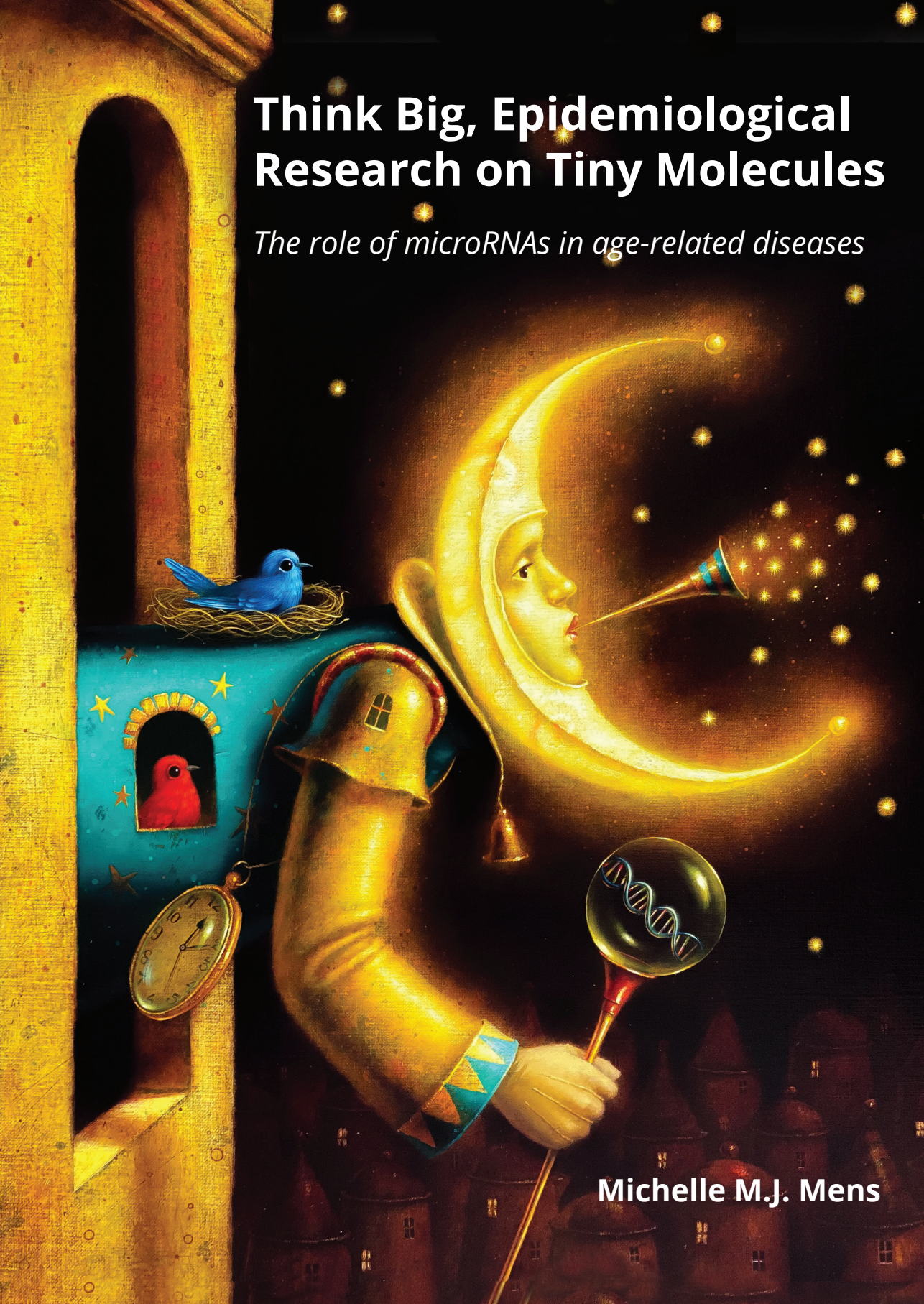
Think Big, Epidemiological Research on Tiny Molecules

The role of microRNAs in age-related diseases



Michelle M.J. Mens

Think Big, Epidemiological Research on Tiny Molecules

The role of microRNAs in age-related diseases

Propositions

1. **Circulatory microRNAs hold great promise as biomarker for metabolic and cardiovascular diseases in the general population - this thesis**
2. **Epigenetic research must go beyond DNA methylation by studying the essential role of other epigenetic markers in particular microRNAs - this thesis**
3. **A signature of plasma-derived microRNAs could help to predict how quickly or slowly an individual ages - this thesis**
4. **The integrative approach to multi-omics data will come with powerful insights about complex diseases, as well as with challenges - this thesis**
5. **Discrepancies in quantification and normalization of microRNA expression data are the obstructive factors why microRNAs fail to take off as biomarkers so far - this thesis**
6. **MicroRNA knocks some sense into senseless - Asher Mullard**
7. **Sure, Big Data Is Great. But So Is Intuition - Steve Lohr**
8. **The outcome measure should not only be exciting, but also the discovery process in which you would like to participate**
9. **Keeping your options open is not the same as having many options, which is largely a state of mind - fueled by curiosity and the ability to think creatively as well as rigorously**
10. **There are no mistakes, only happy little accidents - Bob Ross**
11. **Do not go gentle into that good night - Dylan Thomas**

Think Big, Epidemiological Research on Tiny Molecules

Michelle Machelina Johanna Mens

Acknowledgments

The studies described in thesis were performed within the Rotterdam Study. We gratefully acknowledge the contribution of participants, research staff, data management, and health professionals.

The work presented in this thesis was conducted within the Molecular and Systems Epidemiology group at the Department of Epidemiology of the Erasmus Medical Center Rotterdam, the Netherlands. The Rotterdam Study is supported by the Erasmus Medical Center and the Erasmus University Rotterdam, the Netherlands Organization for Scientific Research (NWO), the Netherlands Organization for Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture, and Science, the Ministry of Health, Welfare and Sports, the European Commission, and the municipality of Rotterdam.

Publication of this thesis was kindly supported by the Department of Epidemiology of the Erasmus Medical Center and Erasmus University Rotterdam.

ISBN:	978-94-6416-685-9
Cover art:	Ronald Companoca for Michelle M.J. Mens (2020)
Lay-out:	Publiss | www.publiss.nl
Print:	Ridderprint | www.ridderprint.nl

Think Big, Epidemiological Research on Tiny Molecules

The role of microRNAs in age-related diseases

Denk groot, epidemiologisch onderzoek naar kleine moleculen

De rol van microRNA's bij leeftijdsgerelateerde ziekten

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. A.L. Bredenoord

and in accordance with the decision of the Doctorate Board.
The public defense shall be held on

Wednesday 13th of October 2021 at 15.30 hours
by

Michelle Machelina Johanna Mens

born in Bergen op Zoom

Erasmus University Rotterdam

Doctoral Committee

Promotor:	Prof.dr. M.A. Ikram
Other members:	Dr. J.B.J. van Meurs Prof.dr. B.T. Heijmans Dr. E.A.C. Wiemer
Copromotor:	Dr. M. Ghanbari
Paranymphs:	Ayda Goharani Rocío López Balarezo

Mille Periculis Supersum

To my parents, Peter & Jacqueline, for their unconditional love and support

Table of contents

Manuscripts that form the basis of this thesis

Chapter 2
Mens MMJ, Ghanbari M. Cell Cycle Regulation of Stem Cells by MicroRNAs. Stem Cell Rev Rep. 2018 Jun;14(3):309-322. doi: 10.1007/s12015-018-9808-y. PMID: 29541978; PMCID: PMC5960494.

Chapter 3
Wu JW, Mens MMJ, Goudsmit J, Ma Y, Liang L, Hofman A, Ikram MA, Ghanbari M. Distinct Plasma MicroRNA Aging Signature Predicts Lifespan but not Health-span. *Submitted*

Chapter 4
Mens MMJ, Mustafa R, Ahmadizar F, Ikram MA, Evangelou M, Kavousi M, Dehghan A, Ghanbari M. MiR-139-5p is a causal biomarker for type 2 diabetes; Results from genome-wide microRNA profiling and Mendelian randomization analysis in a population-based study. *Submitted*

Chapter 5
Zhang X, Mens MMJ, Abozaid YJ, Bos D, Darwish Murad S, de Knegt RJ, Ikram MA, Pan Q, Ghanbari M. Circulatory microRNAs as potential biomarkers for fatty liver disease: the Rotterdam study. Aliment Pharmacol Ther. 2021 Feb;53(3):432-442. doi: 10.1111/apt.16177. Epub 2020 Nov 27. PMID: 33244812; PMCID: PMC7839694.

Chapter 6
Mens MMJ, Maas SCE, Klap J, Weverling GJ, Klatser P, Brakenhoff JPJ, van Meurs JBJ, Uitterlinden AG, Ikram MA, Kavousi M, Ghanbari M. Multi-Omics Analysis Reveals MicroRNAs Associated With Cardiometabolic Traits. Front Genet. 2020 Feb 27;11:110. doi: 10.3389/fgene.2020.00110. PMID: 32174972; PMCID: PMC7056871.

Chapter 7
Mens MMJ, Heshmatollah A, Fani L, Ikram MA, Ikram MK, Ghanbari M. Circulatory MicroRNAs as Potential Biomarkers for Stroke Risk: The Rotterdam Study. Stroke. 2021 Mar;52(3):945-953. doi: 10.1161/STROKEAHA.120.031543. Epub 2021 Feb 10. PMID: 33563011.

Chapter 8
Geurts S, Mens MMJ, Bos MM, Ikram MA, Ghanbari M, Kavousi M. Association between Circulatory microRNAs in Plasma and Atrial Fibrillation among Men and Women from the General Population: the Rotterdam Study. *Submitted*

1

CHAPTER 1

General Introduction

Nothing in biology makes sense except in the light of evolution
- Theodosius Dobzhansky

The history of the Homo Sapiens goes far beyond the hundreds of thousands of years we are walking on planet Earth. As our origin may be relatively novel, the evolutionary timeline, represented as scientific theory goes millions of years back in time. In biology, evolution describes any changes across successive generations in the heritable characteristics of biological populations[1]. This has resulted in diversity at every level of biological organization, from differences in species, to individual organisms, to cellular biology. In this context, we used to think that our genes are the controlling factor that makes each of us unique, both from a physical and health perspective. The human genome consists of approximately 20,000 genes, which are only a tiny part (< 2%) of our DNA[2, 3]. Until recently, the remaining 98% of our DNA was thought to be "Junk DNA", a term coined by Ohno in 1972[4]. This term turned out to be outdated, as scientists soon realized that there are factors within this "Junk DNA", that have major influence on the way genes are controlled. This mechanism was introduced as "epigenetics" [5-7]. Nowadays, epigenetics is considered as at least equivalent, and perhaps more important than our genes themselves.

Epigenetics is the study of heritable changes in organisms caused by modification of gene expression rather than alteration of the DNA itself. In 1969, Griffith and Mahler were the first suggesting that covalent modifications to the DNA may affect the expression of genes[8]. The most studied modification is a methylated cytosine next to a guanine at a 5'-position. This dinucleotide is called a CpG, short for cytosine phosphate guanine. Evidence suggests that cytosines are frequently methylated in promoter regions of silenced genes in comparison to actively transcribed genes[9]. In general, DNA methylation within promoter regions of genes prevents the binding of transcription factors or acts as mediator of gene expression repressors[10, 11]. Besides covalent modifications to the DNA, histone modifications, including acetylation and methylation of histones, affect chromatin structures and thereby transcription of genes[12, 13]. Another phenomena within the epigenetics are non-coding RNAs, which can be characterized into several classes based on their transcript size into small non-coding RNAs (<200 nucleotide) and long non-coding RNAs (>200 nucleotide). In recent years, increasing research has shed light on the potential of non-coding RNAs as post-transcriptional regulators of gene expression.

microRNAs

A mighty flame could begin as a faraway spark
- Dante Alighieri

The most studied small non-coding RNA class are microRNAs (miRNAs), consisting of 20-24 nucleotides. In 1993, miRNAs were first described in *Caenorhabditis elegans* by Ambros and Ruvkun[14, 15]. It was actually during their work on genetic regulation in the roundworm, that they noticed an intriguing complementary relationship between two genes: lin-4 and lin-14. Contrary to their expectations that lin-4 encodes a regulatory gene, they found that lin-4 produces a non-coding string of RNA, with a length of only 22 nucleotides. When they read off the sequences of nucleotides of lin-4 and lin-14, they observed that parts of lin-4 and lin-14 were complementary to each other. Notably, they determined that lin-4 is capable of downregulating the expression of lin-14. In 2000 the second miRNA, let-7, was characterized by Reinhart *et al*.[16]. Soon after, many more miRNAs were discovered and shown to be conserved throughout different species, indicating that miRNAs are widely used as regulatory mechanism and proved the significant evolutionary importance of these small molecules. From 2000 onwards, miRNAs rapidly became an exciting topic of research, reporting the implications of miRNAs in human diseases. According to the miRBase database (v.22), a total number of 2,656 miRNAs have been identified in human[17].

Biogenesis of microRNAs

The genomic location of miRNAs can be discriminated among intergenic and intragenic. Roughly half of the known miRNAs are found to be transcribed from intergenic regions of the genome, suggesting that these miRNAs are transcribed under independent control of regulatory elements of coding genes[18]. The intragenic miRNAs are embedded within sequences of protein-coding genes (host gene), including intronic and exonic regions. If the intragenic miRNA and its host gene share the same promoter, the miRNA is likely to be under control of regulatory elements of the host gene and thus to be co-expressed with the host gene[19]. The maturation process of miRNAs is a complex mechanism that starts in the nucleus, where primary miRNAs (pri-miRNAs) are transcribed from DNA (**Figure 1.1**). These long stem loop-like structures have single-stranded RNA extensions at both ends and are cleaved by enzymes Drosha and DGCR8, yielding precursor miRNAs (pre-miRNAs). These hairpin structure pre-miRNAs are transported to the cytoplasm by Exportin-5 protein. Here, an enzyme called Dicer cleaves the pre-miRNA into a mature miRNA. Subsequently, mature miRNAs can be cleaved by helicase into strands, including 3'-strand and 5'-strand. For clarification of mature miRNAs a nomenclature has been used. MiRNAs originating from the 3′ end or 5′ end of the microRNA gene are denoted with a '-3p' or '-5p' suffix, respectively.

One of the miRNA strands is active and binds to an Ago protein which can then be loaded into a RNA-induced silencing complex (RISC). The miR-RISC complex facilitates base-pairing interaction between miRNAs and 3' untranslated regions (3'UTRs) of target messenger RNA (mRNA). The thermodynamically stable 5' end of a mature miRNA (nucleotides 2-7/8), known as the "seed region", is the essential part in target recognition and interaction. A perfect complementary binding between miRNA and target mRNA can lead to a degradation of the target transcript. If the complementary binding is imperfect, the miRNA can repress the target transcript[20, 21].

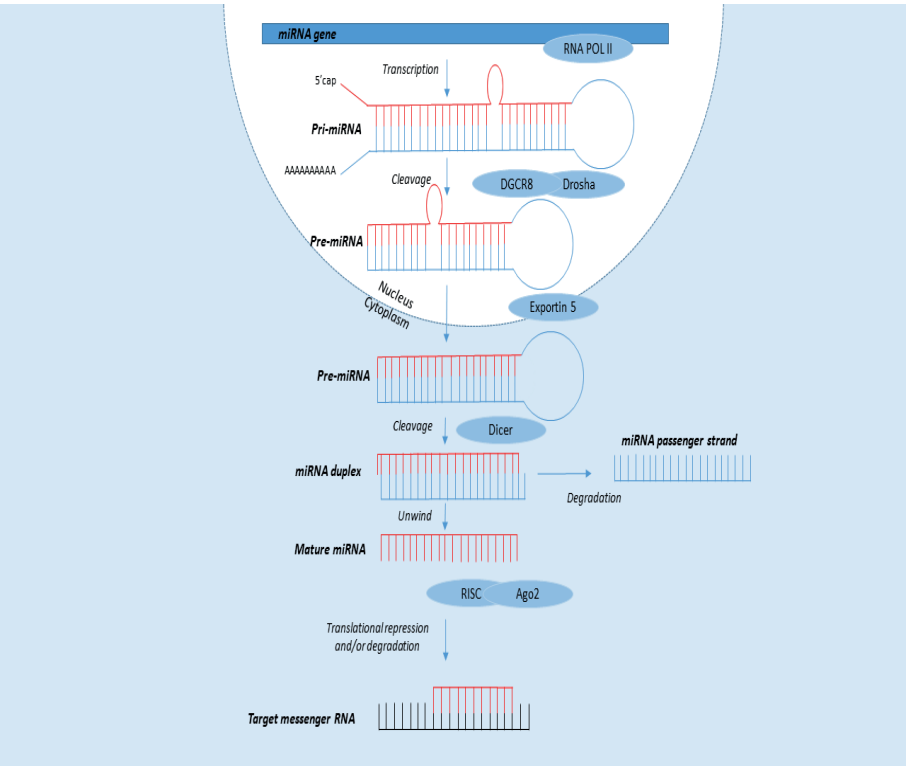


Figure 1.1. The biogenesis of miRNAs. The maturation process of miRNAs begins in the nucleus, where primary miRNAs (pri-miRNAs) are transcribed from double-stranded DNA. Pri-miRNAs have single-stranded RNA extensions at both ends that are cleaved by enzymes Drosha and DGCR8, resulting in precursor miRNAs (pre-miRNAs). These pre-miRNAs are transported to the cytoplasm by Exportin-5 protein. In the cytoplasm, the Dicer enzyme cleaves the pre-miRNA into a mature miRNA. Mature miRNAs can then be cleaved by helicase into strands, including 3 'strand and 5' strand. One of the miRNA strands is active and binds to an Ago protein which can then be loaded into an RNA-induced silencing complex (RISC). The miR-RISC complex facilitates the base pairing interaction between miRNAs and 3 'untranslated regions (3'UTRs) of target messenger RNA (mRNA). Perfect complementary binding between miRNA and target mRNA can lead to degradation of the target transcript. If the complementary binding is not perfect, the miRNA can repress the target messenger RNA transcript.

MicroRNAs as biomarker

MiRNAs have been proposed as instrumental in regulating the interaction of thousands of coding-genes and are involved in a variety of developmental functions, such as cell proliferation and differentiation, as well as pathological conditions[22-24]. Many miRNAs are expressed in a tissue-specific manner[25, 26]. For example, miR-1, miR-133 and miR-208 have cardiac-specific expression patterns and are involved in heart development and disease, including myocardial infarction[27-30]. Another highly tissue-specific miRNA is miR-122, the most abundant hepatic miRNA important for cholesterol and fatty acid synthesis[31] and recognized as marker for hepatocellular carcinoma (HCC)[32], hepatitis C virus (HCV)[33, 34]. Although miRNAs primarily function within cells, previous studies discovered miRNAs circulating in body fluid, including blood[35], saliva[36], serum[37] and urine[38]. Because aberrant expression of miRNAs often results from a pathological process in a specific tissue, miRNAs secreted from these tissues and released into the circulation represent a potentially ongoing pathological process. Indeed, circulatory miRNA profiles differ between healthy donors and individuals with underlying diseases, such as cancer[24, 39] or type 2 diabetes[40]. In addition, given the fact that these miRNAs are easy accessible and remain in a very stable state in the circulation, the interest in using miRNAs as biomarker for disease has grown in recent years[41]. However, miRNAs derived from different types of body fluids were observed to show discrepancy[42]. For example, the levels of miRNAs measured in blood and CSF of Alzheimer's disease patients might be different[43]. In addition, plasma derived miRNAs can show different expression levels as compared to whole blood derived miRNAs. Because whole-blood is cellular, miRNAs are more likely to reflect blood-based features than those caused by disease[44]. This is much less the case with cell-free miRNAs, such as miRNAs derived from plasma. In this thesis, we specially focus on circulatory miRNAs in plasma.

Over the past decade, increasing research has explored the potential of miRNAs as biomarker for common diseases, such as coronary artery disease[45], type 2 diabetes[40] and stroke[46]. To date, the clinical application of the identified biomarkers has not been successful. One of the main reasons for the lack of success is the inability to replicate[47]. Studies are often subject to a small sample size, which is prone to many false positives with inflated effect estimates. Furthermore, miRNAs quantification methods are different, such as qPCR, RNA sequencing and microarray-based hybridization. These methods have all their advantages and disadvantages with regard to sensitivity in identification, quantification, timing and price[48]. Of all methods, next-generation sequencing (NGS)-based RNA sequencing is the most sensitive method that can identify novel miRNAs[48]. In this thesis, we used NGS-RNA sequencing to measure the expression levels of human miRNAs.

Systems biology

Great knowlegde sees all in one. Small knowledge breaks down into the many – Chuang Tzu

Systems biology aims to understand the whole biological system working as a unit, rather than investigating their individual components. This holistic approach deciphers the complexity of biological systems and focuses on the network between different –omics layers, including genomics, epigenomics, transcriptomics, proteomics and metabolomics. The complex interplay between genotype and phenotype is essential for the optimal characterization of diseases. By integrating different omics layers, observed variations in each of these layers can lead to a better understanding of the functional consequences of the pathology of disease. Similar to protein-coding genes, the function of non-coding elements, including miRNAs can be affected by variations on different omics levels.

Genomics: An organism's complete set of DNA is called its genome. The field of genetic epidemiology seeks to define statistical and quantitative analysis of how genetics work in large groups, with as main goal to discover genes involved in diseases. The concept of finding genes for disease by comparing genetic variants between patients and healthy individuals was promising. In genome-wide association studies (GWAS), the observed associations between single nucleotide polymorphisms (SNPs) in, especially common diseases, have led to many novel insights into genetic determinants. Since the first GWAS study published in 2005[49], which was conducted on just over a hundred individuals, the sample sizes of GWAS studies has tremendously increased to over a million individuals in recent studies[50-52]. It has become more usual to combine GWAS results from different studies in order to increase the sample size. These meta-analyses are a collaborative genomic research consortium, often targeting specific characteristics, such as cardiovascular (CARDIoGRAM)[53], diabetic (DIAGRAM)[54] or anthropometrics (GIANT)[55] domains.

