteristics that are correlated to the patient group, such as lifestyle.

The genetic approach may overcomes both challenges. Given that genetic variants are randomly inherited, it is unlikely that other extrinsic factors are associated with them (no confounding bias). In addition, as the genetic information is constant over the life course, reverse causation is refuted. Another advantage of the genetic approach is that diseases with no accessible affected tissues (e.g. eye tissue of AMD patients) can still be studied since genomic DNA is available from other patient tissues such as blood. The genetic approach, however, has some limitations that need to be addressed. One limitation is that the occurrence of common variants in miRNAs is relatively low and because of that the majority (~80%) of genetic variants in miRNA genes are not available in GWAS data. To investigate the association of the rare miRNA variants with human disease, high density genotyping methods or whole-genome sequencing is needed.

In this study, we identified significant association between three miRNA variants and AMD. The first variant, rs4351242, is located in the terminal loop region of miR-3135b. Previous studies have indicated that the structure of the stem loop region of pre-miRNAs is critical for miRNA processing by Drosha and Dicer enzymes [40, 41]. Furthermore, poly-

morphisms in the terminal loop of miRNAs have been shown to block miRNA biogenesis [40, 42]. Rs4351242 could significantly decrease expression of mature miR-3135b. The location of rs4351242, in the center of the terminal loop of miR-3135b, suggests that this variant may affect the binding of RNA-binding proteins that are involved in pre-miRNA processing and consequently reduces miRNA maturation [43, 44]. Next, in order to investigate the downstream effect of miR-3135b in AMD, we performed an in silico search for the miRNA putative targets that are associated with AMD and are expressed in the eye. Among the identified target genes, *C3* is of particular interest because this gene is one of the key AMD genes and its interaction with miR-3135b has been confirmed by CLIP-Seq data [27]. We showed that miR-3135b is expressed in the RPE cells, a retina cell line, and previous studies have also demonstrated *C3* mRNA expression in the retinal pigment epithelium [45, 46]. Higher levels of *C3* contribute to increased risk of AMD [47]. Therefore, reduced expression levels of miR-3135b in rs41275792 minor allele carriers may result in increasing *C3* expression and alter the risk of AMD.

The second variant, rs41292412, is located in a locus in which the forward and reverse DNA strands encode two distinct functional miRNAs, respectively miR-122 and miR-3591. In silico analysis indicated that the mutant allele reduces the stability of the hairpin structures of both pre-miR-122 and pre-miR-3591. Our in vitro experiments confirmed that rs41292412 significantly decreases the expression levels of both miRNAs. MiR-3591-5p is only recently discovered by deep sequencing [48] and we are the first to report the association between this miRNA and a human disease. For miR-122, the pre-miRNA is a precursor of two mature miRNAs, miR-122-3p and miR-122-5p, of which the 5p miRNA is the predominant product [31]. Our results showed that rs41292412 decreases the miR-122-5p expression level. Most likely, this is due to the localization of this variant in the stem region of the pre-miRNA that can interfere with normal biogenesis [35, 49]. Among the miR-122-5p putative targets that are associated with AMD, the interaction between miR-122-5p and *C3* has been confirmed by CLIP-Seq data [27]. However, the expression level of miR-122-5p was not detectable in RPE cells. A number of studies have reported the abundant expression and activation of miR-122-5p in liver and serum [50-52]. *C3* is also predominantly synthesized in the liver and is expressed in serum [46, 53]. Collectively, reduced expression levels of miR-122-5p in carriers of the rs41292412 minor allele may affect miR-122-5p-mediated repression of *C3* in liver or serum. Future studies are needed to investigate the mechanisms at stake in relation to AMD.

The third variant, rs2168518, is located at the seed sequence of miR-4513. The seed sequence is defined as the nucleotide position 2-8 from the 5'-end of a mature miRNA and has to be perfectly complementary to the binding site of a target mRNA to be functional [12]. It is expected that polymorphisms in miRNA seed sequences strongly influence miRNA activity and risk of human disease [37]. Previously, we experimentally showed two mechanisms through which rs2168518 affects miR-4513 function [25]. First, the

minor allele decreases the expression level of miR-4513 by reducing miRNA processing, and second, the minor allele reduces the ability of miR-4513 to interact with its target genes. We have already reported miR-4513 to be associated with cardio-metabolic traits [25]. These data suggest that miR-4513 plays a pleiotropic role in cardiovascular disease and AMD. Pleiotropic roles of miRNAs could be through regulating distinct target genes that have cell type specific expression profiles and are involved in different pathways. Several miRNAs have been reported to function as pleiotropic modulators of physiological and pathological processes, such as miR-206 in cell proliferation, miR-221 in acute liver failure and miR-26a in metabolic regulations [54, 55].

Increasing evidence suggests that 3'UTR of genes contain regulatory elements that play an important role in gene expression [56]. Many studies have recently focused on annotating GWAS hits and functional variants that are located in gene 3'UTRs [28, 57-59]. In this study, we identified 54 3'UTR variants associated with AMD that are located in miRNA-binding sites. Some of these variants had high LD proxies in the protein coding regions of their genes, which hampered interpretation of the miRNA-binding site variants in driving the association with AMD. To gain insight into loci harboring the trait-associated variants identified by GWAS, several bio-informatics strategies have been developed [60-62]. However, these approaches are mainly based on in silico analyses and include only a limited number of functional annotations. Here, we performed various functional annotations and in silico prediction to prioritize the identified variants based on their potential to work through affecting miRNA-target interactions [8, 28]. The top three prioritized variants (rs3217992, rs6857 and rs334348) have previously been confirmed to affect miRNA-mediated regulation of their genes [14, 33]. Among them, rs3217992, the lead variant at the 9p21 locus, disrupts the binding site of miR-138-2-3p within the 3'UTR of *CDKN2B* [28]. *CDKN2B* has not been reported in the original GWAS for AMD, because the lead variant is under the GWAS significance threshold [5]. Polymorphisms in this gene have been shown previously to be associated with open-angle glaucoma, an eye disease that results in damage to the optic nerve and vision loss [63]. Here, we suggest *CDKN2B* as a susceptibility gene for AMD that need to be replicated in future association studies. Further, our results may indicate dysregulation of *CDKN2B* by the miR-132-3p binding site SNP as a functional explanation underlying the association with AMD.

Further, we demonstrated that rs4151672 in the 3'UTR of *CFB* decreases miR-210-5p-mediated repression, resulting in increased expression level of *CFB*. Notably, rs4151672 is in strong LD with rs4151667, a frequently reported AMD-associated missense variant in *CFB*. Smailhodzic et al have shown that the rs4151667 minor allele carriers have reduced levels of *CFB* in serum samples of both AMD cases and controls [64], which seems to be in contrast to our results. In the study of Smailhodzic, however, *CFB* serum levels are analyzed representing the net result of all functional variants that are in LD with the

missense variant. In our study, we investigated the effect of the 3'UTR rs4151672 allele without interference of nearby or distant alleles. In addition, a recent study has shown increased *CFB* protein levels in RPE cells of AMD compared to age-matched normal subjects [65]. Further research is needed to unravel how these two alleles work together in opposite directions to balance the *CFB* expression level [66].

CFB is a critical regulator of the alternative complement pathway [67] and complement has a central role in AMD. We showed that miR-210-5p downregulates *CFB* protein levels. Both miR-210-5p and *CFB* are expressed in ocular tissues [68, 69]. These data suggest a protective role of miR-210-5p in AMD that may be mediated through regulating *CFB* expression. Previous studies have also verified the role of miR-210 in angiogenesis and hypoxia, which are two main mechanisms in the pathogenesis of AMD [70, 71]. Thus, miR-210 could be a promising therapeutic target for AMD that warrant further in vitro and in vivo investigations. A number of complement-targeted therapeutics (e.g. antibodies against complement factors) are currently tested in clinical trials to prevent and treat AMD development [72]. The easy accessibility of the eye by injection makes the eye a suitable organ for miRNA therapeutics. Local delivery of miRNA inhibitors and mimics for treating vascular disorders of the retina has been recently suggested as a potentially clinical approach [16].

This study has strengths and limitations that need be considered in interpreting of the reported results. The main strength of the study is leveraging genetic data from thus far the largest GWAS of AMD which provided us a proper statistical power to detect a number of miRNA-related variants associated with AMD. Moreover, we used a wide range of in silico information to guide our study towards the genetic variants that are likely to be functional in the corresponding loci. Unfortunately, we were not able to perform the functional experiments in a relevant cell line for AMD such as RPE cells, because of slow growth rate and low transfection efficiency of RPE cells in culture. However, to gain insights into the functionality of the identified variants and miRNAs, we performed our experiments using available human cell lines (HEK293 and A549) that are commonly used in these types of studies.

Collectively, this is the first systematic investigation in which miRNAs contributing to AMD are identified using genetic evidence in a population level. Our results showed that aberrant miRNA activities due to genetic variants in miRNA genes as well as miRNA-binding sites within gene 3'UTRs may contribute to an alteration of the AMD risk. Future studies are warranted to dissect the mechanisms of how the identified miRNAs can contribute to the pathogenesis of AMD and their therapeutic potentials. In particular, miR-3135b and miR-210-5p are of interest since they have the potential to regulate *C3* and *CFB*, two genes which play a central role in the alternative complement pathway and contribute to AMD.

Supplement available online at:
http://onlinelibrary.wiley.com/doi/10.1002/humu.23226/suppinfo

Supp. Table S2. List of criteria for prioritization of miRNA-binding site variants

Criterion	Description
SNP functional annotation	Minor allele frequency, Variant position, Predicted function
Strength of the association	The largest GWAS meta-analysis on the studied trait
Haplotype evaluation	Number of proxies, Non-synonymous proxies, Missense variant
Host gene expression	Expression of the gene in tissue relevant with the trait
Host gene significance	The importance of the gene in pathways related to the trait
miRNA conservation	Conservation of the regulatory miRNA in different species
miRNA expression	Expression of the regulatory miRNA in tissues relevant with the trait
Binding site conservation	Conservation of the miRNA binding site in different species
expression-Quantitative Trait Loci	Correlation of 3'UTR variant with its host gene expression (cis-eQTL)
miRNA-target interaction	Evidence supporting the interaction between miRNA and target gene
Context score change	Predicted effect of binding site variant on the miRNA-target interaction

Supplemental Table S5. Association of putative and validated target genes of four identified miRNAs with AMD

miRNA ID	Putative targets (# SNPs)	Associated putative targets	Top SNP	GWAS p-value	Validated targets (# SNPs)	Associated targets
miR-3135b	2,325	*ARMS2*	rs10490924	1.86E-734	245	*C3*
	(220,120)	*CFH*	rs1061147	3.98E-590	(27,880)	
		C3	rs2230199	3.77E-69		
		PVRL2	rs6857	2.42E-33		
		TOMM40	rs2075650	4.73E-26		
		TIMP3	rs5754189	2.15E-14		
		CASP6	rs1800627	4.76E-11		
		C6orf15	rs2233956	4.83E-11		
		HERPUD1	rs9989419	3.30E-10		
		TGFBR1	rs1626340	3.82E-10		
		PILRA	rs7783550	5.62E-09		
miR-122-5p	1,563	*ZBTB41*	rs12724106	1.62E-111	438	*C3*
	(180,300)	*CRB1*	rs10754220	1.27E-99	(57,640)	*HLA-DQA1*
		C3	rs2230199	3.77E-69		
		SLC44A4	rs6907185	1.24E-43		
		HLA-DQA1	rs9272723	1.11E-13		
		CASP6	rs1800627	4.76E-11		
		RORB	rs10781180	2.60E-09		
		CLIC1	rs3131383	2.45E-08		
		SARM1	rs2027993	3.31E-08		
miR-3591-5p	1,174	*ZBTB41*	rs12724106	1.62E-111	81	*SERPINH1*
	(205,400)	*TIMP3*	rs5754227	1.07E-24	(13,157)	
		HERPUD1	rs9989419	3.30E-10		
		TBC1D23	rs2289509	8.19E-09		
		TNFAIP1	rs2073867	2.80E-08		
		FAM124B	rs2098533	2.21E-07		
miR-4513	861	*TIMP3*	rs5754227	1.07E-24	58	*FAM126B*
	(134,490)	*COL8A1*	rs1610258	6.78E-09	(10,265)	
		TNFRSF10A	rs11777697	8.76E-09		
		RLBP1	rs3825991	1.90E-07		
		CDKN2A	rs3218020	3.70E-07		

Shown are target genes of the four identified miRNAs that are associated with AMD. The Bonferroni correction (based on the number of SNPs in target genes of each miRNA) was used to set the significance levels. The SNPs shown in the table are the top SNP in the miRNA target genes associated with AMD using the AMD GWAS (Fritsche et al. 2016). TargetScan v7.0 (Total context++ score > 0.1) was used to extract the predicted target genes of each miRNA and miRTarBase v6.0 was used to extract the validated miRNA-target interactions based on the CLIP-seq data.

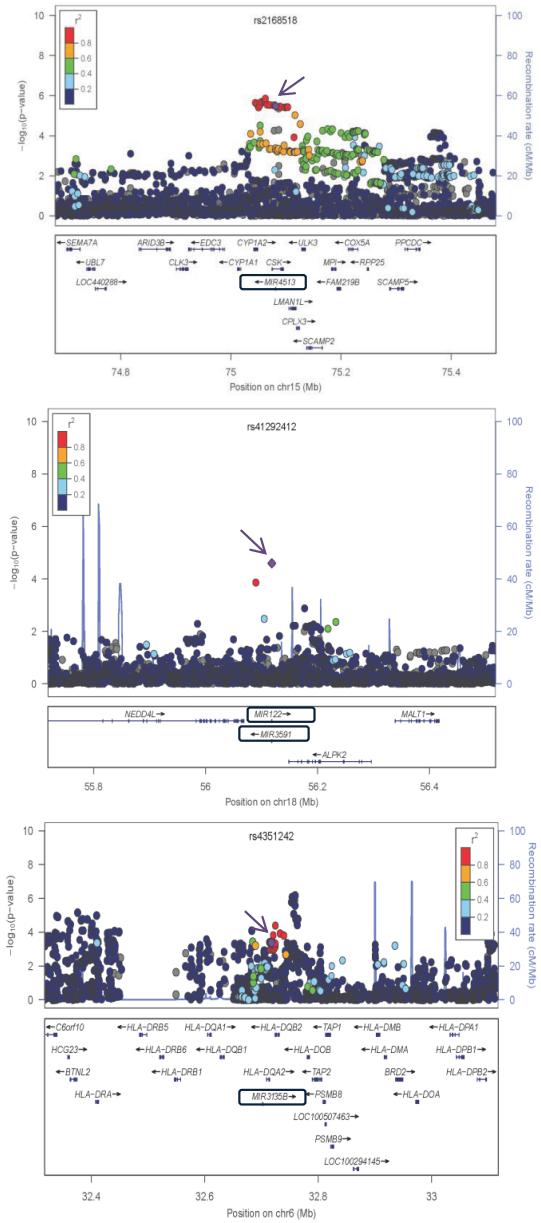
Supplemental Table S11. Prioritization of the 54 miRNA-binding site variants associated with AMD based on different criteria

SNP ID	GWAS Associ.	eQTL	Proxy SNPs (non-syn)	Gene ID	Gene Exp & Sign	Conserved miRNA-BS disruption / creation	miRNAs Exp.	Score
rs6857	✓✓	✓✓	0	PVRL2	✓✓	miR-645/ miR-320e	✓✓	10
rs3217992	✓✓	✓✓	5 (0)	CDKN2B	✓✓	miR-138-3p/ miR-374c	✓✓	9
rs10119	✓✓	✓	0	TOMM40	✓✓	-/ miR-516b-5p	✓	9
rs73203613	✓✓	-	2 (0)	BRAP	✓✓	miR-582-5p/-	✓✓	8
rs4151672	✓✓	✓	20 (3)	CFB	✓✓	miR-210-5p/ miR-188	✓✓	8
rs7674356	✓✓	✓✓	20 (0)	CASP6	✓✓	miR-548at-5p/ miR-628	✓✓	8
rs3087680	✓✓	✓✓	1 (0)	CNN2	✓	miR-378a-5p/ miR-200b	-	8
rs201075702	✓✓	-	0	DENND1B	✓	-/ miR-153-3p	✓✓	8
rs705866	✓✓	✓✓	60 (0)	PILRB	✓✓	miR-374-5p/-	✓✓	8
rs760070	✓✓	✓✓	15 (2)	RDBP	✓✓	miR-548an/ miR-552-5p	✓✓	8
rs334348	✓✓	-	16 (0)	TGFBR1	✓✓	miR-628-5p/ miR-32-3p	✓	8
rs12724106	✓✓	-	0	ZBTB41	✓	miR-551b-5p/ -	✓	8
rs1045216	✓	✓✓	20 (0)	PLEKHA1	✓✓	-/miR-604, miR-647	✓✓	8
rs1055890	✓✓	✓✓	1 (0)	HLA-B	✓✓	-/ miR-6804-5p	-	7
rs2239907	✓	✓✓	20 (0)	SARM1	✓	miR-1285-5p	-	7
rs7405453	✓	✓✓	20 (0)	TSPAN10	✓	miR665/ -	-	7
rs45529339	✓	-	6 (0)	ATP6V1G2	✓✓	miR-520d-5p/ miR-340	✓✓	6
rs12140421	✓	-	1 (0)	CFHR3	✓✓	miR-510/ -	✓✓	6
rs1073	✓	✓	60 (0)	MEPCE	✓	- / miR-340-3p,-942-5p	✓	6
rs17103541	-	✓✓	30 (0)	PLEKHA1	✓✓	miR-371-5p/ -	✓	6
rs522162	✓	✓✓	15 (3)	RDBP	✓✓	-	-	5
rs1128161	✓	-	20 (0)	SLC46A1	✓✓	miR-567/ miR-148,-152	✓✓	6
rs868	✓	-	100 (0)	TGFBR1	✓✓	miR-192/ -	✓✓	5
rs4639796	✓	-	20 (2)	ZBTB41	✓	miR-101/ miR-875	✓✓	5
rs12028827	✓	-	20 (2)	ZBTB41	✓	- / miR-452-3p	✓✓	5
rs2672602	-	✓✓	9 (1)	ARMS2	-	-/ miR-214	✓✓	4
rs390837	-	-	100 (0)	CFHR3	✓✓	miR-182-5p,miR-96/-	✓✓	4
rs9923834	✓	-	50 (0)	CHST6	✓	-/ miR-29b	✓✓	4
rs142641895	✓	-	0	CUX2	-	miR-660-5p/ -	-	4
rs150697472	-	-	0	CYP21A2	✓	-/ miR-455-3p	✓✓	4
rs1057151	✓	-	5 (0)	HLA-B	✓✓	miR-199/ miR-214	✓✓	4
rs11284139	-	✓✓	100 (0)	PILRB	✓	-/ miR-587	-	4
rs143086805	-	-	30 (0)	PLEKHA1	✓✓	-	-	4
rs1128162	✓	-	20 (0)	SLC46A1	✓✓	-/ miR-548an	-	4
rs2239908	✓	-	20 (0)	SLC46A1	✓	-/ miR-20a-3p	✓✓	4
rs2239911	✓	-	20 (0)	SLC46A1	✓	miR-92a-5p/ miR609	✓✓	4

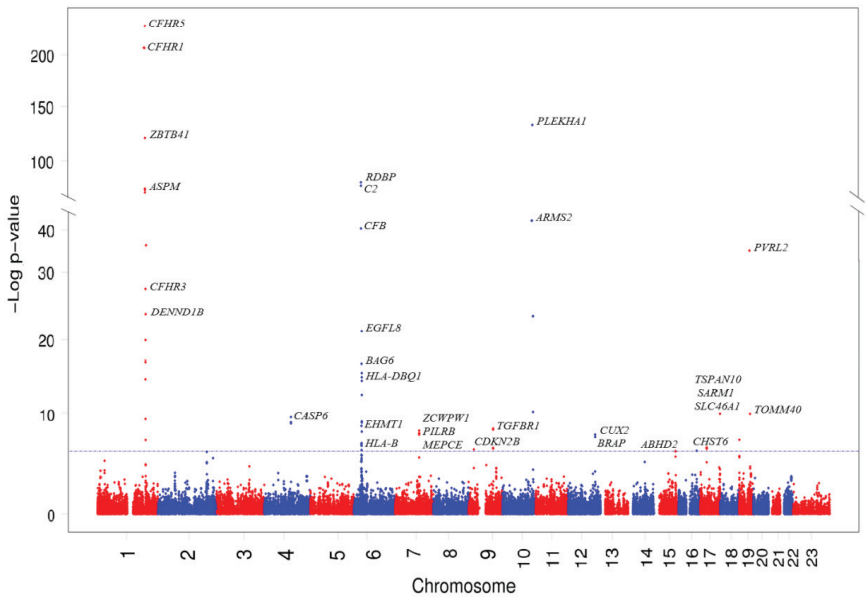
Supplemental Table S11. (continued)

SNP ID	GWAS Associ.	eQTL	Proxy SNPs (non-syn)	Gene ID	Gene Exp & Sign	Conserved miRNA-BS disruption / creation	miRNAs Exp.	Score
rs8079943	✓	-	20 (0)	SLC46A1	✓	miR-296-5p/ miR-361	✓✓	4
rs334349	✓	-	20 (0)	TGFBR1	✓✓	-	-	4
rs1590	✓	-	20 (0)	TGFBR1	✓✓	-	-	4
rs114550651	-	-	10 (0)	ZBTB41	✓	miR-526/ miR-516	✓✓	4
rs2672603	-	✓✓	9 (1)	ARMS2	-	miR-155-5p/miR-331	✓✓	3
rs12677	-	-	100 (5)	ASPM	✓	- / miR-602	✓	3
rs115247800	-	-	4 (0)	BAG6	✓✓	-	-	3
rs402372	-	-	100 (0)	CFHR3	✓✓	-	-	3
rs7383	✓	-	20 (0)	DMBT1	✓	-	-	3
rs1057387	✓	-	5 (0)	HLA-B	✓✓	-	-	3
rs41265225	-	-	0	ASPM	✓	-/ miR-548a-3p	-	2
rs390679	-	-	50 (0)	CFHR1	✓✓	-	-	2
rs10922153	✓	-	5 (0)	CFHR5	-	-	-	2
rs71131768	-	-	1 (0)	DENND1B	✓	-	-	2
rs71750332	-	-	0	EGFL8	✓	-/ miR127-5p	-	2
rs41267086	-	-	0	EHMT1	✓	-	-	1
rs1042891	-	-	100 (1)	CASP6	✓	-	-	1
rs3182325	-	-	100 (0)	CASP6	✓	-	-	1

Prioritization of miRNA-binding site SNPs associated with AMD based on the below criteria. A functional score (between 1 and 10) was calculated for the SNPs by combining the results of related criteria. 1. The strength of association in the GWAS results (✓✓: Lead SNP, ✓: One of the top 10% SNPs, -: Only passed the threshold). 2. eQTL (✓✓: Correlation between SNP and expression of the host gene, ✓: Correlation of SNPs with expression of nearby genes, -: No correlation). 3. Expression and importance of gene in the eye (✓✓: Abundantly expressed / Known gene, ✓: Expressed at low level/ Just association, -: Not expressed / No data). 4. Expression and importance of related miRNA in the eye (✓✓: Abundantly expressed / In vitro evidence, ✓: Expressed at low level / In silico data , -: Not expressed/ No data). In this table only conserved miRNAs are mentioned.



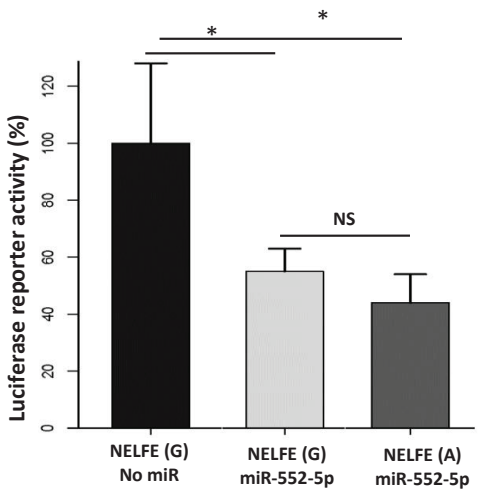
Supplemental Figure S1. Regional plots showing the association of three miRNA-variants with AMD. The *p*-values of variants are plotted (as ×log10 values) against their physical position on corresponding chromosomes. The p-values for the miRNA variants at the GWAS are represented by a purple diamond (depicted by arrow). Estimated recombination rates from the 1000 Genome project (European population) show the local LD structure. The variant's colors indicate LD with the miRNA variants according to a scale from r^2= 0 to r^2= 1 based on pairwise r2 values from the 1000 Genome project. The gene annotations from the UCSC genome browser are shown below.



Supplemental Figure S2. Manhattan plot showing genome-wide p-values of the association between miRNA-binding site variants and AMD. The plot shows the association between 82,051 3'UTR variants located in putative miRNA-binding sites and AMD using data from the recent GWAS of AMD [5]. The dashed line indicates the significance threshold of 6.1×10^{-7}. Genes hosting the binding site variants significantly associated with AMD are highlighted.

```
miR-552-5p          5′ GTTTAACCTTTTGCCTGTTGG 3′
                       |✕||||||     ||     |
 NELFE 3′UTR    3′ ATTTCCGAATTGGGGTTAGGGTTCACGACTT 5′

miR-552-5p          5′ GTTTAACCTTTTGCCTGTTGG 3′
                       ||||||||     ||     |
NELFE 3′UTR     3′ ATTTCCAAATTGGGGTTAGGGTTCACGACTT 5′
```



Supplemental Figure S5. miR-552-5p-mediated regulation of NELFE and the effect of rs760070 on the miRNA-target interaction. Luciferase reporter assays indicating miR-552-5p-mediated repression of NELFE. HEK293 cells were co-transfected with NELFE 3′UTR luciferase reporter vector containing the major allele G and (±) miR-552-5p mimic (15 nM). This experiment shows that the relative luciferase activity of the NELFE 3′UTR construct is significantly decreased in the presence of miR-552-5p. This experiment further showed that Luciferase reporter activity of the NELFE 3′UTR construct containing the rs760070 minor allele A is slightly (~10%), but non-significantly, reduced. * p-value < 0.05 compared with the negative control (no miR) (Students' t-test). NS, non-significant.

REFERENCES

1. Wong WL, Su X, Li X, Cheung CM, Klein R, Cheng CY, Wong TY: Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis. Lancet Glob Health 2014, 2:e106-116.
2. Lim LS, Mitchell P, Seddon JM, Holz FG, Wong TY: Age-related macular degeneration. Lancet 2012, 379:1728-1738.
3. Chen Y, Bedell M, Zhang K: Age-related macular degeneration: genetic and environmental factors of disease. Mol Interv 2010, 10:271-281.
4. Fritsche LG, Chen W, Schu M, Yaspan BL, Yu Y, Thorleifsson G, Zack DJ, Arakawa S, Cipriani V, Ripke S, et al: Seven new loci associated with age-related macular degeneration. Nat Genet 2013, 45:433-439, 439e431-432.
5. Fritsche LG, Igl W, Bailey JN, Grassmann F, Sengupta S, Bragg-Gresham JL, Burdon KP, Hebbring SJ, Wen C, Gorski M, et al: A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. Nat Genet 2016, 48:134-143.
6. Zhang F, Lupski JR: Non-coding genetic variants in human disease. Hum Mol Genet 2015, 24:R102-110.
7. Wagschal A, Najafi-Shoushtari SH, Wang L, Goedeke L, Sinha S, deLemos AS, Black JC, Ramirez CM, Li Y, Tewhey R, et al: Genome-wide identification of microRNAs regulating cholesterol and triglyceride homeostasis. Nat Med 2015, 21:1290-1297.
8. Ryan BM, Robles AI, Harris CC: Genetic variation in microRNA networks: the implications for cancer research. Nat Rev Cancer 2010, 10:389-402.
9. Ebert MS, Sharp PA: Roles for microRNAs in conferring robustness to biological processes. Cell 2012, 149:515-524.
10. Krol J, Loedige I, Filipowicz W: The widespread regulation of microRNA biogenesis, function and decay. Nat Rev Genet 2010, 11:597-610.
11. Shukla GC, Singh J, Barik S: MicroRNAs: Processing, Maturation, Target Recognition and Regulatory Functions. Mol Cell Pharmacol 2011, 3:83-92.
12. Bartel DP: MicroRNAs: target recognition and regulatory functions. Cell 2009, 136:215-233.
13. Ghanbari M, Darweesh SK, de Looper HW, van Luijn MM, Hofman A, Ikram MA, Franco OH, Erkeland SJ, Dehghan A: Genetic Variants in MicroRNAs and Their Binding Sites Are Associated with the Risk of Parkinson Disease. Hum Mutat 2016, 37:292-300.
14. Ghanbari M, Franco OH, de Looper H, Hofman A, Erkeland S, Dehghan A: Genetic Variations in miRNA Binding Sites Affect miRNA-Mediated Regulation of Several Genes Associated with Cardiometabolic Phenotypes. Circ Cardiovasc Genet 2015.
15. Gong J, Tong Y, Zhang HM, Wang K, Hu T, Shan G, Sun J, Guo AY: Genome-wide identification of SNPs in microRNA genes and the SNP effects on microRNA target binding and biogenesis. Hum Mutat 2012, 33:254-263.
16. Zhou Q, Gallagher R, Ufret-Vincenty R, Li X, Olson EN, Wang S: Regulation of angiogenesis and choroidal neovascularization by members of microRNA-23~27~24 clusters. Proc Natl Acad Sci U S A 2011, 108:8287-8292.
17. Murad N, Kokkinaki M, Gunawardena N, Gunawan MS, Hathout Y, Janczura KJ, Theos AC, Golestaneh N: miR-184 regulates ezrin, LAMP-1 expression, affects phagocytosis in human retinal pigment epithelium and is downregulated in age-related macular degeneration. FEBS J 2014, 281:5251-5264.

18. Szemraj M, Bielecka-Kowalska A, Oszajca K, Krajewska M, Gos R, Jurowski P, Kowalski M, Szemraj J: Serum MicroRNAs as Potential Biomarkers of AMD. Med Sci Monit 2015, 21:2734-2742.
19. Wang S, Koster KM, He Y, Zhou Q: miRNAs as potential therapeutic targets for age-related macular degeneration. Future Med Chem 2012, 4:277-287.
20. Gong J, Liu C, Liu W, Wu Y, Ma Z, Chen H, Guo AY: An update of miRNASNP database for better SNP selection by GWAS data, miRNA expression and online tools. Database (Oxford) 2015, 2015:bav029.
21. Bao L, Zhou M, Wu L, Lu L, Goldowitz D, Williams RW, Cui Y: PolymiRTS Database: linking polymorphisms in microRNA target sites with complex traits. Nucleic Acids Res 2007, 35:D51-54.
22. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ: LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 2010, 26:2336-2337.
23. Lorenz R, Bernhart SH, Honer Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL: ViennaRNA Package 2.0. Algorithms Mol Biol 2011, 6:26.
24. Meenhuis A, van Veelen PA, de Looper H, van Boxtel N, van den Berge IJ, Sun SM, Taskesen E, Stern P, de Ru AH, van Adrichem AJ, et al: MiR-17/20/93/106 promote hematopoietic cell expansion by targeting sequestosome 1-regulated pathways in mice. Blood 2011, 118:916-925.
25. Ghanbari M, de Vries PS, de Looper H, Peters MJ, Schurmann C, Yaghootkar H, Dorr M, Frayling TM, Uitterlinden AG, Hofman A, et al: A genetic variant in the seed region of miR-4513 shows pleiotropic effects on lipid and glucose homeostasis, blood pressure, and coronary artery disease. Hum Mutat 2014, 35:1524-1531.
26. Agarwal V, Bell GW, Nam JW, Bartel DP: Predicting effective microRNA target sites in mammalian mRNAs. Elife 2015, 4.
27. Chou CH, Chang NW, Shrestha S, Hsu SD, Lin YL, Lee WH, Yang CD, Hong HC, Wei TY, Tu SJ, et al: miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. Nucleic Acids Res 2016, 44:D239-247.
28. Ghanbari M, Franco OH, de Looper HW, Hofman A, Erkeland SJ, Dehghan A: Genetic Variations in MicroRNA-Binding Sites Affect MicroRNA-Mediated Regulation of Several Genes Associated With Cardio-metabolic Phenotypes. Circ Cardiovasc Genet 2015, 8:473-486.
29. Li Y, Qiu C, Tu J, Geng B, Yang J, Jiang T, Cui Q: HMDD v2.0: a database for experimentally supported human microRNA and disease associations. Nucleic Acids Res 2014, 42:D1070-1074.
30. Ruepp A, Kowarsch A, Schmidl D, Buggenthin F, Brauner B, Dunger I, Fobo G, Frishman G, Montrone C, Theis FJ: PhenomiR: a knowledgebase for microRNA expression in diseases and biological processes. Genome Biol 2010, 11:R6.
31. Kozomara A, Griffiths-Jones S: miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res 2014, 42:D68-73.
32. Krol J, Sobczak K, Wilczynska U, Drath M, Jasinska A, Kaczynska D, Krzyzosiak WJ: Structural features of microRNA (miRNA) precursors and their relevance to miRNA biogenesis and small interfering RNA/short hairpin RNA design. J Biol Chem 2004, 279:42230-42239.
33. Nicoloso MS, Sun H, Spizzo R, Kim H, Wickramasinghe P, Shimizu M, Wojcik SE, Ferdin J, Kunej T, Xiao L, et al: Single-nucleotide polymorphisms inside microRNA target sites influence tumor susceptibility. Cancer Res 2010, 70:2789-2798.
34. Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP: MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol Cell 2007, 27:91-105.
35. Cammaerts S, Strazisar M, De Rijk P, Del Favero J: Genetic variants in microRNA genes: impact on microRNA expression, function, and disease. Front Genet 2015, 6:186.

36. Pritchard CC, Cheng HH, Tewari M: MicroRNA profiling: approaches and considerations. Nat Rev Genet 2012, 13:358-369.
37. Mencia A, Modamio-Hoybjor S, Redshaw N, Morin M, Mayo-Merino F, Olavarrieta L, Aguirre LA, del Castillo I, Steel KP, Dalmay T, et al: Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss. Nat Genet 2009, 41:609-613.
38. Lingzi X, Zhihua Y, Xuelian L, Yangwu R, Haibo Z, Yuxia Z, Baosen Z: Genetic variants in microRNAs predict non-small cell lung cancer prognosis in Chinese female population in a prospective cohort study. Oncotarget 2016, 7:83101-83114.
39. Zhang L, Ye Y, Tu H, Hildebrandt MA, Zhao L, Heymach JV, Roth JA, Wu X: MicroRNA Related Genetic Variants in Iron Regulatory Genes, Dietary Iron Intake, MicroRNAs and Lung Cancer Risk. Ann Oncol 2017.
40. Zhang X, Zeng Y: The terminal loop region controls microRNA processing by Drosha and Dicer. Nucleic Acids Res 2010, 38:7689-7697.
41. Han J, Lee Y, Yeom KH, Nam JW, Heo I, Rhee JK, Sohn SY, Cho Y, Zhang BT, Kim VN: Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. Cell 2006, 125:887-901.
42. Xiong X, Kang X, Zheng Y, Yue S, Zhu S: Identification of loop nucleotide polymorphisms affecting microRNA processing and function. Mol Cells 2013, 36:518-526.
43. Wu M, Jolicoeur N, Li Z, Zhang L, Fortin Y, L'Abbe D, Yu Z, Shen SH: Genetic variations of microRNAs in human cancer and their effects on the expression of miRNAs. Carcinogenesis 2008, 29:1710-1716.
44. Piskounova E, Viswanathan SR, Janas M, LaPierre RJ, Daley GQ, Sliz P, Gregory RI: Determinants of microRNA processing inhibition by the developmentally regulated RNA-binding protein Lin28. J Biol Chem 2008, 283:21310-21314.
45. Luo C, Chen M, Xu H: Complement gene expression and regulation in mouse retina and retinal pigment epithelium/choroid. Mol Vis 2011, 17:1588-1597.
46. Mullins RF, Russell SR, Anderson DH, Hageman GS: Drusen associated with aging and age-related macular degeneration contain proteins common to extracellular deposits associated with atherosclerosis, elastosis, amyloidosis, and dense deposit disease. FASEB J 2000, 14:835-846.
47. Nozaki M, Raisler BJ, Sakurai E, Sarma JV, Barnum SR, Lambris JD, Chen Y, Zhang K, Ambati BK, Baffi JZ, Ambati J: Drusen complement components C3a and C5a promote choroidal neovascularization. Proc Natl Acad Sci U S A 2006, 103:2328-2333.
48. Persson H, Kvist A, Rego N, Staaf J, Vallon-Christersson J, Luts L, Loman N, Jonsson G, Naya H, Hoglund M, et al: Identification of new microRNAs in paired normal and tumor breast tissue suggests a dual role for the ERBB2/Her2 gene. Cancer Res 2011, 71:78-86.
49. Sun G, Yan J, Noltner 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.
50. Tsai WC, Hsu SD, Hsu CS, Lai TC, Chen SJ, Shen R, Huang Y, Chen HC, Lee CH, Tsai TF, et al: MicroRNA-122 plays a critical role in liver homeostasis and hepatocarcinogenesis. J Clin Invest 2012, 122:2884-2897.
51. Bihrer V, Friedrich-Rust M, Kronenberger B, Forestier N, Haupenthal J, Shi Y, Peveling-Oberhag J, Radeke HH, Sarrazin C, Herrmann E, et al: Serum miR-122 as a biomarker of necroinflammation in patients with chronic hepatitis C virus infection. Am J Gastroenterol 2011, 106:1663-1669.
52. Szabo G, Bala S: MicroRNAs in liver disease. Nat Rev Gastroenterol Hepatol 2013, 10:542-552.
53. Yates JR, Sepp T, Matharu BK, Khan JC, Thurlby DA, Shahid H, Clayton DG, Hayward C, Morgan J, Wright AF, et al: Complement C3 variant and the risk of age-related macular degeneration. N Engl J Med 2007, 357:553-561.

54. Fu X, Dong B, Tian Y, Lefebvre P, Meng Z, Wang X, Pattou F, Han W, Wang X, Lou F, et al: MicroRNA-26a regulates insulin sensitivity and metabolism of glucose and lipids. J Clin Invest 2015, 125:2497-2509.
55. Keklikoglou I, Hosaka K, Bender C, Bott A, Koerner C, Mitra D, Will R, Woerner A, Muenstermann E, Wilhelm H, et al: MicroRNA-206 functions as a pleiotropic modulator of cell proliferation, invasion and lymphangiogenesis in pancreatic adenocarcinoma by targeting ANXA2 and KRAS genes. Oncogene 2015, 34:4867-4878.
56. Chatterjee S, Pal JK: Role of 5'- and 3'-untranslated regions of mRNAs in human diseases. Biol Cell 2009, 101:251-262.
57. Skeeles LE, Fleming JL, Mahler KL, Toland AE: The impact of 3'UTR variants on differential expression of candidate cancer susceptibility genes. PLoS One 2013, 8:e58609.
58. Saba R, Medina SJ, Booth SA: A functional SNP catalog of overlapping miRNA-binding sites in genes implicated in prion disease and other neurodegenerative disorders. Hum Mutat 2014, 35:1233-1248.
59. Bulik-Sullivan B, Selitsky S, Sethupathy P: Prioritization of genetic variants in the microRNA regulome as functional candidates in genome-wide association studies. Hum Mutat 2013, 34:1049-1056.
60. Hou L, Zhao H: A review of post-GWAS prioritization approaches. Front Genet 2013, 4:280.
61. Hormozdiari F, Kostem E, Kang EY, Pasaniuc B, Eskin E: Identifying causal variants at loci with multiple signals of association. Genetics 2014, 198:497-508.
62. Yang H, Wang K: Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. Nat Protoc 2015, 10:1556-1566.
63. Ramdas WD, van Koolwijk LM, Lemij HG, Pasutto F, Cree AJ, Thorleifsson G, Janssen SF, Jacoline TB, Amin N, Rivadeneira F, et al: Common genetic variants associated with open-angle glaucoma. Hum Mol Genet 2011, 20:2464-2471.
64. Smailhodzic D, Klaver CC, Klevering BJ, Boon CJ, Groenewoud JM, Kirchhof B, Daha MR, den Hollander AI, Hoyng CB: Risk alleles in CFH and ARMS2 are independently associated with systemic complement activation in age-related macular degeneration. Ophthalmology 2012, 119:339-346.
65. Nashine S, Chwa M, Kazemian M, Thaker K, Lu S, Nesburn A, Kuppermann BD, Kenney MC: Differential Expression of Complement Markers in Normal and AMD Transmitochondrial Cybrids. PLoS One 2016, 11:e0159828.
66. Crispo E, Chapman LJ: Population genetic structure across dissolved oxygen regimes in an African cichlid fish. Mol Ecol 2008, 17:2134-2148.
67. Harboe M, Mollnes TE: The alternative complement pathway revisited. J Cell Mol Med 2008, 12:1074-1084.
68. Xu S, Witmer PD, Lumayag S, Kovacs B, Valle D: MicroRNA (miRNA) transcriptome of mouse retina and identification of a sensory organ-specific miRNA cluster. J Biol Chem 2007, 282:25053-25066.
69. Ragusa M, Caltabiano R, Russo A, Puzzo L, Avitabile T, Longo A, Toro MD, Di Pietro C, Purrello M, Reibaldi M: MicroRNAs in vitreus humor from patients with ocular diseases. Mol Vis 2013, 19:430-440.
70. Liu SC, Chuang SM, Hsu CJ, Tsai CH, Wang SW, Tang CH: CTGF increases vascular endothelial growth factor-dependent angiogenesis in human synovial fibroblasts by increasing miR-210 expression. Cell Death Dis 2014, 5:e1485.
71. Huang X, Ding L, Bennewith KL, Tong RT, Welford SM, Ang KK, Story M, Le QT, Giaccia AJ: Hypoxia-inducible mir-210 regulates normoxic gene expression involved in tumor initiation. Mol Cell 2009, 35:856-867.
72. Ricklin D, Lambris JD: Complement-targeted therapeutics. Nat Biotechnol 2007, 25:1265-1275.

CHAPTER 4.2

Association of miRNA-related genetic variants with primary open-angle glaucoma

Mohsen Ghanbari, Adriana I. Iglesias, Henriët Springelkamp, International Glaucoma Genetics Consortium (IGGC), Arfan Ikram, Cornelia van Duijn, Abbas Dehghan, Stefan J. Erkeland, Caroline C.W. Klaver, Magda A. Meester-Smoor

Manuscript based on this chapter

A genome-wide scan for microRNA-related genetic variants associate with primary open-angle glaucoma

2017, in prepration.

ABSTRACT

Background. Primary open-angle glaucoma (POAG), the most common optic neuropathy, is a complex disease that results from multiple genetic and environmental factors. MicroRNAs (miRNAs) are small non-coding RNA molecules that post-transcriptionally regulate target mRNAs and are frequently implicated in human disease. Genetic variants in miRNAs as well as miRNA-binding sites within gene 3'-untranslated regions (3'UTRs) are expected to contribute to disease risk.

Methods and Results. We investigated the association of genetic variants in miRNAs and miRNA-binding sites with POAG endophenotypes using data from the recent genome-wide association studies (GWAS) of intraocular pressure (IOP), vertical cup-to-disc ratio (VCDR), optic nerve cup and disc area. Two out of 412 variants in miRNA genes available in the GWAS data were significantly associated with VCDR and cup area. The first variant (rs12803915:A/G, chr11:65444508) is located in the terminal loop of pre-miR-612 and has been demonstrated to increase the miRNA expression. The second variant (rs2273626:A/C, chr14:22956973) is located in the seed sequence of miR-4707. Our experiments showed that the variant reduces the interaction between miR-4707-3p and *CARD10*, a known gene associated with VCDR and disc area. Moreover, we investigated 72,052 variants in miRNA-binding sites, of which, 47 variants in the 3'UTR of 21 genes were significantly associated with one or more POAG endophenotypes. The associated binding site variants were subsequently prioritized based on defined criteria for being likely functional. Two of the prioritized variants, rs3217992 and rs1063192, within the CDKN2B 3'UTR have been demonstrated to affect miR-138-3p and miR-323b-5p mediated regulation of CDKN2B.

Conclusion. Our results support the idea that miRNAs may contribute to glaucoma endophenotypes.

INTRODUCTION

Primary open-angle glaucoma (POAG), the most common optic neuropathy, is the leading cause of irreversible blindness, affecting around 60 million individuals worldwide [1, 2]. The disease is characterized by progressive loss of retinal ganglion cells and optic nerve degeneration that can be secondary to elevated intraocular pressure (IOP) [3]. The optic nerve damage is characterized by an increase in cup size which is the central area of the optic disc. This damage can be measured by the vertical cup-to-disc ratio (VCDR), comparing the vertical diameter of the cup with vertical dimeter of the total optic disc [4]. The VCDR ranges from 0 to 1, whereas the ratio above 0.7 or an asymmetry between eyes above 0.2 are considered as suspect for glaucoma in the clinical setting [5]. POAG is presumed to be a complex progressive neurodegenerative disorder caused by multiple genetic as well as environmental factors [2]. Previous genome-wide association studies (GWAS) have revealed a number of susceptibility loci for POAG by studying the disease directly or its endophenotypes including IOP and optic disc parameters (VCDR, cup and disc area) [5, 6]. Majority of the associated variants identified by GWAS are located in non-coding regions of the genome and their contributions to POAG and endophenotypes are poorly understood.

There are strong indications that microRNAs (miRNAs), small non-coding RNAs consisting of 19-22 nucleotides, play a role in the pathogenesis of glaucoma [7-9]. For example, miR-29b and miR-24 that are involved in gene regulation in trabecular meshwork cells [10] or miR-200c, miR-204 and miR-182 that are reported as potential diagnostic biomarkers or therapeutic targets for glaucoma [9, 11]. The biogenesis of miRNAs is a multistep coordinated process and genetic variants in miRNA-encoding sequences can have profound effects on miRNA expression and function [12, 13]. In the nucleus, miRNA genes are initially transcribed as long primary transcripts (pri-miRNA). Further processing and cleavage by the RNase Drosha and Dicer enzymes generate mature miRNAs. The mature miRNAs are subsequently incorporated into the RNA induced silencing complex (RISC) to interact with the 3'-untranslated region (3'UTR) of target mRNAs, resulting in mRNA degradation or translational repression [13, 14]. Genetic variants in miRNAs or miRNA-binding sites within the 3'UTR of target genes are expected to affect miRNA-mediated gene regulation [15-19]. We and others have previously shown that such variants contribute to disease risk [18-21]. In this study, we aimed to identify miRNAs associated with POAG endophenotypes using genetic data. To this end, we used data from the recent GWAS meta-analysis on IOP and optic disc parameters [5], and investigated the association of genetic variants in miRNAs as well as in miRNA-binding sites with POAG endophenotypes. Further, we performed *in silico* and *in vitro* validation studies to determine the functionality of identified variants in relation to POAG.

MATERIAL AND METHODS

Genome-wide association studies on glaucoma endophenotypes

To examine the association of miRNA-related genetic variants with POAG endophenotypes, we used data from the recent GWAS meta-analyses on IOP (n=37,930), VCDR (n=23,899), cup area (n=22,489) and disc area (n=22,504) from individuals with European and Asian ancestry within the International Glaucoma Genetics Consortium (IGGC) [5]. Characteristics of the IGGC has been described elsewhere [5]. These data include approximately 8 million single-nucleotide polymorphisms (SNPs) with minor allele frequency (MAF) > 0.01 and using imputation to the 1000 Genomes (1000G) reference panel.

Identification of genetic variants in miRNAs and miRNA-binding sites

The miRNASNP (v2) online database was used to retrieve variants located in miRNA-encoding sequences [22]. A total number of 2,420 variants in precursor and mature miRNA sequences were retrieved. Of these, we analyzed 412 variants (in 333 miRNAs) that were available in the GWAS data for association with glaucoma endophenotypes. Moreover, we extracted almost 401,000 miRNA-binding site variants (including SNPs and INDELs) in gene 3'UTRs using PolymiRTS online databases (v3) [23]. Of which, 72,052 miRNA-binding site variants available in the GWAS data were included for our analysis. The Bonferroni correction was used to adjust the p-value for the number of variants in each group. The significance threshold was set on 1.21×10^{-4} (0.05 / 412) for variants in miRNAs and 6.94×10^{-7} (0.05 / 72,052) for variants in miRNA-binding sites (**Figure 1**). Regional association plots showing the association of miRNA-related variants with the studied phenotypes were made using LocusZoom tool [24].

The effect of miRNA-variant on the miRNA secondary structure and expression

The secondary structure of pri-miRNA is critical for the miRNA biogenesis [25]. The Vienna RNAfold algorithm (ViennaRNA package 2.0) was used to predict the impact of variants in miRNAs on the hairpin stem-loop structure of pre-miRNAs [26]. This program calculates the changes in minimum free energy (MFE) of the thermodynamic ensemble of the hairpin structure of miRNA (wild type and mutant). Next, to test the functional impact of a miRNA variant on the expression level of the mature miRNA, we cloned the pre-miRNA sequence containing either major or minor allele behind the gene encoding green fluorescent protein (GFP) in the expression plasmid MSCV-BC (Murine Stem Cell Virus-Bar Coded), resulting in GFP-miRNA fusion transcripts as described previously [27]. For cloning purposes, the restriction enzyme sites XhoI and EcoRI were inserted in respectively forward and reverse primers. The primers are listed in **Table S1**. The inserts of all constructs were validated by Sanger sequencing. HEK293 cell transfection, total

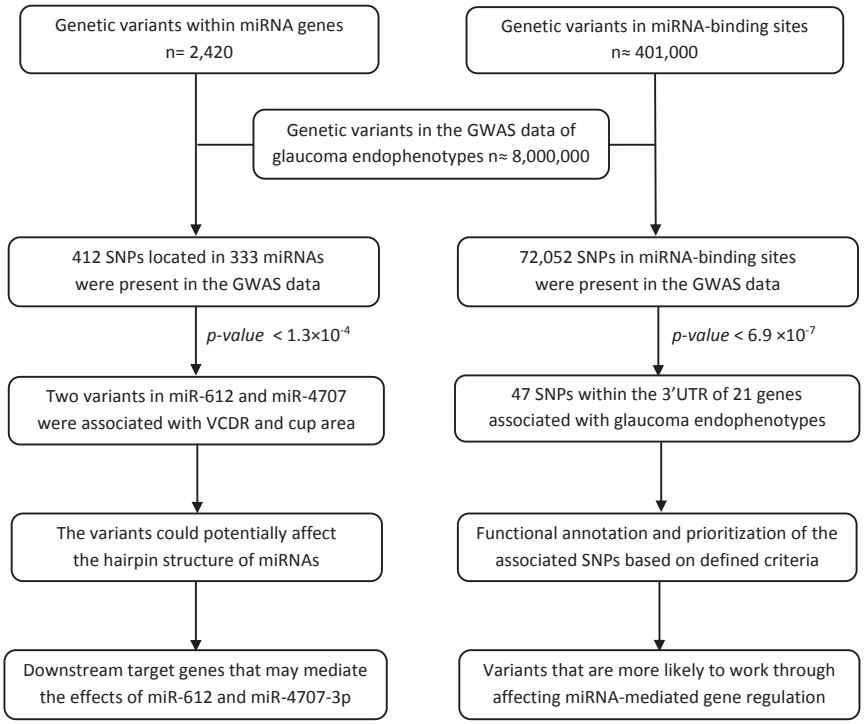


Figure 1. A schematic workflow of our analyses to identify miRNA-related variants associated with glaucoma.