As with protein-coding genes, genetic variants can have profound effects on the miRNA expression and function at different levels, from the regulatory regions such as promoter to (pre)-mature sequences[56]. For example, SNPs located in the promoter region of miRNAs can affect the transcription of miRNAs and thus the expression levels[57]. In addition, SNPs in pri- or pre-miRNA sequences can influence the miRNA maturation process, such as the binding affinity between miRNA and proteins (Drosha, Dicer, RISC)[56]. Overall, this can affect the ability to target mRNAs. Given the fact that a single miRNA can target a plethora of mRNAs, it is not surprising that genetic variants in miRNA genes have been linked to a variety of diseases, such as obesity[58] and Alzheimer's disease[57]. Although the number of disease-associated loci has increased over the years, the identified loci explain only a small proportion of the estimated heritability. In fact, for most complex traits, the majority of the heritability remains unclear.

Epigenomics: Unlike genetics, epigenetics can be influenced by a variety of factors such as the environment, diet, and medication use, making epigenetics a dynamic phenomenon. Epigenetics is therefore recognized as an important link between environment and disease risk. An epigenome-wide association study (EWAS) is an approach to examine the association between changes in DNA methylation patterns and complex traits and diseases[59-62]. Also, factors such as smoking can induce changes in DNA methylation, which in turn can affect risk of diseases such as cancer[63]. DNA methylation at CpG sites in promoter regions of miRNA genes are associated with reductions in miRNA expression. Also, DNA methylation in pre-miRNA coding sequences is reported to affect the miRNA biogenesis[64]. A recent EWAS on a subset of miRNAs found that DNA methylation on CpGs was associated with expression levels of miRNAs that are subsequently causal for disease development[65]. In the latter study, about half of the miRNAs identified were located in intragenic regions, and, more interestingly, whose epigenetic regulation of DNA methylation on miRNA expression was independent of host gene expression. This indicates that DNA methylation is an important regulator of miRNA expression and suggests that underlying mechanisms of diseases can be epigenetically driven.

Transcriptomics: The study of genome-wide levels of gene expression, also known as RNA transcripts, is known as transcriptomics. Both genetic variants and DNA methylation can change the expression of genes. These SNPs are indicated as expression quantitative trait loci (eQTL)[66] and CpGs as methylation expression quantitative trait loci (meQTL)[67]. Gene expression is often tissue specific and to study specific diseases one must focus on the tissue of interest. For example, pancreas and liver are most relevant tissues to diabetes, while the brain is most relevant for studying stroke and Alzheimer's disease. However, tissue availability is a major concern, especially in large observational studies. For that reason, less invasive tissue such as blood is often used[68]. Recent large-scale transcriptome-wide association studies (TWAS) can prioritize genes, that were previously been localized by GWAS, and increase the power in detecting functionally relevant loci[69-71]. In addition to eQTLs, a relative new concept studying the association between genetic variants and expression levels of miRNAs, called miR-eQTLs, added a dimension to the transcriptomics[72, 73].

Objectives

The overall aim of this thesis was to investigate the role of miRNAs in aging and age-related diseases, including various metabolic and cardiovascular diseases, using population-based omics data. In this epidemiological approach, I aimed to better understand the role of miRNAs in pathophysiology and identify circulatory miRNAs as potential biomarkers for these diseases.

Study population

The Rotterdam Study

The studies described in this thesis are performed within a large population based cohort study, the Rotterdam Study, also known as the "Erasmus Rotterdam Gezondheid Onderzoek (ERGO)"[74]. In 1990, residents aged 55 years and older residing in Ommoord, a district of Rotterdam, the Netherlands, were invited to participate in the study. Of 10,215 invited inhabitants, 7,983 agreed to participate in the baseline examinations, forming the Rotterdam cohort I (RS-I). In 2000, the Rotterdam Study was extended by including 3,011 participants who had reached 55 years of age or moved into the study district, forming the Rotterdam Study II (RS-II). In 2006, a further extension of the cohort was started in which 3,932 participants, aged at least 45 years living in Ommoord were included to form the Rotterdam Study III (RS-III). In June 2016, recruitment of the Rotterdam cohort IV (RS-IV) began to enroll participants aged 40 years and over, and is ongoing with a target size of around 3,000 participants. All participants were examined in detail at baseline. Briefly, a home interview was conducted and the participants had an extensive set of examinations in the research center in Ommoord. Participants have been re-examined every 3-4 years and have been followed for multiple diseases and determinants. MiRNA profiling has been conducted in 2,000 randomly selected individuals from the fourth round of Rotterdam Study I (RS-I-4) and the second round of Rotterdam Study II (RS-II-2) between January 2002 and December 2005. In addition, the miRNA expression was profiled in 754 randomly selected individuals from Rotterdam Study IV. DNA methylation was measured in a random sample of 1,454 participants from the third round of Rotterdam Study II and first round of Rotterdam Study III. Genotyping was performed in 11,496 individuals from Rotterdam Study I, II and III. **Figure 1.2** illustrates omics layers across different cohorts of the Rotterdam Study.

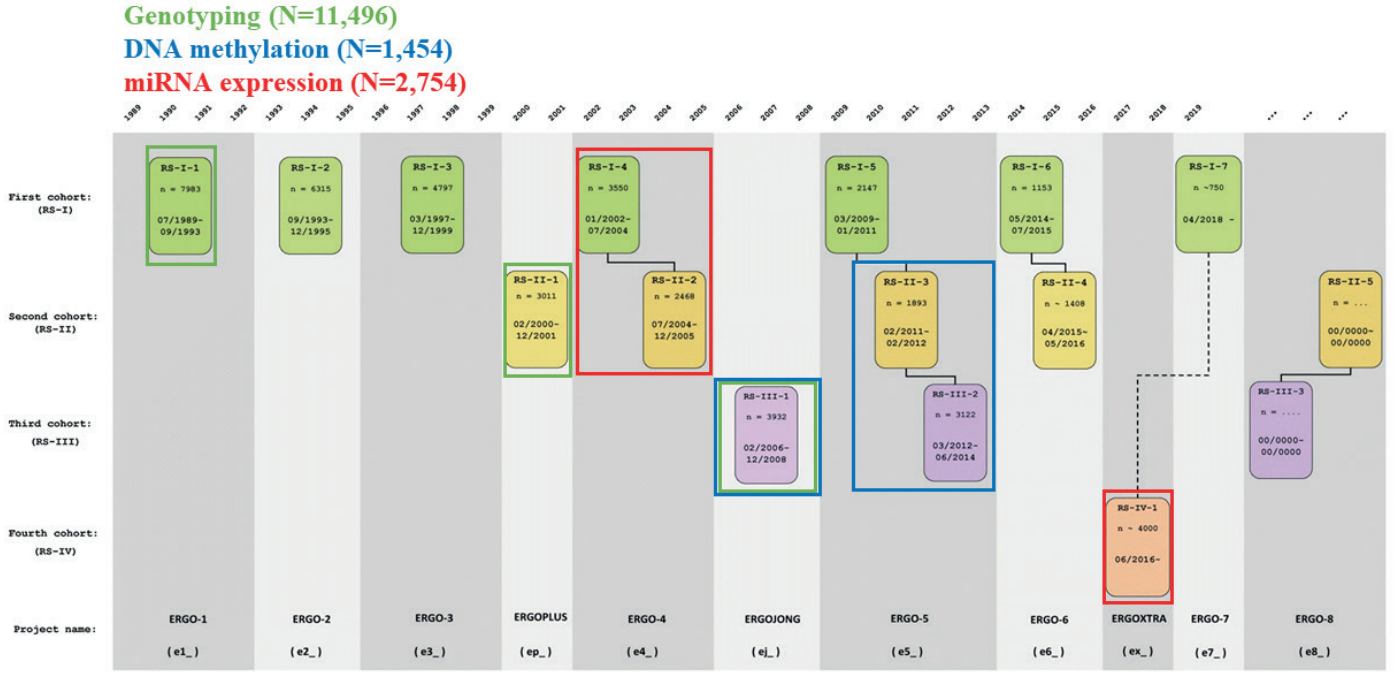


Figure 1.2. Different omics datasets of the Rotterdam Study. Green indicates cohorts in which genotyping was conducted (RS-I-1, RS-II-1, RS-III-1). Blue indicates cohorts in which DNA methylation was conducted (RS-III-1, RS-II-2, RS-III-2). Red indicates cohorts in which miRNA expression was conducted (RS-I-4, RS-II-2, RS-IV-1).

Thesis outline

In **Part I**, I focused on elucidating the role of miRNAs on biological aging. In **chapter 2**, the current knowledge of miRNA regulatory roles in cell cycle progression in different types of stem cells is discussed. **Chapter 3** presents a microRNA aging signature describing the ability to predict lifespan.

In **Part II**, I aimed to investigate miRNAs as potential diagnostic or prognostic biomarkers, and elaborated on possible underlying mechanisms to link the identified miRNAs to the diseases of interest. In **chapter 4,** a genome-wide profiling of plasma miRNA levels was performed and the causal association of miRNAs with type 2 diabetes was studied. In **chapter 5**, I have described the association between circulatory miRNAs and liver enzymes, hepatic steatosis and liver fibrosis at the population level**.** In **chapter 6,** I conducted a multi-omics analysis to identify miRNAs associated with cardiometabolic risk factors and diseases. To do that, I used genetic (GWAS), epigenetic (EWAS), and miRNA expression data. In **chapter 7**, I investigated the association of plasma levels of miRNAs with incident stroke and their potential as biomarker for the disease. Finally, **Chapter 8** presents the association and sex-specific role of miRNAs in atrial fibrillation.

In **Chapter 9** the main findings of the studies described in this thesis, their applications and future directions are discussed together.

References

1. Forbes AAK, B. A. . Evolution Is Change in the Inherited Traits of a Population through Successive Generations. Nature Education Knowledge 2010;3(10):6.
2. Salzberg SL. Open questions: How many genes do we have? BMC Biology 2018;16(1):94.
3. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature 2001;409(6822):860-921.
4. Ohno S. So much "junk" DNA in our genome. Brookhaven Symp Biol 1972;23:366-70.
5. Waddington CH. The Strategy of the Genes; a Discussion of Some Aspects of Theoretical Biology. Allen & Unwin, London 1957.
6. Waddington CH. The Epigenotype. Endeavour 1942:18-20.
7. Russo VEA, Martienssen RA, Riggs AD. Epigenetic mechanisms of gene regulation: Cold Spring Harbor Laboratory Press; 1996.
8. Griffith JS, Mahler HR. DNA Ticketing Theory of Memory. Nature 1969;223(5206):580-582.
9. Naveh-Many T, Cedar H. Active gene sequences are undermethylated. Proc Natl Acad Sci U S A 1981;78(7):4246-50.
10. Boyes J, Bird A. DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. Cell 1991;64(6):1123-1134.
11. Bird A. DNA methylation patterns and epigenetic memory. Genes Dev 2002;16(1):6-21.
12. Shen X, Gorovsky MA. Linker histone H1 regulates specific gene expression but not global transcription in vivo. Cell 1996;86(3):475-83.
13. Grunstein M. Histone acetylation in chromatin structure and transcription. Nature 1997;389 (6649):349-352.
14. 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(5):843-54.
15. 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(5):855-862.
16. Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, et al. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature 2000;403(6772): 901-906.
17. Chen Y, Wang X. miRDB: an online database for prediction of functional microRNA targets. Nucleic Acids Res 2020;48(D1):D127-D131.
18. França GS, Vibranovski MD, Galante PAF. Host gene constraints and genomic context impact the expression and evolution of human microRNAs. Nature communications 2016;7:11438-11438.
19. Lutter D, Marr C, Krumsiek J, Lang EW, Theis FJ. Intronic microRNAs support their host genes by mediating synergistic and antagonistic regulatory effects. BMC Genomics 2010;11(1):224.
20. Shukla GC, Singh J, Barik S. MicroRNAs: Processing, Maturation, Target Recognition and Regulatory Functions. Mol Cell Pharmacol 2011;3(3):83-92.
21. Huntzinger E, Izaurralde E. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. Nature Reviews Genetics 2011;12(2):99-110.
22. Sims RJ, Nishioka K, Reinberg D. Histone lysine methylation: a signature for chromatin function. Trends in Genetics 2003;19(11):629-639.

23. Shenoy A, Blelloch RH. Regulation of microRNA function in somatic stem cell proliferation and differentiation. Nat Rev Mol Cell Biol 2014;15(9):565-76.

24. Peng Y, Croce CM. The role of MicroRNAs in human cancer. Signal Transduction and Targeted Therapy 2016;1(1):15004.

25. Sempere LF, Freemantle S, Pitha-Rowe I, Moss E, Dmitrovsky E, Ambros V. Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation. Genome Biology 2004;5(3):R13.

26. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, et al. MicroRNA expression profiles classify human cancers. Nature 2005;435(7043):834-838.

27. Chistiakov DA, Orekhov AN, Bobryshev YV. Cardiac-specific miRNA in cardiogenesis, heart function, and cardiac pathology (with focus on myocardial infarction). J Mol Cell Cardiol 2016;94:107-121.

28. Yang B, Lin H, Xiao J, Lu Y, Luo X, Li B, et al. The muscle-specific microRNA miR-1 regulates cardiac arrhythmogenic potential by targeting GJA1 and KCNJ2. Nat Med 2007;13(4):486-91.

29. van Rooij E, Sutherland LB, Qi X, Richardson JA, Hill J, Olson EN. Control of stress-dependent cardiac growth and gene expression by a microRNA. Science 2007;316(5824):575-9.

30. Sayed D, Hong C, Chen I-Y, Lypowy J, Abdellatif M. MicroRNAs Play an Essential Role in the Development of Cardiac Hypertrophy. Circulation Research 2007;100(3):416-424.

31. Esau C, Davis S, Murray SF, Yu XX, Pandey SK, Pear M, et al. miR-122 regulation of lipid metabolism revealed by in vivo antisense targeting. Cell Metabolism 2006;3(2):87-98.

32. Tsai WC, Hsu PW, Lai TC, Chau GY, Lin CW, Chen CM, et al. MicroRNA-122, a tumor suppressor microRNA that regulates intrahepatic metastasis of hepatocellular carcinoma. Hepatology 2009;49(5):1571-82.

33. Jopling CL, Yi M, Lancaster AM, Lemon SM, Sarnow P. Modulation of Hepatitis C Virus RNA Abundance by a Liver-Specific MicroRNA. Science 2005;309(5740):1577.

34. Janssen HLA, Reesink HW, Lawitz EJ, Zeuzem S, Rodriguez-Torres M, Patel K, et al. Treatment of HCV Infection by Targeting MicroRNA. New England Journal of Medicine 2013;368(18):1685-1694.

35. Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, et al. Circulating microRNAs as stable blood-based markers for cancer detection. Proceedings of the National Academy of Sciences 2008;105(30):10513.

36. Park NJ, Zhou H, Elashoff D, Henson BS, Kastratovic DA, Abemayor E, et al. Salivary microRNA: discovery, characterization, and clinical utility for oral cancer detection. Clin Cancer Res 2009;15(17):5473-7.

37. Gilad S, Meiri E, Yogev Y, Benjamin S, Lebanony D, Yerushalmi N, et al. Serum MicroRNAs Are Promising Novel Biomarkers. PLOS ONE 2008;3(9):e3148.

38. Hanke M, Hoefig K, Merz H, Feller AC, Kausch I, Jocham D, et al. A robust methodology to study urine microRNA as tumor marker: microRNA-126 and microRNA-182 are related to urinary bladder cancer. Urol Oncol 2010;28(6):655-61.

39. Esquela-Kerscher A, Slack FJ. Oncomirs - microRNAs with a role in cancer. Nat Rev Cancer 2006;6(4):259-69.

40. Jiménez-Lucena R, Camargo A, Alcalá-Diaz JF, Romero-Baldonado C, Luque RM, van Ommen B, et al. A plasma circulating miRNAs profile predicts type 2 diabetes mellitus and prediabetes: from the CORDIOPREV study. Experimental & Molecular Medicine 2018;50(12):1-12.

41. Xi Y, Nakajima G, Gavin E, Morris CG, Kudo K, Hayashi K, et al. Systematic analysis of microRNA expression of RNA extracted from fresh frozen and formalin-fixed paraffin-embedded samples. RNA 2007;13(10):1668-74.

42. Weber JA, Baxter DH, Zhang S, Huang DY, Huang KH, Lee MJ, et al. The microRNA spectrum in 12 body fluids. Clin Chem 2010;56(11):1733-1741.

43. Sørensen SS, Nygaard A-B, 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;5:6-6.

44. Pritchard CC, Kroh E, Wood B, Arroyo JD, Dougherty KJ, Miyaji MM, et al. Blood cell origin of circulating microRNAs: a cautionary note for cancer biomarker studies. Cancer Prev Res (Phila) 2012;5(3):492-497.

45. Fichtlscherer S, De Rosa S, Fox H, Schwietz T, Fischer A, Liebetrau C, et al. Circulating microRNAs in patients with coronary artery disease. Circ Res 2010;107(5):677-84.

46. Sonoda T, Matsuzaki J, Yamamoto Y, Sakurai T, Aoki Y, Takizawa S, et al. Serum MicroRNA-Based Risk Prediction for Stroke. Stroke 2019;50(6):1510-1518.

47. Tiberio P, Callari M, Angeloni V, Daidone MG, Appierto V. Challenges in using circulating miRNAs as cancer biomarkers. Biomed Res Int 2015;2015:731479-731479.

48. Krepelkova I, Mrackova T, Izakova J, Dvorakova B, Chalupova L, Mikulik R, et al. Evaluation of miRNA detection methods for the analytical characteristic necessary for clinical utilization. Biotechniques 2019;66(6):277-284.

49. Klein RJ, Zeiss C, Chew EY, Tsai J-Y, Sackler RS, Haynes C, et al. Complement Factor H Polymorphism in Age-Related Macular Degeneration. Science 2005;308(5720):385.

50. Jansen PR, Watanabe K, Stringer S, Skene N, Bryois J, Hammerschlag AR, et al. Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. Nature Genetics 2019;51(3):394-403.

51. Evangelou E, Warren HR, Mosen-Ansorena D, Mifsud B, Pazoki R, Gao H, et al. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. Nature genetics 2018;50(10):1412-1425.

52. Karlsson Linnér R, Biroli P, Kong E, Meddens SFW, Wedow R, Fontana MA, et al. Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. Nat Genet 2019;51(2):245-257.

53. Nikpay M, Goel A, Won H-H, Hall LM, Willenborg C, Kanoni S, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nature genetics 2015;47(10):1121-1130.

54. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segrè AV, Steinthorsdottir V, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nature Genetics 2012;44(9):981-990.

55. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature 2015;518(7538):197-206.

56. 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;6:186-186.

57. Ghanbari M, Munshi ST, Ma B, Lendemeijer B, Bansal S, Adams HH, et al. A functional variant in the miR-142 promoter modulating its expression and conferring risk of Alzheimer disease. Hum Mutat 2019;40(11):2131-2145.

58. Ghanbari M, Sedaghat S, de Looper HW, Hofman A, Erkeland SJ, Franco OH, et al. 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(2):495-503.

59. Mendelson MM, Marioni RE, Joehanes R, Liu C, Hedman Å K, Aslibekyan S, et al. Association of Body Mass Index with DNA Methylation and Gene Expression in Blood Cells and Relations to Cardiometabolic Disease: A Mendelian Randomization Approach. PLoS Med 2017;14(1):e1002215.

60. Braun KVE, Dhana K, de Vries PS, Voortman T, van Meurs JBJ, Uitterlinden AG, et al. Epigenome-wide association study (EWAS) on lipids: the Rotterdam Study. Clin Epigenetics 2017;9:15-15.

61. Hedman Å K, Mendelson MM, Marioni RE, Gustafsson S, Joehanes R, Irvin MR, et al. Epigenetic Patterns in Blood Associated With Lipid Traits Predict Incident Coronary Heart Disease Events and Are Enriched for Results From Genome-Wide Association Studies. Circ Cardiovasc Genet 2017;10(1):e001487.

62. Cardona A, Day FR, Perry JRB, Loh M, Chu AY, Lehne B, et al. Epigenome-Wide Association Study of Incident Type 2 Diabetes in a British Population: EPIC-Norfolk Study. Diabetes 2019;68(12):2315.

63. Roby J, Allan CJ, Riccardo EM, Luke CP, Lindsay MR, Pooja RM, et al. Epigenetic Signatures of Cigarette Smoking. Circulation: Cardiovascular Genetics 2016;9(5):436-447.

64. Glaich O, Parikh S, Bell RE, Mekahel K, Donyo M, Leader Y, et al. DNA methylation directs microRNA biogenesis in mammalian cells. Nature Communications 2019;10(1):5657.

65. Huan T, Mendelson M, Joehanes R, Yao C, Liu C, Song C, et al. Epigenome-wide association study of DNA methylation and microRNA expression highlights novel pathways for human complex traits. Epigenetics 2020;15(1-2):183-198.

66. Rockman MV, Kruglyak L. Genetics of global gene expression. Nature Reviews Genetics 2006;7(11):862-872.

67. Gibbs JR, van der Brug MP, Hernandez DG, Traynor BJ, Nalls MA, Lai S-L, et al. Abundant Quantitative Trait Loci Exist for DNA Methylation and Gene Expression in Human Brain. PLOS Genetics 2010;6(5):e1000952.