RNA isolation and quantitative PCRs of EGFP and miRNAs were performed as previously described [27]. The expression levels of mutant and wild type miRNA relative to EGFP was calculated. All experiments were performed in triplicate, and repeated at least three times.

Association of target genes of the identified miRNAs with glaucoma endophenotypes

The biological role of miRNAs is dictated through regulating expression of their target mRNAs. We aimed to identify target genes that may mediate the downstream effect of the associated miRNAs in relation to POAG endophenotypes. Since such genes should be associated with the phenotype, we used the GWAS data in a candidate gene approach and searched among the putative target genes of a miRNA for those that were associated with POAG endophenotypes. We extracted the miRNA putative target genes using TargetScan v7.0 (total context++ score > 0.1) (http://www.TargetScan. org/) [28]. The association of variants located in these target genes with the glaucoma endophenotypes were subsequently examined using the GWAS data [5]. The Bonferroni correction

was used to calculate a significance level based on the number of tested variants in the target genes of each miRNA.

Interaction between a miRNA and its target genes using the Rotterdam Study data

The external validity of association between a miRNA variant and associated phenotype improves when the variant affects the miRNA target genes which are known to be involved in the phenotype. We tested the interaction between the associated miRNA variants and the most significant variant in the miRNA target genes in relation to POAG endophenotype using the Rotterdam Study data. The design of the Rotterdam Study has been described in detail elsewhere [19]. The baseline characteristics of the subjects are shown in **Table S2**. To examine the interaction, we introduced the product of miRNA variant and the target gene variant in the logistic regression model, adjusted for age and sex: POAG endophenotype ~ age + sex + miRNA variant + target gene variant + (miRNA variant × target gene variant).

Luciferase reporter assays

Luciferase reporter assay system was used to examine the binding of a miRNA to the 3'UTR of its target gene. In addition, this experiment was used to determine the impact of a miRNA variant in the interaction between miRNA-target gene. Primers were designed to amplify the 3'UTR sequence of target gene and included restriction enzyme sites *XbaI* for the forward primer and *ApaI* for the reverse. The 3'UTR sequences (wild type and mutated), containing putative binding site of the miRNA, were amplified and cloned into the pGL3 Luciferase reporter vector (Promega) downstream of the Luciferase open reading frame[27]. All the primers are shown in **Table S3**. The inserts of all constructs were confirmed by Sanger sequencing. HEK293 cells (n=10,000) were plated into 96-well plates and co-transfected with 1µg of pGL3 containing the 3'UTR with either the major or minor allele, miRNA mimic (mirVanaTM Mimics) and a plasmid expressing the Renilla Luciferase which served as transfection control, with Lipofectamine^{R} RNAiMAX (Invitrogen). Luciferase activity was measured with the Dual-Glo Luciferase Assay System according to manufacturer's protocol (Promega). Renilla activity was used for normalization of the data. All experiments were performed in triplicate and repeated three independent experiments.

Functional annotation of miRNA-binding site variants associated with glaucoma endophenotypes

Specific criteria have been suggested to establish whether 3'UTR variants located in miRNA-binding sites are likely to be functional [16, 21]. These include a strong association between the variant and the phenotype, the expression of the host gene and

regulatory miRNA in a relevant tissue, an allele-specific expression of the host gene. We prioritized miRNA-binding site variants associated with POAG endophenotypes using these criteria to identify the variants that are more likely to affect miRNA-mediated gene regulation. To this end, we retrieved all proxies in high linkage disequilibrium (LD) (R^2 threshold > 0.8, limit distance 100 kb, and population panel CEU) with the identified variants and checked their effects on protein structure, gene regulation and splicing using the HaploRegv4.1 (http://www.broadinstitute.org/mammals/haploreg/haploreg.php). The HaploReg 4.1 was used to study the LD variants and the functional potential of the binding site variants associated with glaucoma endophenotypes. To scan the correlation between the identified variants and expression levels of the host transcripts, we used GTExV6 (http://www.gtexportal.org/home/) and HaploReg v4.1 which provide information on expression quantitative trait loci (cis-eQTL) in different tissues. To check whether miRNAs are expressed in human eye and other tissues, several databases were screened, including miRmine (http://guanlab.ccmb.med.umich.edu/mirmine/help.html), HMDD [29], PhenomiR [30] and miR2disease [31]. Other miRNA information, including miRNA conservation in different species was obtained from miRBase (release 20) [32]. In addition, the Ocular Tissue Database (https://genome.uiowa.edu/otdb/) was used to examine the expression of miRNA target genes across eye tissues. **Table S4** shows a list of web tools and databases that we used for our analyses.

RESULTS

Two miRNA variants were associated with POAG endophenotypes

We examined the association of 412 miRNA variants (in 333 miRNAs) available in the GWAS data with IOP, VCDR, cup and disc area (**Figure 1**). Two miRNA variants passed the Bonferroni corrected significance threshold of 1.3×10^{-4}. These variants include rs12803915:A/G (Chr11:65444508) in the pre-miR-612 sequence associated with VCDR (p-value= 4.6×10^{-9}, beta= -0.009) and cup area (p-value= 1.2×10^{-7}, beta= -0.014) and rs2273626:A/C (Chr14:22956973) in the seed region of miR-4707 associated with VCDR (p-value= 9.5×10^{-5}, beta= -0.005) and cup area (p-value= 9.9×10^{-5}, beta= -0.008). **Table 1** shows three miRNA-variants with the most significant association with each of the studied phenotypes.

The impact of miRNA variants on the miRNA secondary structure and expression

The impact of miRNA variants on biogenesis of the miRNAs measured by performing structural analysis of the miRNA hairpin using Vienna RNAfold algorithm [26]. This *in silico* analysis showed 0.8 kcal/mol difference for pre-miR-612 and 5.8 kcal/mol difference for

pre-miR-4707 in MFE of the thermodynamic ensemble of the miRNA hairpin structure, suggesting that the variants may affect the miRNA maturation (**Figure 2**). Previously, the minor allele A of rs12803915 in pre-miR-612 sequence has been experimentally demonstrated to increasethe expression of mature miR-612 [33]. For rs2273626, we cloned miR-4707 precursor sequence either with the major allele C or the minor allele A behind

Table 1. Top variants in miRNAs associated with glaucoma endophenotypes

Phenotype	SNP ID	A1/ A2	MAF	miRNA ID	SNP position	MFE	*p*-value	Beta	SE
VCDR	rs12803915	A/G	0.14	miR-612	Stem region	0.8	**4.6×10^{-9}**	-0.009	0.0015
	rs2273626	A/C	0.48	miR-4707	Seed	5.8	**9.5×10^{-5}**	-0.005	0.0012
	rs58834075	C/A	0.09	miR-656	Loop	0.6	2.2×10^{-3}	-0.014	0.0047
Cup area	rs12803915	A/G	0.2	miR-612	Stem region	0.8	**1.2×10^{-7}**	-0.014	0.0028
	rs2273626	A/C	0.5	miR-4707	Seed	5.8	**9.9×10^{-5}**	-0.008	0.0021
	rs13299349	A/G	0.22	miR-3152	Mature	4.9	1.9×10^{-4}	-0.009	0.0024
Disc area	rs76118570	A/T	0.02	miR-6130	Pre-mature	0.0	5.4×10^{-4}	-0.288	0.0831
	rs72631820	A/G	0.015	miR-339	Mature	0.6	3.2×10^{-3}	0.143	0.0485
	rs12512664	A/G	0.61	miR-4274	Pre-mature	-0.3	4.6×10^{-3}	-0.013	0.005
IOP	rs2289030	C/G	0.11	miR-492	Pre-mature	1.7	4.2×10^{-3}	0.12	0.042
	rs515924	A/G	0.18	miR-548al	Seed	0.0	4.5×10^{-3}	0.086	0.03
	rs9877402	A/G	0.13	miR-5682	Pre-mature	-1.4	5.0×10^{-3}	0.156	0.056

Shown are three variants in miRNAs with the most significant association with each of the four glaucoma endophenotypes. MAF, minor allele frequency; Chr, Chromosome; A, allele; MFE, Change in minimum free energy; SE, Standard error. Significant associations are shown in bold.

GFP in the pMSCV-BC vector. HEK293 cells were transfected with the vectors expressing transcripts with EGFP and pre-miRNA. We observed no significant difference between mature miR-4707 expression levels in cells transfected with pre-miRNA sequences containing the minor allele compared to the major allele (**Figure S1**). This suggests that the effect of rs2273626 on the miR-4707 biogenesis is minor and not detectable with this experimental setup.

Target genes of miR-612 and miR-4707 associated with VCDR and cup area

In order to identify target genes that may mediate the downstream effect of miR-612 and miR-4707, we studied the association of genetic variants in their putative target genes with VCDR and cup area using the GWAS data [5]. **Table S5** shows miR-612 and miR-4707 target genes that are associated with these phenotypes. **Table S6** displays expression of these genes across eye tissues using the Ocular Tissue Database. Using the Rotterdam Study data we further tested the interaction between the variants in miR-612 and miR-4707 and the top variant (with the most significant association) in their associated target

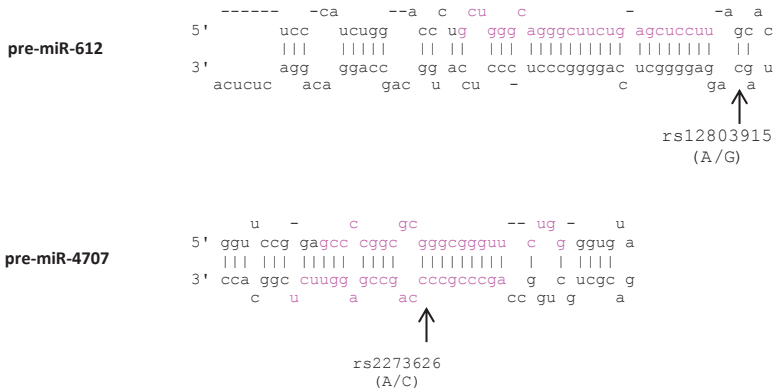


Figure 2. The predicted hairpin structure of pre-miR-612 and pre-miR-4707 and the position of two identified variants associated with VCDR and cup area.

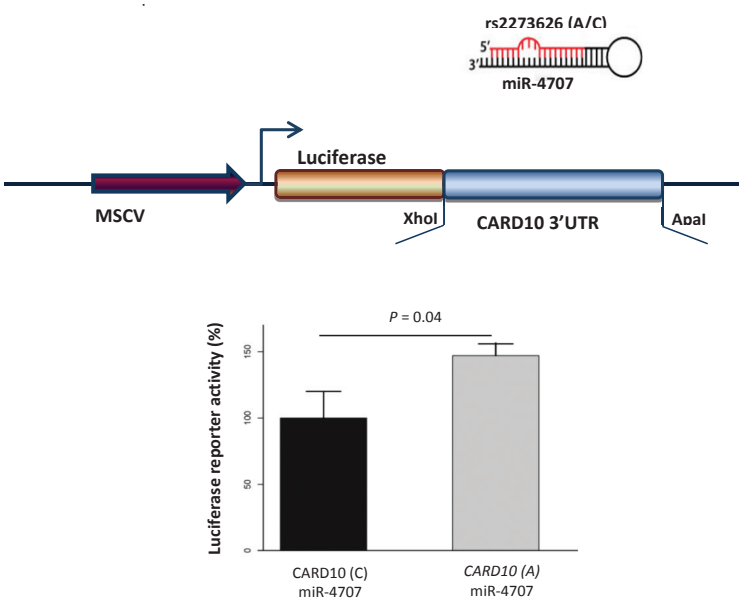


Figure 3. miR-4707-mediated regulation of *CARD10* and the effect of rs2273626 seed variant on the miRNA-mRNA interaction.

genes in relation to VCDR and cup area. This analysis showed a significant interaction between rs2273626 in miR-4707 and the top variant in *CARD10* (**Table S7**). Further, Luciferase reporter assays demonstrated that miR-4707-3p down-regulates the expression of CARD10 (**Figure 3**). This experiment also showed a significant difference between miR-4707-3p mediated regulation of *CARD10* reporters containing the major allele C and the

Table 2. Characteristics of the top ten miRNA-binding site variants associated with glaucoma endophenotypes

	rs1063192	rs1132554	rs10216533	rs1057001	rs6660601	rs7344	rs1052990	rs1050285	rs1043207	rs8176751
GWAS result	CA	CA	VCDR	CA	IOP	CA	IOP	IOP	VCDR	IOP
p-value	5.3×10^{-33}	5.5×10^{-7}	2.8×10^{-8}	2.3×10^{-7}	4.2×10^{-11}	4.0×10^{-7}	1.9×10^{-9}	2.9×10^{-7}	6.5×10^{-7}	1.1×10^{-8}
effect allele	G (-)	C (-)	A (-)	T (+)	C (+)	C (+)	T (+)	C (-)	A (-)	T (+)
Expression										
Gene ID	*CDKN2B*	*PFAS*	*PSCA*	*TRIB2*	*TMCO1*	*SRSF3*	*CAV2*	*AAGAB*	*RBM23*	ABO
Exp in eye	16.5 - 43	30 - 40	22 - 36	31 - 227	119 -225	179 - 559	19.5 - 40	37 - 148	37 - 56	20 - 36
Haplotype										
Proxy SNPs	29	5	48	10	33	64	11	74	1	7
Non-syn	0	0	0	0	0	0	0	0	0	1
Cis-eQTL										
Tissue	Whole blood	Multiple tissues	Multiple tissues	Whole blood	Multiple tissues	Multiple issues	Multiple tissues	Multiple tissues	Multiple tissues	Multiple tissues
miR-BS effect										
Dis/Cre	miR-323-5p	miR-149-3p	miR-342-5p	miR-877-5p	miR-296-3p	Let-7f-2-3p	miR-224-5p	miR-329-5p	miR-190a-3p	miR-855-3p
Δscore	Cre (-0.12)	Dis (-0.17)	Dis (-0.24)	Cre (-0.09)	NA	Dis (-0.01)	Dis (-0.2)	Dis (0.02)	Dis (-0.39)	Dis (-0.09)

Shown are top 10 miRNA-binding site variants associated with glaucoma endophenotypes that are more likely to be functional in the corresponding loci. The table summarizes functional characteristics of the variants including the most significant association from the GWAS, Expression from the OTDB, haplotype information and cis-eQTL from the HaploReg v4 and miRNA binding site effect from the PolymiRTS v3. For variants that are predicted to affect more than one miRNA-binding site, only the highly conserved miRNA has mentioned in the table. CA, cup area; IOP, intraocular pressure; Eye Exp., Expression in retina. Non-syn, Non-Synonymous proxy SNPs; Multiple, associated with host gene expression in several tissues. Δscore, Context score change; Dis, Disruption; Cre, Creation.

minor allele A at SNP site rs2273626, demonstrating that the minor allele decreases the binding of miR-4707-3p to *CARD10* (**Figure 3**).

Multiple 3'UTR variants in miRNA-binding sites were associated with POAG endophenotypes

In the second part of study, we examined the associations of 72,052 miRNA-binding site variants available in the GWAS data with four POAG endophenotypes [34]. In total, 47 SNPs (within the 3'UTR of 21 genes) passed the Bonferroni corrected significance threshold of 6.94×10^{-7} that are associated with one or more of the endophenotypes. These SNPs are predicted to affect the interaction between a number of miRNAs and their target genes that are shown in **Table S8**. To identify the miRNA-binding site variants that are more likely to be functional in their loci, we prioritized them based on the suggested criteria (e.g. the strength of association, LD, eQTL analysis, co-expression of the miRNA and target gene). Haploreg v4.1 results showed that most of the variants (n=37) have no non-synonymous proxy variants in strong LD ($R^2 > 0.8$) (**Table S9**). Through cis-eQTL analysis, we found that 33 of the binding site variants are correlated with expression levels of their host genes in different tissues (**Table S9**). The Ocular Tissue Database showed that all 21 genes hosting the 47 miRNA-binding site variants are expressed in the eye (**Table S10**). Using the miRNA expression databases, we found evidence for the expression of several miRNAs in our collection in the eye (**Tables S11**). Integrating the results of our in silico analyses for the 47 associated variants (**Tables S11**), we prioritized the variants that are more likely to affect their miRNA-binding sites (**Table 2**).

DISCUSSION

In this study, we performed a genome-wide scan for miRNAs associated with glaucoma using GWAS data. We identified two variants in miR-612 and miR-4707 that are significantly associated with VCDR and cup area. The minor allele of first variant has been shown to increase the expression level of miR-612. The second variant does not influence the expression of miR-4707, but the variant disrupts the regulatory interaction between the miRNA and its target gene *CARD10*, a known gene associated with VCDR and disc area. Furthermore, we identified 47 variants in miRNA-binding sites (within the 3'UTR of 21 genes) that are significantly associated with POAG endophenotypes. We prioritized these binding site variants based on predefined criteria and highlighted the variants that are more likely to affect miRNA-binding sites. Of these, rs3217992 and rs1063192 have been confirmed to affect miR-138-3p and miR-323b-5p mediated regulation of *CDKN2B*.

Previous studies have shown that miRNAs are involved in the pathogenesis of glaucoma [7-9]. These studies are mainly focused on differentially expressed miRNAs in a small number of samples, which makes it difficult to interpret their results to the general population. Here, we used a genome-wide approach to identify miRNAs associated with glaucoma endophenotypes using GWAS data. In this approach, the starting point is a linkage between miRNA-related variants and the disease. We and others have proven previously that the genetic approach is efficient and successful for identification of miRNAs involved in complex traits [16, 18-21]. The main advantage of the genetic approach compared to miRNA profiling is that when a miRNA variant is associated with disease risk, it supports the idea that the miRNA has a primary effect in the disease mechanisms [35].

We found two miRNA variants that are significantly associated with POAG endophenotypes. The first variant, rs12803915, is located in the terminal loop of miR-612 and its minor allele shows a protective effect on VCDR and cup area. In silico analysis indicated that the variant leads to alteration in free energy or conformation change of the pre-miR-612 secondary structure. The structure of the stem loop region of pre-miRNA is critical for miRNA processing by Drosha and Dicer enzymes [36, 37]. Polymorphisms located in the terminal loop of miRNAs have been shown to alter miRNA biogenesis [35, 36, 38, 39]. Rs12803915 minor allele A has been demonstrated previously to increase the expression of mature miR-612 in vitro. The location of variant in the terminal loop of miR-612 makes it likely to improve the binding of RNA-binding proteins that are involved in the pre-miRNA processing, and consequently increasing the miRNA expression [40, 41]. Here, we also highlighted one of the putative miR-612 target genes, *SIX4*, which is significantly associated with VCDR and cup area. The interaction between miR-612 and *SIX4* has been confirmed by CLIP-Seq data [42]. This gene encodes a member of the homeobox family, subfamily *SIX*, and is required for eye development [43]. The association of rs12803915 on VCDR and cup area may be explained by increased levels of miR-612. Future studies are warranted to examine the expression of miR-612 and its target genes in eye tissues and determine the this miRNA in the pathogenesis of glaucoma.

The second variant, rs2273626 is located in the seed sequence of miR-4707. The seed sequence, nucleotides 2-7/8 from the 5'-end of a mature miRNA, is defined as the most crucial part of a miRNA for target recognition that has to be perfectly complementary to the binding site of a target mRNA [14]. Polymorphism in miRNA seed sequence is expected to strongly influence the miRNA activity and risk of disease[16, 44]. We tested two mechanisms through which rs2273626 may affect miR-4707 function. First, we examined the effect of rs2273626 on the production of miR-4707 and observed no significant differences between the miRNA expression levels from the pre-miRNA sequences containing the major or minor allele, suggesting that the effect of rs2273626 on the miR-4707 biogenesis is minor and not substantial. Second, we tested whether

the variant reduces the ability of miR-4707-3p to bind to its target genes. We found five putative target genes of miR-4707-3p that are significantly associated with VCDR and cup area. Among them, *CARD10* is of special interest because it has been reported in previous epidemiologic and experimental studies to be involved in the development of glaucoma [6, 45, 46]. Furthermore, the activity of *CARD10* has been suggested to offer neuroprotection in glaucomatous optic neuropathy [45]. Our interaction analysis using the Rotterdam Study data showed an interaction between miR-4707 and *CARD10*. Our *in vitro* experiments further showed that miR-4707-3p down-regulates *CARD10* expression and rs2273626 reduces the miRNA-mRNA interaction, resulting increased *CARD10* levels. These data suggest that an impaired function of miR-4707-3p in the rs2273626 minor allele carriers may result in higher levels of *CARD10* and explain, in part, protective effect of rs2273626 on VCDR.

In this study we further investigated the association of miRNA-binding site variants with POAG endophenotypes and identified 47 associated variants within the 3'UTR of 21 genes. Since GWAS are based on tagging variants that are in high LD with functional variant(s) in the region, it is challenging to determine the exact localization of the variant that cause the associations. We prioritized the identified binding site variants based on a defined set of criteria that are suggested to determine the functional variants in miRNA-binding sites (e.g. the strength of association in GWAS, haplotype evaluation, eQTL analysis, co-expression of the miRNA and target gene) [16, 21]. Two of the prioritized variants, rs1063192 and rs3217992, reside in the 3'UTR of *CDKN2B* and do not display strong LD ($R^2 > 0.8$) with any other known *CDKN2B* variants. This gene encodes a cyclin-dependent kinase inhibitor and with its antisense (*CDKN2B-AS1*) lie in a well-known glaucoma associated locus on Chr.9p21 [5, 6]. The minor allele (G) at SNP site rs1063192 is predicted to create a binding site for miR-323b-5p and the rs3217992 minor allele (T) is expected to disrupt the existing binding site of miR-138-2-3p within the 3'UTR of *CDKN2B*. Horswell et al., have previously exhibited that these miRNAs control both CDKNB2 mRNA and protein levels. Moreover, they have shown the correlation of rs1063192 minor allele with lower *CDKN2B* mRNA levels and conversely the correlation of rs3217992 minor allele with higher *CDKN2B* mRNA levels in adipose tissue [47]. Furthermore, the functional consequences of both variants on the miRNA-binding sites have been confirmed by Luciferase reported assays [21, 47]. Both miR-323b-5p and miR-138-2-3p are expressed in the eye. These data may indicate allelic-specific miR-232b-5p and miR-138-2-3p regulation of *CDKN2B* as a potential explanation for the observed association between *CDKN2B* (rs1063192 and rs3217992) and glaucoma.

We also highlighted several other 3'UTR variants in miRNA-binding sites that are one of the top variants in their loci and have no non-synonymous proxy variants in high LD ($R^2 > 0.8$). These miRNA-binding site variants have the potential to be functional variants

in their loci by affecting miRNA-mediated regulation of their genes and warrant further investigations. For example, rs1052990 in the 3'UTR of *CAV2*, which is a known gene for glaucoma. The minor allele of rs1052990 is predicted to disrupt the binding site of miR-224-5p, expecting to increase the transcript levels of *CAV2*. Both miR-224-5p and *CAV2* are expressed in the eye [48, 49]. Furthermore, the minor allele of rs1052990 has been previously shown to increase the *CAV2* expression[50]. These data suggest that allele-specific miR-224-5p regulation of *CAV2* as a functional mechanism underlying the GWAS association.

Collectively, we investigated the association of miRNA-related genetic variants with POAG endophenotypes. We found two variants in miR-612 and miR-4707 that their minor alleles are negatively associated with VCDR and cup are. We showed that these variants may affect expression or function of the miRNAs. Further, we identified 47 3'UTR variants in miRNA-binding sites associated with POAG endophenotypes. We provide evidence showing that some of these variants have the potential to positively or negatively affect miRNA-mediated regulation of genes involved in POAG endophenotypes. These findings support the contribution of miRNAs in the pathogenesis of glaucoma endophenotypes.

Supplementary Table S5. Shown are putative target genes of miR-612 and miR-4707-3p that are associated with VCDR and cup area

miRNA ID	Phenotype	Putative target (# SNPs)	Associated targets	Top SNP	GWAS p-value
miR-612	VCDR	175 (34,854)	SIX4	rs17834412	6.63E-08
miR-612	Cup area	175 (34,877)	SIX4	rs17834412	4.0E-17
miR-4707-3p	VCDR	844 (88,078)	MYPN	rs10733840	4.43E-31
			CARD10	rs6000755	3.52E-12
			LTBP3	rs12270054	5.69E-09
			H1F0	rs5756825	1.54E-08
			FRMD8	rs1626021	2.27E-07
miR-4707-3p	Cup area	844 (87,991)	MYPN	rs10733840	2.57E-09
			FRMD8	rs1626021	3.93E-07
			LTBP3	rs4244811	4.84E-07

We used a Bonferroni correction (based on the number of variants in target genes of each miRNA) to set the significance level. The SNPs shown in the table are the most significant variants located in each target gene associated with VCDR and cup area in our GWAS data. TargetScan v7.0 (Total context++ score > 0.1) was used to extract miRNA target genes.

Table S11. Prioritization of 47 miRNA-binding site variants associated with glaucoma endophenotypes

SNP ID	Top in GWAS	Cis-eQTL	Minor allele (effect)	Proxy SNPs (# non-syn)	Gene ID	Gene exp in eye	Known for eye disease	MiRNA-binding site (disruption/ creation)	miRNA Exp & Fun	Score
rs3217992	✓	✓	*T (+)*	4 (0)	*CDKN2B*	✓	✓✓	miR-138-2-3p	✓✓	9
rs1063192	✓	✓	*G (-)*	29 (0)	*CDKN2B*	✓	✓✓	miR-323-5p	✓✓	9
rs1052990	✓	✓✓	*G (-)*	11 (0)	*CAV2*	✓	✓✓	miR-224-5p	✓	8
rs1050285	✓	✓	*C (-)*	74 (0)	*AAGAB*	✓	✓	miR-329-5p	✓	8
rs10216533	✓	✓✓	*A (+)*	48 (0)	*PSCA*	✓	✓	miR-342-5p, miR-1273h-3p	✓	8
rs1132554	✓	✓	*T (-)*	5 (0)	*PFAS*	✓	✓	miR-149-3p, miR-423-5p	✓	8
rs1057001	✓	✓✓	*T (-)*	10 (0)	*TRIB2*	✓✓	✓✓	miR-877-5p	✓	8
rs1043207	✓	✓✓	*A (-)*	1 (0)	*RBM23*	✓	✓	miR-190a-3p	✓	8
rs8176751	✓	✓✓	*T (+)*	7 (1)	*ABO*	✓	✓	miR-885-3p	✓	8
rs7344	✓	✓✓	*C (+)*	64 (0)	*SFRS3*	✓✓	✓	Let-7f-2-3p	✓	7
rs3781568	-	✓✓	*A (+)*	81 (0)	*RUFY2*	✓✓	✓✓	miR-1298-3p	✓	7
rs2281698	-	✓✓	*C (+)*	81 (0)	*RUFY2*	✓✓	✓✓	miR-132-5p	✓	7
rs10998098	-	✓✓	*A (+)*	81 (0)	*RUFY2*	✓✓	✓✓	miR-3613-3p	✓	7
rs3781567	-	✓✓	*T (+)*	81 (0)	*RUFY2*	✓✓	✓✓	miR-219b-3p	✓	7
rs1045547	✓	✓✓	*G (+)*	48 (0)	*PSCA*	✓	✓✓	miR-597-5p	✓	7
rs10249656	✓	✓✓	*T (-)*	10 (1)	*CAV2*	✓	✓✓	miR-575	✓	7
rs6660601	✓	✓✓	*C (-)*	33 (0)	*TMCO1*	✓✓	✓	miR-323-5p, miR-410-5p	✓	7
rs2281697	-	✓✓	*G (+)*	81 (0)	*RUFY2*	✓✓	✓✓	miR-548, miR-559	✓	7
rs1045605	✓	✓✓	*G (+)*	48 (0)	*PSCA*	✓	✓	miR-485-5p	✓	7
rs2278675	-	✓✓	*G (+)*	11 (0)	*MYPN*	✓	✓✓	miR-1271-5p, miR-96-5p	✓	6
rs2976396	✓	✓✓	*A (+)*	48 (0)	*PSCA*	✓	✓✓	miR-676-3p	✓	6
rs2305083	✓	✓✓	*A (+)*	30 (1)	*PBLD*	✓	✓✓	miR-1252-5p	✓	6
rs881976	-	✓✓	*T (+)*	11 (0)	*MYPN*	✓	✓✓	miR-548g-3p	✓	6
rs1045574	✓	✓✓	*A (+)*	48 (0)	*PSCA*	✓	✓	miR-296-5p	✓	6
rs10278782	✓	✓✓	*G (-)*	10 (1)	*CAV2*	✓	✓✓	miR-1228-3p	✓	6
rs1044269	-	✓	*G (+)*	30 (0)	*CELF1*	✓	✓✓	miR-1273c	✓	5
rs3026401	✓	-	-	4 (0)	*PAX6*	✓✓	✓✓	miR-664a-3p	✓	5
rs1506	✓	-	-	4 (0)	*PAX6*	✓✓	✓✓	miR-370-3p	✓	5
rs2976393	✓	✓✓	*G (+)*	48 (0)	*PSCA*	✓	✓	miR-4677-3p	-	5
rs948577	-	✓✓	*C (-)*	20 (0)	*MAP3K11*	✓	✓	miR-520a, miR-525	✓	5
rs7524755	✓	✓✓	*T (-)*	33 (0)	*TMCO1*	✓✓	✓	miR-548q	✓	5
rs731063	-	✓✓	*T (+)*	30 (1)	*PBLD*	✓	✓✓	miR-543	✓	5
rs595139	-	✓✓	*A (+)*	60 (2)	*FOLH1*	✓	✓	miR-186-5p	✓	5

Table S11. (continued)

SNP ID	Top in GWAS	Cis-eQTL	Minor allele (effect)	Proxy SNPs (# non-syn)	Gene ID	Gene exp in eye	Known for eye disease	MiRNA-binding site (disruption/ creation)	miRNA Exp & Fun	Score
rs3171656	-	✓✓	*A (-)*	61 (2)	*TRIOBP*	✓	✓	miR-30a-3p	✓	5
rs1807347	-	✓✓	*A (+)*	30 (1)	*PBLD*	✓	✓✓	miR-133a-5p	✓	5
rs17297439	-	✓✓	*A (+)*	81 (0)	*RUFY2*	✓✓	✓✓	miR-4719	✓	5
rs12125947	-	✓✓	*C (-)*	15 (0)	*CDC7*	✓	✓	miR-579-3p	✓	5
rs12421026	-	✓✓	*T (-)*	3 (0)	*PAX6*	✓✓	✓✓	miR-3135b	✓	5
rs3199937	-	-	-	81 (0)	*HNRNPH3*	✓✓	✓✓	miR-361-5p	✓	4
rs876489	✓	-	-	22 (0)	*MYPN*	✓	✓✓	miR-133a, miR-320	✓	4
rs4350264	✓	✓	*A (-)*	60 (0)	*SLC25A16*	✓	✓✓	miR-4422	-	4
rs2976395	✓	✓✓	*A (+)*	48 (0)	*PSCA*	✓	✓	miR-3160-3p	-	4
rs2976394	✓	✓✓	*T (+)*	48 (0)	*PSCA*	✓	✓	miR-3934-5p	-	4
rs1136645	-	✓	*C (-)*	26 (0)	*SLC25A16*	✓	✓✓	miR-564	✓	4
rs1177364	-	-	-	30 (1)	*RUFY2*	✓✓	✓✓	miR-32-3p	✓	3
rs1162756	-	-	-	30 (1)	*HNRNPH3*	✓✓	✓✓	miR-489-3p	✓	3
rs881975	-	-	-	22 (0)	*MYPN*	✓	✓✓	Let-7a-2-3p, miR-486-5p	✓	3

Forty-seven miRNA-binding site SNPs associated with POAG endophenotypes are prioritized based on the below criteria. A functional score (between 1 and 10) was calculated for the SNPs by combining the results of related criteria.

1. The strength of association in the GWAS results (✓✓: Lead SNP, ✓: One of the top 10% SNPs, -: Only passed the threshold).

2. eQTL (✓✓: Correlation between SNP and expression of the host gene, ✓: Correlation of SNPs with expression of nearby genes, -: No correlation).

3. Expression and importance of gene in the eye (✓✓: Abundantly expressed / Known gene, ✓: Expressed at low level/ Just association, -: Not expressed / No data).

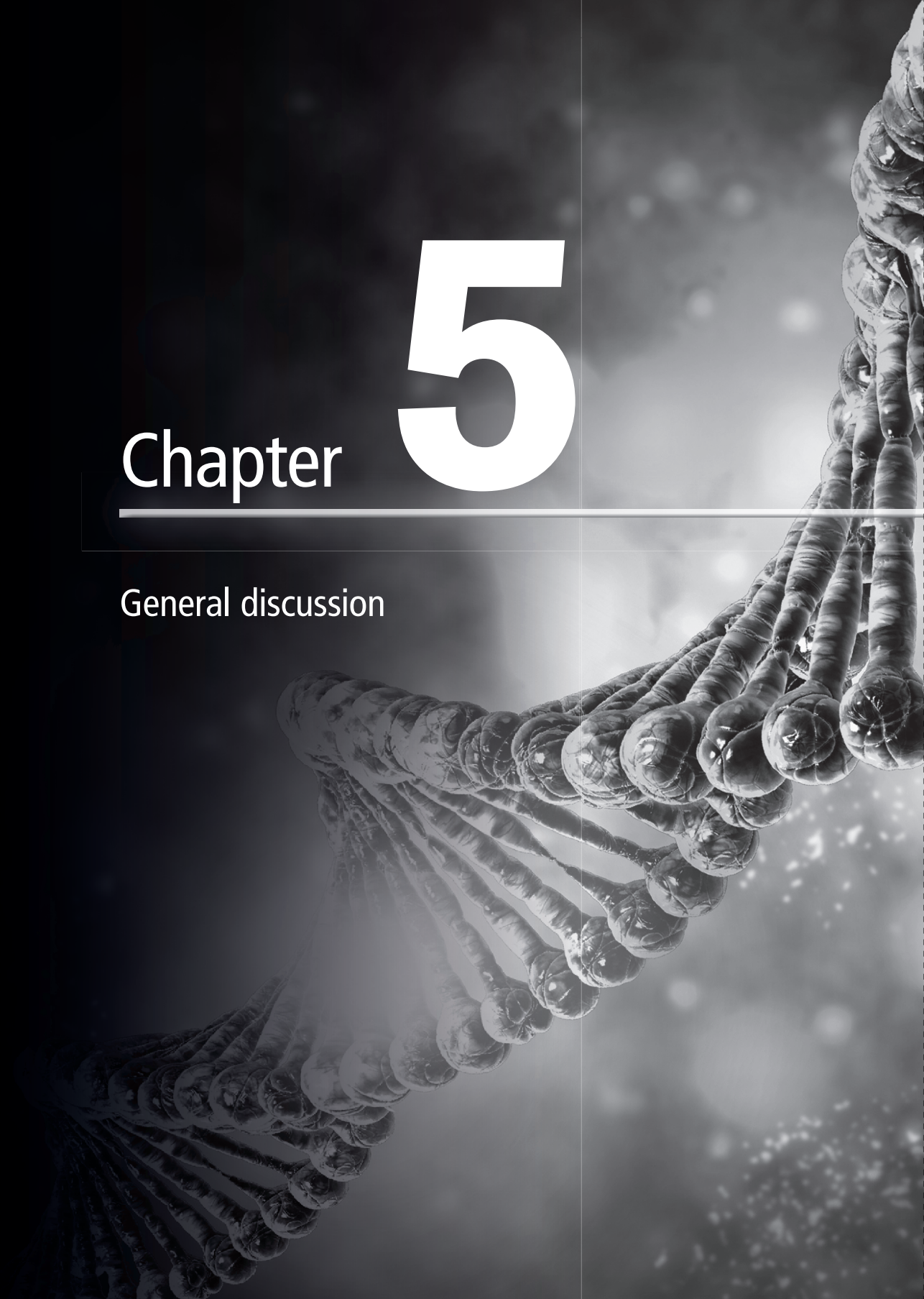
4. Expression and importance of related miRNA in the eye (✓✓: Abundantly expressed / *In vitro* evidence, ✓: Expressed at low level / *In silico* data , -: Not expressed/ No data). In this table only conserved miRNAs are mentioned.

REFERENCES

1. Quigley HA, Broman AT: The number of people with glaucoma worldwide in 2010 and 2020. Br J Ophthalmol 2006, 90:262-267.
2. Liu Y, Allingham RR: Molecular genetics in glaucoma. Exp Eye Res 2011, 93:331-339.
3. Kwon YH, Fingert JH, Kuehn MH, Alward WL: Primary open-angle glaucoma. N Engl J Med 2009, 360:1113-1124.
4. Charlesworth J, Kramer PL, Dyer T, Diego V, Samples JR, Craig JE, Mackey DA, Hewitt AW, Blangero J, Wirtz MK: The path to open-angle glaucoma gene discovery: endophenotypic status of intraocular pressure, cup-to-disc ratio, and central corneal thickness. Invest Ophthalmol Vis Sci 2010, 51:3509-3514.
5. Springelkamp H, Iglesias AI, Mishra A, Hohn R, Wojciechowski R, Khawaja AP, Nag A, Wang YX, Wang JJ, Cuellar-Partida G, et al: New insights into the genetics of primary open-angle glaucoma based on meta-analyses of intraocular pressure and optic disc characteristics. Hum Mol Genet 2017.
6. Springelkamp H, Hohn R, Mishra A, Hysi PG, Khor CC, Loomis SJ, Bailey JN, Gibson J, Thorleifsson G, Janssen SF, et al: Meta-analysis of genome-wide association studies identifies novel loci that influence cupping and the glaucomatous process. Nat Commun 2014, 5:4883.
7. Jayaram H, Cepurna WO, Johnson EC, Morrison JC: MicroRNA Expression in the Glaucomatous Retina. Invest Ophthalmol Vis Sci 2015, 56:7971-7982.
8. Kong N, Lu X, Li B: Downregulation of microRNA-100 protects apoptosis and promotes neuronal growth in retinal ganglion cells. BMC Mol Biol 2014, 15:25.
9. Molasy M, Walczak A, Szaflik J, Szaflik JP, Majsterek I: MicroRNAs in glaucoma and neurodegenerative diseases. J Hum Genet 2017, 62:105-112.
10. Luna C, Li G, Qiu J, Epstein DL, Gonzalez P: Cross-talk between miR-29 and transforming growth factor-betas in trabecular meshwork cells. Invest Ophthalmol Vis Sci 2011, 52:3567-3572.
11. Gonzalez P, Li G, Qiu J, Wu J, Luna C: Role of microRNAs in the trabecular meshwork. J Ocul Pharmacol Ther 2014, 30:128-137.
12. Krol J, Loedige I, Filipowicz W: The widespread regulation of microRNA biogenesis, function and decay. Nat Rev Genet 2010, 11:597-610.
13. Shukla GC, Singh J, Barik S: MicroRNAs: Processing, Maturation, Target Recognition and Regulatory Functions. Mol Cell Pharmacol 2011, 3:83-92.
14. Bartel DP: MicroRNAs: target recognition and regulatory functions. Cell 2009, 136:215-233.
15. Hughes AE, Bradley DT, Campbell M, Lechner J, Dash DP, Simpson DA, Willoughby CE: Mutation altering the miR-184 seed region causes familial keratoconus with cataract. Am J Hum Genet 2011, 89:628-633.
16. Ryan BM, Robles AI, Harris CC: Genetic variation in microRNA networks: the implications for cancer research. Nat Rev Cancer 2010, 10:389-402.
17. Duan R, Pak C, Jin P: Single nucleotide polymorphism associated with mature miR-125a alters the processing of pri-miRNA. Hum Mol Genet 2007, 16:1124-1131.
18. Ghanbari M, de Vries PS, de Looper H, Peters MJ, Schurmann C, Yaghootkar H, Dorr M, Frayling TM, Uitterlinden AG, Hofman A, et al: A genetic variant in the seed region of miR-4513 shows pleiotropic effects on lipid and glucose homeostasis, blood pressure, and coronary artery disease. Hum Mutat 2014, 35:1524-1531.

19. Ghanbari M, Sedaghat S, de Looper HW, Hofman A, Erkeland SJ, Franco OH, Dehghan A: The association of common polymorphisms in miR-196a2 with waist to hip ratio and miR-1908 with serum lipid and glucose. Obesity (Silver Spring) 2015, 23:495-503.
20. Ghanbari M, Darweesh SK, de Looper HW, van Luijn MM, Hofman A, Ikram MA, Franco OH, Erkeland SJ, Dehghan A: Genetic Variants in MicroRNAs and Their Binding Sites Are Associated with the Risk of Parkinson Disease. Hum Mutat 2016, 37:292-300.
21. Ghanbari M, Franco OH, de Looper H, Hofman A, Erkeland S, Dehghan A: Genetic Variations in miRNA Binding Sites Affect miRNA-Mediated Regulation of Several Genes Associated with Cardiometabolic Phenotypes. Circ Cardiovasc Genet 2015.
22. Gong J, Liu C, Liu W, Wu Y, Ma Z, Chen H, Guo AY: An update of miRNASNP database for better SNP selection by GWAS data, miRNA expression and online tools. Database (Oxford) 2015, 2015:bav029.
23. Bhattacharya A, Ziebarth JD, Cui Y: PolymiRTS Database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. Nucleic Acids Res 2014, 42:D86-91.
24. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ: LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 2010, 26:2336-2337.
25. Krol J, Sobczak K, Wilczynska U, Drath M, Jasinska A, Kaczynska D, Krzyzosiak WJ: Structural features of microRNA (miRNA) precursors and their relevance to miRNA biogenesis and small interfering RNA/short hairpin RNA design. J Biol Chem 2004, 279:42230-42239.
26. Lorenz R, Bernhart SH, Honer Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL: ViennaRNA Package 2.0. Algorithms Mol Biol 2011, 6:26.
27. Meenhuis A, van Veelen PA, de Looper H, van Boxtel N, van den Berge IJ, Sun SM, Taskesen E, Stern P, de Ru AH, van Adrichem AJ, et al: MiR-17/20/93/106 promote hematopoietic cell expansion by targeting sequestosome 1-regulated pathways in mice. Blood 2011, 118:916-925.
28. Agarwal V, Bell GW, Nam JW, Bartel DP: Predicting effective microRNA target sites in mammalian mRNAs. Elife 2015, 4.
29. Caussy C, Charriere S, Marcais C, Di Filippo M, Sassolas A, Delay M, Euthine V, Jalabert A, Lefai E, Rome S, Moulin P: An APOA5 3′ UTR variant associated with plasma triglycerides triggers APOA5 downregulation by creating a functional miR-485-5p binding site. Am J Hum Genet 2014, 94:129-134.
30. Ruepp A, Kowarsch A, Schmidl D, Buggenthin F, Brauner B, Dunger I, Fobo G, Frishman G, Montrone C, Theis FJ: PhenomiR: a knowledgebase for microRNA expression in diseases and biological processes. Genome Biol 2010, 11:R6.
31. Jiang Q, Wang Y, Hao Y, Juan L, Teng M, Zhang X, Li M, Wang G, Liu Y: miR2Disease: a manually curated database for microRNA deregulation in human disease. Nucleic Acids Res 2009, 37:D98-104.
32. Kozomara A, Griffiths-Jones S: miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res 2014, 42:D68-73.
33. Consortium CAD, Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, Thompson JR, Ingelsson E, Saleheen D, Erdmann J, et al: Large-scale association analysis identifies new risk loci for coronary artery disease. Nat Genet 2013, 45:25-33.
34. Fritsche LG, Igl W, Bailey JN, Grassmann F, Sengupta S, Bragg-Gresham JL, Burdon KP, Hebbring SJ, Wen C, Gorski M, et al: A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. Nat Genet 2016, 48:134-143.

35. Cammaerts S, Strazisar M, De Rijk P, Del Favero J: Genetic variants in microRNA genes: impact on microRNA expression, function, and disease. Front Genet 2015, 6:186.
36. Zhang X, Zeng Y: The terminal loop region controls microRNA processing by Drosha and Dicer. Nucleic Acids Res 2010, 38:7689-7697.
37. Han J, Lee Y, Yeom KH, Nam JW, Heo I, Rhee JK, Sohn SY, Cho Y, Zhang BT, Kim VN: Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. Cell 2006, 125:887-901.
38. Gottwein E, Cai X, Cullen BR: A novel assay for viral microRNA function identifies a single nucleotide polymorphism that affects Drosha processing. J Virol 2006, 80:5321-5326.
39. Xiong X, Kang X, Zheng Y, Yue S, Zhu S: Identification of loop nucleotide polymorphisms affecting microRNA processing and function. Mol Cells 2013, 36:518-526.
40. Wu M, Jolicoeur N, Li Z, Zhang L, Fortin Y, L'Abbe D, Yu Z, Shen SH: Genetic variations of microRNAs in human cancer and their effects on the expression of miRNAs. Carcinogenesis 2008, 29:1710-1716.
41. Piskounova E, Viswanathan SR, Janas M, LaPierre RJ, Daley GQ, Sliz P, Gregory RI: Determinants of microRNA processing inhibition by the developmentally regulated RNA-binding protein Lin28. J Biol Chem 2008, 283:21310-21314.
42. Chou CH, Chang NW, Shrestha S, Hsu SD, Lin YL, Lee WH, Yang CD, Hong HC, Wei TY, Tu SJ, et al: miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. Nucleic Acids Res 2016, 44:D239-247.
43. Iglesias AI, Springelkamp H, van der Linde H, Severijnen LA, Amin N, Oostra B, Kockx CE, van den Hout MC, van Ijcken WF, Hofman A, et al: Exome sequencing and functional analyses suggest that SIX6 is a gene involved in an altered proliferation-differentiation balance early in life and optic nerve degeneration at old age. Hum Mol Genet 2014, 23:1320-1332.
44. Mencia A, Modamio-Hoybjor S, Redshaw N, Morin M, Mayo-Merino F, Olavarrieta L, Aguirre LA, del Castillo I, Steel KP, Dalmay T, et al: Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss. Nat Genet 2009, 41:609-613.
45. Khor CC, Ramdas WD, Vithana EN, Cornes BK, Sim X, Tay WT, Saw SM, Zheng Y, Lavanya R, Wu R, et al: Genome-wide association studies in Asians confirm the involvement of ATOH7 and TGFBR3, and further identify CARD10 as a novel locus influencing optic disc area. Hum Mol Genet 2011, 20:1864-1872.
46. Wang L, Guo Y, Huang WJ, Ke X, Poyet JL, Manji GA, Merriam S, Glucksmann MA, DiStefano PS, Alnemri ES, Bertin J: Card10 is a novel caspase recruitment domain/membrane-associated guanylate kinase family member that interacts with BCL10 and activates NF-kappa B. J Biol Chem 2001, 276:21405-21409.
47. Horswell SD, Fryer LG, Hutchison CE, Zindrou D, Speedy HE, Town MM, Duncan EJ, Sivapackianathan R, Patel HN, Jones EL, et al: CDKN2B expression in adipose tissue of familial combined hyperlipidemia patients. J Lipid Res 2013, 54:3491-3505.
48. Kato T, Meguro A, Nomura E, Uemoto R, Nomura N, Ota M, Kashiwagi K, Mabuchi F, Iijima H, Kawase K, et al: Association study of genetic variants on chromosome 7q31 with susceptibility to normal tension glaucoma in a Japanese population. Eye (Lond) 2013, 27:979-983.
49. Wiggs JL, Kang JH, Yaspan BL, Mirel DB, Laurie C, Crenshaw A, Brodeur W, Gogarten S, Olson LM, Abdrabou W, et al: Common variants near CAV1 and CAV2 are associated with primary open-angle glaucoma in Caucasians from the USA. Hum Mol Genet 2011, 20:4707-4713.
50. Kim S, Kim K, Heo DW, Kim JS, Park CK, Kim CS, Kang C: Expression-associated polymorphisms of CAV1-CAV2 affect intraocular pressure and high-tension glaucoma risk. Mol Vis 2015, 21:548-554.



Chapter 5

General discussion

MiRNAs are important post-transcriptional regulators of gene expression, and deregulation of gene expression is a mechanism in the pathophysiology of human diseases [1]. Given the broad clinical and therapeutic potentials of miRNAs, it is of interest to investigate whether miRNAs may play a role in age-related disorders, such as cardiovascular, neurodegenerative and ophthalmic diseases. Previously, the expression and function of several miRNAs have been shown to be dysregulated in these complex disorders [2-7]. Further, a number of miRNAs have been suggested as potential diagnostic biomarkers and therapeutic targets for each of the diseases [8-11]. Nevertheless, previous studies have mainly focused on differentially expressed miRNAs in a small number of samples, which is difficult to be generalized to the general population. In recent years, technological advances have made it possible to apply high-throughput omics data (e.g. genomics and transcriptomics) in large scale to epidemiological studies. For instance, genome-wide association studies (GWAS) have been implemented in past decade and led to the discovery of a large number of loci contributing to complex traits and diseases [12-15]. The availability of these omics data encourages researchers to develop new methods to understand the complexity of the genomic mechanisms behind diseases. In addition, the reliance of miRNA biogenesis to sequences to determine the required secondary structure for correct processing and target recognition increases the possibility that variants in miRNAs may have the potential to affect functionality of miRNAs, and consequently alter disease risk [16-18]. In this thesis, we aimed to identify miRNAs that are involved in age-related disorders using genetic data followed by various *in silico* and *in vitro* validation studies.