68. Võsa U, Claringbould A, Westra H-J, Bonder MJ, Deelen P, Zeng B, et al. Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. bioRxiv 2018:447367.

69. Mancuso N, Gayther S, Gusev A, Zheng W, Penney KL, Kote-Jarai Z, et al. Large-scale transcriptome-wide association study identifies new prostate cancer risk regions. Nature Communications 2018;9(1):4079.

70. Lappalainen T, Sammeth M, Friedländer MR, 't Hoen PAC, Monlong J, Rivas MA, et al. Transcriptome and genome sequencing uncovers functional variation in humans. Nature 2013;501(7468):506-511.

71. Barbeira AN, Dickinson SP, Bonazzola R, Zheng J, Wheeler HE, Torres JM, et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. Nature Communications 2018;9(1):1825.

72. Nikpay M, Beehler K, Valsesia A, Hager J, Harper M-E, Dent R, et al. Genome-wide identification of circulating-miRNA expression quantitative trait loci reveals the role of several miRNAs in the regulation of cardiometabolic phenotypes. Cardiovascular Research 2019;115(11):1629-1645.

73. Huan T, Rong J, Liu C, Zhang X, Tanriverdi K, Joehanes R, et al. Genome-wide identification of microRNA expression quantitative trait loci. Nature Communications 2015;6(1):6601.

74. Ikram MA, Brusselle G, Ghanbari M, Goedegebure A, Ikram MK, Kavousi M, et al. Objectives, design and main findings until 2020 from the Rotterdam Study. European Journal of Epidemiology 2020;35(5):483-517.

PART I

MicroRNAs in aging

2

CHAPTER 2

Cell Cycle Regulation of Stem Cells by MicroRNAs

Manuscript based on this chapter: Mens MMJ, Ghanbari M. Cell Cycle Regulation of Stem Cells by MicroRNAs. Stem Cell Rev Rep. 2018 Jun;14(3):309-322. doi: 10.1007/s12015-018-9808-y

Abstract

MicroRNAs (miRNAs) are a class of small non-coding RNA molecules involved in the regulation of gene expression. They are involved in the fine-tuning of fundamental biological processes such as proliferation, differentiation, survival and apop- tosis in many cell types. Emerging evidence suggests that miRNAs regulate critical pathways involved in stem cell function. Several miRNAs have been suggested to target transcripts that directly or indirectly coordinate the cell cycle progression of stem cells. Moreover, previous studies have shown that altered expression levels of miRNAs can contribute to pathological conditions, such as cancer, due to the loss of cell cycle regulation. However, the precise mechanism underlying miRNA- mediated regulation of cell cycle in stem cells is still incompletely understood. In this review, we discuss current knowledge of miRNAs regulatory role in cell cycle progression of stem cells. We describe how specific miRNAs may control cell cycle associated molecules and checkpoints in embryonic, somatic and cancer stem cells. We further outline how these miRNAs could be regulated to influence cell cycle progression in stem cells as a potential clinical application.

Introduction

Stem cells and cell cycle regulation

Stem cells are characterized by their unlimited ability to self-renew and capability to differentiate into multiple cell lineages [1]. In this end, stem cells undergo an asymmetric cell division during which only one of the two daughter cells differentiates. This is a complex mechanism in which different transcription factors, epigenetic modifications and hormones are involved. There are two broad types of stem cells including embryonic stem cells (ESCs), which are solely present at the earliest stages of development, and various types of somatic (or adult) stem cells, which appear during fetal development and remain throughout life. ESCs are pluripotent and therefore have the capacity to differentiate into all the possible cell types of the three germ layers. Somatic stem cells, however, are multipotent and can only differentiate into cell types of the specific tissue or organ from which they originate. It is also suggested that a certain type of stem-like cells is responsible for the initiation of cancer, so-called cancer stem cells (CSCs). It is thought that CSCs arise from either differentiated cancer cells or somatic stem cells [2].

In eukaryotes, the cell division cycle includes four discrete phases: Gap 1 (G1), Synthesis (S), Gap 2 (G2) and Mitosis (M). During the G1 phase, which is known as the first interphase, the cell synthesizes proteins that are needed for DNA replication and continuous growth. DNA replication takes place during the S phase and is followed by the G2 phase, which is known as the second interphase, where the DNA integrity is checked. At this point, the cell is growing and preparing for cell division. During the M phase, the cell divides into two daughter cells. After the mitotic phase, the daughter cells re-enter the G1 phase or go into the quiescent state. This is defined as a state of reversible cell cycle arrest and is known as the G0 phase [3]. The quiescent state is important for cellular homeostasis, meaning that it has the ability to either stop proliferating or to re-enter the cell cycle and self-renew when needed [4, 5].

The duration of the cell cycle and the transition from one phase to the next is highly variable between different cell types. For example, while the cell cycle duration in murine somatic cells is relatively long (>16h), the duration in murine ESCs (mESCs) is much faster (8-10h). A reduced G1 phase and prolonged S phase in ESCs are causes of this difference. In addition, human ESCs (hESCs) spend only 3h in the G1 phase, compared to human somatic cells that spend 10h in this phase [6]. The difference in cell cycle duration between ESCs and somatic stem cells is remarkable, an explanation could be that somatic stem cells are predominantly in a quiescent state compared to the fast dividing ESCs. Previous studies have indicated that the G1 phase is the most variable phase and that its duration contributes to cell fate determination [7-9].

When a cell enters the G1 phase, a protein called cyclin D increases in response to mitogenic stimuli. Cyclin D proteins bind to enzymes called CDK4/6 and together they form heterodimers. These complexes subsequently phosphorylate proteins of the retinoblastoma (*RB*) family. The *E2F* family is a group of genes encoding for transcription factors *E2F-1*, *E2F-2* and *E2F-3*, which are downstream targets of the *RB* family. The central member of the *RB* family, the *RB* tumor suppressor protein (pRb), is a negative regulator of the *E2F* genes. When pRb is hypophosphorylated, it inactivates *E2F* transcription factors, which results in the inhibition of transition from G1 to S phase. Hyperphosphorylation of pRb leads to dissociation of *E2F* from the E2F/pRb complex and contributes to the G1/S transition from. Recent findings show the importance of the E2F/pRb activity in relation to ESCs self-renewal and differentiation [10-12].

Cyclin dependent kinase proteins (CDK) tightly regulate the progression of the cell cycle. A CDK binds to its regulatory cyclin protein partner to control the different cell cycle phases. Progression through S phase is regulated by the cyclin E-CDK2 complex, while the G2/M transition is under control of cyclin A and cyclin B-CDK1 complexes. Cyclin dependent kinase inhibitor (CDKI) proteins including p21/Cip1, p27/Kip1 and p57/Kip2, block the activity of cyclin E-CDK2 and cyclin A-CDK1 [13]. Furthermore, proteins of the *INK4* family, including p16/INK4A, p15/INK4B, p18/INK4C and p19/INK4D inhibit the cyclin D-CDK4/6 activity. These mechanisms can lead to cell cycle arrest and are of major importance to regulate tissue homeostasis and prevent tumorigenesis. The p53-p21 signaling pathway is also involved in the transition of G1 to S phase and G2 to M phase. It is well established that loss of p53 is the main reason for genomic instability as the p53-null cells have disrupted the G1/S checkpoint [14-17]. In addition, the expression levels of p53 and p21 in ESCs are important for the maintenance of pluripotency [18].

Biogenesis of microRNAs

Epigenetic features, such as the activity of microRNAs (miRNAs), modulate the expression of cell cycle-associated genes [19-23]. MiRNAs are a conserved class of endogenously expressed small non-coding RNAs (spanning 20-24 nucleotides), that have been widely implicated in fine-tuning various biological processes. Since the discovery of the first miRNA in 1993 [24], the knowledge on miRNAs has been rapidly increased. MiRNAs are ubiquitously expressed in plants, animals and viruses, indicating the evolutionary importance of these small molecules. According to the miRBase database (v.21), 1881 miRNAs have been identified with confidence in human [25]. These miRNAs are suggested to regulate the expression of more than 60% of all protein-coding genes. Previous research has investigated the functional role of miRNAs in diverse mechanisms including cell proliferation, apoptosis, and differentiation. Additionally, alteration in the expression of miRNAs contribute to human diseases such as cancer and cardiovascular disease [26-33].

MiRNA maturation is a complex biological process that is subjected to tight molecular regulation. In the nucleus, miRNAs are initially transcribed as 800-3000nt long primary transcripts (pri-miRNA). These pri-miRNAs are subsequently cleaved by Drosha, RNaseII, endonuclease III, and Pasha/DGCR8 proteins to generate ~70nt hairpin precursor miRNAs (pre-miRNAs). Following this initial process, pre-miRNAs are transported to the cytoplasm by Exportin 5. Subsequently, the hairpin precursor is cleaved in a ~22nt double-stranded miRNA by the ribonuclease III enzyme called Dicer together with TRBP/ PACT proteins. The guide strand (5' end) then associates with members of the Argonaute family and is been incorporated into the RNA-induced silencing complex (RISC). The miR-RISC complex facilitates base-pairing interaction between miRNA and the 3' untranslated region (3'UTR) of target mRNA. The core of a mature miRNA, called the 'seed' region, includes nucleotides 2-7/8 from the 5' end of the miRNA and plays a critical role in target recognition and interaction. Binding of the miRNA seed region to the complementary site in target mRNA leads to translational repression or degradation of the target transcript.

The first studies investigating miRNA function in cell cycle regulation were published two decades ago, where two independent studies revealed that miRNAs lin-4 and let-7 induce cell cycle arrest in the nematode, *C. elegans* [24, 34]. Since then, several studies have demonstrated the importance of miRNAs in cell cycle regulation in different cell types including stem cells [21, 35, 36]. The role of miRNAs in stem cell proliferation was initially observed in knockout mice lacking Dicer and Dgcr8, which are key components of the miRNA biogenesis [37]. Dicer knockout mice were embryonic lethal and ESCs from Dicer-deficient mice exhibited defects in cell cycle progression [38]. Similarly, ESCs derived from Dgcr8-deficient mice exhibited delay in the cell cycle progression due to downregulation of genes involved in regulation of self-renewal [37]. These initial studies indicated that miRNAs are crucial for cell cycle regulation of stem cells. Then, other studies demonstrated that miRNAs are involved in the cell cycle progression of stem cells by direct or indirect targeting of different cell cycle-associated genes (e.g. Cyclins, CDKs and CDKIs). Understanding the tightly regulated networks of the cell cycle in which miRNAs are interacting, will enhance our knowledge in the development of both healthy and disease state of the human body. In the following, we will discuss the recent advances on the functions of miRNAs in cell cycle regulation of stem cells. In addition, a promising therapeutic potential of miRNAs in controlling somatic and cancer stem cells self-renewal and proliferation will be discussed.

MiRNAs and cell cycle regulation of stem cells

Embryonic Stem Cells (ESCs)

The duration of the cell cycle is variable between different types of stem cells. ESCs have a shorter cell cycle compared to somatic stem cells, which is due to a significantly abbreviated G1 phase and a prolonged S phase [39-41]. Previous studies have explored on the phosphorylation status of pRb as a regulator for the length of G1 phase. Since mESCs lack cyclin D-CDK4-as well as cyclin E-CDK2, pRb will not be phosphorylated and therefore not stimulating the cyclin E-CDK2 activity [42]. Therefore, the time spent in G1 phase compared to S phase may be a key feature of the pluripotency fate [12]. Moreover, DNA damage response pathways, which are activated in the G1 phase, are reduced or absent in both hESCs and mESCs [43]. Several negative regulators of cell cycle progression, including p53, p16/INK4A, p19/ARF and p21/Cip1, are expressed at low levels in ESCs, while DNA repair and replication regulators are expressed at high levels [6, 43].

Previous studies have shown the distinct expression pattern of miRNAs in ESCs. These studies demonstrate that ESCs express a set of miRNAs, of which a few are abundantly expressed at 60,000 or more copies per cell. The most abundantly expressed miRNAs in ESCs are miR-290-295, miR-302, miR-17-92, miR-106b-25 and miR-106a-363 clusters, which provide approximately 70% of the total miRNA molecules in ESCs [20, 44-46]. These miRNAs are expressed in homologous clusters, so-called polycistronic loci, which contribute to the same cis-regulatory elements [47]. The miR-290-295 cluster and miR-302 share a highly conserved seed-sequence 'AAGUGCU', while miR-17-92, miR-106b-25 and miR-106a-363 clusters share the seed-sequence 'AAAGUGC' [20]. These miRNAs are called the regulators of the embryonic stem cell cycle (ESCC), because of the ability in rescuing cell cycle progression in Dgcr8 knockout ESCs [20, 44, 48-50]. A schematic overview of the functionality of ESCC miRNAs is illustrated in **Figure 2.1**. In general, ESCC miRNAs facilitate the G1/S transition mainly through suppressing the expression of RB proteins [44]. In addition, these miRNAs have been demonstrated to directly regulate the expression of p21/Cip1 and cyclin E-CDK2 regulatory molecules in mESCs, including *RB*, *RBL1, RBL2*, and *LATS2*,[21, 48-50].

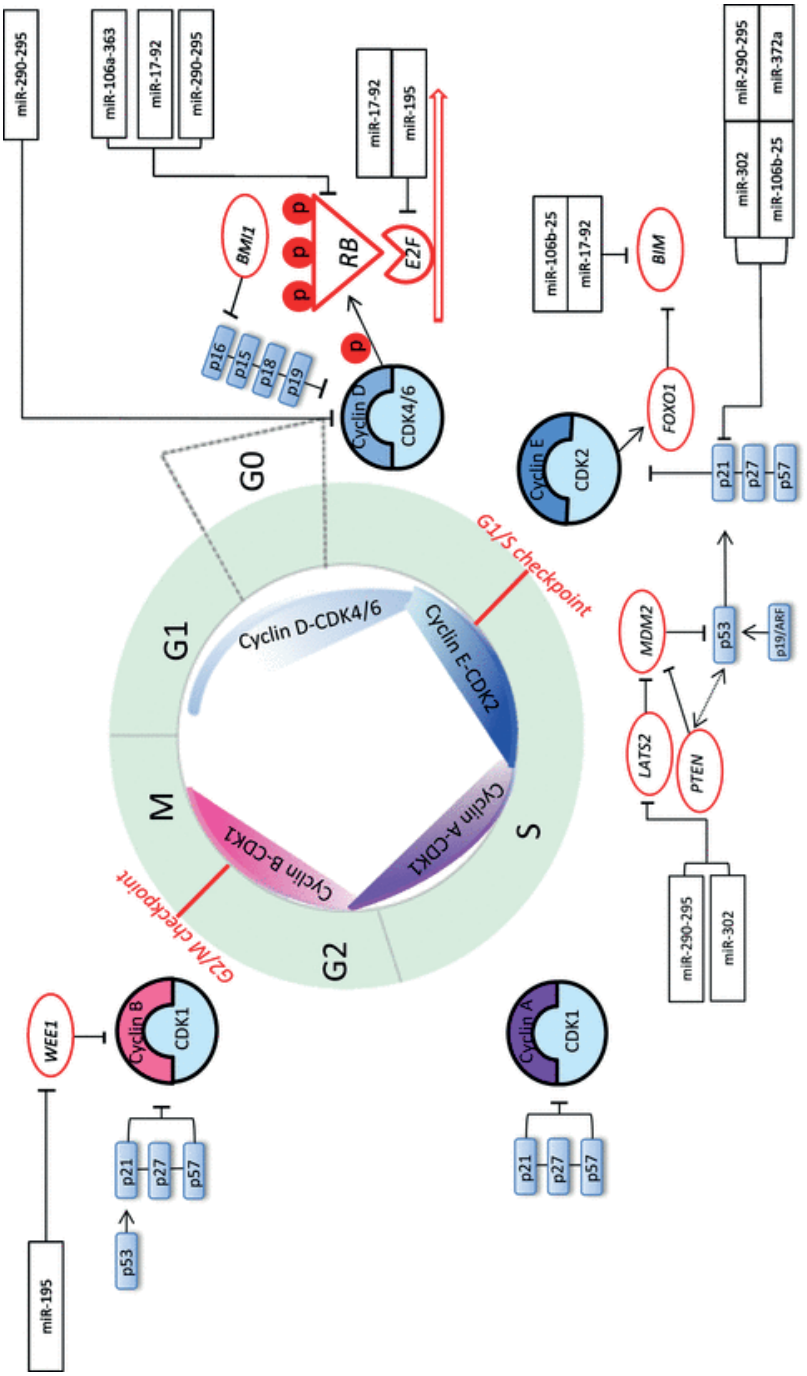


Figure 2.1. An overview of cell cycle regulation of ESCs by miRNAs. The figure illustrates the regulatory process that contributes to the cell cycle progression in embryonic stem cells (ESCs). Multiple key regulatory elements including cyclins, CDKs and cell cycle associated genes are forming a network that progress cells through the four different phases of cell cycle. miRNAs affect this progress by directly or indirectly targeting the cell cycle associated components (e.g. *RB*, p53, p21, *LATS2*, *PTEN*, cyclin

D, cyclin E). Several cluster and single miRNAs are involved in the regulation of cell cycle in ESCs, among them miR-17-92, miR-290-295, miR-302, miR-106b-25 and miR-106a-363 are abundantly expressed in ESCs. For example, through inhibition of *E2F* by miR-92 and miR-195, transcription of multiple transcription factors and proteins (e.g. *E2F1*, *E2F2*, *E2F3*, *CDK2*, *CDC25A*) will decrease, resulting in a reduction of G1 duration. Further, the expression of main G1/S and G2/M checkpoint regulator p53 is decreased via indirect targeting of miR-290-295 and miR-302 in ESCs. This contributes to facilitating the transition to S phase, which duration is prolonged. Moreover, p21 expression is reduced via miR-290-295, miR-372a, miR-302 and miR-106b-25 in a direct manner. This will inhibit cyclin E-CDK2 activity and therefore facilitating the G1/S transition. Additionally, pro-apoptotic gene *BIM* will not be activated, resulting in a reduced fraction of cells going into apoptosis.

The miR-290 cluster, consisting of miR-291a-3p, miR-291b-3p, miR-294, and miR-295, is upregulated in undifferentiated ESCs, but is rapidly downregulated during differentiation [21, 50, 51]. It has been shown that members of this miRNA cluster promote the G1/S transition. Cells can relatively quick enter the S phase, because members of the miR-290-295 cluster directly target cyclin D-CDK4/6 and indirectly downregulate the cyclin E-CDK2 complex (**Figure 2.1**). MiR-290-295 downregulates diverse inhibitors of the cell cycle, including *RB, RBL1, RBL2*, p21 and *LATS2,*which change the distribution of ECS in each cell cycle phase [47]. Furthermore, the miR-290-295 cluster enhances the somatic reprogramming by increasing the expression of pluripotent transcription factors *OCT4, SOX2, KLF4, LIN28, MYC* and *NANOG* [47, 52]. Also, miR-290-295 is shown to be directly involved to suppress apoptosis by targeting *Caspase 2* [53]. This leads to a reduced percentage of ESCs in G1 phase and an increased fraction of cells in S or G2/M phases. Due to the enhanced proliferation, the metabolism of ESCs rather rely on glycolysis than aerobic respiration. This metabolism is similar to the Warburg effect that is known in cancer cells [44, 47, 48]. Therefore, glycolysis-associated genes, such as *MYC, LIN28* and *HIF1,* have been promoted by the miR-290-295 cluster [44, 47]. Moreover, members of this miRNA cluster could affect epigenetic pathways including DNA methylation, histone acetylation and activation of Polycomb proteins, which inactivates genes involved in differentiation [54, 55].

The miR-17-92 cluster consists of miR-17, miR-18a, miR-19a, miR-19b, miR-20a and miR-92a. This miRNA cluster is crucial in early mammalian development by supporting cellular reprogramming and tumorigenesis [44]. In particular, miR-17-92 is a regulator of the *MYC* oncogene [56, 57]. *MYC* inhibits the expression of chromatin regulatory genes including *SIN3B*, *HBP1*, and *BTG1*, via miR-17-92. Through epigenetic mechanisms including reduced recruitment of histone deacetylase (*HADC)* via *HBP1*, miR-17-92 controls the chromatin stage of cell cycle related genes (**Figure 2.1**) [56]. *MYC* through miR-17-92 contributes to the euchromatin formation of specific gene expression involved in DNA replication and repair mechanisms that goes along with a shift in the percentage of cells in a proliferating state [56]. Likewise, miR-106b, which share a similar seed sequence, has been reported to promote G1/S transition by directly targeting p21, which results in a higher portion of cells in S phase compared to G1 phase [58].

The miR-302-367 cluster, consisting of miR-302a, b, c, d, and miR-367, has also been shown to play a crucial role in the proliferation of ESCs. Members of the miR-302-367 cluster are highly expressed in early stages of embryonic development [59]. This miRNA cluster targets genes that are involved in epigenetic mechanisms. For example, the miRNA cluster downregulates lysine demethylases and CpG binding proteins MECP1-p66 and MECP2 [59]. This facilitates the transcription of pluripotent genes and thereby contributes to the sustenance of pluripotency in mammalian ESCs [59]. Furthermore, it has been demonstrated that the promoter of miR-302-367 is activated when bound by *OCT4, SOX2*, which are core transcription factors directly involved in the maintenance of ESCs [59, 60]. It has been also shown that this cluster promotes pluripotency in ESCs by targeting the SMAD signaling pathway and the PI3K/PKB signaling molecules. MiR-302 inhibits the expression of transforming growth factor beta-receptor 2 (*TGFBR2*) and *RAS* homolog gene family member C (*RHOC*), which leads to a reduction of epithelial-mesenchymal transition [59, 61, 62]. In addition, the miR-302 cluster has been suggested to negatively regulate p21 and *LATS2* activity in both hESCs and mESCs [63, 64]. These described molecular mechanisms enlighten the important role of the miR-302-367 cluster with respect to pluripotency and cell cycle modulations.