MAIN FINDINGS AND INTERPRETATION

Genetic variants of miRNA sequences

MiRNA seed sequence variants. The seed sequence, nucleotides 2-7/8 from the 5′ end of a mature miRNA, is the core of a miRNA for target gene recognition and interaction [19]. Genetic variation in this critical region of miRNA may alter both the miRNA expression level and the miRNA interaction with its target transcripts [20, 21]. Our analyses in this thesis showed that the occurrence of genetic variants in miRNA seed sequences is relatively rare and the majority of them (> 80%) are not available in the current GWAS data. Since the miRNA seed sequence variant render a strong effect on the miRNA function, they are expected to be present at a low frequency in the human genome. This is in line with previous *in silico* analyses of genetic variants in miRNA sequences that show miRNAs have lower variant densities than their flanking regions and the human genome [22-25]. In addition, it was shown that within the miRNA-encoding sequence, the mature miRNA sequence has a lower SNP density than the precursor, and the seed sequence has the

lowest SNP density [22]. In this thesis we found two variants in miRNA seed sequences that were significantly associated with age-related disorders. In **Chapter 2.1** we showed that rs2168518 in the miR-4513 seed is associated with a number of cardiometabolic traits and diseases. We demonstrated two mechanisms through which rs2168518 may affect miR-4513 function. First, the SNP minor allele significantly reduces the expression of miR-4513 compared to the major allele, which could be through interfering with the miRNA processing efficiency [26]. Second, the SNP minor allele decreases the ability of miR-4513 to interact and repress its target genes expression. We were the first study to implicate miR-4513 in human diseases. In a second study, Li et al recently investigated the association of rs2168518 with clinical outcome in coronary heart disease (CHD) [27]. In 1,004 patients with angiographic CHD, they found that miR-4513 was associated with even free survival and mortality, further demonstrating the importance of miR-4513 in cardiovascular disease. In **Chapter 4.1** we also found the association of rs2168518 in miR-4513 with age-related macular degeneration (AMD). These data suggest a pleiotropic role for miR-4513 in cardiometabolic traits and AMD that could be through regulating distinct target genes involved in the associated disorders. In **Chapter 4.2** we identified another seed sequence variant, rs2273626, located in miR-4707 which was significantly associated with two glaucoma endophenotypes, vertical cup to disk ratio (VCDR) and cup area. Our *in silico* analysis predicted a large alteration in the minimum free energy of miR-4707, suggesting that the variant may affect miRNA's secondary structure. However, we observed no significant difference between mature miR-4707 levels in cells transfected with vectors containing pre-miR-4707 sequence with the minor allele and the major allele at the SNP site. We then examined the second possibility that the seed sequence variant affect miR-4707 function by disrupting the interaction between the miRNA and its target genes. Our *in vitro* experiments showed that rs2273626 reduces the ability of miR-4707 to repress expression of its candidate target gene *CARD10*, which is a known glaucoma-associated gene [28, 29]. These findings support the functional importance of miRNA seed sequence variants in affecting miRNA function, which depends on the position of a variant in the seed sequence and can affect both miRNA expression and targeting or only one of these (**Figure 1**).

Pre-miRNA sequence variants. Genetic variants in pre-miRNA transcripts can affect miRNA expression levels by influencing the miRNA maturation process. These variants can lead to change the binding affinity of the miRNA hairpin to biogenesis enzymes (*Drosha* and *Dicer*) or lead to altered strand loading bias into RISC complex accessory proteins [22]. These changes could result in altered expression of the canonical miRNA, and consequently lead to deregulation of target genes. In **Chapters 2-4** we found ten variants in miRNA-encoding sequences, beyond the seed sequence, that are associated with different age-related disorders. These variants fall in different parts of the hairpin precursor of miRNAs, ranging from the stem region to terminal-loop (**Figure 1**). It has

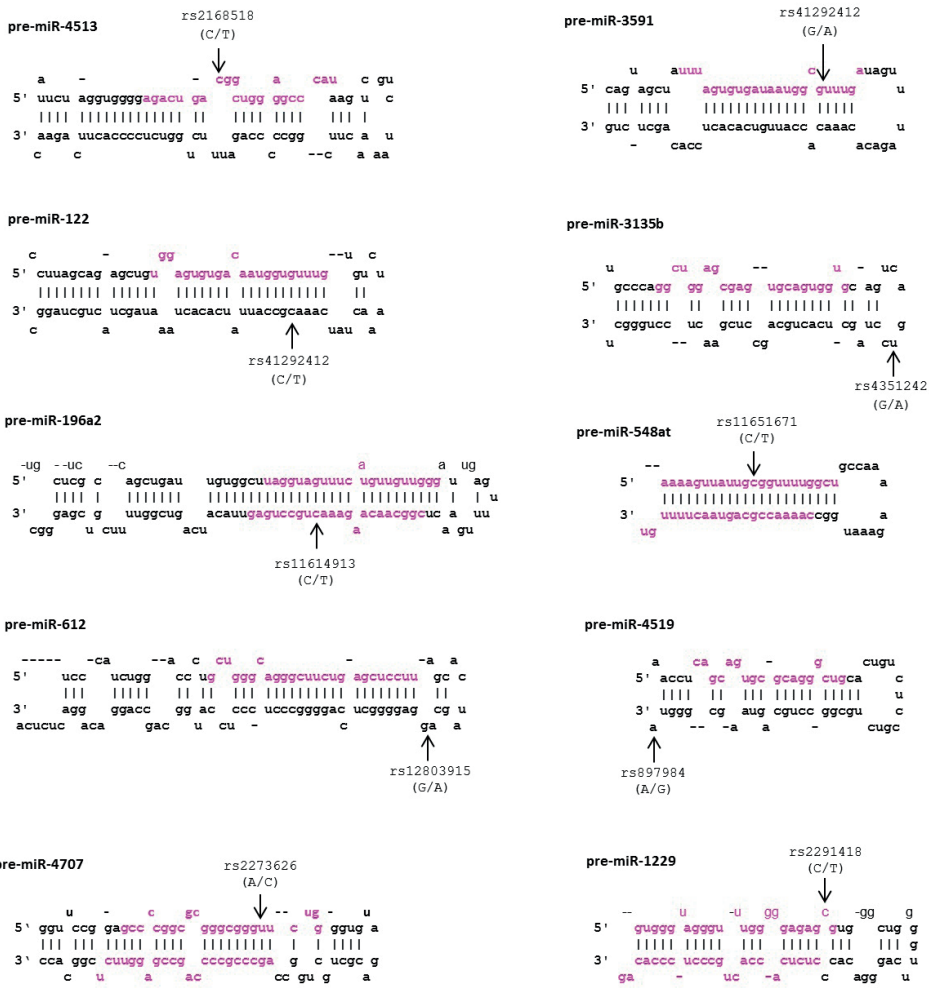


Figure 1. Predicted hairpin structures of the ten identified miRNAs hosting the variants associated with age-related disorders. The mature miRNA sequences are shown in purple and the position of variants are depicted by arrows.

been suggested that the effect of variant on the miRNA biogenesis can be predicted by performing structural analysis of the pre-miRNA sequence [30]. Our *in silico* analyses showed two of the variants located in pre-miR-1229 and pre-miR-3135b had no alterations in minimum free energy (MFE) of the thermodynamic ensemble of their pre-miRNAs structure. This leads to the assumption that these variants would not affect the miRNA maturation. Our *in vitro* experiments, however, showed that one of the SNPs significantly decreases miR-3135b expression and the other SNP significantly increases the expression of miR-1229. Thus, miRNA variants may affect the mature miRNA expression even in

the absence of predicted effects on the miRNA secondary structure. It has been shown before for miRNA let-7 that rs41275792 (19 nt downstream of the miRNA precursor) reduces the mature miRNA levels *in vivo*, although the miRNA secondary structure was not predicted to be changed [31]. The processing and maturation of miRNAs *in vivo* are thereby more complex and subtle than that can be currently predicted.

Target genes mediating downstream effects of the identified miRNAs. In order to identify the downstream targets that may mediate the effect of identified miRNAs in age-related disorders, we performed various *in silico* and *in vitro* experiments. These analyses highlighted a number of target genes associated with the studied traits as potential mediators of the miRNA effects (**Figure 2**). As an example, in **Chapter 3.1** we found the association of miR-1229, a brain expressed miRNA, with Alzheimer's disease (AD). Performing *in silico* analyses, we filtered the miR-1229-3p target genes that are associated with AD and are expressed in the brain. We experimentally showed that miR-1229-3p controls expression of *SORL1*, which is the top AD-associated target gene of miR-1229-3p [32, 33]. We proposed that higher expression levels of miR-1229-3p in rs2291418 minor allele carriers may increase risk of AD, at least in part, through repression of *SORL1* expression. While our work was under revision, miR-1229 was also reported by another group to be dysregulated across the brain regions of Alzheimer's patients (in 87 AD and 74 non-demented controls) [34]. These data convincingly implicate miR-1229 to AD. To gain insight into the mechanisms that the altered expression levels of miR-1229 equate to change in the biological function and contribute to the pathology of AD further experimental studies are needed. The next step would be to inhibit (LNA-antagomiR) and overexpress (mimic) the miRNA in the right cellular context (e.g. iPSC-derived neurons) or brain tissue and assess the expression levels of the highlighted target genes (at mRNA and protein levels). Further, generating an animal model in which miR-1229 is deleted (e.g. knockout mouse) followed by assessment of the expected phenotypic changes and expression levels of the miRNA target genes in the brain sections of animal models can provide great insight into the function and biological role of miR-1229 in relation to AD. Such *in vitro* and *in vivo* experiments would help to be able to fully grasp the regulation and the genesis of the identified miRNAs and their highlighted target genes in the pathogenesis of age-related disorders.

miRNAs as a source for pleiotropic effect. Genetic studies have shown in many instances that a certain gene could be related to more than one phenotype, that so-called pleiotropic gene [35-37]. GWAS have also identified several genes that are associated with multiple age-related disorders. For instance, *ANKRD26* is a gene associated with cardiovascular and neurodegenerative diseases [35]. Understanding the extent of pleiotropy is important as it may reveal information regarding the underlying pathways of disease. In this thesis we found at least two miRNAs that show pleiotropic effects. In **Chapter 2.1** and **Chapter 4.1** we showed that miR-4513, hosting the seed sequence

Figure 2. Venn diagram shows miRNAs and target genes found to be associated with age-related disorders using the GWAS data (bold) which the effect of their variants on the miRNA-mediated gene regulation experimentally validated in this study.

variant rs2168518, is associated with fasting glucose, LDL and total cholesterol, blood pressure, coronary artery disease and age-related macular degeneration. In **Chapter 2.2** we found that miR-196a2, hosting the SNP rs11614913, is associated with waist to hip ratio (WHR). This miRNA variant has also been reported to contribute to an increased risk of breast and lung cancer [38, 39]. In addition, a meta-analysis of GWAS on bone mineral density (BMD) further demonstrated that the minor allele of rs11614913 is associated with lumbar spine and femoral neck BMD [40]. In these chapters we highlighted a number of target genes of miR-4513 and miR-196a2 that are involved in different pathways belonging to the associated phenotypes and that may be dysregulated as a consequence of the identified variants. By their nature, miRNAs can bind to multiple target genes and are potentially involved in different biological pathways. Thus, the pleiotropic effect of miRNAs can be explained by regulating distinct target mRNAs that have cell type and even cell developmental stage-specific expression profiles. Several miRNAs have been reported previously to function as pleiotropic modulators of physiological and pathological processes, such as miR-206 in cell proliferation, miR-221 in acute liver failure and miR-26a in metabolic regulations [41-43].

Genetic variants in miRNA-binding sites

The untranslated regions at the 3' end of mRNA transcripts (3'UTRs) are involved in multiple levels of regulation [44, 45]. Although increasing evidence indicate that 3'UTRs contain regulatory elements that play important roles in gene expression [7, 8], they are not extensively mined for genetic variant discovery so far. The importance of variants in these regions for human diseases can be exemplified by the range of phenotypes affected by these variants [16, 46, 47]. In this thesis our data support the notion that 3'UTR variants can both positively and negatively influence gene expression by affecting miRNA-mediated gene regulation. In **Chapter 2.3** we showed that approximately 1% of common variants are expected to be located in miRNA-binding sites. Moreover, we demonstrated that GWAS findings for cardiometabolic disorders are enriched in miRNA-binding sites, suggesting the importance of such variants in genetic associations to disease. Likewise in **Chapters 3** and **4** we showed a number of potentially functional 3'UTR variants associated with neurodegenerative or ophthalmic diseases that are likely to influence the interaction between miRNAs and their target genes, resulting in altred gene expression. To provide direct evidence that genetic variants in miRNA-binding sites and deregulated gene expression in disease are linked, detailed functional validation is required. Performing functional experiments for all disease-associated variants in putative miRNA binding sites is costly and labor-intensive. Furthermore, 3'UTR variants in miRNA-binding sites in many cases have high LD proxy variants in the coding regions of genes, which are more likely to drive the observed associations. To gain insight into loci harboring the associated variants in GWAS, researchers usually employ bioinformatics strategies to prioritize the variants based on their potential function [24, 48, 49]. These in silico approaches are focused mainly on coding sequence variants and include only a limited number of functional annotations to non-coding regions. In **Chapter 2.3** we used a prioritization method to determine miRNA-binding site variants that are more likely to be functional in their loci. This approach incorporates diverse biological features and functional annotations to conclude whether a miRNA-binding site variant has the potential to be functional [16]. In **Chapter 2.3** we found 192 cardiovascular-associated SNPs that are located in miRNA-binding sites. Our prioritization approach enabled us to rank these SNPs and select ten of them as the most likely functional variants for experimental studies. Our *in vitro* experiments then confirmed that 7 of the variants decrease miRNA-mediated regulation of the genes and two of them enhance the binding of miRNAs to the host genes (**Table 1**), may indicating our approach is a valid strategy to detect the functional 3'UTR variants in miRNA-binding sites. In **Chapters 3** and **4** we successfully used this approach and highlighted a number of likely functional miRNA-binding site variants associated with neurodegenerative and ophthalmic diseases. Thesehighlighted miRNA-binding site variants may be of clinical importance and warrant further investigations. For instance, in **Chapter 3.3** we found that rs4151672 associated with

age-related macular degeneration (AMD) is located in miR-210-5p binding site within the 3'UTR of *CFB*. We experimentally showed that, while miR-210-5p downregulates *CFB* expression, the rs4151672 minor allele decreases miR-210-5p-mediated repression of *CFB*. *CFB* is a key factor involved in activation of the complement cascade and higher levels of the gene is associated with increased risk of AMD [50]. Our findings suggest that miR-210-5p may contribute to reduced risk of AMD through decreasing *CFB* expression. We demonstrated that miR-210-5p down-regulates CFB protein levels. In addition, both miR-210-5p and *CFB* are expressed in the eye [51, 52]. These data suggest miR-210-5p as a promising therapeutic target for AMD that warrant further investigations. Moreover, a number of complement-based therapeutics (e.g. antibodies against *CFB* protein) are currently tested in clinical trials to prevent and treat AMD development [53]. The easy accessibility of the eye by injection also makes the eye a suitable organ for miRNA therapeutics. In this context, local delivery of miRNA inhibitors and mimics for treating vascular disorders of the retina has been shown recently as a potentially clinical approach [54]. Thus, future studies (e.g. testing the role of miR-210-5p *in vivo* in knockout mice) are warranted to dissect the mechanisms that miR-210-5p may contribute to AMD risk and its potential in therapeutics.

Table 1. A summary of 10 experimentally validated variants in miRNA-binding sites associated with age-related disorders

Variant ID	Host gene	Associated Phenotypes	miRNA ID	Score change	Effect on miRNA-mRNA
rs3217992	*CDKN2B*	CAD, AMD, VCDR	miR-138-2-3p	-0.19	Decrease
rs174545	*FADS1*	Lipids, FG	miR-181a-2	-0.13	Decrease
rs1059611	*LPL*	HDL,TG	miR-136	-0.17	Decrease
rs13702	*LPL*	HDL,TG	miR-410	NA	Decrease
rs11735092	*HSD17B13*	TG	miR-375	-0.2	Decrease
rs7956	*MKRN2*	TC	miR-154-5p	-0.06	Decrease
rs1046875	*FN3KRP*	HbA1C	miR-34a	-0.13	Decrease
rs4151672	*CFB*	AMD	miR-210-5p	-0.27	Decrease
rs6857	*PVRL2*	Lipids, Alzheimer	miR-320e	-0.19	Increase
rs907091	*IKZF3*	HDL	miR-326	NA	Increase

Score change, change in the minimum free energy for the miRNA-target interaction; NA, Not available; CAD, Coronary artery disease; AMD, Age-related macular degeneration; VCDR, Vertical cup to disk ratio; FG, Fasting glucose; HDL, High Density Lipoprotein; TG, Triglyceride; TC, Total cholesterol.

Genetic variants in long non-coding RNAs

lncRNAs are a heterogeneous group of transcripts with over 200 nt in length that exhibit no coding potential [55, 56]. Over the past few years, it has become increasingly evident that lncRNAs act as important players in the epigenetic, post-transcriptional,

and translational coordination of gene expression [57-59]. Nevertheless, except some well-characterized lncRNAs, such as *ANRIL* and *Xist,* little is known about the general features of most lncRNAs and their possible molecular mechanisms in human diseases [57, 60-62]. In recent years, GWAS have enabled the discovery of a large number of genetic variants at multiple loci associated with age-related disorders. However, the majority of the identified variants map to non-coding genomic regions, including lncRNAs, that their biological relevance to the diseases remain poorly understood. In **Chapter 2.4** we investigated the association of genetic variants in lncRNAs with cardiovascular risk factors and disease. We found 55 cardiovascular-associated SNPs, either the lead SNP or in strong LD with the lead SNP in their loci, that are located in lncRNAs. Our *in silico* predictions and functional annotations for these SNPs as well as DNA methylation and expression analyses for their lncRNAs highlighted the a number of SNPs that fulfill our predefined criteria for being functional and act through lncRNAs. It has been shown that lncRNAs can post-transcriptionally regulate gene expression by binding to miRNAs and preventing specific miRNAs from binding to their target mRNAs. We found that some of the highlighted SNPs are associated with the expression levels of (nearby or distant) coding genes and have the potential to affect the interaction between lncRNAs and miRNAs. Further, we highlighted a number of loci harboring the associated lncRNA SNPs for which the role of lncRNA in the association have been overlooked in the current literature. For instance, we found several pieces of evidence suggesting *LOC157273* to be involved in regulating serum lipid-cholesterol, while the GWAS association has been annotated to *PPP1R3B* [12]. In particular, we found that rs4841132 in the second exon and cg17371580 in the promoter region of *LOC157273* are associated with lipids, the lncRNA is exclusively expressed in liver and regulates expression of *PPP1R3B*, a coding gene which is 175 kb away from the lead variant. We posited that *LOC157273* may be the key regulator that links the identified variant with *PPP1R3B* activity and with lipid profile.

In **Chapter 3.3** we investigated the association of lncRNA variants with Alzheimer's disease and revealed a new susceptibility locus at chromosome 17q22 for AD. The lead variant in this locus is rs2632516 which overlaps the intron of anti-sense lncRNA *BZRAP1-AS1* and the promoter region of miR-142 host gene in the opposite strand. Our *in silico* analysis posited that the SNP is more likely to act through affecting the promoter activity of miR-142. The impact of variants in promoter regions on miRNA expression have been shown before, for example rs57095329 in the miR-146a promoter region associated with systemic lupus erythematosus [63] and rs4938723 in the miR-34 promoter region associated with hepatocellular carcinoma [64]. Given the potential impact of variants in miRNA promoter regions on altering transcription rate, it is not surprising that such variants can be casual for or associated with human diseases. Moreover, we found evidence showing that both miR-142-3p and miR-142-5p are upregulated in hippocampus and blood of AD patients [65]. A large number of putative target genes of

miR-142 are also expressed in the brain and they are enriched in neuronal pathways. The regulatory interaction between miR-142 and a number of these AD-associated target genes, such as *SIRTL1*, have been validated previously *in vitro* [66]. Together, these data strongly suggest miR-142 as a promising biomarker or even therapeutic target for AD that warrant further investigations.

METHODOLOGICAL CONSIDERATIONS

Previous studies have suggested two strategies to study if and which miRNAs are involved in human diseases, including miRNA expression profiling and genetic approach. In the first approach, affected tissue of patients is subjected to miRNA expression profiling and compared to expression profiles of the same tissue of healthy individuals. In the genetic approach, the starting point is a linkage or an association analysis of miRNA-related variants between patients and controls. In this thesis we used an integrative approach to identify miRNAs involved in age-related disorders using genetic data followed by various *in silico* and *in vitro* validation studies (**Figures 3** and **4**). In this approach, we examine the association of miRNA-related genetic variants with the trait of interest using publicly available GWAS data. Subsequently, we perform different *in silico* analyses (e.g. RNA structure analysis, LD evaluation, pathway analysis, and miRNA-target interaction analysis) and biological studies (e.g. miRNA and gene expression profiles, eQTL and methylation analysis) for the associated variants to gain some insight into the potential function of the associated miRNAs and their target genes in regard to the studied trait. Next, we perform experimental studies (e.g. transient transfection of the identified miRNAs and target genes, Luciferase reporter assay, and Western blot) to test functional consequences of the identified variants on the miRNA-mediated gene regulation. This approach is proven to be efficient and successful by using it for multiple traits. In comparison with miRNA expression profiling approach, our strategy has several advantages for the identification of miRNAs associated with human diseases. The main advantage of this approach compared to miRNA profiling is that when a miRNA variant is associated with disease risk, it supports the idea that the miRNA has a primary effect in the disease mechanisms [22]. While differential expression of miRNAs is not providing evidence for a causal role of the miRNAs in relation to the disease. Expression profiling studies are subject to confounding bias or reverse causation. For instance, differential expression of miRNAs can be a secondary effect of disturbed pathways due to the disease or other characteristics that are correlated to the patient group, such as lifestyle. The genetic approach may overcomes both challenges. Given that genetic variants are randomly inherited, it is unlikely that other extrinsic factors are associated with them (no confounding bias). In addition, as the genetic information is constant over the life

course, reverse causation is refuted. Another advantage of the genetic approach is that diseases with no accessible affected tissues (e.g. eye tissue of AMD patients or brain tissue of AD patients) can still be studied since genomic DNA is available from other tissues such as blood. Finally, in the our approach we use the summary statistics from GWAS which are publicly available. The secondary use of published data is very efficient and indicates that our strategy is implemented in the right time when large waves of GWAS are completed. In addition, the results are generalizable to general population due to using population-level data for the association studies.

Our approach, however, has some limitations that need to be addressed. First, although there are 1,080 mature miRNAs that host genetic variants (n=2,347), the studies described in this thesis showed that the occurrence of common variants in miRNAs is relatively rare. Therefore, majority of the variants (~80%) are not available in GWAS data and cannot be tested using the genetic approach with the data available at this moment. Previous *in silico* analyses of variants in miRNAs have also shown that miRNA genes have lower variant densities than their flanking regions or the human genome [22-25]. To investigate the association of the rare miRNA variants with human disease, high density genotyping methods or whole-genome sequencing is needed. Second, to detect target genes of the associated miRNAs that mediate the downstream effects of miRNAs in the associated phenotypes, we tested the association of their target genes with the phenotype using GWAS data. Although this strategy enabled us to identify those genes that are associated with the phenotype, this method may overlook certain target genes of the miRNA and might miss some other important targets. The optimal approach would be to check the co-expression patterns of the miRNA and its putative targets in relevant cells/target tissues by RNA-sequencing and select the deregulated target genes. We can further filter the deregulated target genes based on their associations with the phenotype or pathway analysis to select the best candidates for additional *in vitro* or *in vivo* studies. To provide functional evidence for the binding of miRNAs to target mRNAs, and to show the effect of variants on the miRNA-target gene interactions, we used the well-accepted *in vitro* Luciferase reporter assays. This experiment provides strong support for the functionality of miRNA variants, however, the chain of evidence would be more complete if we examine the influences of variants on the post-transcriptional regulation of miRNA target genes at the protein level as well in the right cellular context, though this may not be possible in all cases. In spite of the above-mentioned limitations, we believe that our approach is valid and helps for both identification of miRNAs for complex disorders and annotation of GWAS findings located in miRNAs and miRNA-binding sites within gene 3'UTRs.

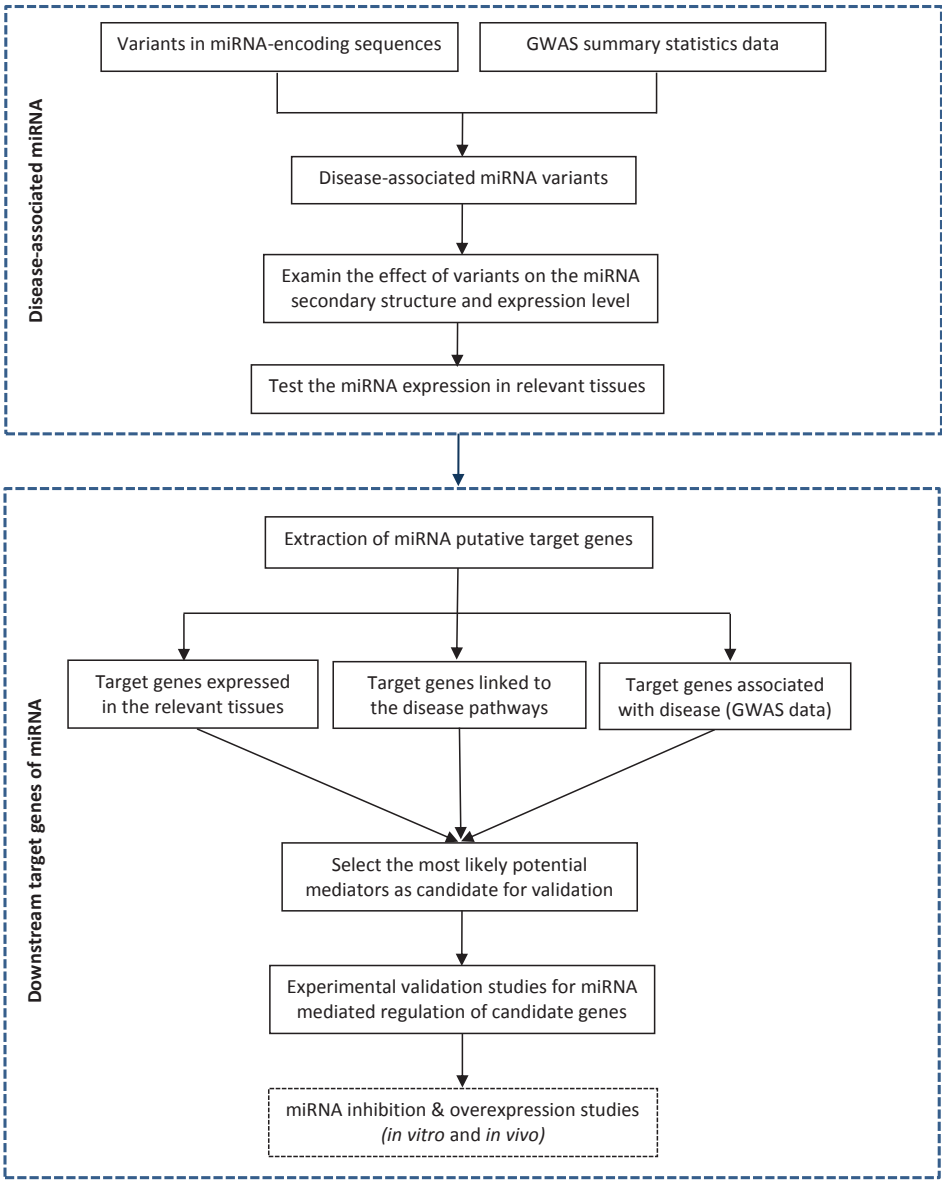


Figure 3. Identification and functional analyses of variants in miRNAs associated with complex disorders.