Another well-known miRNA family that is involved in the regulation of cell cycle progression is the let-7 family, which consist of let-7a-1, a-2, a-3, b, c, d, e, f-1, f-2, g, i and miR-98. Members of this miRNA family affect the G1/S transition of ESCs differently than the above-described ESCC miRNAs. While most of the ESCC miRNAs are related to promote self-renewal, the let-7 miRNAs suppress self-renewal [35, 51]. The mechanism underlying this antagonistic effect remains unclear. However, it has been suggested that the ESCC miRNAs positively regulate the expression of *LIN28*, which through a negative feedback loop suppress the let-7 maturation [65, 66].

Two other miRNAs known to affect the regulation of ESCs are miR-195 and miR-372a. Both miRNAs are highly enriched in hESCs compared to differentiated cells and their function also relies on maintaining the proliferative capacity of hESCs [67]. For example, ectopic expression of miR-195 results in reduced expression of the G2/M cell cycle checkpoint kinase *WEE1* and an enhancement of BrdU incorporation [67, 68]. Ectopic expression of miR-372 has been shown to reduce the p21 expression levels in Dicer-knockdown hESCs [67].

Human ESCs have the therapeutic potential to treat a myriad of disorders by cell replacement. In theory, ESCs could be used in regenerative medicine, drugs discovery and disease modeling. However, the usage of ESCs as clinical application is limited because of high tumorigenicity and ethical restrictions. A miRNA-based therapy that use induced pluripotent stem cells (iPSC) might overcome these limitations. In this regard, ectopic expression of ESCC miRNAs may contribute to expansion of stem cells for regenerative medicine purposes [12, 20, 44].

Somatic Stem Cells

An extensive body of research has revealed the role of miRNAs in the cell cycle regulation of somatic stem cells [45, 69, 70]. In particular, studies with tissue specific Dicer-knockout or Dgcr8-deficient mice have demonstrated that miRNAs are essential regulators of proliferation, survival and differentiation in somatic stem cells [71]. In the following paragraphs, the role of miRNAs in the cell cycle regulation of hematopoietic and mesenchymal stem cells, which are two commonly investigated somatic stem cells, will be discussed. The associations of miRNAs with other somatic stem cells are summarized in **Table 2.1.**

Table 2.1. miRNAs associated with the cell cycle regulation in somatic stem cells

Stem cell	miRNA ID	Potential target gene(s)	Reference
Epidermal	miR-205	PI3K-AKT	[148]
	miR-203	*SNAI2*, p63, *SNAP2*	[149]
	miR-34	p63	[150]
	miR-184	*NOTCH*, p63, *FIH1*	[151]
	miR-214	*WNT/β-catenin*	[152]
Neural	miR-9	*TLX, BAF53A*	[153]
	miR-137	*TLX*	[154]
	miR-184	*MBD1*	[155]
	miR-195	*MBD1*	[156]
	miR-124	*SOX-2, PTBP1, SCP1*	[157-159]
	miR-302	*p53, OCT4, SOX2, NANOG*	[160]
	miR-148b	*WNT/β-catenin*	[161]
	miR-138	*TRIP6*	[162]
Muscle	miR-27	*PAX3*	[163]
	miR-322	*CDC25A*	[164]
	miR-206	*HDAC4, PAX7*	[165, 166]
	miR-1	*HDAC4, PAX7*	[166]
	miR-133	*SRF, MALAT1*	[167]
	miR-221	*PI3K-AKT*	[168]
	miR-143	*IGFBP5, ERK1/2*	[169]
	miR-486	*PAX7*	[170]

Hematopoietic stem cell (HSC) development has been characterized by several mechanisms leading to generate multiple cell lineages. Adult HSCs are predominantly quiescent (in the Go phase) compared to fetal HSCs [4]. Well established is the self-renewal function of the *LIN28* gene, which is highly expressed in fetal HSCs compared

to adult HSCs (**Figure 2.2b**) [72,73]. This process based on a form-feedback loop includes the downregulation of let-7 through *LIN28*, which subsequently downregulates *HMGA2*. Given that *HMGA2* enhances the self-renewal capacity, the LIN28-HMGA2 pathway is crucial in stem cell development [74]. Most of the previous research has focused on determining the expression of miRNAs in hematopoietic stem and progenitor cells during lineage differentiation [75]. Several studies have also reported differential miRNA expressions between HSCs, hematopoietic progenitor cells and both myeloid and lymphoid linages (e.g. T cell, B cell, Granulocyte, Monocyte, Erythrocyte), demonstrating that miRNAs are involved in the differentiation of specific hematopoietic lineages [72,76-78]. Although the conventional model suggests that hematopoietic lineages are derived from a common HSC, more recent research revealed that a rather large number of progenitor cells are the main drivers behind steady-state hematopoiesis and clonal diversity [79]. In this regards, short-term HSC could support the heterogeneous range of progeny [79]. Taken the functional role of miRNAs into consideration, both progenitor cells and diverse miRNAs may be equally important for clonal expansion and hematopoiesis.

For example, miRNAs are differentially expressed between long term hematopoietic stem cells (LT-HSCs) and short term HSCs, which are defined by a combination of cell surface markers such as c-Kit$^+$/Sca-1$^+$/Lin$^-$ (KSL). Based on the expression levels of cell surface markers includingCD34, Flk-2, CD150, CD48, CD224, c-Kit, Sca-1, and Lin the heterogeneous population of HSCs differ in proliferation and differentiation capacity [80]. The transition of HSCs into progenitor cells is related with a switch from quiescent into rapid proliferating cells, and subsequently an alteration in expression of surface makers (**Figure 2.2c**).

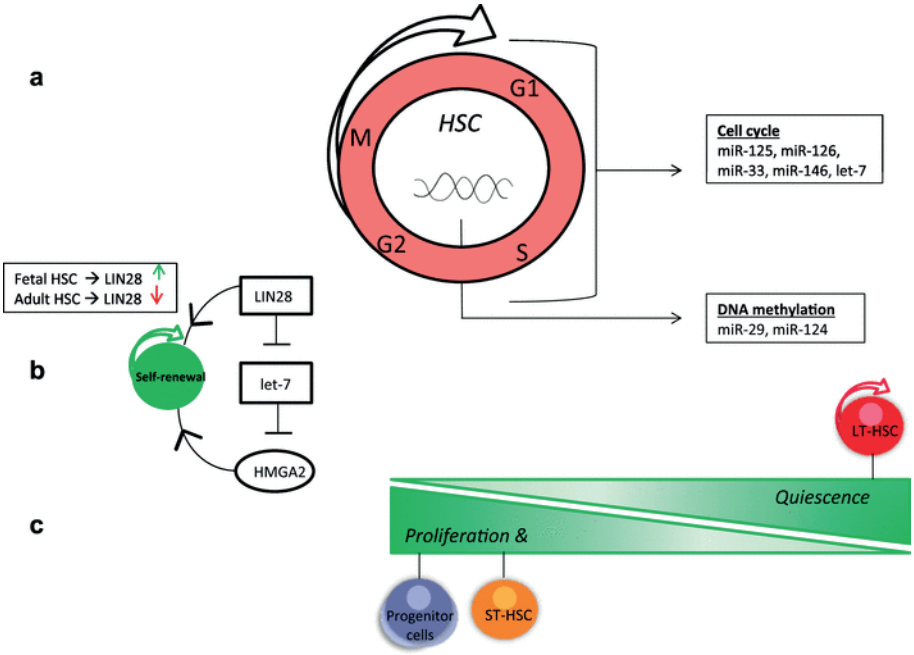


Figure 2.2. miRNA mediated regulation of cell cycle in HSCs. A) The schematic describes the miRNAs that have critical roles in the cell cycle regulation in adult HSC by directly targeting one of the cell cycle components, including miR-125, miR-126, miR-33, miR-146 and let-7. Furthermore, miR-29 and miR-124 target components involved in DNA methylation, whereby the chromatin architecture has been affected and thereby indirectly influences the expression of genes involved in cell cycle. **B)** The LIN28-HMGA2 feed-forward loop is among the most important mechanisms that drive fetal HSC self-renewal. *LIN28* is highly expressed in fetal HSC compared to adult HSC. This indicates the important role of let-7 upon stem cell differentiation, as *LIN28* directly inhibits let-7 expression. Decreased level of let-7 has resulted in higher expression of *HMGA2*, which induces self-renewal. Additionally, *LIN28* independently of let-7, contributes to self-renewal [73, 74]. **C)** Adult HSCs are a heterogeneous population that differs in self-renewal and differentiation capacity based on surface markers. Long term HSC (LT-HSC) are predominantly quiescent (c-kit$^+$ Sca-1$^+$ Lin$^-$Flk-2$^-$CD34$^-$)[147]. However, a large fraction of short term-HSC (c-kit$^+$ Sca-1$^+$ Lin$^-$Flk-2$^-$ CD34$^+$) gives rise to the differentiated progeny, and also shows greater cell proliferation capacity than LT-HSCs [80, 147]. Progenitor cells are associated with proliferation and differentiation into hematopoietic lineages. KSL (c-kit$^+$ Sca-1$^+$ Lin$^-$) with high CD150$^+$ expression may give predominant rise to myeloid linages, whereas KSL-CD150$^-$ are more likely to a lymphoid outcome [81]. Moreover, several studies demonstrated that specific miRNAs are differentially expressed among HSC and progenitor cells.

Therefore, the expression of cell cycle related miRNAs in exclusively progenitor cells is likely to be involved in the alteration of cell cycle duration [70]. One of the enhanced expressed miRNAs in LT-HSC is the miR-125 cluster (miR-125a, miR-125b1, miR-125b2). The expression of miR-125 has been shown to be associated with self-renewal and expansion of the stem cell population in vivo [81-83]. Furthermore, miR-29a has been

revealed to regulate the G1/S transition in hematopoietic progenitor stem cells. MiR-29a promotes the self-renewal capacity by targeting a subset of genes that are involved in cell cycle progression, including *CDC42EP2* and *HBP1* [84]. Recently, Lechman et al. demonstrated that miR-126 can control the cell cycle progression by targeting the PI3K/AKT/MTOR pathway [85]. They showed that overexpression of miR-126 results in an increased percentage of quiescent cells, whereas a knockdown of miR-126 lead to enhanced proliferation and differentiation of HSCs [85-87].

Additionally, previous studies have suggested miR-125 and miR-126 as potential target treatment for acute myeloid leukemia (AML) [88, 89]. An indication for the potential therapeutic function is based on the alternated expression of these miRNAs between $CD34^+$ $CD38^-$ HSC and $CD34^+$ $CD38^-$ leukemic stem cells. A reduction of miR-126 stimulates the PI3K/AKT/MTOR pathway in HSCs and will result in an increased number of HSCs, while this effect decreases the self-renewal capacity in $CD34^+$ $CD38^-$ leukemic stem cells [88]. However, this miRNA-based treatment holds promising capacity in vivo experiments, issues with respect to toxicity and delivery need to be solved before application in AML patients [88].

Mesenchymal stem cells (MSCs) are multipotent cells that originate from bone marrow stroma, but are present in various tissues such as adipose tissue, bone, skeletal muscle, cartilage and tendon [90]. Evidence suggests that miRNAs are closely involved in the regulation of MSC differentiation into specific cell lineages [78, 91-93]. The role of miRNAs in proliferation and cell cycle regulation of human MSCs has been investigated through Drosha and Dicer knockdown studies [94]. These studies have shown a significant increase in the number of cells in G1 phase and a reduced proliferation rate of MSCs [94]. In the same study, Drosha knockdown in MSCs resulted in a decrease of pRb and an increase in p16 and p15 levels [94]. Other studies have been implicated miR-16 and miR-143 in the regulation of MSC proliferation and differentiation. In this regard, miR-16 has been shown to inhibit MSC proliferation and induce cell cycle arrest by targeting cyclin E [95]. Likewise, miR-143 targets *ERK5* (member of MAPK family), which itself decreases the expression of cyclin D and CDK6. This reduces cell entry into S phase, suggesting miR-143 to be a negative regulator of the cell cycle progression [96, 97].

Moreover, a number of miRNAs have determined to control the differentiation into specific linages, such as osteoblasts [98]. For example, Peng et al. demonstrated that miRNAs promote the osteogenic differentiation of MSCs via *BMP, WNT/β-catenin* and *NOTCH* signaling pathways. Among them, miR-27 promotes differentiation by targeting *APC*, which modulates the G2/M transition [98, 99]. On the other hand, miR-27 expression is shown to be downregulated upon adipocyte differentiation [100, 101]. Several cell cycle associated genes, including *ERK1/2, ERK5, TGF-β1* and *KLF5* are related to adipocyte differentiation, which is explained by miRNA regulation [102]. Notably, miR-143, miR-448 and miR-375 have been reported as negative regulators and miR-21 as positive regulator of adipocyte differentiation [102].

Cancer Stem Cells (CSCs)

Altered expression and molecular abnormalities of the cell-cycle-regulatory proteins, such as pRB, p53, CDKs, CDKIs and cyclins, play a central role in cancer initiation and progression [17, 103-105]. Notably, it has been suggested that a class of cancer cells with characteristics of stem cells, so-called cancer stem cells (CSCs), are responsible for tumor initiation, invasion, metastasis and chemoresistance [106, 107]. As discussed previously in this review, miRNAs have the ability to suppress apoptosis and promote proliferation by interplaying with the cell cycle components. Therefore, miRNAs and CSCs share common properties with respect to tumorigenesis. The transcriptional levels of several miRNAs have shown to vary between normal stem cells and CSCs [108]. Furthermore, associations between either cell cycle components including cyclins and transcription factors or miRNA expression and specific CSC markers have been investigated [109, 110]. MiRNAs as regulators of CSCs have gained attention in recent years in multiple fields of research [107, 109, 111, 112]. The associations between miRNAs expression and various cancers are summarized in **Table 2.2**. In the following paragraph, some of the main CSC-related miRNAs will be discussed.

The miR-17-92 cluster affects the cell cycle by targeting *E2F1* and cyclin D as well as it cooperates with the oncogene *MYC* to prevent apoptosis in CSCs [113-116]. Li et al. investigated the miR-17-92 target genes involved in the *MYC* suppression. They demonstrated that the functionalities of the miR-17-92 target genes rely on multiple DNA replication, cell cycle regulation, chromosome organization, RNA transcription or protein metabolism [56]. Similarly, this miRNA cluster is shown to coordinate the timing of cell cycle progression by modulating expression of *BMI1, PTEN, RBL2* and p21 [117-121].

Table 2.2. miRNAs associated with the cell cycle progression in cancer stem cells

Cancer type	miRNA ID	Potential target gene(s)	Exp. of miRNA	Reported biological effect	Reference
Breast	let-7	*LIN28*	Downregulated	Upregulation of *LIN28* results in supporting *RAS*, *MYC* and *HMGA2*	170
	miR-21	*PTEN*	Upregulated	Promote PI3K/AKT signaling activation through directly inhibiting *PTEN* expression	171
	miR-221/222	*PTEN*	Upregulated	Promote AKT/NF-κβ/COX-2 pathway by targeting *PTEN*	172
	miR-93	*JAK1, SOX4, STAT3, AKT, EZH1, HMGA2*	Upregulated	Regulate CSC proliferation	173
	miR-34	*CDK4, CDK6, NOTCH1*	Downregulated	Regulate p53	132
	miR-16	*BMI1*	Upregulated	Inhibit DNA repair by repressing *BMI1*	174
	miR-200	*ZEB1, ZEB2, WNT-signaling*	Downregulated	Reduction of EMT	175
	miR-494-3p	*PAK1*	Downregulated	Inhibit proliferation via MAPK by targeting *PAK1*	176
Liver (HCC)	miR-34	*Cyclin D1, BCL2*	Downregulated	Regulate p53	177
	miR-365	*BCL2*	Upregulated	Apoptosis	178
	miR-31	*HDCA2, CDK2*	Downregulated	Induction of p16 and p21. Repression of cyclin D, CDK4, CDK2	133
	miR-26a	*EZH2*	Upregulated	Reduction of EMT	179
	miR-150	*GAB1*	Downregulated	Suppress proliferation and invasion via MAPK pathway by targeting *GAB1* and *ERK1/2*	180
Head and Neck	let-7	*ABCB1*	Downregulated	Reduced of cell proliferation	181
Pancreatic	let-7	*LIN28*	Downregulated	Inhibit EMT, induces cell cycle arrest when *LIN28* is reduced	182
	miR-21	*PTEN, PDCD4*	Upregulated	Promote metastasis	183
	miR-203	*ZEB1, ZEB2*	Downregulated	Reduction of EMT	184

Cancer type	miRNA ID	Potential target gene(s)	Exp. of miRNA	Reported biological effect	Reference
	miR-34	*BCL2, NOTCH1/2*	Downregulated	Regulate p53	112
	miR-17-92	*p21, p57, TBX3*	Downregulated	Maintain stemness characteristics in pancreatic CSC. Downregulation of *MYC*.	121
Prostate	let-7	*LIN28*	Upregulated	Upregulating cell cycle via cyclin D1	185
	miR-100	*CDK6, RB1, mTOR*	Downregulated	Regulation of cell growth	186
	miR-34	*Cyclin D1, CDK4, CDK6, c-MET, CD44*	Downregulated	Mediating p53. Tumor metastasis	187
	miR-221/222	*p27/Kip1*	Upregulated	Regulate activation of cyclin E and cyclin D	188
Glioblastoma	miR-124	*CDK6*	Upregulated	Inhibit cell proliferation	189
	miR-137	*CDK6*	Upregulated	Inhibit cell proliferation	190
	miR-128	*BMI1*	Upregulated	Decreasing cell proliferation in *IDH1* mutant glioma	191
	miR-23b	*HMGA2*	Upregulated	Cell cycle arrest and proliferation inhibition	192
	miR-125b	*CDK6, E2F3, CDC25A*	Downregulated	Induce G1/S cell cycle arrest	193
	miR-34	*BCL2, NOTCH1*	Downregulated	Targeting p53. Anti-apoptotic, increase cell proliferation	194
Lung	miR-605	*LATS2*	Upregulated	Promote cell proliferation, migration and invasion	195
	let-7	*KRAS, MYC, CDK6 HMGA2, TGFBR2*	Downregulated	Suppression of multiple oncogenic members	196
	miR-21	*MDM4*	Upregulated	Repress *MDM4* to activate p53	127
	miR-15a/ miR-16	*RB*	Downregulated	Cell cycle arrest	140

Other important regulators of CSCs are the members of the let-7 family. Evidence suggests that let-7 is among the most important miRNAs involved in tumor progression and chemoresistance [107, 122]. The expression of the let-7 family is reduced in various types of tumor cells, including breast, head and neck squamous (HNSCC), lung, pancreatic, neuroblastoma cells, among others [107, 109, 123, 124]. Accordingly, decreased expression of let-7 has resulted in overexpression of oncogenes *MYC, RAS, HMGA2* and *BLIMP1* [91, 122, 125]. Furthermore, members of the let-7 family have been recognized as negative regulators of *PTEN* that inactivate the PI3K/AKT/MTOR pathway. The let-7 family has also shown to be involved in suppressing the epithelial-to-mesenchymal transition (EMT), which is related to metastasis and chemoresistance and therefore a characteristic of CSCs [107, 122]. Multiple genes involved in cell cycle progression are suggested to be targets for the let-7 family. The latter include cyclin D, cyclin A, *CDK1*, *CDK2*, *CDK4*, *CDK6*, *CDK8* and *CDC25A* [91, 122, 125]. Also, it has been shown that the RNA binding protein *LIN28* inhibits let-7 by stimulating cellular proliferation via cyclin D, *CDK2* and *CDC25A* and thereby contribute to the maintenance of stemness characteristics of CSCs [46, 126]. *LIN28* has been recognized as an oncogene, as it promotes tumor progression by repressing let-7 [122]. Previous studies based on let-7 expression and tumor progression display that ectopic expression of let-7 was sufficient enough to inhibit proliferation and clonal expansion in vitro and tumor recurrence in prostate cancer cells in vivo [117].

The next miRNA family, consisting of miR-34a, b, and c, is well-studied regarding to cell cycle progression and its expression is downregulated in several types of cancer cells including lung adenocarcinomas, colon cancer and liver cancer (HCC) [127-132]. MiR-34a induces both G1/S cell cycle arrest and cell senescence [127]. Reduced expression of miR-34 has been associated with enhanced levels of *BCL2* and *NOTCH*, which are target genes for tumor suppressor gene p53 [107, 111, 127]. Similarity, miR-34 promotes apoptosis via *Caspase 3*, and therefore increases sensitivity for anti-cancer treatment [111]. By regulating CDK6, cyclin D1 and *E2F*, miR-34 negatively affects cell cycle progression in colon cancer cells [107, 130, 131]. In addition, miR-34 represses pluripotency genes inclusive of *NANOG, SOX2* and *MYC* [111]. Thus, overexpression of this miRNA family may cause an accumulated percentage of cells in the G0/G1 phase and significantly reduces the population of cells in the S phase.

MiR-31 has also shown to be inversely correlated with metastasis, since high expression in liver cancer is linked with a poor prognosis in patients. Kim et al. showed that ectopic expression of miR-31 evokes an overexpression of CDK2 and *HDAC2* [133]. They have demonstrated that through abnormal expression of *HDAC2*, negative cell cycle regulators p16/INK4A, p19/INK4D and p21/Cip1 are induced.