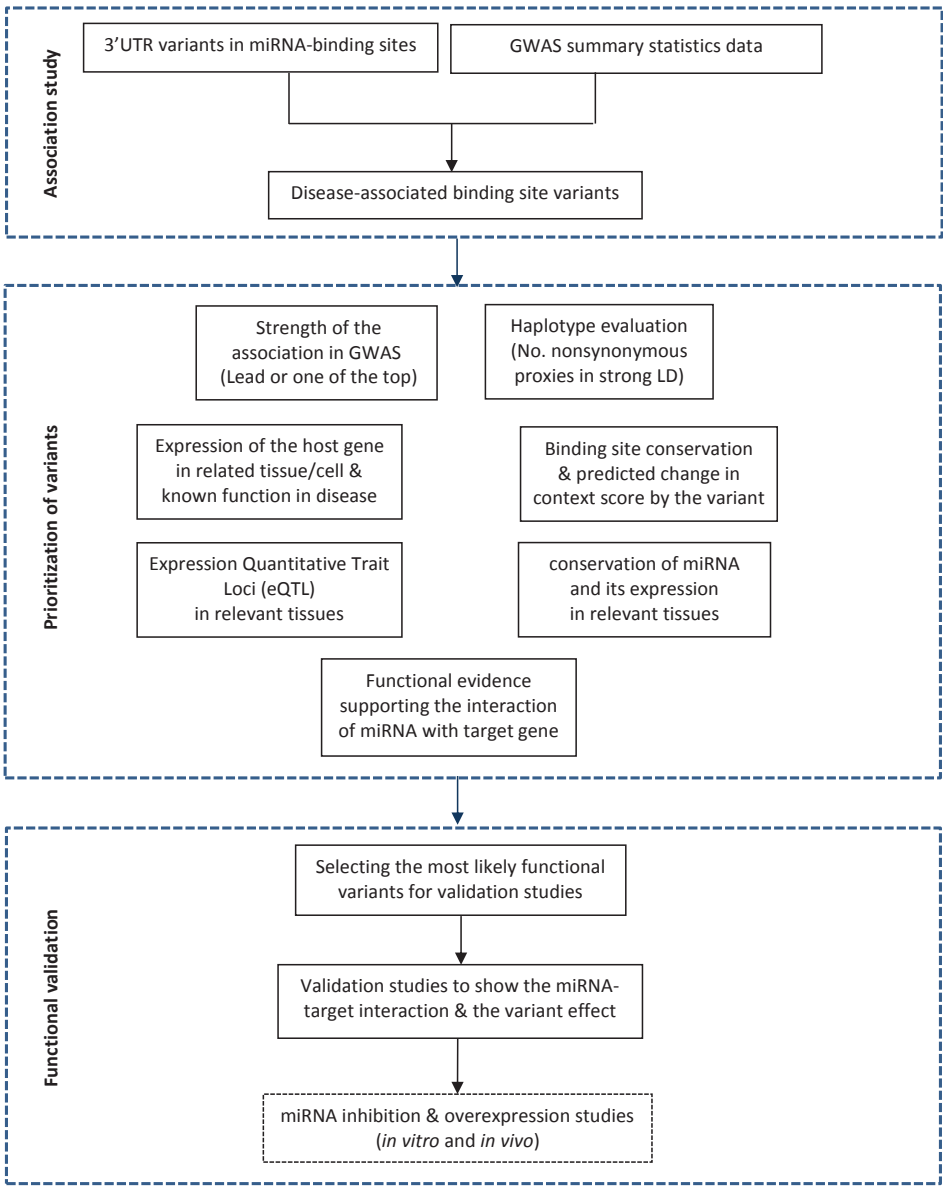


Figure 4. Identification and functional analyses of miRNA-binding site variants associated with complex disorders.

FUTURE DIRECTIONS

In this thesis, we identified a number of miRNAs that may contribute to the induction or development of age-related disorders (**Figure 2**). We focused our analyses on known genetic variants in miRNAs. As the miRNAome evolves, it is likely that new miRNAs and variants will emerge that their associations with age-related disorders need to be investigated in future studies. Moreover, miRNA variants are rare and majority of them are absent in the current GWAS. To investigate the association of these rare and sometimes population-specific variants, high density genotyping methods or whole-genome sequencing (to increase the coverage of variants) and GWAS in larger sample sizes (to increase the power of study) are needed.

Genetic variants within miRNA-processing machinery genes (such as *Drosha*, *DGCR8, Dicer1, XPO5, TRBP,* and *AGO2*) have the potential to globally affect processing of small RNAs [16, 22]. Recent studies have suggested that such variants, for example, rs3742330 in *Dicer* and rs3757 in *DGCR8*, are linked to complex diseases [16, 67, 68]. Since the global repression of miRNA maturation affect cells at several points, most of variants in miRNA-processing machinery genes may have deleterious effects on the miRNAome. Therefore, they are expected to be very rare and not available in the current GWAS. Accordingly, genetic variants can occur in promoter regions of miRNAs and result in an altered transcription rate. In silico approaches have been used to predict the effect of variants in miRNA gene promoters [69, 70]. Nevertheless, experimental investigation of this type of variants can be complicated by the presence of an independent miRNA promoter in addition to the host gene promoter (for intronic miRNAs). In **Chapter 3.3** we showed one example of such variants which is located in the promoter region of miR-142 and is significantly associated with AD. We found several pieces of evidence showing that the variant may affect miR-142 expression and consequently alter AD risk. In addition, genetic variants in splice site of miRNA host genes (for intronic miRNAs) or of the polycistron (for clustered miRNAs) could also result in aberrant miRNA expression patterns [71]. Future studies are needed to systematically investigate the association of variants in miRNA machinery genes and miRNA regulatory regions with age-related disorders and uncover how these variants may influence the function of miRNAs and alter the expression of genes involved in age-related disorders..

Another point of attention is the complexity of miRNA-mediated gene regulation [72]. miRNAs can regulate multiple target genes in the same or different biological pathways. Further, several miRNAs can target the same gene and interact with each other creating miRNA-miRNA co-targeting networks. Oversimplification of miRNA functions, based on the traditional view of linear signaling pathways, led to ignorance the reality of biological complexity. The complexity of the miRNA network is further intensified by recognition of the presence of isomiRs (sequences that have variations with respect to

the reference miRNA, which may not be found yet) and alternative miRNA processing, which bypasses *Drosha* and instead uses a splicing technique to generate pre-miRNAs from short intronic sequences [16]. These sequences are not mapped to the genome and are lost in the analysis of RNA-seq data. How genetic variants affect these alternative pathways and miRNA forms need to be investigated in future studies.

CLINICAL IMPLICATIONS

MiRNA as therapeutic targets. Age-related disorders are the most common health threats in developed countries. Despite the success of broadly prescribed drugs for these complex disorders, they continue to increase in prevalence worldwide, indicating the need for deeper insights into the disease mechanisms and innovative therapeutic strategies. In recent years, miRNAs have become a source of great excitement as regulators of complex disorders because they provide new insights into disease mechanisms and can be therapeutically targeted. However, much remains to be learned about the precise mechanisms of miRNA action within human disease pathways and whether such modest regulators can, indeed, be modulated in the setting of disease. In this thesis we identified a number of miRNAs associated with age-related disorders that may be of clinical importance (**Figure 2**). The potential therapeutic role of miRNAs has extensively studied in cardiovascular disease, such as ongoing phase II clinical trials of miRNA inhibitory molecules for miR-208a and miR-33 [73]. The administration of miRNA-targeted therapies for long-term risk modification warrants the consideration of several potential adverse effects [73]. While we are making important steps forward in developing inhibitory and mimic miRNA chemistries as a novel therapy, numerous obstacles and questions remain on the path towards the development of miRNA-based therapeutics in general [73]. For instance, only a few miRNAs have a very cell or tissue-specific expression pattern (e.g. miR-122 in liver), but many miRNAs are broadly expressed and may have multiple effects in different tissues. One undesirable side effect of chronically administered miRNA therapeutics could be diminished by modulation of miRNA expression in a specific tissue. This could be achieved by targeted delivery of miRNA therapeutics or specifically targeting miRNAs that show limited expression in a tissue of interest or specific regulation of a biochemical process. These therapeutic strategies may minimize adverse off-tissue effects but does not address the potential adverse on-target effects of chronic miRNA antagonism in the cell type of interest. Further optimization of the inhibitors or mimics for the identified miRNAs, as well as the search for techniques for efficient and safe delivery *in vivo*, will trigger advancement of miRNA therapeutics.

MiRNA as diagnostic biomarkers. miRNAs have been detected in a range of biological fluids such as blood plasma and serum, urine, saliva and cerebrospinal fluid (CSF) [74-76]. Many studies have reported that miRNAs are stable in body fluids and their expression profiles can change under different physiological and pathological conditions [74, 77]. An increasing number of studies have also suggested miRNAs as putative noninvasive biomarkers in diagnosis, prognosis and response to treatment [74, 77]. Circulating cell-free miRNAs can be usually packaged in micro-vesicles, exosomes or apoptotic bodies, as well be associated with Argonaut protein family or lipoprotein complexes, and thereby are resistant to RNases present in extracellular environments [78, 79]. Although serum miRNAs remain stable after being exposed to severe conditions, their low enrichment in fluids, such as plasma and CSF, make isolation and quantification of circulating cell-free miRNAs difficult [77, 80]. There are different methods available to detect the circulating cell-free miRNAs [81]. Currently, the most frequently adopted methods have been the reverse transcription real-time quantitative PCR (RT-qPCR) and the next generation sequencing (NGS). In order to consider the identified miRNAs in this thesis as diagnostic biomarkers for age-related disorders, the miRNAs have to be validated, for instance by individual RT-qPCR assays, in body fluids of patients from large epidemiologic studies with validated clinical data.

REFERENCES

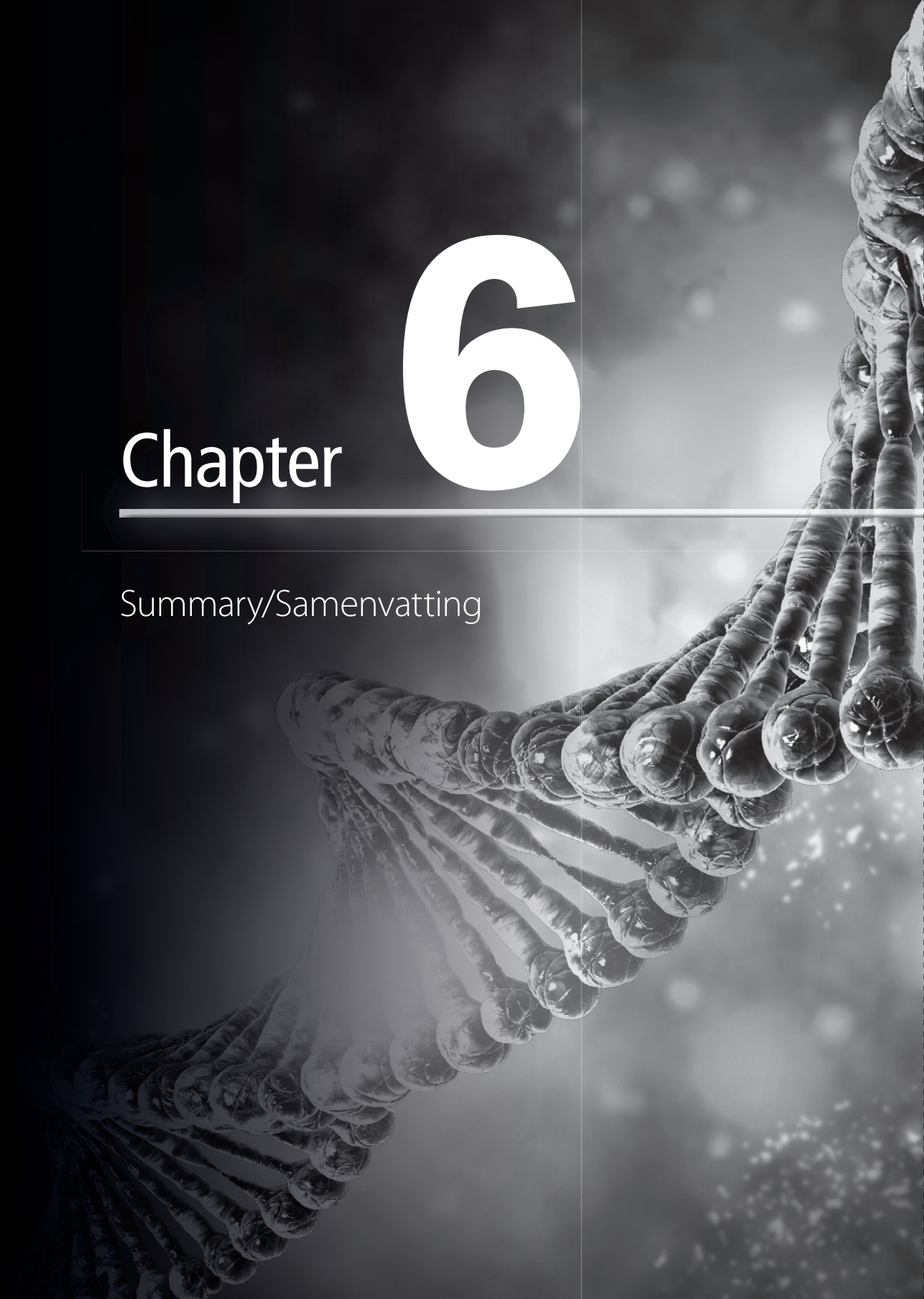
1. Pasquinelli AE: MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. Nat Rev Genet 2012, 13:271-282.
2. Abe M, Bonini NM: MicroRNAs and neurodegeneration: role and impact. Trends Cell Biol 2013, 23:30-36.
3. Geekiyanage H, Chan C: MicroRNA-137/181c regulates serine palmitoyltransferase and in turn amyloid beta, novel targets in sporadic Alzheimer's disease. J Neurosci 2011, 31:14820-14830.
4. Long JM, Lahiri DK: MicroRNA-101 downregulates Alzheimer's amyloid-beta precursor protein levels in human cell cultures and is differentially expressed. Biochem Biophys Res Commun 2011, 404:889-895.
5. Patel N, Hoang D, Miller N, Ansaloni S, Huang Q, Rogers JT, Lee JC, Saunders AJ: MicroRNAs can regulate human APP levels. Mol Neurodegener 2008, 3:10.
6. Small EM, Frost RJ, Olson EN: MicroRNAs add a new dimension to cardiovascular disease. Circulation 2010, 121:1022-1032.
7. Richardson K, Nettleton JA, Rotllan N, Tanaka T, Smith CE, Lai CQ, Parnell LD, Lee YC, Lahti J, Lemaitre RN, et al: Gain-of-function lipoprotein lipase variant rs13702 modulates lipid traits through disruption of a microRNA-410 seed site. Am J Hum Genet 2013, 92:5-14.
8. Gupta SK, Bang C, Thum T: Circulating microRNAs as biomarkers and potential paracrine mediators of cardiovascular disease. Circ Cardiovasc Genet 2010, 3:484-488.
9. Shin VY, Chu KM: MiRNA as potential biomarkers and therapeutic targets for gastric cancer. World J Gastroenterol 2014, 20:10432-10439.
10. Tan L, Yu JT, Tan L: Causes and Consequences of MicroRNA Dysregulation in Neurodegenerative Diseases. Mol Neurobiol 2015, 51:1249-1262.
11. Zi Y, Yin Z, Xiao W, Liu X, Gao Z, Jiao L, Deng L: Circulating MicroRNA as Potential Source for Neurodegenerative Diseases Biomarkers. Mol Neurobiol 2015, 52:1494-1503.
12. Global Lipids Genetics C, Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, Ganna A, Chen J, Buchkovich ML, et al: Discovery and refinement of loci associated with lipid levels. Nat Genet 2013, 45:1274-1283.
13. International Consortium for Blood Pressure Genome-Wide Association S, Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, Chasman DI, Smith AV, Tobin MD, Verwoert GC, et al: Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature 2011, 478:103-109.
14. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, DeStafano AL, Bis JC, Beecham GW, Grenier-Boley B, et al: Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet 2013, 45:1452-1458.
15. Nalls MA, Pankratz N, Lill CM, Do CB, Hernandez DG, Saad M, DeStefano AL, Kara E, Bras J, Sharma M, et al: Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. Nat Genet 2014, 46:989-993.
16. Ryan BM, Robles AI, Harris CC: Genetic variation in microRNA networks: the implications for cancer research. Nat Rev Cancer 2010, 10:389-402.
17. Bhattacharya A, Ziebarth JD, Cui Y: PolymiRTS Database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. Nucleic Acids Res 2014, 42:D86-91.

18. Gong J, Liu C, Liu W, Wu Y, Ma Z, Chen H, Guo AY: An update of miRNASNP database for better SNP selection by GWAS data, miRNA expression and online tools. Database (Oxford) 2015, 2015:bav029.
19. Petersen CP, Bordeleau ME, Pelletier J, Sharp PA: Short RNAs repress translation after initiation in mammalian cells. Mol Cell 2006, 21:533-542.
20. Mencia A, Modamio-Hoybjor S, Redshaw N, Morin M, Mayo-Merino F, Olavarrieta L, Aguirre LA, del Castillo I, Steel KP, Dalmay T, et al: Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss. Nat Genet 2009, 41:609-613.
21. Richardson K, Lai CQ, Parnell LD, Lee YC, Ordovas JM: A genome-wide survey for SNPs altering microRNA seed sites identifies functional candidates in GWAS. BMC Genomics 2011, 12:504.
22. Cammaerts S, Strazisar M, De Rijk P, Del Favero J: Genetic variants in microRNA genes: impact on microRNA expression, function, and disease. Front Genet 2015, 6:186.
23. Gong J, Tong Y, Zhang HM, Wang K, Hu T, Shan G, Sun J, Guo AY: Genome-wide identification of SNPs in microRNA genes and the SNP effects on microRNA target binding and biogenesis. Hum Mutat 2012, 33:254-263.
24. Jia P, Zhao Z: Network.assisted analysis to prioritize GWAS results: principles, methods and perspectives. Hum Genet 2014, 133:125-138.
25. Saunders MA, Liang H, Li WH: Human polymorphism at microRNAs and microRNA target sites. Proc Natl Acad Sci U S A 2007, 104:3300-3305.
26. Krol J, Sobczak K, Wilczynska U, Drath M, Jasinska A, Kaczynska D, Krzyzosiak WJ: Structural features of microRNA (miRNA) precursors and their relevance to miRNA biogenesis and small interfering RNA/short hairpin RNA design. J Biol Chem 2004, 279:42230-42239.
27. Agarwal V, Bell GW, Nam JW, Bartel DP: Predicting effective microRNA target sites in mammalian mRNAs. Elife 2015, 4.
28. Iglesias AI, Springelkamp H, Ramdas WD, Klaver CC, Willemsen R, van Duijn CM: Genes, pathways, and animal models in primary open-angle glaucoma. Eye (Lond) 2015, 29:1285-1298.
29. Zhou T, Souzeau E, Sharma S, Siggs OM, Goldberg I, Healey PR, Graham S, Hewitt AW, Mackey DA, Casson RJ, et al: Rare variants in optic disc area gene CARD10 enriched in primary open-angle glaucoma. Mol Genet Genomic Med 2016, 4:624-633.
30. Cammaerts S, Strazisar M, Dierckx J, Del Favero J, De Rijk P: miRVaS: a tool to predict the impact of genetic variants on miRNAs. Nucleic Acids Res 2016, 44:e23.
31. Wu M, Jolicoeur N, Li Z, Zhang L, Fortin Y, L'Abbe D, Yu Z, Shen SH: Genetic variations of microRNAs in human cancer and their effects on the expression of miRNAs. Carcinogenesis 2008, 29:1710-1716.
32. Liu F, Ikram MA, Janssens AC, Schuur M, de Koning I, Isaacs A, Struchalin M, Uitterlinden AG, den Dunnen JT, Sleegers K, et al: A study of the SORL1 gene in Alzheimer's disease and cognitive function. J Alzheimers Dis 2009, 18:51-64.
33. Vardarajan BN, Zhang Y, Lee JH, Cheng R, Bohm C, Ghani M, Reitz C, Reyes-Dumeyer D, Shen Y, Rogaeva E, et al: Coding mutations in SORL1 and Alzheimer disease. Ann Neurol 2015, 77:215-227.
34. Puthiyedth N, Riveros C, Berretta R, Moscato P: Identification of Differentially Expressed Genes through Integrated Study of Alzheimer's Disease Affected Brain Regions. PLoS One 2016, 11:e0152342.
35. He L, Kernogitski Y, Kulminskaya I, Loika Y, Arbeev KG, Loiko E, Bagley O, Duan M, Yashkin A, Ukraintseva SV, et al: Pleiotropic Meta-Analyses of Longitudinal Studies Discover Novel Genetic Variants Associated with Age-Related Diseases. Front Genet 2016, 7:179.

36. Sivakumaran S, Agakov F, Theodoratou E, Prendergast JG, Zgaga L, Manolio T, Rudan I, McKeigue P, Wilson JF, Campbell H: Abundant pleiotropy in human complex diseases and traits. Am J Hum Genet 2011, 89:607-618.
37. Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW: Pleiotropy in complex traits: challenges and strategies. Nat Rev Genet 2013, 14:483-495.
38. Hu Z, Liang J, Wang Z, Tian T, Zhou X, Chen J, Miao R, Wang Y, Wang X, Shen H: Common genetic variants in pre-microRNAs were associated with increased risk of breast cancer in Chinese women. Hum Mutat 2009, 30:79-84.
39. Tian T, Shu Y, Chen J, Hu Z, Xu L, Jin G, Liang J, Liu P, Zhou X, Miao R, et al: A functional genetic variant in microRNA-196a2 is associated with increased susceptibility of lung cancer in Chinese. Cancer Epidemiol Biomarkers Prev 2009, 18:1183-1187.
40. Estrada K, Styrkarsdottir U, Evangelou E, Hsu YH, Duncan EL, Ntzani EE, Oei L, Albagha OM, Amin N, Kemp JP, et al: Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. Nat Genet 2012, 44:491-501.
41. Keklikoglou I, Hosaka K, Bender C, Bott A, Koerner C, Mitra D, Will R, Woerner A, Muenstermann E, Wilhelm H, et al: MicroRNA-206 functions as a pleiotropic modulator of cell proliferation, invasion and lymphangiogenesis in pancreatic adenocarcinoma by targeting ANXA2 and KRAS genes. Oncogene 2015, 34:4867-4878.
42. Szabo G, Bala S: MicroRNAs in liver disease. Nat Rev Gastroenterol Hepatol 2013, 10:542-552.
43. Fu X, Jin L, Wang X, Luo A, Hu J, Zheng X, Tsark WM, Riggs AD, Ku HT, Huang W: MicroRNA-26a targets ten eleven translocation enzymes and is regulated during pancreatic cell differentiation. Proc Natl Acad Sci U S A 2013, 110:17892-17897.
44. Hesketh J: 3'-Untranslated regions are important in mRNA localization and translation: lessons from selenium and metallothionein. Biochem Soc Trans 2004, 32:990-993.
45. Skeeles LE, Fleming JL, Mahler KL, Toland AE: The impact of 3'UTR variants on differential expression of candidate cancer susceptibility genes. PLoS One 2013, 8:e58609.
46. Chen K, Song F, Calin GA, Wei Q, Hao X, Zhang W: Polymorphisms in microRNA targets: a gold mine for molecular epidemiology. Carcinogenesis 2008, 29:1306-1311.
47. Sethupathy P, Collins FS: MicroRNA target site polymorphisms and human disease. Trends Genet 2008, 24:489-497.
48. Edwards SL, Beesley J, French JD, Dunning AM: Beyond GWASs: illuminating the dark road from association to function. Am J Hum Genet 2013, 93:779-797.
49. Kichaev G, Yang WY, Lindstrom S, Hormozdiari F, Eskin E, Price AL, Kraft P, Pasaniuc B: Integrating functional data to prioritize causal variants in statistical fine-mapping studies. PLoS Genet 2014, 10:e1004722.
50. Xu S, Witmer PD, Lumayag S, Kovacs B, Valle D: MicroRNA (miRNA) transcriptome of mouse retina and identification of a sensory organ-specific miRNA cluster. J Biol Chem 2007, 282:25053-25066.
51. Smailhodzic D, Klaver CC, Klevering BJ, Boon CJ, Groenewoud JM, Kirchhof B, Daha MR, den Hollander AI, Hoyng CB: Risk alleles in CFH and ARMS2 are independently associated with systemic complement activation in age-related macular degeneration. Ophthalmology 2012, 119:339-346.
52. Ragusa M, Caltabiano R, Russo A, Puzzo L, Avitabile T, Longo A, Toro MD, Di Pietro C, Purrello M, Reibaldi M: MicroRNAs in vitreus humor from patients with ocular diseases. Mol Vis 2013, 19:430-440.
53. Black JR, Clark SJ: Age-related macular degeneration: genome-wide association studies to translation. Genet Med 2016, 18:283-289.

54. Zhou Q, Gallagher R, Ufret-Vincenty R, Li X, Olson EN, Wang S: Regulation of angiogenesis and choroidal neovascularization by members of microRNA-23~27~24 clusters. Proc Natl Acad Sci U S A 2011, 108:8287-8292.
55. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL: Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev 2011, 25:1915-1927.
56. Ponting CP, Oliver PL, Reik W: Evolution and functions of long noncoding RNAs. Cell 2009, 136:629-641.
57. Bhan A, Mandal SS: Long noncoding RNAs: emerging stars in gene regulation, epigenetics and human disease. ChemMedChem 2014, 9:1932-1956.
58. Huarte M: The emerging role of lncRNAs in cancer. Nat Med 2015, 21:1253-1261.
59. Wapinski O, Chang HY: Long noncoding RNAs and human disease. Trends Cell Biol 2011, 21:354-361.
60. Broadbent HM, Peden JF, Lorkowski S, Goel A, Ongen H, Green F, Clarke R, Collins R, Franzosi MG, Tognoni G, et al: Susceptibility to coronary artery disease and diabetes is encoded by distinct, tightly linked SNPs in the ANRIL locus on chromosome 9p. Hum Mol Genet 2008, 17:806-814.
61. Pasmant E, Sabbagh A, Vidaud M, Bieche I: ANRIL, a long, noncoding RNA, is an unexpected major hotspot in GWAS. FASEB J 2011, 25:444-448.
62. Yan B, Yao J, Liu JY, Li XM, Wang XQ, Li YJ, Tao ZF, Song YC, Chen Q, Jiang Q: lncRNA-MIAT regulates microvascular dysfunction by functioning as a competing endogenous RNA. Circ Res 2015, 116:1143-1156.
63. Luo X, Yang W, Ye DQ, Cui H, Zhang Y, Hirankarn N, Qian X, Tang Y, Lau YL, de Vries N, et al: A functional variant in microRNA-146a promoter modulates its expression and confers disease risk for systemic lupus erythematosus. PLoS Genet 2011, 7:e1002128.
64. Chen LL, Shen Y, Zhang JB, Wang S, Jiang T, Zheng MQ, Zheng ZJ, Chen CX: Association between polymorphisms in the promoter region of pri-miR-34b/c and risk of hepatocellular carcinoma. Genet Mol Res 2016, 15.
65. Sorensen SS, Nygaard AB, Christensen T: miRNA expression profiles in cerebrospinal fluid and blood of patients with Alzheimer's disease and other types of dementia - an exploratory study. Transl Neurodegener 2016, 5:6.
66. Chaudhuri AD, Yelamanchili SV, Marcondes MC, Fox HS: Up-regulation of microRNA-142 in simian immunodeficiency virus encephalitis leads to repression of sirtuin1. FASEB J 2013, 27:3720-3729.
67. Clague J, Lippman SM, Yang H, Hildebrandt MA, Ye Y, Lee JJ, Wu X: Genetic variation in MicroRNA genes and risk of oral premalignant lesions. Mol Carcinog 2010, 49:183-189.
68. Zhou Y, Wang J, Lu X, Song X, Ye Y, Zhou J, Ying B, Wang L: Evaluation of six SNPs of MicroRNA machinery genes and risk of schizophrenia. J Mol Neurosci 2013, 49:594-599.
69. Marsico A, Huska MR, Lasserre J, Hu H, Vucicevic D, Musahl A, Orom U, Vingron M: PROmiRNA: a new miRNA promoter recognition method uncovers the complex regulation of intronic miRNAs. Genome Biol 2013, 14:R84.
70. Saini HK, Griffiths-Jones S, Enright AJ: Genomic analysis of human microRNA transcripts. Proc Natl Acad Sci U S A 2007, 104:17719-17724.
71. Janas MM, Khaled M, Schubert S, Bernstein JG, Golan D, Veguilla RA, Fisher DE, Shomron N, Levy C, Novina CD: Feed-forward microprocessing and splicing activities at a microRNA-containing intron. PLoS Genet 2011, 7:e1002330.
72. Olson EN: MicroRNAs as therapeutic targets and biomarkers of cardiovascular disease. Sci Transl Med 2014, 6:239ps233.

73. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, Powell C, Vedantam S, Buchkovich ML, Yang J, et al: Genetic studies of body mass index yield new insights for obesity biology. Nature 2015, 518:197-206.
74. Katoh M: Cardio-miRNAs and onco-miRNAs: circulating miRNA-based diagnostics for non-cancerous and cancerous diseases. Front Cell Dev Biol 2014, 2:61.
75. Kumar P, Dezso Z, MacKenzie C, Oestreicher J, Agoulnik S, Byrne M, Bernier F, Yanagimachi M, Aoshima K, Oda Y: Circulating miRNA biomarkers for Alzheimer's disease. PLoS One 2013, 8:e69807.
76. Sheinerman KS, Umansky SR: Circulating cell-free microRNA as biomarkers for screening, diagnosis and monitoring of neurodegenerative diseases and other neurologic pathologies. Front Cell Neurosci 2013, 7:150.
77. Batistela MS, Josviak ND, Sulzbach CD, de Souza RL: An overview of circulating cell-free microRNAs as putative biomarkers in Alzheimer's and Parkinson's Diseases. Int J Neurosci 2016:1-12.
78. Turchinovich A, Weiz L, Burwinkel B: Extracellular miRNAs: the mystery of their origin and function. Trends Biochem Sci 2012, 37:460-465.
79. Turchinovich A, Weiz L, Langheinz A, Burwinkel B: Characterization of extracellular circulating microRNA. Nucleic Acids Res 2011, 39:7223-7233.
80. Wang WT, Chen YQ: Circulating miRNAs in cancer: from detection to therapy. J Hematol Oncol 2014, 7:86.
81. Moldovan L, Batte KE, Trgovcich J, Wisler J, Marsh CB, Piper M: Methodological challenges in utilizing miRNAs as circulating biomarkers. J Cell Mol Med 2014, 18:371-390.



Chapter 6

Summary/Samenvatting

ENGLISH SUMMARY

MiRNAs are small non-coding RNA regulators of gene expression that play important roles in diverse biological processes and are frequently shown to be implicated in human diseases. In this thesis we performed genome-wide scans to identify miRNAs involved in age-related disorders (including cardiovascular, neurodegenerative and ophthalmic disorders) using genetic evidence followed by various *in silico* and *in vitro* validation studies.

In **Chapter 2.1** we examined the association of genetic variants in miRNA seed sequences with different cardiovascular risk factors and disease. We found a variant in the seed sequence of miR-4513 which is significantly associated with multiple cardiometabolic traits. Performing *in silico* and experimental studies, we showed that the variant affects both miR-4513 expression and function. Our results further indicated that miR-4513 may play a pleiotropic role in cardiovascular disease. In **Chapter 2.2** we examined the association between genetic variants in miRNA genes, beyond the seed sequence, and cardiometabolic disorders. We identified two associated polymorphisms including a variant in pre-miR-196a2 associated with waist to hip ratio (WHR) and a variant in pre-miR-1908 associated with serum lipid and glucose. The variant in miR-196a2 has been shown to reduce the mature miRNA expression. We experimentally showed that miR-196a2 regulates the expression of *SFMBT1* and *HOXC8*, which are two known genes involved in WHR. In **Chapter 2.3** we extracted genetic variants in miRNA-binding sites within gene 3'UTRs and investigated their associations with cardiovascular risk factors and disease. We identified 195 associated miRNA-binding site variants. We prioritized the associated variants based on a defined set of criteria for being functional and appointed 10 of them for validation studies. We experimentally demostrated the functional impact of the appointed variants on miRNA-mediated gene regulation. In **Chapter 2.4** we investigated the association of variants in lncRNAs with ten cardiovascular risk factors and disease. We identified 179 cardiovascular-associated variants that are located in lncRNAs. Of these, we selected 55 SNPs, either the sentinel SNP or in strong LD with the sentinel SNP, for further investigations. We found a number of lncRNA SNPs that are likely to be functional variants in their loci. Our results in this chapter further showed a number of loci (hosting the identified functional variants) for which the role of lncRNA may have been overlooked in the literature.

In **Chapter 3.1** we sought to identify miRNA-related variants associated Alzheimer's disease (AD). We found a variant in miR-1229 associated with AD that increase the miRNA expression. Our results highlighted four miR-1229-3p putative targets that are associated with AD and might mediate the downstream effect of miR-1229-3p in relation to AD. Of these, we experimentally demonstrated miR-1229-3p-mediated regulation of *SORL1*, which is a known gene involved in the pathogenesis of AD. Further, we found

multiple 3'UTR variants in miRNA-binding sites that are associated with AD and could potentially affect miRNA-mediated gene regulation in AD. In **Chapter 3.2** we identified two variants in miR-4519 and miR-548at significantly associated with Parkinson's disease (PD). Our analyses showed that these variants affect expression levels of the miRNAs. We then highlighted target genes that may mediate the function of these miRNAs in PD. Of which, we experimentally showed that *NSF* is a direct target of miR-4519. Moreover, we found a number of 3'UTR variants associated with PD that are likely to act through affecting miRNA-mediated regulation of their genes involved in PD. In **Chapter 3.3** our systematic investigation of ncRNA variants associated with Alzheimer's disease determined a new susceptibility locus (17q22) for AD. The lead variant overlaps BZRAP1-AS1 and miR-142 host gene. Our subsequent *in silico* and *in vitro* studies showed that upregulation of miR-142 is implicated in increased risk of AD.

In **Chapter 4.1** we performed a genome-wide scan for miRNAs contributing to Age-related macular degeneration (AMD). We identified three miRNA-variants associated with AMD that decrease expression levels of four miRNAs (miR-4513, miR-122/miR-3591 and miR-3135b). We highlighted the target genes of these miRNAs that are associated with AMD and have the potential to mediate the miRNA functions in AMD. Moreover, we found a number of variants associated with AMD in miRNA-binding sites. We prioritized the associated variants based on their functional potential to affect miRNA-mediated regulation of the host genes. Among them, we experimentally showed a *CFB* 3'UTR variant in the binding site of miR-210-5p decreases the miRNA-mediated regulation and results in increased *CFB* expression. Finally, in **Chapter 4.2** we examined the association of miRNA-related variants with four glaucoma endophenotypes, aiming to identify miRNAs that are involved in developing Primary Open-Angle Glaucoma (POAG). We found two variants in miR-612 and miR-4707 to be significantly associated with POAG endophenotypes. The first variant has been shown to increase the expression level of miR-612. We showed that, while miR-4707 down-regulates the expression of *CARD10*, the second variant decreases the miRNA-mRNA interaction, resulting in increased *CARD10* expression. In addition, we found several miRNA-binding site variants associated with POAG endophenotypes that have the potential to affect miRNA-mediated regulation of their genes involved in associated phenotypes.

NEDERLANDSE SAMENVATTING

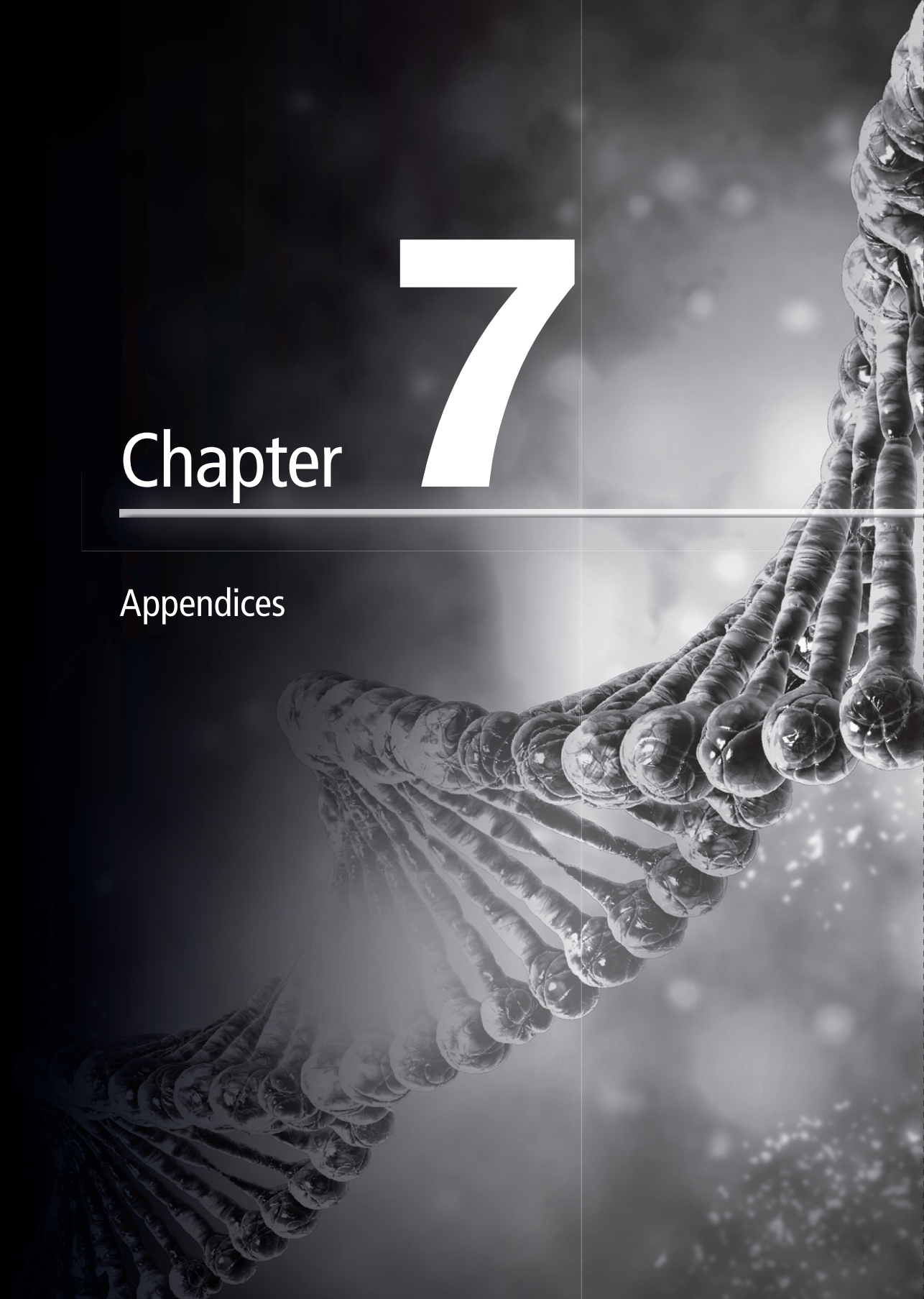
MiRNAs zijn kleine niet-coderende RNA regulatoren van genexpressie die een belangrijke rol spelen in diverse biologische processen. Gezien de brede klinische en therapeutische mogelijkheden van miRNAs is het van belang om te onderzoeken of en welke miRNAs een rol spelen in leeftijdsgerelateerde aandoeningen. In dit proefschrift hebben we genoom-brede scans uitgevoerd om miRNAs te identificeren die betrokken zijn bij hart en vaatziekten, neurodegeneratieve en oogheelkundige aandoeningen met behulp van genetische gegevens, gevolgd door verschillende in silico en in vitro validatie studies.

In **hoofdstuk 2.1** hebben we een dataset gemaakt van genetische variaties in miRNA-coderende sequenties. We hebben gekeken naar de associatie van variaties in miRNA seed sequenties met cardiovasculaire risicofactoren en ziekte. We vonden een variantie in de seed sequentie van miR-4513 die significant geassocieerd is met meerdere cardiometabole eigenschappen. Met het uitvoeren van in silico en experimentele analyses toonden we aan dat de variatie zowel de expressie als de functie van miR-4513 beïnvloed. Onze resultaten suggereren dat miR-4513 een pleiotrope rol speelt in hart en vaatziekten. In **hoofdstuk 2.2** onderzochten we de associatie tussen genetische varianties in miRNA genen, verder dan alleen de seed sequentie, en cardiometabole eigenschappen en ziekten. We identificeerden twee geassocieerde polymorfieën , waaronder een variant in de pre-miR-196a2 geassocieerd met een taille tot heup ratio (WHR) en een andere variate in de pre-miR-1908 geassocieerd de serum lipiden en glucose. Het is aangetoond dat de variatie in miR-196a2 de expressie van mature miRNA vermindert. We hebben experimenteel aangetoond dat miR-196a2 de expressie van *SFMBT1* en *HOXC8*, twee genen betrokken bij WHR, reguleert. In **hoofdstuk 2.3** hebben we een omvangrijke dataset gemaakt van 3'UTR variaties die zich vermoedelijk bevinden in miRNA bindingsplaatsen en hun associatie onderzocht met cardiovasculaire risicofactoren en ziekte. We identificeerden 195 variaties in miRNA-bindingsplaatsen die geassocieerd zijn met de onderzochte factoren . We gaven prioriteit aan varianties op basis van een gedefinieerde set van criteria voor functionaliteit en benoemde tien van hen voor de validatie studies. We hebben de impact van deze 10 geselecteerde variaties op de miRNA-gemedieerde regulatie van hun gastheer genen experimenteel bevestigd. In **hoofdstuk 2.4** onderzochten we de associatie tussen variaties in lncRNAs met tien cardiovasculaire risicofactoren en ziekte. We identificeerden 179 cardiovasculair geassocieerde variaties in lncRNAs. Hiervan selecteerden we 55 SNPs, ofwel de sentinel SNP of in sterke LD met de sentinel SNP, voor verder onderzoek. We toonden aan dat veel van deze geselecteerde lncRNA SNPs waarschijnlijk functioneel zijn in hun loci. Onze resultaten in dit hoofdstuk betwisten het conventionele model voor de aanwijzing van de met het kenmerk geassocieerde variatie, waarbij niet-coderende variaties omstanders zijn in de verwijzing naar LD blokken van coderende variaties.

In **hoofdstuk 3.1** zochten we naar nieuwe miRNA-gerelateerde variaties die geassocieerd zijn met de ziekte van Alzheimer (AD) . We identificeerde een functionele variatie in miR-1229 dat de miRNA expressie verhoogd. We selecteerde vier mogelijke doelwitten van miR-1229-3p die geassocieerd zijn met AD en mogelijk een downstream effect kunnen hebben voor miR-1229-3p in AD. We hebben de miR-1229-3p gemedieerde regulering van *SORL1*, een bekend gen betrokken bij de pathogenese van AD, experimenteel bevestigd. Verder vonden we meerdere variaties in miRNA bindingsplaatsen geassocieerd met AD die mogelijk van invloed kunnen zijn op de miRNA-gemedieerde regulatie van hun genen geassocieerd met AD. In **hoofdstuk 3.2** identificeerden we twee variaties in miR-4519 en miR-548at die significant geassocieerd zijn met de ziekte van Parkinson (PD). Het uitvoeren van in silico en in vitro analyses toonden aan dat de variaties invloed hebben op expressieniveaus van de miRNAs. Vervolgens hebben we targetgenen geselecteerd die de functie van de miRNAs kan bemiddelen bij PD. We stelde experimenteel vast dat *NSF* een direct doelwit is van miR-4519. Bovendien vonden we een aantal 3'UTR variaties geassocieerd met PD die waarschijnlijk werken via miRNA gemedieerde regulatie van hun genen. In **hoofdstuk 3.3** in ons systematisch onderzoek naar ncRNA variaties geassocieerd met de ziekte van Alzheimer is mogelijk een nieuwe locus (17q22) voor AD geïdentificeerd. De lead variant overlapt met *BZRAP1-AS1* en miR-142 host-gen. Onze *in silico* en *in vitro* studies toonden vervolgens dat miR-142 betrokken is bij AD.

In **hoofdstuk 4.1** voerde we een genoomwijde scan uit voor miRNAs die bijdragen aan leeftijdsgebonden maculadegeneratie (AMD). We identificeerden drie miRNA-variaties geassocieerd met AMD die invloed hebben op de expressie niveaus van vier miRNAs (miR-4513, miR-122 / miR-3591 en miR-3135b). We selecteerde de target genen van deze miRNAs die geassocieerd zijn met AMD met de potentie om de miRNA functie te medieren in AMD. Bovendien vonden we een aantal variaties geassocieerd met AMD in miRNA bindingsplaatsen. We selecteerde geassocieerde variaties op basis van hun functionele mogelijkheden om miRNA-gemedieerde regulering van de gastheer genen te beïnvloeden. We toonde experimenteel aan dat een *CFB* 3'UTR variatie in de bindingsplaats van miR-210-5p de miRNA-gemedieerde regulering vermindert en resulteert in verhoogde expressie van *CFB*. Tenslotte wordt in **hoofdstuk 4.2** de associatie onderzocht van miRNA-gerelateerde variaties met vier glaucoom endofenotypes, met als doel miRNAs en genen te identificeren die betrokken zijn bij de ontwikkeling van primair open-hoek glaucoom (POAG). We vonden twee variaties in miR-612 en miR-4707 die significant geassocieerd zijn met POAG. We selecteerde targetgenen die kunnen worden gedereguleerd als gevolg van deze variaties, waaronder *CARD10* voor miR-4707. We toonden aan dat miR-4707 de expressie van CARD10 down-reguleert. Daarnaast hebben we verschillende 3'UTR variaties geassocieerd met POAG in miRNA-

bindingsplaatsen met de potentie om miRNA-gemedieerde regulering van hun genen die geassocieerd zijn met POAG.



Chapter 7

Appendices

ACKNOWLEDGEMENTS

This thesis is the result of five long years which passed in a blink of an eye. I feel very grateful to have been given the opportunity to work in such an inspiring scientific environment, enjoy constructive academic debates in our meetings and the most important, I have made lifelong friendships. With no doubt, completion of this thesis would have not been possible without the impact of many people who contributed in different stages of my PhD project. I take the opportunity to express my gratitude as follows:

First of all, I wish to express my special gratitude to my promotor Prof. Oscar H. Franco. Dear Oscar, you gave me the chance to come in your department and continue my work. The completion of my PhD is the result of your trust in me. Thank you for believing in me. You are a true source of motivation.

Secondly, I would like to extend my gratitude to my first co-promotor, Dr. Abbas Dehghan who bridged softly my first contact with genetic epidemiology. I want to specially thank you for your patience with me during the first six months of PhD period when I had to bother you almost every day with basic questions about the field. Your tutorials on genetics and epidemiological methods in Farsi set the foundations of my expertise.

I wish to sincerely thank my second co-promotor Dr. Stefan Erkeland, who was always provoking and bringing me into the depth of biological reasoning of my projects. I want to specially thank you for your guidance and technical advice to perform the experimental work of my PhD. It was a pleasure working under your mentorship, demanding but highly effective supervision.

I would like to also thank my third co-promotor Dr. Magda Meester-Smoor for giving me the chance to extend my PhD project for one more year. You joined the team in the last year of my PhD, however, I learned a lot from your expertise in eye-related phenotypes. Thank you for your trust on me and your support during this year.

During these years, I feel lucky to have shared my office with great roommates. From room 2901 (the 2901_club): Paul, I want to specially thank you for your great help and patience with me to learn R and work with big data during my first months of my PhD. Symen, thank you for sharing your thoughts and suggestions on my projects; your multidimensional perspective on science was definitely interesting. Sanaz, thank you for your help during my PhD especially in the beginning when you were sharing with me your SPSS expertise. Your friendship beyond work has always been valuable to me. Layal, it was always a pleasure to discuss with you about science in a thoughtful and practical perspective. Also, I so much enjoyed our socio-political discussion after a hard day at work. From the 2901_club_new: Jana, you have been an active presence during the last one and half year. It is very interesting to have such an active colleague as a roommate to chat about scientific and non-scientific topics during our daily breaks. Also, I really appreciate your help and support for my defense. Adela, you have been a calm-

ing presence on our furious discussions about genetics in the office, but I can't say the same in other topics. From room 2905: Taulant, you are always a super active and smart person with new ideas for epidemiological research in different fields. Oscar, I enjoyed so much our conversations about life and I really hope to read your name at NEJM. It was a pleasure for me being an office mate with you guys and I will never forget our memories during these years.

I would also like to thank all my old roommates at the Department of Epidemiology from the Valkenburg library (Anna, Myrte, Klodian, Lisanne, Charlotte, and Thirsa), who made it a pleasant place after hard days of moving from department of hematology to epidemiology. I want to specially thank Anna who was my first neighbor at the Department of Epidemiology and helped me a lot with her tips and smiles in the early days of moving to the new environment. I want to thank other members of the cardiovascular group such as Maryam, Janine, Jessica, Trudy, Paula, Eliana, Kim, Loes, Ester, Audry, Hoyan, Silvana, Irma, Vincent, Jelena, Carolina and all others who I missed their names here for providing a scientific and collaborative environment during these years, it was a pleasure working with you. I also very grateful Mirjam for her great help with administrative procedures, Frank for data management support and Nano for the computer support.