Furthermore, an oncogenic role has been reported for the miR-15a/16 family in chronic lymphocytic leukemia (CLL), pituitary adenomas, and gastric cancer [134, 135]. On the other hand, this miRNA family has shown to act as a tumor suppressor in a subset of B cell lymphoma, where deletion of this miRNA family in a subset of B cell lymphomas resulted in chronic lymphocytic leukemia in mice [136]. In fact, miR-15a and miR-16 display an anti-proliferative potential in this type of cancer stem cell by silencing *BCL2* and activating the intrinsic apoptosis pathway [137, 138]. In addition, some studies revealed the miR-15a/16 family as regulator of various cyclins, including cyclins D1 and D2 and cyclin E1, and pRb [125, 139, 140].

An additional miRNA that has been suggested as an oncomiR, through targeting multiple signaling pathways, is miR-21 [33]. Upregulation of this miRNA has an oncogenic potential in a wide range of tumors including lung, breast, pancreatic, brain and colon cancers, through downregulation of p21 and tumor suppressor genes *PTEN* and *PDCD4* [33, 141-143]. MiR-26a has also mentioned to be a negative regulator of cancer cell proliferation by targeting cyclins D2 and E2, and CDK6. It has been established that overexpression of miR-26a results in cell cycle arrest in human liver cancer cells in vitro [144, 145].

Concluding remarks and future prospects

A growing body of evidence has addressed the potential role of miRNAs in cell cycle regulation of stem cells. In light of recent discoveries about the role miRNAs in self-renewal, proliferation and differentiation, it is crucial to unravel the complex mechanisms and molecular interactions within this field of research. In this review, we outlined the most established miRNAs involved in the cell cycle progression of stem cells. We highlighted several clusters and single miRNAs that may control self-renewal and maintenance of the pluripotency status in ESCs. These include but are not limited to ESCC miRNAs (let-7, miR-290-295, miR-302, miR-17-92, miR-106b-25 and miR-106a-363), which are functionally upregulated to suppress negative regulators and to enhance pluripotent transcription factors such as *NANOG* and *MYC* in an epigenetic manner [45].

Furthermore, specific profiles of miRNA expression in distinct somatic stem cell lineages are linked with developmental control by keeping several multipotent stem cells (e.g. HSCs) in a quiescent state. Previous research based on Dicer-knockout and Dgcr8-deficient mice have elucidated that miRNAs are expressed temporally and spatially among somatic stem cells and precursor cells [37]. It is crucial for somatic stem cells like HSCs to keep a balance between quiescent state and proliferating state. To accomplish that, a complex network of miRNAs exists that inhibit positive cell cycle regulators such as cyclins, as well as miRNAs modulating anti-apoptotic properties. This review set out to summarize distinct miRNAs involved in differentiation of several somatic stem cells. Complex interactions between miRNAs, transcription factors and cell cycle-mediated components may control the gene expression upon differentiation of multipotent stem cells into progenitor cells and mature cells.

It is clear that abnormalities in the cell cycle are related to tumorigenesis and previous studies have highlighted the significant importance of miRNAs in the regulation of CSCs [108]. Since CSC features are linked to metastasis, invasion and therapeutic resistance, it is of main clinical relevance to unravel the interactive properties between CSC-related miRNAs and cell cycle components. From the data available so far it appears that there is a great overlapping role between ESCC miRNAs that are expressed in both ESCs and CSCs. However, a subset of miRNAs is characterized as tumor suppressor genes as they are expressed regarding anti-proliferating features by targeting oncogenic pathways including *MYC*. Those miRNAs, including let-7, miR-34, miR-31 and miR-17-92 family, are of major interest since they are associated with a good prognosis in cancer patients. Future research should focus on targeting the CSC-related miRNAs involved in oncogenic pathways since they will provide a more effective approach to exterminate CSCs. Subsequently, a miRNA based method for cancer treatment is highly target driven as it interferes with specific abnormalities in the cell cycle within the tumor microenvironment.

Collectively, this review marks several noteworthy insights into the cell cycle regulation of stem cells by miRNAs. Understanding the tightly regulated molecular networks in which miRNAs are interacting, will greatly enhance our knowledge in the development of both healthy and disease states of the human body.

References

1. Draper JS, Moore HD, Ruban LN, Gokhale PJ, Andrews PW. Culture and characterization of human embryonic stem cells. *Stem Cells Dev*. 2004;13:325-336
2. Xie X, Teknos TN, Pan Q. Are all cancer stem cells created equal? *Stem Cells Transl Med*. 2014;3:1111-1115
3. Harashima H, Dissmeyer N, Schnittger A. Cell cycle control across the eukaryotic kingdom. *Trends Cell Biol*. 2013;23:345-356
4. Wilson A, Laurenti E, Oser G, van der Wath RC, Blanco-Bose W, Jaworski M, et al. Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. *Cell*. 2008;135:1118-1129
5. Blomen VA, Boonstra J. Cell fate determination during g1 phase progression. *Cell Mol Life Sci*. 2007;64:3084-3104
6. She S, Wei Q, Kang B, Wang YJ. Cell cycle and pluripotency: Convergence on octamerbinding transcription factor 4 (review). *Mol Med Rep*. 2017
7. Pauklin S, Vallier L. The cell-cycle state of stem cells determines cell fate propensity. *Cell*. 2013;155:135-147
8. Dalton S. Linking the cell cycle to cell fate decisions. *Trends Cell Biol*. 2015;25:592-600
9. Coronado D, Godet M, Bourillot PY, Tapponnier Y, Bernat A, Petit M, et al. A short g1 phase is an intrinsic determinant of naive embryonic stem cell pluripotency. *Stem Cell Res*. 2013;10:118-131
10. Conklin JF, Baker J, Sage J. The rb family is required for the self-renewal and survival of human embryonic stem cells. *Nat Commun*. 2012;3:1244
11. Yeo HC, Beh TT, Quek JJ, Koh G, Chan KK, Lee DY. Integrated transcriptome and binding sites analysis implicates e2f in the regulation of self-renewal in human pluripotent stem cells. *PLoS One*. 2011;6:e27231
12. Hindley C, Philpott A. The cell cycle and pluripotency. *Biochem J*. 2013;451:135-143
13. Lu Z, Hunter T. Ubiquitylation and proteasomal degradation of the p21(cip1), p27(kip1) and p57(kip2) cdk inhibitors. *Cell Cycle*. 2010;9:2342-2352
14. Fabbro M, Savage K, Hobson K, Deans AJ, Powell SN, McArthur GA, et al. Brca1-bard1 complexes are required for p53ser-15 phosphorylation and a g1/s arrest following ionizing radiation-induced DNA damage. *J Biol Chem*. 2004;279:31251-31258
15. Hyun SY, Jang YJ. P53 activates g(1) checkpoint following DNA damage by doxorubicin during transient mitotic arrest. *Oncotarget*. 2015;6:4804-4815
16. Lanni JS, Jacks T. Characterization of the p53-dependent postmitotic checkpoint following spindle disruption. *Mol Cell Biol*. 1998;18:1055-1064
17. Bieging KT, Mello SS, Attardi LD. Unravelling mechanisms of p53-mediated tumour suppression. *Nat Rev Cancer*. 2014;14:359-370
18. Jain AK, Allton K, Iacovino M, Mahen E, Milczarek RJ, Zwaka TP, et al. P53 regulates cell cycle and micrornas to promote differentiation of human embryonic stem cells. *PLoS Biol*. 2012;10:e1001268
19. Bueno MJ, Malumbres M. Micrornas and the cell cycle. *Biochim Biophys Acta*. 2011;1812:592-601
20. Wang Y, Blelloch R. Cell cycle regulation by micrornas in embryonic stem cells. *Cancer Res*. 2009;69:4093-4096

21. Wang Y, Blelloch R. Cell cycle regulation by micrornas in stem cells. *Results Probl Cell Differ*. 2011;53:459-472

22. Yu Z, Li Y, Fan H, Liu Z, Pestell RG. Mirnas regulate stem cell self-renewal and differentiation. *Front Genet*. 2012;3:191

23. Braun CJ, Zhang X, Savelyeva I, Wolff S, Moll UM, Schepeler T, et al. P53-responsive micrornas 192 and 215 are capable of inducing cell cycle arrest. *Cancer Res*. 2008;68:10094-10104

24. 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

25. Kozomara A, Griffiths-Jones S. Mirbase: Annotating high confidence micrornas using deep sequencing data. *Nucleic Acids Res*. 2014;42:D68-73

26. Esquela-Kerscher A, Slack FJ. Oncomirs - micrornas with a role in cancer. *Nat Rev Cancer*. 2006;6:259-269

27. van Rooij E, Olson EN. Microrna therapeutics for cardiovascular disease: Opportunities and obstacles. *Nat Rev Drug Discov*. 2012;11:860-872

28. Ghanbari M, de Vries PS, de Looper H, Peters MJ, Schurmann C, Yaghootkar H, 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

29. van Rooij E, Sutherland LB, Liu N, Williams AH, McAnally J, Gerard RD, et al. A signature pattern of stress-responsive micrornas that can evoke cardiac hypertrophy and heart failure. *Proc Natl Acad Sci U S A*. 2006;103:18255-18260

30. Goren Y, Kushnir M, Zafrir B, Tabak S, Lewis BS, Amir O. Serum levels of micrornas in patients with heart failure. *Eur J Heart Fail*. 2012;14:147-154

31. Karolina DS, Tavintharan S, Armugam A, Sepramaniam S, Pek SL, Wong MT, et al. Circulating mirna profiles in patients with metabolic syndrome. *J Clin Endocrinol Metab*. 2012;97:E2271-2276

32. 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

33. Malumbres M. Mirnas and cancer: An epigenetics view. *Mol Aspects Med*. 2013;34:863-874

34. Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, et al. The 21-nucleotide let-7 rna regulates developmental timing in caenorhabditis elegans. *Nature*. 2000;403:901-906

35. Shim J, Nam JW. The expression and functional roles of micrornas in stem cell differentiation. *BMB Rep*. 2016;49:3-10

36. Huang XA, Lin H. The mirna regulation of stem cells. *Wiley Interdiscip Rev Membr Transp Signal*. 2012;1:83-95

37. Wang Y, Medvid R, Melton C, Jaenisch R, Blelloch R. Dgcr8 is essential for microrna biogenesis and silencing of embryonic stem cell self-renewal. *Nat Genet*. 2007;39:380-385

38. Bernstein E, Kim SY, Carmell MA, Murchison EP, Alcorn H, Li MZ, et al. Dicer is essential for mouse development. *Nat Genet*. 2003;35:215-217

39. Dalton S. Exposing hidden dimensions of embryonic stem cell cycle control. *Cell Stem Cell*. 2009;4:9-10

40. Stein GS, van Wijnen AJ, Stein JL, Lian JB, Montecino M, Zaidi SK, et al. An architectural perspective of cell-cycle control at the g1/s phase cell-cycle transition. *J Cell Physiol*. 2006;209:706-710

41. Sengupta S, Nie J, Wagner RJ, Yang C, Stewart R, Thomson JA. Microrna 92b controls the g1/s checkpoint gene p57 in human embryonic stem cells. *Stem Cells*. 2009;27:1524-1528

42. Kareta MS, Gorges LL, Hafeez S, Benayoun BA, Marro S, Zmoos AF, et al. Inhibition of pluripotency networks by the rb tumor suppressor restricts reprogramming and tumorigenesis. *Cell Stem Cell*. 2015;16:39-50

43. Becker KA, Ghule PN, Therrien JA, Lian JB, Stein JL, van Wijnen AJ, et al. Self-renewal of human embryonic stem cells is supported by a shortened g1 cell cycle phase. *J Cell Physiol*. 2006;209:883-893

44. Hao J, Duan FF, Wang Y. Micrornas and rna binding protein regulators of micrornas in the control of pluripotency and reprogramming. *Curr Opin Genet Dev*. 2017;46:95-103

45. Li N, Long B, Han W, Yuan S, Wang K. Micrornas: Important regulators of stem cells. *Stem Cell Res Ther*. 2017;8:110

46. Lichner Z, Pall E, Kerekes A, Pallinger E, Maraghechi P, Bosze Z, et al. The mir-290-295 cluster promotes pluripotency maintenance by regulating cell cycle phase distribution in mouse embryonic stem cells. *Differentiation*. 2011;81:11-24

47. Yuan K, Ai WB, Wan LY, Tan X, Wu JF. The mir-290-295 cluster as multi-faceted players in mouse embryonic stem cells. *Cell Biosci*. 2017;7:38

48. Wang Y, Baskerville S, Shenoy A, Babiarz JE, Baehner L, Blelloch R. Embryonic stem cell-specific micrornas regulate the g1-s transition and promote rapid proliferation. *Nat Genet*. 2008;40:1478-1483

49. Calabrese JM, Seila AC, Yeo GW, Sharp PA. Rna sequence analysis defines dicer's role in mouse embryonic stem cells. *Proc Natl Acad Sci U S A*. 2007;104:18097-18102

50. Houbaviy HB, Murray MF, Sharp PA. Embryonic stem cell-specific micrornas. *Dev Cell*. 2003;5:351-358

51. Melton C, Judson RL, Blelloch R. Opposing microrna families regulate self-renewal in mouse embryonic stem cells. *Nature*. 2010;463:621-626

52. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*. 2006;126:663-676

53. Zheng GX, Ravi A, Calabrese JM, Medeiros LA, Kirak O, Dennis LM, et al. A latent pro-survival function for the mir-290-295 cluster in mouse embryonic stem cells. *PLoS Genet*. 2011;7:e1002054

54. Kanellopoulou C, Gilpatrick T, Kilaru G, Burr P, Nguyen CK, Morawski A, et al. Reprogramming of polycomb-mediated gene silencing in embryonic stem cells by the mir-290 family and the methyltransferase ash1l. *Stem Cell Reports*. 2015;5:971-978

55. Richly H, Aloia L, Di Croce L. Roles of the polycomb group proteins in stem cells and cancer. *Cell Death Dis*. 2011;2:e204

56. Li Y, Choi PS, Casey SC, Dill DL, Felsher DW. Myc through mir-17-92 suppresses specific target genes to maintain survival, autonomous proliferation, and a neoplastic state. *Cancer Cell*. 2014;26:262-272

57. Aguda BD, Kim Y, Piper-Hunter MG, Friedman A, Marsh CB. Microrna regulation of a cancer network: Consequences of the feedback loops involving mir-17-92, e2f, and myc. *Proc Natl Acad Sci U S A*. 2008;105:19678-19683

58. Ivanovska I, Ball AS, Diaz RL, Magnus JF, Kibukawa M, Schelter JM, et al. Micrornas in the mir-106b family regulate p21/cdkn1a and promote cell cycle progression. *Mol Cell Biol*. 2008;28:2167-2174

59. Kuo CH, Deng JH, Deng Q, Ying SY. A novel role of mir-302/367 in reprogramming. *Biochem Biophys Res Commun*. 2012;417:11-16

60. Card DA, Hebbar PB, Li L, Trotter KW, Komatsu Y, Mishina Y, et al. Oct4/sox2-regulated mir-302 targets cyclin d1 in human embryonic stem cells. *Mol Cell Biol*. 2008;28:6426-6438

61. Anokye-Danso F, Trivedi CM, Juhr D, Gupta M, Cui Z, Tian Y, et al. Highly efficient mirna-mediated reprogramming of mouse and human somatic cells to pluripotency. *Cell Stem Cell*. 2011;8:376-388

62. Lipchina I, Studer L, Betel D. The expanding role of mir-302-367 in pluripotency and reprogramming. *Cell Cycle*. 2012;11:1517-1523

63. Dolezalova D, Mraz M, Barta T, Plevova K, Vinarsky V, Holubcova Z, et al. Micrornas regulate p21(waf1/cip1) protein expression and the DNA damage response in human embryonic stem cells. *Stem Cells*. 2012;30:1362-1372

64. Liang Y, Li Y, Li Z, Liu Z, Zhang Z, Chang S, et al. Mechanism of folate deficiency-induced apoptosis in mouse embryonic stem cells: Cell cycle arrest/apoptosis in g1/g0 mediated by microrna-302a and tumor suppressor gene lats2. *Int J Biochem Cell Biol*. 2012;44:1750-1760

65. Fu Y, Gao L, Shi Z, You F, Zhang J, Li W. Characterization and expression of lin-28a involved in lin28/let-7signal pathway during early development of p. Olivaceus. *Fish Physiol Biochem*. 2017

66. Rybak A, Fuchs H, Smirnova L, Brandt C, Pohl EE, Nitsch R, et al. A feedback loop comprising lin-28 and let-7 controls pre-let-7 maturation during neural stem-cell commitment. *Nat Cell Biol*. 2008;10:987-993

67. Qi J, Yu JY, Shcherbata HR, Mathieu J, Wang AJ, Seal S, et al. Micrornas regulate human embryonic stem cell division. *Cell Cycle*. 2009;8:3729-3741

68. Bhattacharya A, Schmitz U, Wolkenhauer O, Schonherr M, Raatz Y, Kunz M. Regulation of cell cycle checkpoint kinase wee1 by mir-195 in malignant melanoma. *Oncogene*. 2013;32:3175-3183

69. Shenoy A, Blelloch RH. Regulation of microrna function in somatic stem cell proliferation and differentiation. *Nat Rev Mol Cell Biol*. 2014;15:565-576

70. Arnold CP, Tan R, Zhou B, Yue SB, Schaffert S, Biggs JR, et al. Microrna programs in normal and aberrant stem and progenitor cells. *Genome Res*. 2011;21:798-810

71. Andersson T, Rahman S, Sansom SN, Alsio JM, Kaneda M, Smith J, et al. Reversible block of mouse neural stem cell differentiation in the absence of dicer and micrornas. *PLoS One*. 2010;5:e13453

72. Mehta A, Baltimore D. Micrornas as regulatory elements in immune system logic. *Nat Rev Immunol*. 2016;16:279-294

73. Copley MR, Babovic S, Benz C, Knapp DJ, Beer PA, Kent DG, et al. The lin28b-let-7-hmga2 axis determines the higher self-renewal potential of fetal haematopoietic stem cells. *Nat Cell Biol*. 2013;15:916-925

74. Zhou Y, Li YS, Bandi SR, Tang L, Shinton SA, Hayakawa K, et al. Lin28b promotes fetal b lymphopoiesis through the transcription factor arid3a. *J Exp Med*. 2015;212:569-580

75. Bissels U, Bosio A, Wagner W. Micrornas are shaping the hematopoietic landscape. *Haematologica*. 2012;97:160-167

76. Monticelli S, Ansel KM, Xiao C, Socci ND, Krichevsky AM, Thai TH, et al. Microrna profiling of the murine hematopoietic system. *Genome Biol*. 2005;6:R71

77. Chen CZ, Lodish HF. Micrornas as regulators of mammalian hematopoiesis. *Semin Immunol*. 2005;17:155-165

78. Gangaraju VK, Lin H. Micrornas: Key regulators of stem cells. *Nat Rev Mol Cell Biol*. 2009;10:116-125

79. Sun J, Ramos A, Chapman B, Johnnidis JB, Le L, Ho YJ, et al. Clonal dynamics of native haematopoiesis. *Nature*. 2014;514:322-327

80. Crisan M, Dzierzak E. The many faces of hematopoietic stem cell heterogeneity. *Development*. 2016;143:4571-4581

81. Wojtowicz EE, Lechman ER, Hermans KG, Schoof EM, Wienholds E, Isserlin R, et al. Ectopic mir-125a expression induces long-term repopulating stem cell capacity in mouse and human hematopoietic progenitors. *Cell Stem Cell*. 2016;19:383-396

82. Guo S, Lu J, Schlanger R, Zhang H, Wang JY, Fox MC, et al. Microrna mir-125a controls hematopoietic stem cell number. *Proc Natl Acad Sci U S A*. 2010;107:14229-14234

83. Ooi AG, Sahoo D, Adorno M, Wang Y, Weissman IL, Park CY. Microrna-125b expands hematopoietic stem cells and enriches for the lymphoid-balanced and lymphoid-biased subsets. *Proc Natl Acad Sci U S A*. 2010;107:21505-21510

84. Han YC, Park CY, Bhagat G, Zhang J, Wang Y, Fan JB, et al. Microrna-29a induces aberrant self-renewal capacity in hematopoietic progenitors, biased myeloid development, and acute myeloid leukemia. *J Exp Med*. 2010;207:475-489

85. Lechman ER, Gentner B, Ng SWK, Schoof EM, van Galen P, Kennedy JA, et al. Mir-126 regulates distinct self-renewal outcomes in normal and malignant hematopoietic stem cells. *Cancer Cell*. 2016;29:602-606

86. Lechman ER, Gentner B, van Galen P, Giustacchini A, Saini M, Boccalatte FE, et al. Attenuation of mir-126 activity expands hsc in vivo without exhaustion. *Cell Stem Cell*. 2012;11:799-811

87. de Leeuw DC, Denkers F, Olthof MC, Rutten AP, Pouwels W, Schuurhuis GJ, et al. Attenuation of microrna-126 expression that drives cd34+38- stem/progenitor cells in acute myeloid leukemia leads to tumor eradication. *Cancer Res*. 2014;74:2094-2105

88. Martianez Canales T, de Leeuw DC, Vermue E, Ossenkoppele GJ, Smit L. Specific depletion of leukemic stem cells: Can micrornas make the difference? *Cancers (Basel)*. 2017;9

89. Testa U, Pelosi E. Micrornas expressed in hematopoietic stem/progenitor cells are deregulated in acute myeloid leukemias. *Leuk Lymphoma*. 2015;56:1466-1474

90. Krampera M, Franchini M, Pizzolo G, Aprili G. Mesenchymal stem cells: From biology to clinical use. *Blood Transfus*. 2007;5:120-129

91. Guo L, Zhao RC, Wu Y. The role of micrornas in self-renewal and differentiation of mesenchymal stem cells. *Exp Hematol*. 2011;39:608-616

92. Li JP, Zhuang HT, Xin MY, Zhou YL. Mir-214 inhibits human mesenchymal stem cells differentiating into osteoblasts through targeting beta-catenin. *Eur Rev Med Pharmacol Sci*. 2017;21:4777-4783

93. Hodges WM, O'Brien F, Fulzele S, Hamrick MW. Function of micrornas in the osteogenic differentiation and therapeutic application of adipose-derived stem cells (ascs). *Int J Mol Sci*. 2017;18

94. Oskowitz AZ, Penfornis P, Tucker A, Prockop DJ, Pochampally R. Drosha regulates hmscs cell cycle progression through a mirna independent mechanism. *Int J Biochem Cell Biol*. 2011;43:1563-1572

95. Wang Y, Fan H, Zhao G, Liu D, Du L, Wang Z, et al. Mir-16 inhibits the proliferation and angiogenesis-regulating potential of mesenchymal stem cells in severe pre-eclampsia. *FEBS J*. 2012;279:4510-4524

96. Lai VK, Ashraf M, Jiang S, Haider K. Microrna-143 is a critical regulator of cell cycle activity in stem cells with co-overexpression of akt and angiopoietin-1 via transcriptional regulation of erk5/cyclin d1 signaling. *Cell Cycle*. 2012;11:767-777

97. He B, Xu Z, Chen J, Zheng D, Li A, Zhang LS. Upregulated microrna-143 inhibits cell proliferation in human nasopharyngeal carcinoma. *Oncol Lett*. 2016;12:5023-5028

98. Peng S, Gao D, Gao C, Wei P, Niu M, Shuai C. Micrornas regulate signaling pathways in osteogenic differentiation of mesenchymal stem cells (review). *Mol Med Rep*. 2016;14:623-629

99. Park MG, Kim JS, Park SY, Lee SA, Kim HJ, Kim CS, et al. Microrna-27 promotes the differentiation of odontoblastic cell by targeting apc and activating wnt/beta-catenin signaling. *Gene*. 2014;538:266-272

100. You L, Pan L, Chen L, Gu W, Chen J. Mir-27a is essential for the shift from osteogenic differentiation to adipogenic differentiation of mesenchymal stem cells in postmenopausal osteoporosis. *Cell Physiol Biochem*. 2016;39:253-265

101. Zhu Y, Zhang X, Ding X, Wang H, Chen X, Zhao H, et al. Mir-27 inhibits adipocyte differentiation via suppressing creb expression. *Acta Biochim Biophys Sin (Shanghai)*. 2014;46:590-596

102. Hamam D, Ali D, Kassem M, Aldahmash A, Alajez NM. Micrornas as regulators of adipogenic differentiation of mesenchymal stem cells. *Stem Cells Dev*. 2015;24:417-425

103. Bendris N, Lemmers B, Blanchard JM. Cell cycle, cytoskeleton dynamics and beyond: The many functions of cyclins and cdk inhibitors. *Cell Cycle*. 2015;14:1786-1798

104. Manning AL, Dyson NJ. Rb: Mitotic implications of a tumour suppressor. *Nat Rev Cancer*. 2012;12:220-226

105. Thangavel C, Boopathi E, Liu Y, Haber A, Ertel A, Bhardwaj A, et al. Rb loss promotes prostate cancer metastasis. *Cancer Res*. 2017;77:982-995

106. Bai X, Zhou Y, Chen P, Yang M, Xu J. Microrna-142-5p induces cancer stem cell-like properties of cutaneous squamous cell carcinoma via inhibiting pten. *J Cell Biochem*. 2017

107. Bao B, Li Y, Ahmad A, Azmi AS, Bao G, Ali S, et al. Targeting csc-related mirnas for cancer therapy by natural agents. *Curr Drug Targets*. 2012;13:1858-1868

108. Lavon I, Zrihan D, Granit A, Einstein O, Fainstein N, Cohen MA, et al. Gliomas display a microrna expression profile reminiscent of neural precursor cells. *Neuro Oncol*. 2010;12:422-433

109. Vira D, Basak SK, Veena MS, Wang MB, Batra RK, Srivatsan ES. Cancer stem cells, micrornas, and therapeutic strategies including natural products. *Cancer Metastasis Rev*. 2012;31:733-751

110. Xia W, Lo CM, Poon RYC, Cheung TT, Chan ACY, Chen L, et al. Smad inhibitor induces csc differentiation for effective chemosensitization in cyclin d1- and tgf-beta/smad-regulated liver cancer stem cell-like cells. *Oncotarget*. 2017;8:38811-38824

111. Garg M. Emerging role of micrornas in cancer stem cells: Implications in cancer therapy. *World J Stem Cells*. 2015;7:1078-1089

112. Xu YF, Hannafon BN, Ding WQ. Microrna regulation of human pancreatic cancer stem cells. *Stem Cell Investig*. 2017;4:5

113. O'Donnell KA, Wentzel EA, Zeller KI, Dang CV, Mendell JT. C-myc-regulated micrornas modulate e2f1 expression. *Nature*. 2005;435:839-843

114. Sylvestre Y, De Guire V, Querido E, Mukhopadhyay UK, Bourdeau V, Major F, et al. An e2f/mir-20a autoregulatory feedback loop. *J Biol Chem*. 2007;282:2135-2143

115. Pickering MT, Stadler BM, Kowalik TF. Mir-17 and mir-20a temper an e2f1-induced g1 checkpoint to regulate cell cycle progression. *Oncogene*. 2009;28:140-145

116. Yu Z, Wang C, Wang M, Li Z, Casimiro MC, Liu M, et al. A cyclin d1/microrna 17/20 regulatory feedback loop in control of breast cancer cell proliferation. *J Cell Biol*. 2008;182:509-517

117. Babashah S, Soleimani M. The oncogenic and tumour suppressive roles of microrna s in cancer and apoptosis. *Eur J Cancer*. 2011;47:1127-1137

118. Mu P, Han YC, Betel D, Yao E, Squatrito M, Ogrodowski P, et al. Genetic dissection of the mir-17~92 cluster of microrna s in myc-induced b-cell lymphomas. *Genes Dev*. 2009;23:2806-2811

119. Olive V, Jiang I, He L. Mir-17-92, a cluster of mirnas in the midst of the cancer network. *Int J Biochem Cell Biol*. 2010;42:1348-1354

120. Wong P, Iwasaki M, Somervaille TC, Ficara F, Carico C, Arnold C, et al. The mir-17-92 microrna polycistron regulates mll leukemia stem cell potential by modulating p21 expression. *Cancer Res*. 2010;70:3833-3842

121. Cioffi M, Trabulo SM, Sanchez-Ripoll Y, Miranda-Lorenzo I, Lonardo E, Dorado J, et al. The mir-17-92 cluster counteracts quiescence and chemoresistance in a distinct subpopulation of pancreatic cancer stem cells. *Gut*. 2015;64:1936-1948

122. Balzeau J, Menezes MR, Cao S, Hagan JP. The lin28/let-7 pathway in cancer. *Front Genet*. 2017;8:31

123. Jakymiw A, Patel RS, Deming N, Bhattacharyya I, Shah P, Lamont RJ, et al. Overexpression of dicer as a result of reduced let-7 microrna levels contributes to increased cell proliferation of oral cancer cells. *Genes Chromosomes Cancer*. 2010;49:549-559

124. Li XX, Gao SY, Wang PY, Zhou X, Li YJ, Yu Y, et al. Reduced expression levels of let-7c in human breast cancer patients. *Oncol Lett*. 2015;9:1207-1212

125. Bueno MJ, Gomez de Cedron M, Gomez-Lopez G, Perez de Castro I, Di Lisio L, Montes-Moreno S, et al. Combinatorial effects of micrornas to suppress the myc oncogenic pathway. *Blood*. 2011;117:6255-6266

126. Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB. Prediction of mammalian microrna targets. *Cell*. 2003;115:787-798

127. Okada N, Lin CP, Ribeiro MC, Biton A, Lai G, He X, et al. A positive feedback between p53 and mir-34 mirnas mediates tumor suppression. *Genes Dev*. 2014;28:438-450

128. Bommer GT, Gerin I, Feng Y, Kaczorowski AJ, Kuick R, Love RE, et al. P53-mediated activation of mirna34 candidate tumor-suppressor genes. *Curr Biol*. 2007;17:1298-1307

129. He L, He X, Lim LP, de Stanchina E, Xuan Z, Liang Y, et al. A microrna component of the p53 tumour suppressor network. *Nature*. 2007;447:1130-1134

130. Tazawa H, Tsuchiya N, Izumiya M, Nakagama H. Tumor-suppressive mir-34a induces senescence-like growth arrest through modulation of the e2f pathway in human colon cancer cells. *Proc Natl Acad Sci U S A*. 2007;104:15472-15477

131. Sun F, Fu H, Liu Q, Tie Y, Zhu J, Xing R, et al. Downregulation of ccnd1 and cdk6 by mir-34a induces cell cycle arrest. *FEBS Lett*. 2008;582:1564-1568

132. Engkvist ME, Stratford EW, Lorenz S, Meza-Zepeda LA, Myklebost O, Munthe E. Analysis of the mir-34 family functions in breast cancer reveals annotation error of mir-34b. *Sci Rep*. 2017;7:9655

133. Kim HS, Lee KS, Bae HJ, Eun JW, Shen Q, Park SJ, et al. Microrna-31 functions as a tumor suppressor by regulating cell cycle and epithelial-mesenchymal transition regulatory proteins in liver cancer. *Oncotarget*. 2015;6:8089-8102

134. Calin GA, Cimmino A, Fabbri M, Ferracin M, Wojcik SE, Shimizu M, et al. Mir-15a and mir-16-1 cluster functions in human leukemia. *Proc Natl Acad Sci U S A*. 2008;105:5166-5171

135. Bottoni A, Piccin D, Tagliati F, Luchin A, Zatelli MC, degli Uberti EC. Mir-15a and mir-16-1 down-regulation in pituitary adenomas. *J Cell Physiol*. 2005;204:280-285

136. Klein U, Lia M, Crespo M, Siegel R, Shen Q, Mo T, et al. The dleu2/mir-15a/16-1 cluster controls b cell proliferation and its deletion leads to chronic lymphocytic leukemia. *Cancer Cell*. 2010;17:28-40

137. Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, et al. Frequent deletions and down-regulation of micro- rna genes mir15 and mir16 at 13q14 in chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A*. 2002;99:15524-15529

138. Cimmino A, Calin GA, Fabbri M, Iorio MV, Ferracin M, Shimizu M, et al. Mir-15 and mir-16 induce apoptosis by targeting bcl2. *Proc Natl Acad Sci U S A*. 2005;102:13944-13949

139. Linsley PS, Schelter J, Burchard J, Kibukawa M, Martin MM, Bartz SR, et al. Transcripts targeted by the microrna-16 family cooperatively regulate cell cycle progression. *Mol Cell Biol*. 2007;27:2240-2252

140. Bandi N, Zbinden S, Gugger M, Arnold M, Kocher V, Hasan L, et al. Mir-15a and mir-16 are implicated in cell cycle regulation in a rb-dependent manner and are frequently deleted or down-regulated in non-small cell lung cancer. *Cancer Res*. 2009;69:5553-5559

141. Asangani IA, Rasheed SA, Nikolova DA, Leupold JH, Colburn NH, Post S, et al. Microrna-21 (mir-21) post-transcriptionally downregulates tumor suppressor pdcd4 and stimulates invasion, intravasation and metastasis in colorectal cancer. *Oncogene*. 2008;27:2128-2136

142. Frankel LB, Christoffersen NR, Jacobsen A, Lindow M, Krogh A, Lund AH. Programmed cell death 4 (pdcd4) is an important functional target of the microrna mir-21 in breast cancer cells. *J Biol Chem*. 2008;283:1026-1033

143. Meng F, Henson R, Wehbe-Janek H, Ghoshal K, Jacob ST, Patel T. Microrna-21 regulates expression of the pten tumor suppressor gene in human hepatocellular cancer. *Gastroenterology*. 2007;133:647-658

144. Chen L, Zheng J, Zhang Y, Yang L, Wang J, Ni J, et al. Tumor-specific expression of microrna-26a suppresses human hepatocellular carcinoma growth via cyclin-dependent and -independent pathways. *Mol Ther*. 2011;19:1521-1528

145. Zhu Y, Lu Y, Zhang Q, Liu JJ, Li TJ, Yang JR, et al. Microrna-26a/b and their host genes cooperate to inhibit the g1/s transition by activating the prb protein. *Nucleic Acids Res*. 2012;40:4615-4625

146. Liu L, Papa EF, Dooner MS, Machan JT, Johnson KW, Goldberg LR, et al. Homing and long-term engraftment of long- and short-term renewal hematopoietic stem cells. *PLoS One*. 2012;7:e31300

147. Wang D, Zhang Z, O'Loughlin E, Wang L, Fan X, Lai EC, et al. Microrna-205 controls neonatal expansion of skin stem cells by modulating the pi(3)k pathway. *Nat Cell Biol*. 2013;15:1153-1163

148. Jackson SJ, Zhang Z, Feng D, Flagg M, O'Loughlin E, Wang D, et al. Rapid and widespread suppression of self-renewal by microrna-203 during epidermal differentiation. *Development*. 2013;140:1882-1891

149. Antonini D, Russo MT, De Rosa L, Gorrese M, Del Vecchio L, Missero C. Transcriptional repression of mir-34 family contributes to p63-mediated cell cycle progression in epidermal cells. *J Invest Dermatol*. 2010;130:1249-1257

150. Nagosa S, Leesch F, Putin D, Bhattacharya S, Altshuler A, Serror L, et al. Microrna-184 induces a commitment switch to epidermal differentiation. *Stem Cell Reports*. 2017

151. Ahmed MI, Alam M, Emelianov VU, Poterlowicz K, Patel A, Sharov AA, et al. Microrna-214 controls skin and hair follicle development by modulating the activity of the wnt pathway. *J Cell Biol*. 2014;207:549-567

152. Zhao C, Sun G, Li S, Shi Y. A feedback regulatory loop involving microrna-9 and nuclear receptor tlx in neural stem cell fate determination. *Nat Struct Mol Biol*. 2009;16:365-371

153. Sun G, Ye P, Murai K, Lang MF, Li S, Zhang H, et al. Mir-137 forms a regulatory loop with nuclear receptor tlx and lsd1 in neural stem cells. *Nat Commun*. 2011;2:529

154. Liu C, Teng ZQ, Santistevan NJ, Szulwach KE, Guo W, Jin P, et al. Epigenetic regulation of mir-184 by mbd1 governs neural stem cell proliferation and differentiation. *Cell Stem Cell*. 2010;6:433-444

155. Liu C, Teng ZQ, McQuate AL, Jobe EM, Christ CC, von Hoyningen-Huene SJ, et al. An epigenetic feedback regulatory loop involving microrna-195 and mbd1 governs neural stem cell differentiation. *PLoS One*. 2013;8:e51436

156. Cheng LC, Pastrana E, Tavazoie M, Doetsch F. Mir-124 regulates adult neurogenesis in the subventricular zone stem cell niche. *Nat Neurosci*. 2009;12:399-408

157. Makeyev EV, Zhang J, Carrasco MA, Maniatis T. The microrna mir-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mrna splicing. *Mol Cell*. 2007;27:435-448

158. Visvanathan J, Lee S, Lee B, Lee JW, Lee SK. The microrna mir-124 antagonizes the anti-neural rest/scp1 pathway during embryonic cns development. *Genes Dev*. 2007;21:744-749

159. Liu Z, Zhang C, Skamagki M, Khodadadi-Jamayran A, Zhang W, Kong D, et al. Elevated p53 activities restrict differentiation potential of microrna-deficient pluripotent stem cells. *Stem Cell Reports*. 2017;9:1604-1617

160. Wang J, Chen T, Shan G. Mir-148b regulates proliferation and differentiation of neural stem cells via wnt/beta-catenin signaling in rat ischemic stroke model. *Front Cell Neurosci*. 2017;11:329

161. Wang J, Li J, Yang J, Zhang L, Gao S, Jiao F, et al. Microrna1385p regulates neural stem cell proliferation and differentiation in vitro by targeting trip6 expression. *Mol Med Rep*. 2017;16:7261-7266

162. Crist CG, Montarras D, Pallafacchina G, Rocancourt D, Cumano A, Conway SJ, et al. Muscle stem cell behavior is modified by microrna-27 regulation of pax3 expression. *Proc Natl Acad Sci U S A*. 2009;106:13383-13387

163. Sarkar S, Dey BK, Dutta A. Mir-322/424 and -503 are induced during muscle differentiation and promote cell cycle quiescence and differentiation by down-regulation of cdc25a. *Mol Biol Cell*. 2010;21:2138-2149

164. Chen JF, Tao Y, Li J, Deng Z, Yan Z, Xiao X, et al. Microrna-1 and microrna-206 regulate skeletal muscle satellite cell proliferation and differentiation by repressing pax7. *J Cell Biol*. 2010;190:867-879

165. Dai Y, Wang YM, Zhang WR, Liu XF, Li X, Ding XB, et al. The role of microrna-1 and microrna-206 in the proliferation and differentiation of bovine skeletal muscle satellite cells. *In Vitro Cell Dev Biol Anim*. 2016;52:27-34

166. Han X, Yang F, Cao H, Liang Z. Malat1 regulates serum response factor through mir-133 as a competing endogenous rna in myogenesis. *FASEB J*. 2015;29:3054-3064

167. Liu B, Shi Y, He H, Cai M, Xiao W, Yang X, et al. Mir-221 modulates skeletal muscle satellite cells proliferation and differentiation. *In Vitro Cell Dev Biol Anim*. 2017

168. Zhang WR, Zhang HN, Wang YM, Dai Y, Liu XF, Li X, et al. Mir-143 regulates proliferation and differentiation of bovine skeletal muscle satellite cells by targeting igfbp5. *In Vitro Cell Dev Biol Anim*. 2017;53:265-271

169. Dey BK, Gagan J, Dutta A. Mir-206 and -486 induce myoblast differentiation by downregulating pax7. *Mol Cell Biol*. 2011;31:203-214

170. Peng F, Li TT, Wang KL, Xiao GQ, Wang JH, Zhao HD, et al. H19/let-7/lin28 reciprocal negative regulatory circuit promotes breast cancer stem cell maintenance. *Cell Death Dis*. 2017;8:e2569

171. Liu X, Feng J, Tang L, Liao L, Xu Q, Zhu S. The regulation and function of mir-21-foxo3a-mir-34b/c signaling in breast cancer. *Int J Mol Sci*. 2015;16:3148-3162

172. Li B, Lu Y, Yu L, Han X, Wang H, Mao J, et al. Mir-221/222 promote cancer stem-like cell properties and tumor growth of breast cancer via targeting pten and sustained akt/nf-kappab/cox-2 activation. *Chem Biol Interact*. 2017;277:33-42

173. Liu S, Patel SH, Ginestier C, Ibarra I, Martin-Trevino R, Bai S, et al. Microrna93 regulates proliferation and differentiation of normal and malignant breast stem cells. *PLoS Genet*. 2012;8:e1002751

174. Patel N, Garikapati KR, Pandita RK, Singh DK, Pandita TK, Bhadra U, et al. Mir-15a/mir-16 down-regulates bmi1, impacting ub-h2a mediated DNA repair and breast cancer cell sensitivity to doxorubicin. *Sci Rep*. 2017;7:4263

175. Janaki Ramaiah M, Naushad SM, Lavanya PON, Gayatri A, Vaishnave S, Pal-Bhadra M. Epigenetic regulation of mir-200 as the potential strategy for the therapy against triple-negative breast cancer. *Gene*. 2017

176. Zhan MN, Yu XT, Tang J, Zhou CX, Wang CL, Yin QQ, et al. Microrna-494 inhibits breast cancer progression by directly targeting pak1. *Cell Death Dis*. 2017;8:e2529

177. Xiao Z, Li CH, Chan SL, Xu F, Feng L, Wang Y, et al. A small-molecule modulator of the tumor-suppressor mir34a inhibits the growth of hepatocellular carcinoma. *Cancer Res*. 2014;74:6236-6247

178. Li M, Yang Y, Kuang Y, Gan X, Zeng W, Liu Y, et al. Mir-365 induces hepatocellular carcinoma cell apoptosis through targeting bcl-2. *Exp Ther Med*. 2017;13:2279-2285

179. Ma DN, Chai ZT, Zhu XD, Zhang N, Zhan DH, Ye BG, et al. Microrna-26a suppresses epithelial-mesenchymal transition in human hepatocellular carcinoma by repressing enhancer of zeste homolog 2. *J Hematol Oncol*. 2016;9:1

180. Sun W, Zhang Z, Wang J, Shang R, Zhou L, Wang X, et al. Microrna-150 suppresses cell proliferation and metastasis in hepatocellular carcinoma by inhibiting the gab1-erk axis. *Oncotarget*. 2016;7:11595-11608

181. Yu CC, Chen YW, Chiou GY, Tsai LL, Huang PI, Chang CY, et al. Microrna let-7a represses chemoresistance and tumourigenicity in head and neck cancer via stem-like properties ablation. *Oral Oncol*. 2011;47:202-210

182. Wang Y, Li J, Guo S, Ouyang Y, Yin L, Liu S, et al. Lin28b facilitates the progression and metastasis of pancreatic ductal adenocarcinoma. *Oncotarget*. 2017;8:60414-60428

183. Wang P, Zhuang L, Zhang J, Fan J, Luo J, Chen H, et al. The serum mir-21 level serves as a predictor for the chemosensitivity of advanced pancreatic cancer, and mir-21 expression confers chemoresistance by targeting fasl. *Mol Oncol*. 2013;7:334-345