An interesting part of my PhD project has been collaborating with other groups within our department and outside. From the Department of Hematology, I want to acknowledge the support received from Farshid who helped me a lot to learn practical issues working in a scientific laboratory. I also thank Hans for his great contribution and help in the experimental part of my projects. From the eye group, I am very grateful to Prof. Caroline Klaver and Dr. Magda Meester-Smoor who gave me the chance to extend my PhD project for one year and do some projects on eye diseases. Thank you for your trust on me and your support during this year. From the Neurology group I would like to thank Prof. Arfan Ikram. I have really enjoyed collaborating with you on projects about the role of miRNAs in Alzheimer and Parkinson diseases which resulted in interesting publications. Dear Arfan, I would also like to thank you for the advice you gave me when I had some difficult carrier choices to make and for supporting part of my PhD project. From the genetics lab, Department of Internal medicine, I would like to thank Prof. Andre Uitterlinden, Dr. Joyce van Meurs, and Dr. Fernando Rivadeneira for their suggestions and supports for the projects in this thesis. From the Department of Immunology, I would like to thank Steven, Roos, Inge, Marion, Chris and Jenna. I would also like to thank PhDs in the aforementioned groups for their contributions to the projects that we did together during in the last five years, Carolina, Jeroen, Cindy, Annamarie, Milly, Peter, Sirwan, Marjolein, Adriana and all others who I missed their names.

I would like to thank my other Iranian friends in the Netherlands, inside and outside Erasmus MC, Raha, Setareh, Fatemeh, Marjan, Nima, Yaser, Parham, Ronak, Behnam and

all others who I missed their names here, for your support and all the scientific and non-scientific discussion during these years. I specially want to thank my precious paranymphs, Sam and Lea, and also Frazin and Sima for all their help, advice and support.

Naturally, there are also people in my personal sphere who I want to take this chance to thank. My wife Sara, thank you for your support, care, encouragement and smiles. I can't articulate how valuable it has been to have you when coming home every night, allowing me to forget about stress during my PhD. I want to thank you for listening to my never-ending genetic talks, for bearing with my frustration regarding work or when my papers were rejected and many other moments when I was tired of work. Thank you for the lovely life we share together with Shahrad. Dear Shahrad, when you came into my life everything just started making more sense. My purpose in life felt fulfilled when I wore the proud crown of "Babi" that you gave me. You are the prime source of my inspiration and an unlimited source of joy. My best of love to both of you, Sara and Shahrad!

و اما تشکر از خانواده‌ام: پدر عزیزم که سنبل سختکوشی و اراده است و مادرم عزیزم که نمونه واقعی یه مادر صبور و مهربونه، هر چی دارم از برکت وجود شما هاست. همچنین خواهرها و برادرای عزیزم که همیشه حمایتهای شما انگیزه بوده واسه ادامه این راه طولانی. از همتون ممنونم.

PHD PORTFOLIO

Name PhD student: Mohsen Ghanbari
Erasmus MC department: Epidemiology
PhD period: September 2011 – December 2012 (Hematology), December 2012 - December 2016 (Epidemiology)
Promotor: Prof.dr. Oscar H. Franco
Copromotors: Dr. Abbas Dehghan, Dr. Stefan Erkeland, Dr. Magda Meester-Smoor

PhD training	Year	Workload (ECTS/ Hours)
Master of Science (MSc), Genetic Epidemiology, NIHES, Erasmus Medical Center, Rotterdam, the Netherlands	2014-2015	
Courses	**Year**	**ECTS**
Biomedical Research Techniques	2011	1.6
Basic Human Genetics	2011	0.5
SNPs and Human Diseases	2012	1.4
Research Management for PhD-students	2012	1.0
Gene Expression Analyzing by R and Web tools	2012	2.0
Laboratory Animal Science (Article 9, except exam)	2012	2.0
Basic Introduction Course on SPSS	2012	0.8
First Encounter with Next Generation Sequencing Data	2013	1.4
Genome Wide Association Analysis	2013	1.4
Genomics in Molecular Medicine	2013	1.4
Advances in Genomic Research	2013	0.4
Principals of Genetic Epidemiology	2013	0.7
Methods of Public Health Research	2013	0.7
Study Design	2013	4.3
Bio-statistical Methods I	2013	5.7
Linux for Scientists	2013	0.6
Genetic-Epidemiology Research Methods	2013	5.1
Advances in Genome-Wide Association Studies	2013	1.4
Family-based Genetic Analysis	2013	1.4
Bio-statistical Methods II	2014	4.3
Introduction to R Software	2014	0.7
Principals of Research in Medicine	2014	0.7
Health Economics	2014	0.7
Introduction to Writing Methods	2014	1.1
The practice of Epidemiological Analysis	2015	0.7
Logistic Regression	2015	1.4
Casual Mediation Analysis	2015	0.7
Molecular Diagnostic	2015	1.0
Scientific Integrity	2017	0.3
Ensembl workshop	2017	0.5

International Conferences and Meetings		
CHARGE meeting, Rotterdam, the Netherlands	2013	Attendance
European microRNA Meeting, Cambridge, UK	2013	Poster
European Human Genetics Conference, Millan, Italy	2014	Poster
Ingenuity Pathway Analysis workshop, Nijmegen, the Netherlands	2014	Attendance
Mutation in the Genome, Leiden, the Netherlands	2015	Oral
American Human Genetics Meeting, Baltimore, USA	2015	Poster
International Society of Genetic Epidemiology Meeting, Baltimore, USA	2015	Poster
European Human Genetics Conference, Barcelona, Spain	2016	Poster
The non-coding Genome, Leuven, Belgium	2016	Attendance
European Human Genetics Conference, Copenhagen, Denmark	2017	Poster
Internal Presentations		
Cardiovascular group meeting (5x presentations)	2013-2017	1.0
Epidemiology Department 2020 meeting (1x presentations)	2014-2016	0.25
Molecular Epidemiology meeting (2x presentations)	2014-2016	0.5
Genetic epidemiology meeting (1x presentation)	2014	0.25
Internal medicine meeting (2x presentations)	2014-2015	0.5
Hematology Department meeting (2x presentations)	2011-2012	0.5
Work Discussions and Group Meetings		
Work report and discussion with promoter (Monthly)	2013-2017	2.0
Work discussion with co-promoters (Bi-weekly)	2013-2017	2.0
Group meetings (Bi-weekly)	2013-2017	2.0
Cardiovascular group meetings (Bi-weekly)	2013-2017	2.0
Molecular Epidemiology meetings (Bi-weekly)	2013-2017	2.0
Epidemiology Department seminar series (Bi-weekly)	2013-2017	2.0
Internal medicine meetings	2014-2015	0.5
Genetic epidemiology meetings	2014-2015	0.5
Teaching (Lecturer)		
General Epidemiology (Assistant, NIHES course ESP1)	2016	0.5
Advances in Genome-Wide Association Studies (NIHES course GE03)	2017	1.0
Cardiovascular Epidemiology Course (NIHES course EP20)	2017	1.0
Genomics in Molecular Medicine (NIHES course ESP57)	2017	1.0
Supervising Students		
Paula Batista, DSc (NIHES) and PhD	2014-2017	2.0
Silvana Mass, MSc (NIHES) and PhD	2015-2017	1.0
Arthur de Almedia Vieira, MSc (NIHES)	2015-2016	1.0
Irma Karabegovic, MSc (NIHES)	2016-2017	1.0
Mariana Almeida, MSc (MolMed)	2017-2018	1.0
Michelle Mens (University of Amsterdam)	2017-2018	1.0
Peer Reviews for Scientific Journals		
Scientific Reports	2016	1
International journal of Alzheimer	2016	1
Human Mutation	2016	1

1 ECTS (European Credit Transfer System) equal to workload of 28 hrs

LIST OF PUBLICATIONS

PUBLICATIONS BASED ON THIS THESIS

1. **Ghanbari M**, Erkeland S, Xu L, Colijn J, Hofman A, Uitterlinden A, Franco O.H, Dehghan A, Klaver C.C.W, Meester-Smoor M.A. Genetic variants in microRNAs and their target sites within gene 3'UTRs associate with susceptibility to age-related macular degeneration (AMD), *Human Mutation*, 2017.
2. **Ghanbari M**, Ikram MA, de Looper HW, Hofman A, Erkeland SJ, Franco OH, Dehghan A. Genome-wide identification of microRNA-related variants associated with risk of Alzheimer's disease. *Scientific Reports*, 2016.
3. **Ghanbari M**, Darweesh SK, de Looper HW, van Luijn MM, Hofman A, Ikram MA, Franco OH, Erkeland SJ, Dehghan A. Genetic Variants in MicroRNAs and Their Binding Sites Are Associated with the Risk of Parkinson Disease. *Human Mutation*, 2016.
4. **Ghanbari M**, Franco OH, de Looper H, Hofman A, Erkeland S, Dehghan A. Genetic variations in miRNA binding sites affect miRNA-mediated regulation of several genes associated with cardiometabolic phenotypes. *Circulation Cardiovascular Genetic*, 2015.
5. **Ghanbari M**, Sedaghat S, de Looper H, Hofman A, Erkeland S, Franco OH, Dehghan A. The association of common polymorphisms in miR-196a2 with waist to hip ratio and miR-1908 with serum lipid and glucose. *Obesity*, 2015.
6. **Ghanbari M**, de Vries PS, de Looper H, Peters MJ, Schurmann C, Yaghootkar H, Dörr M, Frayling TM, Uitterlinden AG, Hofman A, van Meurs JB, Erkeland SJ, Franco OH, Dehghan A. A Genetic Variant in the Seed Region of miR-4513 Shows Pleiotropic Effects on Lipid and Glucose Homeostasis, Blood Pressure, and Coronary Artery Disease. *Human Mutation*, 2014.
7. **Ghanbari M**, Iglesias A, Springelkamp H, Erkeland S, Dehghan A, Van Duijn C.M, Franco O.H, Klaver C.C.W, Meester M.A. Identification of microRNAs associated with primary open-angle glaucoma using GWAS data, Submitted.
8. **Ghanbari M**, Peters M, de Vries S, Boer C, van Rooij J, Lee Y, Kumar V, BIOS Consortium, Uitterlinden A, Hofman A, Wijmenga C, Ordovas J, Smith C, van Meurs J, Erkeland S, Franco O.H, Dehghan A. Integrative genome-wide analysis highlights the role of long non-coding RNAs in cardiometabolic disorders, Submitted.
9. **Ghanbari M**, Munshi S.T, Adams H.H, Hofman A, Kushner S, Erkeland S, Ikram M.A. Genome-wide analysis of long non-coding RNAs reveals BZRAP1-AS1/microRNA-142 at 17q22 as a new susceptibility locus for Alzheimer's disease, In preparation.

OTHER PUBLICATIONS AND MANUSCRIPTS

10. Nano J, **Ghanbari M**, Wang W, de Vries P, Dhana K, Muka T, Uitterlinden A, van Meurs J, Hofman A, BIOS consortium, Franco O, Pan Q, Darwish Murad S, Dehghan A. An epigenome-wide association study identifies methylation sites associated with liver enzymes and hepatic steatosis. Gastroentrology, 2017.
11. Medina-Gomez M.C, Kemp J, Dimou N, Kreiner-Møller E, Chesi A, Zemel B, Bønnelykke K, Boer C, Ahluwalia T, Bisgaard H, Evangelou E, Heppe D, Bonewald L, Gorski J, **Ghanbari M**, Demissie S, Duque G, Maurano M, Kiel D, van der Eerden B, Ackert-Bicknell C, Reppe S, Gautvik K, Raastad T, Karasik D, van der Peppel J, Jaddoe V, Uitterlinden A, Tobias J, Bagos P, Grant S, Evans D, Rivadeneira F. Bivariate genome-wide association analysis implicates pleiotropic effects at the SREBF1/TOM1L2 locus on bone mineral density and lean mass in children. Nature Communications, 2017.
12. Brahimaj A, Ligthart S, **Ghanbari M**, Ikram MA, Hofman A, Franco OH, Kavousi M, Dehghan A. Novel inflammatory markers for incident pre-diabetes and type 2 diabetes: the Rotterdam Study. Eur J Epidemiol, 2017.
13. van Rooij FJ, Qayyum R, Smith AV, Zhou Y, Trompet S, Tanaka T, Keller MF, Chang LC, Schmidt H, Yang ML, Chen MH, Hayes J, Johnson AD, Yanek LR, Mueller C, Lange L, Floyd JS, **Ghanbari M**, Zonderman AB, Jukema JW, Hofman A, van Duijn CM, Desch KC, Saba Y, Ozel AB, Snively BM, Wu JY, Schmidt R, Fornage M, Klein RJ, Fox CS, Matsuda K, Kamatani N, Wild PS, Stott DJ, Ford I, Slagboom PE, Yang J, Chu AY, Lambert AJ, Uitterlinden AG, Franco OH, Hofer E, Ginsburg D, Hu B, Keating B, Schick UM, Brody JA, Li JZ, Chen Z, Zeller T, Guralnik JM, Chasman DI, Peters LL, Kubo M, Becker DM, Li J, Eiriksdottir G, Rotter JI, Levy D, Grossmann V, Patel KV, Chen CH; BioBank Japan Project., Ridker PM, Tang H, Launer LJ, Rice KM, Li-Gao R, Ferrucci L, Evans MK, Choudhuri A, Trompouki E, Abraham BJ, Yang S, Takahashi A, Kamatani Y, Kooperberg C, Harris TB, Jee SH, Coresh J, Tsai FJ, Longo DL, Chen YT, Felix JF, Yang Q, Psaty BM, Boerwinkle E, Becker LC, Mook-Kanamori DO, Wilson JG, Gudnason V, O'Donnell CJ, Dehghan A, Cupples LA, Nalls MA, Morris AP, Okada Y, Reiner AP, Zon LI, Ganesh SK. Genome-wide Trans-ethnic Meta-analysis Identifies Seven Genetic Loci Influencing Erythrocyte Traits and a Role for RBPMS in Erythropoiesis. American Journal Human Genetics, 2017.
14. De Vries PS, Chasman DI, Sabater-Lleal M, Chen MH, Huffman JE, Steri M, Tang W, Teumer A, Marioni RE, Grossmann V, Hottenga JJ, Trompet S, Müller-Nurasyid M, Zhao JH, Brody JA, Kleber ME, Guo X, Wang JJ, Auer PL, Attia JR, Yanek LR, Ahluwalia TS, Lahti J, Venturini C, Tanaka T, Bielak LF, Joshi PK, Rocanin-Arjo A, Kolcic I, Navarro P, Rose LM, Oldmeadow C, Riess H, Mazur J, Basu S, Goel A, Yang Q, **Ghanbari M**, Willemsen G, Rumley A, Fiorillo E, de Craen AJ, Grotevendt A, Scott R, Taylor KD, Delgado GE, Yao J, Kifley A, Kooperberg C, Qayyum R, Lopez LM, Berentzen TL, Räikkönen K, Mangino

M, Bandinelli S, Peyser PA, Wild S, Trégouët DA, Wright AF, Marten J, Zemunik T, Morrison AC, Sennblad B, Tofler G, de Maat MP, de Geus EJ, Lowe GD, Zoledziewska M, Sattar N, Binder H, Völker U, Waldenberger M, Khaw KT, Mcknight B, Huang J, Jenny NS, Holliday EG, Qi L, Mcevoy MG, Becker DM, Starr JM, Sarin AP, Hysi PG, Hernandez DG, Jhun MA, Campbell H, Hamsten A, Rivadeneira F, Mcardle WL, Slagboom PE, Zeller T, Koenig W, Psaty BM, Haritunians T, Liu J, Palotie A, Uitterlinden AG, Stott DJ, Hofman A, Franco OH, Polasek O, Rudan I, Morange PE, Wilson JF, Kardia SL, Ferrucci L, Spector TD, Eriksson JG, Hansen T, Deary IJ, Becker LC, Scott RJ, Mitchell P, März W, Wareham NJ, Peters A, Greinacher A, Wild PS, Jukema JW, Boomsma DI, Hayward C, Cucca F, Tracy R, Watkins H, Reiner AP, Folsom AR, Ridker PM, O'Donnell CJ, Smith NL, Strachan DP, Dehghan A.. A meta-analysis of 120,246 individuals identifies 19 new loci for fibrinogen concentration using 1000 Genomes Project imputation. Human Molecular Genetics, 2015.

15. Rahimi M, **Ghanbari M**, Fazeli Z, Rouzrokh M, Omrani S, Mirfakhraie R, Omrani MD. Association of SRD5A2 gene mutations with risk of hypospadias in the Iranian population. J Endocrinol Invest, 2016.
16. Bautista-Nino P, Matej D, Danser A.H, de Vries R, Musterd-Bhaggoe U, Meima M, Kavousi M, **Ghanbari M**, Hoeijmakers J.H.J, O'Donnell C, Franceschini N, Janssen G, De Mey J, Liu Y, Shanahan C, Franco OH, Dehghan A, Roks A. Phosphodiesterase 1 regulation is a key mechanism in vascular aging. Clinical Science, 2015.
17. Shirdel A, Azarpazhooh MR, Sahebari M, **Ghanbari M**, Mirfeizi SZ, Hutchinson I, Ziaee A, Rafatpanah H. Association of IL-10 Gene Polymorphisms and Human T Lymphotropic Virus Type I-Associated Myelopathy/tropical Spastic Paraparesis in North-East of Iran (Mashhad). Iran J Basic Med Sci, 2013.
18. Azarpazhooh MR, Hasanpour K, **Ghanbari M**, Rezaee SA, Mashkani B, Hedayati-Moghaddam MR, Valizadeh N, Farid Hosseini R, Foroghipoor M, Soltanifar A, Sahebari M, Azadmanesh K, Hassanshahi G, Rafatpanah H. Human T-lymphotropic virus type 1 prevalence in northeastern Iran, Sabzevar: an epidemiologic-based study and phylogenetic analysis. AIDS Res Hum Retroviruses, 2012.
19. Haber M, Platt DE, Ashrafian Bonab M, Youhanna SC, Soria-Hernanz DF, Martínez-Cruz B, Douaihy B, Ghassibe-Sabbagh M, Rafatpanah H, **Ghanbari M**, Whale J, Balanovsky O, Wells RS, Comas D, Tyler-Smith C, Zalloua PA; Genographic Consortium. Afghanistan's Ethnic Groups Share a Y-Chromosomal, Heritage Structured by Historical Events. PLoS ONE, 2012.
20. Mohammadnejad Z, **Ghanbari M**, Ganjali R, Afshari JT, Heydarpour M, Taghavi SM, Fatemi S, Rafatpanah H. Association between Vitamin D Receptor Gene Polymorphisms and type 1 diabetes mellitus in Iranian population. Molecular Biology Report, 2012.

21. Ahmadnia H, **Ghanbari M**, Moradi MR, Khaje-Dalouee. Effect of cigarette Smoke on Spermatogenesis in Rats. Urology Journal, 2009.
22. Nazarabadi MH, Arghami A, **Ghanbari M**. The relation between consanguineous marriage and mental retardation of the offspring. MJIRI, 2008.
23. Portilla E, **Ghanbari M**, van Meurs J. Danser J, Franco O, Muka T, Roks A, Dehghan A. Association between autophagy and cardiometabolic disorders: the Rotterdam Study. Submitted.
24. H. Adams, K. Wittfeld, N. Jahanshad, G. Roshchupkin, B. Mazoyer, A. Beiser, L. Launer, J. Bis, X. Jian, M. Luciano, S. Trompet, E. Hofer, L. Yanek, P. Kolovos, A. Mahfouz, C. van Duijn, A. Teumer, D. Hibar, G. Chauhan, C. Satizabal, A. Smith, J. Becker, R. Gottesman, M. Bastin, E. Slagboom, S. Seiler, D. Vaidya, A. Uitterlinden, S. Langner, G. Perchey, K. Davis, A. Zijdenbos, J. Rotter, J. Storrs, S. Hagenaars, D. Stott, **M. Ghanbari**, H. Hakonarson, W. Huizinga, P. Sleiman, V. Van Deerlin, N. Amin, W. Hoffmann, P. Parva, D. Knopman, I. Ford, S. Li, Q. Yang, G. Homuth, N. Hosten, M. Nauck, D. Liewald, B. Windham, B. Psaty, J. Romero, D. Janowitz, A. van der Lugt, D. Becker, H. Schmidt, J. van der Grond, I. Deary, T. Mosley, O. Lopez, S. Sigurdsson, C. DeCarli, P. Amouyel, J. Cheung, H. Völzke, M. Vernooij, B. Lelieveldt, R. Poot, P. Nyquist, R. Schmidt, W. Jukema, J. Wardlaw0, M. Fornage, W. Longstreth, V. Gudnason, S. Seshadri, S. Debette, W. Niessen, P. Thompson, H. Grabe, A. Ikram. Genetic architecture of the human anterior commissure. Submitted.
25. Medina-Gomez C, Kemp J, Trajanoska K, Luan J, Chesi A, Ahluwalia T, Mook-Kanamori D, Ham A, Hartwig F, Evans D, Joro R, Nedeljkovic I, Zheng H, Zhu K, Atalay M, Liu C, Nethander M, Broer L, Porleifsson G, Mullin B, Handelman S, Heppe D, Richards B, Wang C, Schraut K, Amin N, Wareham N, Karasik D, Van der Velde N, Ikram A, Zemel B, Zhou Y, Harris T, McGuigan F, Boer C, Bønnelykke K, Ralston S, Robbins J, Walsh J, Zillikens C, Langenberg C, Li-Gao R, Williams F, Akesson K, Jackson R, Sigurdsson G, den Heijer M, Spector T, Garcia M, Pennell C, Horta B, Felix J, Zhao J, Price R, Wilson S, de Mutsert R, Bisgaard h, Nielson C, Jaddoe V, Lakka T, Stefansson K, Scott R, Grant S, van Duijn C, Wilson J, Styrkársdóttir U, Psatty B, Orwoll E, Kiel D, Ohlsson C, Forgetta V, Ntzani E, **Ghanbari M**, Williams G, Bassett D, Evangelou E, Uitterlinden A, Ackert-Bicknell C, Tobias J, Evans D & Rivadeneira F. Genome-Wide Association Study Meta-analysis of TB-BMD Across the Life-course Identifies Thirty-Six Novel Loci and Age-specific. Submitted.
26. **Ghanbari M**, Alemdehy F, de Looper H, Buitenhuis M, Erkeland S. MicroRNAs are involved in the regulation of cell cycle in mouse and human hematopoietic stem cells. In preparation.
27. Jaspers L**, Ghanbari M**, Uitterlinden AG, van Meurs JB, Franco OH, Kavousi M. Healthy ageing score and measures of biological age. In preparation.

28. Karabegovic I, Medinagomez C, Rivadeneira F, Franco OH, **Ghanbari M**. The association of miR-196a2 with bone mineral density. In preparation.
29. CHARGE Consortium Nutrition Working Group. Identification of miRNA-mediated genetic variants for blood lipids; Meta-analyses of functional genome-wide association and interaction with macronutrient intake in the CHARGE consortium. In preparation.
30. **Ghanbari M**, Tedja M, Dehghan A, Klaver C.C.W, Meester-Smoor M.A. MicroRNAs: regulators in the development of refractive errors and myopia. In preparation.

CITATION INDICES:

Google Scholar Citations (July 2017)	
Total number of citations	255
h-index	10
http://scholar.google.nl/citations?user=pwJpmOQAAAAJ&hl=en	Hyperlink

ABOUT THE AUTHOR

Mohsen Ghanbari was born in Mashhad, Iran on 5th May 1978. He completed his secondary education at the Shariati high school in 1999. In 2000, he passed the national university entrance exam of Iran with a high score and received a scholarship from the Ministry of Health to study medicine at his home town at Mashhad University of Medical School (MUMS), one of the top three medical universities in Iran. After graduating as a medical doctor (MD) in 2007, he worked for 4 years as an academic staff and a research associate at MUMS, where he conducted laboratory based research and supervised students till 2011. He then decided to pursue a career in academia. He successfully passed the national exam and got a full scholarship to do his PhD abroad from the Ministry of Health in Iran and MUMS. In order to gain further training in research, in September 2011, he moved to the Netherlands where he was admitted to work as a PhD scientist at the Department of Hematology, Erasmus MC. He learned about the different aspects of basic molecular biology and various laboratory techniques during almost one and a half years working on a challenging project entitled "Functional dissection of microRNAs in hematopoietic stem cells". In December 2012, after his supervisor had to move to another institution, he also moved to the Department of Epidemiology, Erasmus MC to pursue his PhD. Thereafter, he continued his work with a project about "The role of microRNAs in age-related disorders" at the Department of Epidemiology, culminating in this PhD thesis under the supervision of Prof. O.H. Franco and co-supervision of Dr. A. Dehghan, Dr. S. Erkeland and Dr. M. Meester-Smoor. In 2015, he obtained a Master degree (MSc) in genetic epidemiology from the Netherlands Institute for Health Sciences (NIHES), Rotterdam, the Netherlands. In 2016, in collaboration with Dr. M. Meester-Smoor from the eye group at the department of Epidemiology, he received a grant to investigate the role of microRNAs in eye disorders. From January 2017, he started working as a postdoctoral research fellow and group leader in the field of molecular epidemiology and genetics, at the Department of Epidemiology, Erasmus MC. Mohsen is married to Samaneh (Sara) Mokhber and they have one son, Shahrad.