184. Karamitopoulou E, Haemmig S, Baumgartner U, Schlup C, Wartenberg M, Vassella E. Microrna dysregulation in the tumor microenvironment influences the phenotype of pancreatic cancer. *Mod Pathol*. 2017;30:1116-1125

185. Kang M, Lee KH, Lee HS, Jeong CW, Ku JH, Kim HH, et al. Concurrent treatment with simvastatin and nf-kappab inhibitor in human castration-resistant prostate cancer cells exerts synergistic anti-cancer effects via control of the nf-kappab/lin28/let-7 mirna signaling pathway. *PLoS One*. 2017;12:e0184644

186. Nabavi N, Saidy NRN, Venalainen E, Haegert A, Parolia A, Xue H, et al. Mir-100-5p inhibition induces apoptosis in dormant prostate cancer cells and prevents the emergence of castration-resistant prostate cancer. *Sci Rep*. 2017;7:4079

187. Liu C, Kelnar K, Liu B, Chen X, Calhoun-Davis T, Li H, et al. The microrna mir-34a inhibits prostate cancer stem cells and metastasis by directly repressing cd44. *Nat Med*. 2011;17:211-215

188. Galardi S, Mercatelli N, Giorda E, Massalini S, Frajese GV, Ciafre SA, et al. Mir-221 and mir-222 expression affects the proliferation potential of human prostate carcinoma cell lines by targeting p27kip1. *J Biol Chem*. 2007;282:23716-23724

189. Silber J, Hashizume R, Felix T, Hariono S, Yu M, Berger MS, et al. Expression of mir-124 inhibits growth of medulloblastoma cells. *Neuro Oncol*. 2013;15:83-90

190. Tamim S, Vo DT, Uren PJ, Qiao M, Bindewald E, Kasprzak WK, et al. Genomic analyses reveal broad impact of mir-137 on genes associated with malignant transformation and neuronal differentiation in glioblastoma cells. *PLoS One*. 2014;9:e85591

191. Nie X, Lin, Y., Yang, X., Guo, L., Que, S., Li, X., Ge, J., Wang, G., Xiong, W., Guo, P. and Qiu, Y. Idh1r132h decreases the proliferation of u87 glioma cells through upregulation of microrna-128a. *Mol Med Rep*. 2015;2015;12:6695-6701

192. Geng J, Luo H, Pu Y, Zhou Z, Wu X, Xu W, et al. Methylation mediated silencing of mir-23b expression and its role in glioma stem cells. *Neurosci Lett*. 2012;528:185-189

193. Shi L, Wan Y, Sun G, Gu X, Qian C, Yan W, et al. Functional differences of mir-125b on the invasion of primary glioblastoma cd133-negative cells and cd133-positive cells. *Neuromolecular Med*. 2012;14:303-316

194. Rupaimoole R, Slack FJ. A role for mir-34 in colon cancer stem cell homeostasis. *Stem Cell Investig*. 2016;3:42

195. Ye Y, Zhuang J, Wang G, He S, Ni J, Xia W, et al. Microrna-605 promotes cell proliferation, migration and invasion in non-small cell lung cancer by directly targeting lats2. *Exp Ther Med*. 2017;14:867-873

196. Yang G, Zhang W, Yu C, Ren J, An Z. Microrna let-7: Regulation, single nucleotide polymorphism, and therapy in lung cancer. *J Cancer Res Ther*. 2015;11 Suppl 1:C1-6

3

CHAPTER 3

Distinct Plasma MicroRNA Aging Signature Predicts Lifespan but not Health-span

Manuscript based on this chapter: Wu JW, Mens MMJ, Goudsmit J, Ma Y, Liang L, Hofman A, Ikram MA, Ghanbari M. Distinct Plasma MicroRNA Aging Signature Predicts Lifespan but not Health-span. *Submitted*

Abstract

Background: MicroRNAs (miRNAs) are small non-coding RNAs that post-transcriptionally regulate gene expression. Evidences have shown differential expression of miRNAs in relation to age but plasma-based miRNA aging signature has not been well examined in large population studies. Here we hypothesized that a plasma-based miRNA signature of age is predictive of actual age and can be used as a biomarker of all-cause mortality.

Methods: We determined plasma expression levels of 2,083 miRNAs by RNA-sequencing in 1,930participants from the Rotterdam Study with a mean age of 72 years, followed from 2002-2012. Elastic net regression and cross-validation were used to build an age prediction model adjusted for sex. Cox proportional hazard models were applied to assess the association between plasma miRNA Delta Age (miRNA predicted age minus chronological age) and risk of mortality as well as first major morbidity, with adjustment for confounders.

Findings: Of 1,930 participants, 411 died during a median follow-up of over 7 years. Of the 591 well expressed miRNAs, we identified 291 miRNAs that were differentially expressed in relation to age ($p<8.5x10^{-5}$). The predicted miRNA-Age was correlated with chronological age (r=0.60). Each additional year increase in the miRNA Delta Age was associated with 9% elevated risk (Hazard Ratio (HR)= 1.09, 95%CI 1.05-1.13) of mortality, after adjusting for age, APOE and sex. We observed no association between miRNA Delta Age and risk of first major morbidity (HR=1.01, 95%CI 0.96-1.05).

Interpretation: Together, this study described a distinct plasma miRNA-Age signature that predicts lifespan but not health-span, shedding lights on the complicated process of human aging.

Introduction

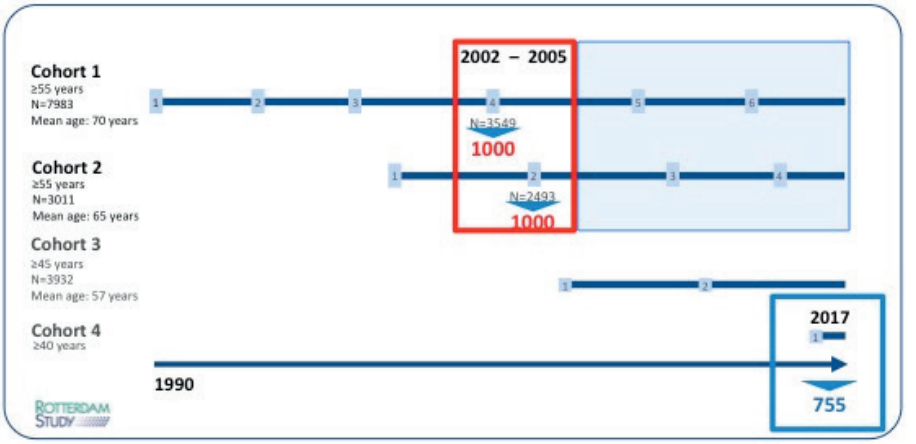
Population aging presents a serious public health burden, as age is one of the strongest risk factors for chronic-illnesses and deaths[1]. Prevention of age-related diseases and promotion of healthy aging interventions are thus of paramount importance. It has long been observed that there is person-to-person variation in terms of the pace of aging-associated functional decline[2-3]. Understanding this variation at cellular and molecular levels elucidates the underlying complex process of human aging. The concept of biological age combines the information for several easily measurable biomarkers to determine the actual health status of an individual[4]. The integrated information is expressed at an age scale. This value can be similar to the actual chronological age of the individual, higher or lower than the chronological age, reflecting increased or decreased risks of mortality. This concept of biological age, or aging clock, has been implemented and confirmed for a panel of biochemical and physical biomarkers[5-7], plus epigenetic marks using differentially methylated cytosine phosphate guanine sites (CpGs) in DNA[8-10], as well as transcriptome data including differential expression levels of messenger- RNAs[11].

MicroRNAs (miRNAs) are small non-coding RNAs that regulate gene expression at the posttranscriptional level. The high reliability and stability as well as accessibility in blood make miRNAs important candidates as diagnostic and prognostic biomarkers in human diseases[12] Evidences from studies have shown differentially expressed miRNAs in relation to age[13-14]. In a previous study with sub-set of miRNAs, age-associated miRNA expression in whole blood measured by qPCR method from more than 5,000 Framingham Heart Study (FHS) participants was found to be associated with all-cause mortality and age-related traits[15]. Biological age as a concept was not implemented and evaluated on circulatory miRNAs in plasma. It has been shown that circulating microRNA from cellular and extracellular Sources had discordant expression[15b]. It remains unclear whether cell-free miRNA expression in plasma captures similar age-signatures that predict risk of all-cause mortality and/or morbidities. Plasma is one of the standard types for circulating sample storage in large biobank cohorts. Using the more robust quantification of RNA-seq than qPCR, we aim to investigate whether a plasma-based miRNA signature of age is predictive of actual chronological age and can be used as a biomarker of risk for all-cause mortality and/or aging-related morbidities in a large human advanced-age cohort.

Methods

Study Settings

This study included participants from the Rotterdam Study, a prospective population-based cohort study[16]. In 1990, residents aged 55 years and older residing in Ommoord, a district of Rotterdam, the Netherlands, were invited to participate in the study. Of 10,215 invited inhabitants, 7,983 agreed to participate in the baseline examinations, forming the Rotterdam cohort I (RS-I). In 2000, 3,011 participants (of 4,472 invitees) who had reached 55 years of age or moved into the study district since the start of the study were added, forming the Rotterdam cohort II (RS-II). In 2006, a further extension of the cohort was started in which 3,932 participants, of 6,057 invited, aged at least 45 years living in Ommoord were included to form the Rotterdam cohort III (RS-III). In June 2016, recruitment of the Rotterdam cohort IV (RS-IV) began to enroll participants aged 40 years and over, and is ongoing with a target size of around 3,000 participants. Follow-up examinations take place every 3-4 years. The Rotterdam Study was approved by the Institutional Review Board at the Erasmus University Medical Center and received informed consent forms from all participants. For the purpose of this study, 2,000 individuals were randomly selected from the fourth round of Rotterdam Study-I (RS-I-4) and second round of Rotterdam Study-II (RS-II-2) between 2002-2005. The criteria for inclusion included the availability of informed consent and valid plasma samples that were both available at the selected visit in the Rotterdam Study. Data analyzed in this study concern 1,930 participants with physiological, clinical and miRNA assay information available (**Figure 3.1**). We focus on the RS-I and RS-II study participants (n=1,930) for miRNA age signature identification as well as development and validation of miRNA age given that they were the only ones with both miRNA profiling and clinical data. In addition, RS-IV study participants (n=754), a much younger cohort, were used for sensitivity analyses comparing miRNA age signatures between the old vs. young groups.



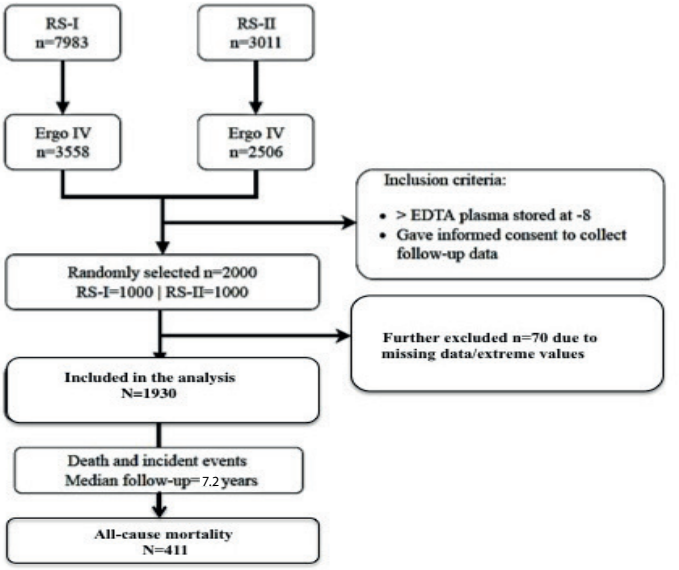


Figure 3.1a-b. Study flow chart. Illustration of the study population selection: 2,000 individuals were randomly selected from the fourth round of Rotterdam Study-I (RS-I-4) and second round of Rotterdam Study-II (RS-II-2) between 2002-2005. The criteria for inclusion included the availability of informed consent and valid plasma samples at the visits. Data analyzed in this study included 1,930 participants with physiological, clinical and miRNA assay information available. miRNA data of RS-IV were used for sensitivity analyses.

MiRNAs expression profiling and normalization

Expression levels of 2,083 plasma miRNAs were determined using the HTG EdgeSeq miRNA Whole Transcriptome Assay (HTG Molecular Diagnostics, Tuscon, AZ, USA) by the Illumina NextSeq 500 sequencer (Illumina, San Diego, CA, USA). The bioinformatics workflow consisted of two parallel paths, one for the QC checks and one for actual results processing. For the QC checks, first the average amount of ANT probe signal was tested relative to the total signal of the sample. If the relative ANT probe signal was too high, a sign that the signal of the sample was too low, the sample was either re-tested (if re-test material was available) or otherwise rejected. For the actual results processing, we used counts per million (CPM) to quantify miRNA expression. For the purpose of standardization of adjustment for total reads per sample, Log2 transformation of CPM was used. We then applied a cut-off of 1.0 so that Log2 CPM < 1.0 were considered as non-expression in the samples. Finally, we used the lower limit of quantification (LLOQ) method for normalization to select well-expressed miRNAs, by modelling the relation between mean and standard deviation of Log2 CPM among the 1,930 participants with a monotonic smooth spline fit. The miRNAs with 50% expression values above LLOQ were defined as well-expressed in plasma. Out of the 2,083 detected miRNAs, 591 miRNAs were considered well-expressed.

Differentially expressed miRNAs in relation to chronological age

To identify age-related miRNA expression in plasma, we used linear regression to model individual miRNA expression as the dependent variable and age as an explanatory variable, with adjustment of sex. We applied FDR adjustments for p-values. We constructed a volcano plot (with –log10(pvalue) on the y-axis and fold change of mean expression level per 40 years on the x-axis) to display the significance and magnitude of each bivariate association.

MiRNA age prediction

Accounting for correlation between the individual miRNAs, we used an elastic net regression model (implemented in the R package glmnet)[17] to regress chronological age on miRNAs. We cross-validated with a one-time 60:40 split of training/validation sets. We trained the prediction algorithm of miRNA age in the discovery set (n=1158), with adjustment for sex. Regularization parameters for elastic net modeling were optimized through 10-fold cross-validation. The final alpha parameter of glmnet was set to 0.90 and the lambda value was set to 0.19. The elastic net modeling, a combination of traditional Lasso and ridge regression methods, automatically selected miRNAs for building an age predictor. The predicted age resulted from the final model was used to calculate miRNA age in the validation set. MiRNA Delta Age was then defined as miRNA age minus chronological age.

Association between miRNA Delta Age and risk of all-cause mortality or morbidities

We used Cox proportional hazard models to assess the association between miRNA Delta Age and risk of all-cause mortality in the final validation dataset (n=720), controlling for chronological age, APOE status and sex. We did not include modifiable factors such as smoking and BMI because we set to ask the question whether miRNA age can serve as a biological aging surrogate and be capable of predicting risks of mortality/morbidities, with adjustment of major non-modifiable factors. Follow-up was truncated on January 1st 2012 in the Cox-models. Hazard ratios were expressed as annual risk of death over the follow up period. The association between the set of miRNAs and morbidities was also analyzed in a similar way. Prevalent cases of the outcome of interest was excluded from the respective analyses. The morbidities included: diabetes mellitus, stroke, CHD, dementia, COPD and cancer. First morbidity was defined as first occurrence of any of these major morbidities.

Ascertainment of outcomes

The outcome measures for this analysis were all-cause mortality and morbidities. Outcome analyses included all deaths/morbidity endpoints that occurred prior to January 1st 2012. Information on vital status of participants was obtained on a weekly basis via municipal population registries and through general practitioners' and hospitals' databases. Events were coded according to the International Classification of Diseases 10th version (ICD-10) by two independent research physicians. All-cause mortality is defined as participants who died from any cause during the total follow-up period, which was completed until January 1st 2012. More details on ascertainment of morbidities are provided in **Supplementary Table 3.1**.

Results

Study population

In total 1,930 participants from RS-I and RS-II were followed-up between 2002-2012, with a mean follow-up time of 7.21 (SD=1.8) years, with ascertainment of 411 all-cause mortalities. The mean age for this study sample was 71.6 years (SD=7.6), representing an advanced age population. The basic characteristics of the study population are summarized in **Table 3.1**. There was no significant difference with regards to these characteristics comparing the training and validation sets.

Table 3.1. Descriptive characteristics of the study population (n=1930)

Characteristics	Study Population (n=1930)
Age, mean (SD)	71.60 (7.6)
Gender, female (%)	1099 (56.9%)
Smoking status	
Current (%)	280 (14.51%)
Former (%)	1072 (55.54%)
Never (%)	578 (29.9%)
BMI (kg/m2), mean (SD)	27.64(4.1)
Total cholesterol, mean (SD)	5.64(0.99)
High-density lipoprotein, mean (SD)	1.45(0.40)
Glucose (mmol/L), mean (SD)	5.83(1.4)
Log(C-reactive protein), mean (SD)	3.05(5.6)
Systolic blood pressure (mmHg), mean (SD)	148.1±21.0
Diastolic blood pressure (mmHg), mean (SD)	79.5±10.9
Prevalent type 2 diabetes (%)	264 (13.7%)
Prevalent stroke (%)	79 (4.1%)
Prevalent coronary heart disease (%)	210 (10.9%)
Prevalent dementia (%)	11 (0.57%)
Prevalent COPD (%)	171 (8.9%)
Prevalent any cancer (%)	294 (15.2%)
Prevalent lung cancer (%)	4 (2.1%)

Identifying plasma-based miRNA-Age signature

Of the 591 well-expressed miRNAs remained after normalization using the RS-I & II cohorts (n=1,930), we identified 291 miRNAs that were differentially expressed in relation to chronological age, at $p < 8.5x10^{-5}$ (Bonferroni-corrected, 0.05/591), adjusted for sex. **Table 3.2** summarizes the top 20 miRNAs among them, and **Figure 3.2** is a volcano plot illustrating the significance level as well as magnitude of the difference in relation to chronological age for all the 591 miRNAs. The top three significant miRNAs were miR-19b-3p, miR-3141 and miR-197-5p. **Figure 3.3** plots the top ten significant miRNA expression values by age at the Log2 CPM scale.

Table 3.2. Plasma miRNA aging signature according to chronological age

miRNA	Mean Expression (Log2)	Coefficient (Log2)	Fold Change /40 Years	P-Value	FDR-adjusted P-Value
miR_1 9b_3p	11.174	-0.039	0.336	4.269E-79	2.523E-76
miR_3141	10.195	0.026	2.030	1.331E-66	3.932E-64
miR_197_5p	12.121	0.044	3.404	1.226E-62	2.415E-60
miR_19a_3p	9.224	-0.023	0.531	5.635E-59	8.326E-57
miR_93_5p	10.384	-0.022	0.538	9.209E-59	1.089E-56
miR_658	9.855	0.021	1.814	1.625E-57	1.601E-55
miR_17_3p	8.228	-0.020	0.580	3.219E-57	2.718E-55
miR_6887_5p	7.657	0.029	2.215	5.096E-54	3.765E-52
miR_106b_5p	9.754	-0.017	0.633	1.112E-49	7.301E-48
miR_146a_5p	10.879	-0.039	0.339	1.343E-49	7.936E-48
miR_185_5p	10.723	-0.032	0.415	1.390E-48	7.466E-47
miR_92a_3p	11.136	-0.018	0.599	5.121E-47	2.522E-45
miR_4447	8.285	0.029	2.236	1.856E-45	8.439E-44
miR_150_5p	12.194	-0.020	0.576	3.533E-45	1.392E-43
miR_425_5p	9.068	-0.020	0.574	3.488E-45	1.392E-43
miR_324_3p	8.958	-0.018	0.606	4.042E-45	1.493E-43
miR_1287_5p	12.118	0.031	2.343	1.457E-44	5.064E-43
miR_6722_3p	7.645	0.021	1.806	6.145E-44	2.018E-42
miR_25_3p	9.720	-0.018	0.603	4.573E-43	1.423E-41
miR_652_3p	8.473	-0.018	0.600	5.250E-41	1.551E-39

The top 20 miRNAs that were most significantly associated with age, after gender adjustment, among the 591 well-expressed miRNAs in 1,930 participants from the Rotterdam Study. We used linear regression to model individual miRNA expression as the dependent variable and age as an explanatory variable, with adjustment of gender.

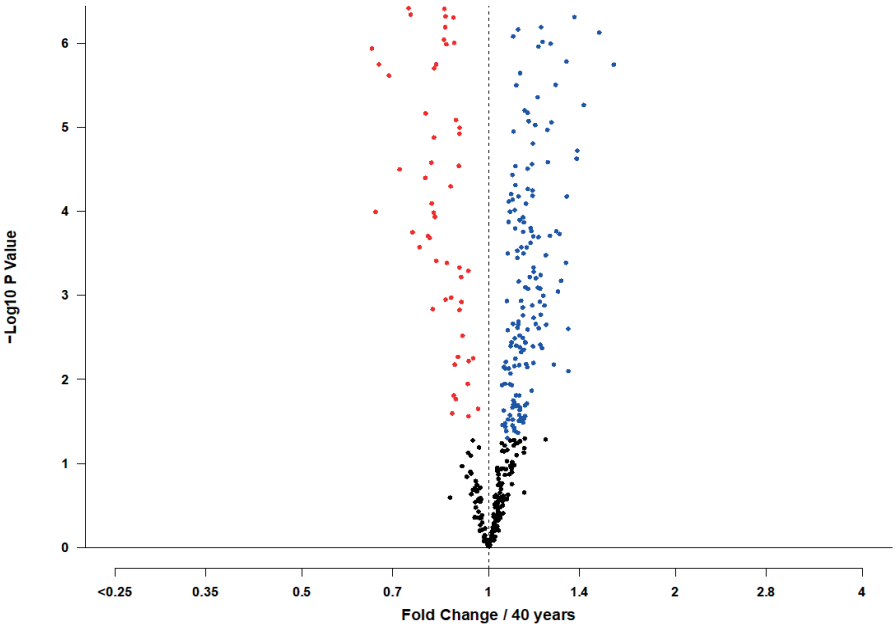
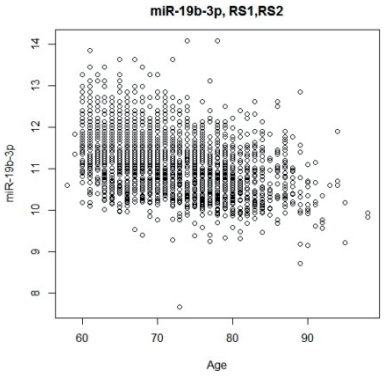
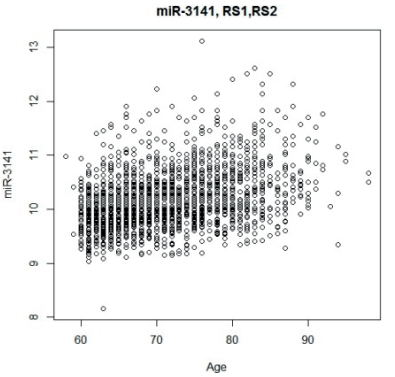
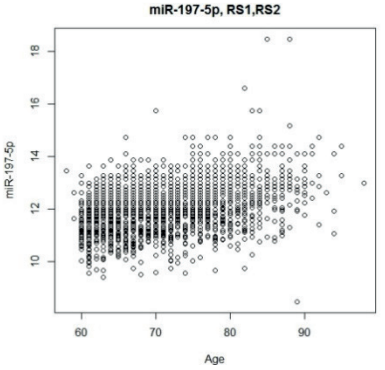
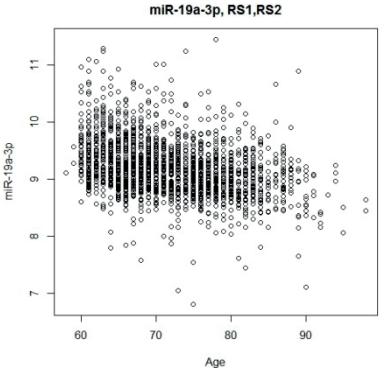


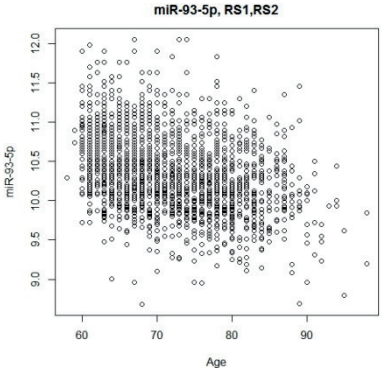
Figure 3.2. Plasma miRNA age signature. The figure illustrates univariate association of the 591 well-expressed miRNAs and chronological age, with adjustment for gender, in 1,930 participants from the Rotterdam Study. Black: insignificant (p-value>0.05); Blue: significant (p-value<=0.05. effect size positive); Red: significant (p-value<=0.05. effect size negative).

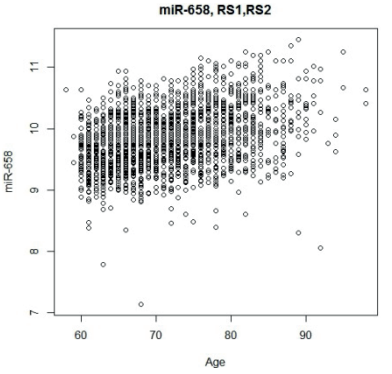

miR-19b-3p, RS1,RS2
miR-19b-3p
Age


miR-3141, RS1,RS2
miR-3141
Age


miR-197-5p, RS1,RS2
miR-197-5p
Age


miR-19a-3p, RS1,RS2
miR-19a-3p
Age


miR-93-5p, RS1,RS2
miR-93-5p
Age


miR-658, RS1,RS2
miR-658
Age

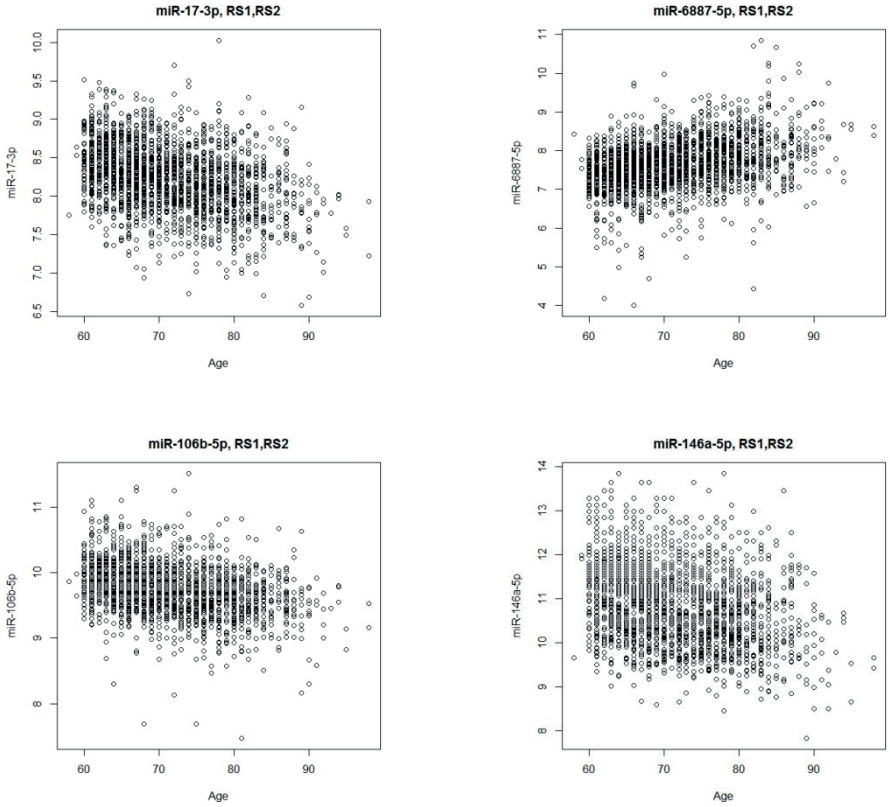


Figure 3.3. Top 10 age-related plasma miRNA expression levels by age, among the 1,930 participants from the Rotterdam Study.

MiRNA-Age signature in plasma vs. whole blood: sensitivity analyses

Intriguingly, the top 20 significant miRNA list did not match well with what Huan et al. identified previously based on whole blood samples in the Framingham Heart Study[15]. Out of the top twenty age-related whole-blood miRNAs they reported, only seven overlapped with the top age-related plasma miRNAs we identified among the RS-I and RS-II cohorts (n=1,930) (**Supplementary Table 3.2**). This discrepancy could be due to a more advanced age structure in the Rotterdam Study, miRNA levels measurement (qPCR vs RNA-Seq), differential bioinformatics processes, or sample type difference of plasma vs. whole blood. To answer this important question, we conducted a series of sensitivity analyses. First, to assess to what extent miRNA-Age signature differs across age groups, we carried out the same analyses using a younger cohort within the Rotterdam Study, the RS-IV group (mean age 54.6 ±11.5 years, n=754), and compared the miRNA aging signature between young (RS-IV) vs. advanced (RS-I & RS-II) age groups within the Rotterdam Study. Around 80% miRNAs that were differentially expressed in relation to age in the young group were overlapped with those identified in the old group, which may rule out more advanced age structure as the reason of the largely non-overlapping significant age-related miRNAs. Second, to investigate whether bioinformatics process played a role in this discrepancy, we repeated the same analyses varying the cut-off value for defining well-expressed miRNAs from 50% to 25% and 0% expression values above LLOQ. The resulting top age-related miRNA lists were almost identical, indicating that normalization was not likely the main source of the difference.

MiRNA age prediction

We used elastic net regression modeling and cross-validation to build an age prediction model selecting miRNAs while adjusting for sex. We employed cross-validations with a one-time 60:40 split of training/validation sets on the RS data (n=1,930). In the training set we developed the age prediction model; in the validation set (n=720) we calculated miRNA predicted age as well as miRNA Delta Age (miRNA predicted age minus chronological age). MiRNA age was correlated with chronological age (r=0.60) (**Figure 3.4a**). MiRNA Delta Age, another way of summarizing biological aging based on our hypothesis of capturing the magnitude of accelerated aging (for a positive Delta Age) or slower aging (for a negative Delta Age), had a normal distribution (**Figure 3.4b**).

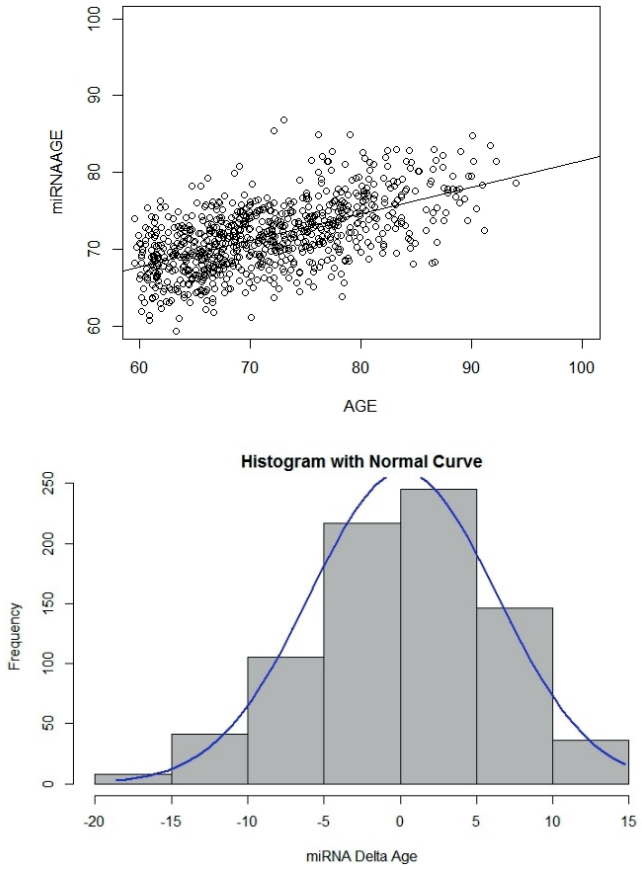


Figure 3.4. Relationship between miRNA Age and chronological age, in the validation set (n=720), the Rotterdam Study, 2000–2012. The figure shows **(a)** miRNA Age was highly correlated with chronological age (r=0.60x). **(b)** miRNA Delta Age was normally distributed.

MiRNA Delta Age predictive of all-cause mortality but not first morbidity

With the newly developed miRNA Delta Age, we then tested whether it predicts additional risks of all-cause mortality in the validation set, for people with fixed chronological age. **Figure 3.5** shows the association between miRNA Delta Age and risk of all-cause mortality in the validation set (n=720). Our results showed that each additional year increase in miRNA Delta Age was associated with 9% elevated risk of mortality, after adjusting for chronological age, APOE status and sex. On the other hand, when we assessed the association of miRNA Delta Age and the risk of first morbidity in the same validation set, no evidence was found indicating any such association (**Figure 3.5**).

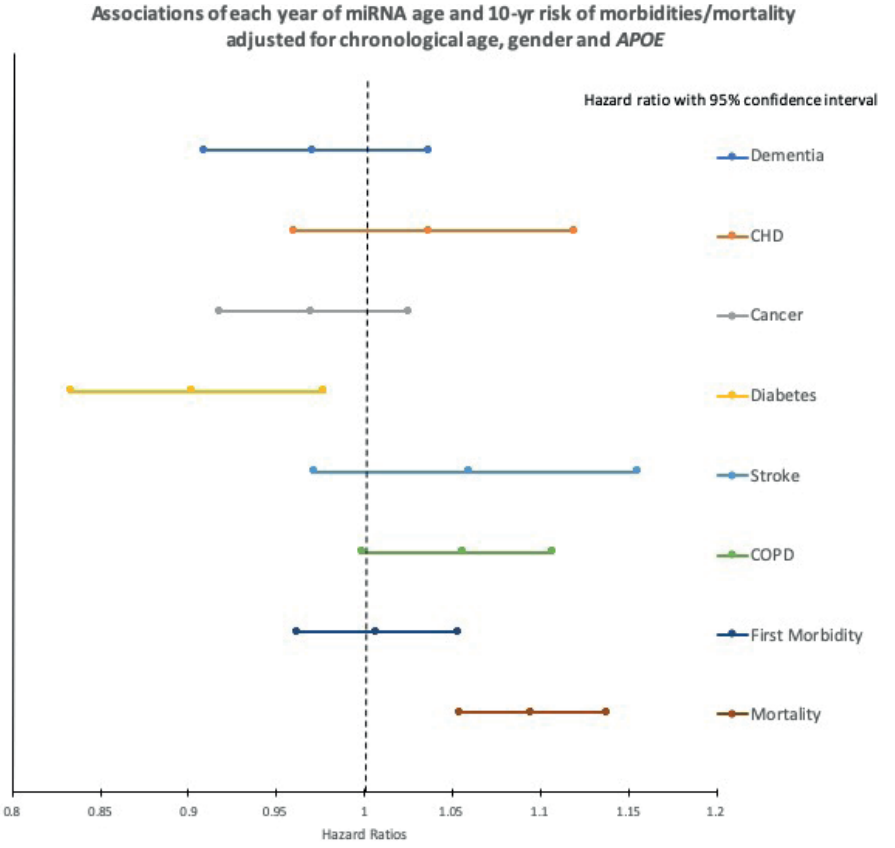


Figure 3.5. Association of miRNA Delta Age and morbidities/mortality. Each additional year increase in miRNA Delta Age was associated with 9% elevated risk of mortality, after adjusting for chronological age, APOE status and gender, in the validation set (n=720), the Rotterdam Study, 2002–2012. On the other hand, when we assessed the association of miRNA Delta Age and the risk of first morbidity in the same validation set, no evidence was found indicating any such association.

Discussion

In an advanced-age cohort we assessed the levels of circulatory miRNAs in plasma by RNAsequencing and identified 291 age-associated miRNAs. We described the panels of miRNAs of which expression levels in plasma were associated with age, and developed a plasma-based miRNA-predicted age proxy. We found that the difference between our developed plasma miRNA age and chronological age predicted additional risks of all-cause mortality but not allcause morbidity.

To the best of our knowledge, we are the first to systematically describe a distinct plasma miRNA age signature in an advanced-age cohort. This is based on a novel RNA-seq method (the Whole Transcriptome Assay technology) in a large cohort of comparable size to assay over 2,000 plasma miRNAs. Our findings are consistent with previous literature in confirming the age-relatedness of miRNA expression. The plasma miRNA predicted age as a biological age clock performed similarly to the whole blood miRNA aging clock, with a comparable correlation coefficient to chronological age. On the other hand, the specific panels of age-related miRNA signature differ drastically from the ones identified based on the Framingham Heart Study[15]. There were some key differences between our study and the Huan et al. paper: i. sample type (plasma vs whole blood), ii. age structure (older vs. younger), iii. the number of miRNAs, and iv. QC/normalization method. A series of sensitivity analyses showed that this difference could not be explained by the advanced age structure in the RS, nor could it be due to the QC/normalization method, as we varied these processes and the results were consistent. A more likely explanation for this discrepancy thereby pointed to the sample type. This sample-based divergence is consistent with a previous large-scale demonstration on discordant expression of circulating miRNA from cellular and extracellular sources[15], as well as evidences from a few smaller scale studies[19-20]. The finding of a distinct plasma miRNA aging signature, in comparison to the whole blood miRNAs expression, has important implications on designing future studies and developing miRNA-based biomarkers for aging and aging-related diseases. When comparing the plasma miRNA aging clock to other plasma-based molecular aging clock, we noticed that the correlation coefficient with chronological age was somewhat smaller for miRNA aging clock in comparison to, for example, DNA methylation clock or physiological aging clock. Both DNA methylation and physiological aging clocks predicted lifespan and health span[5,9], whereas the plasma miRNA aging clock only predicted lifespan and not health span. Whole-blood miRNA aging clock predicted lifespan and cardiometabolic traits; it was unclear whether it predicted other morbidities. Future studies are warranted to investigate why the age associated plasma miRNA predict only lifespan and not health span, and determine the underlying biological mechanisms. Both plasma and whole blood samples contain circulating RNAs[21]. Notably, whole blood includes cellular RNA from cells such as platelet, erythrocyte and white blood cells. Studies have shown that cellular-based

derived particles engage in genetic information transferring and gene expression regulation in distance, and could potentially contain more information for disease therapeutics[22]. On the other hand, high levels of whole blood based cellular miRNAs may mask disease-specific expression[23]. Circulatory miRNAs in plasma may have less of that issue but may be less informative for a given disease with local intracellular function only[18]. Pritchard et al suggested that circulating miRNA biomarkers might be more likely to reflect blood-based mechanisms than phenomena caused by diseases[24]. We sought to explore this mechanism by examining the top significant age-associated plasma miRNAs we identified, in particular miR-19b. Cell cycle regulation research based on this miRNA cluster is crucial in stem cell cycle regulation for supporting cellular reprogramming[25]. Some of the other miR-17-92 cluster members miR-19a, miR-92a and miR- 17 were also presented in the top significant list we identified, indicating a potential link of the miR-17-92 cluster involvement in the plasma miRNA aging mechanism. It might be possible that other specific diseases have more localized intracellular miRNAs changes that are less well captured by plasma miRNA expression, in comparison to DNA methylation or physiological biomarkers. Similarly, miR-106b that shares a high sequence homology with the members of the latter mentioned cluster, including miR-17, is involved in cell proliferation by targeting p21, a protein involved in the cell-cycle regulation[26]. We found miR-106b-5p among the top 10 age associated miRNAs, suggesting its role in aging.

The results of this study indicate that the difference between plasma miRNA age and chronological age predicts no additional risk for all-cause morbidity. Notably, we observed a smaller risk for diabetes mellitus (HR=0.90, 95%CI 0.83-0.98). Prior studies have investigated the role of circulatory miRNAs in diabetes patients. For example, Jiménez-Luena et al. found depressed levels of miR-145 established in diabetes patients, resulting in a lower glucose metabolism[27]. This is consistent with a sensitivity analysis performed in the Rotterdam Study showing that miR-145-5p was significantly lower expressed in diabetes patients compared to non-patients. Also, miR-145-5p was negatively associated with miRNA age. Out of the 291 age associated miRNAs, 64 miRNAs were at least nominally associated with diabetes in the Rotterdam Study (data not shown). Interestingly, more than half of these miRNAs were negatively associated with diabetes and most were in the same direction as the observed association with age. A skewed distribution of diabetes-associated miRNAs might be the reason for the seemingly smaller observed risk for diabetes. This is further supported by our finding that a top negatively age-associated miRNA, miR-146a-5p, decreased in diabetes patients and negatively associated with CRP and glucose levels, indicating the potential involvement of miR- 146a-5p in underlying pathways. This is consistent with findings by Balasubramanyam *et al.* of a correlation between down-regulation of miR-146a and pro-inflammatory markers and cytokines, such as *TRAF6*, NFκβ, TNFα and IL-6, indicating its role inflammatory changes in diabetes patients[28].

A major strength of our study is the usage of a large community-based follow-up study data with a novel RNA-sequencing method to assay the levels of a wide array of miRNAs in stored plasma samples. The mortality and morbidity outcomes were well-ascertained. Nevertheless, there were several limitations with our study. We have not used an independent cohort outside of the RS to replicate our results even though we have conducted cross-validation across three different RS cohorts and established internal validity. Secondly, due to data unavailability we were unable to assess the correlation of miRNA age to age prediction from other omics data such as messenger-RNAs and DNA methylation at the same visit point. Last, longitudinal miRNA data were unavailable and we could not assess the pace of miRNA aging in the cohorts and its determinants. Future studies are needed to further understand the complex biological aging process that is manifested by patterns of miRNA as well as other molecular and physiological profiles.

Conclusion

To the best of our knowledge, this is the first systematic study to describe a distinct plasma miRNA age signature in an advanced-age human cohort. The plasma-based miRNA age predictor was associated with lifespan but not health-span, shedding lights on the complicated process of human aging and aging-related outcomes. The finding of a distinct plasma miRNA aging signature in comparison to the whole blood miRNAs has important implications on designing future studies and developing miRNA-based biomarkers for aging and aging-related diseases. Further work is warranted to further develop the multi-biomarker-based biological age construct, for the purposes of monitoring and prevention of aging-related diseases.

References

1. Mokdad AH, Ballestros K, Echko M, et al. The state of US health, 1990-2016: burden of diseases, injuries, and risk factors among US states. Jama. 2018;319(14):1444-72.
2. Burch JB, Augustine AD, Frieden LA, et al. Advances in geroscience: impact on healthspan and chronic disease. Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences. 2014;69(Suppl_1):S1-S3.
3. Kennedy BK, Berger SL, Brunet A, et al. Geroscience: linking aging to chronic disease. Cell. 2014; 159:709–13.
4. Zierer J, Menni C, Kastenmuller G, Spector TD (2015) Integration of 'omics' data in aging research: from biomarkers to systems biology. Aging Cell 14(6), 933–44.
5. Levine ME (2013) Modeling the rate of senescence: can estimated biological age predict mortality more accurately than chronological age? J. Gerontol. A Biol. Sci. Med. Sci. 68, 667–674.
6. Sanders JL, Minster RL, Barmada MM, Matteini AM, Boudreau RM, Christensen K, Mayeux R, Borecki IB, Zhang Q, Perls T, Newman AB (2014) Heritability of and mortality prediction with a longevity phenotype: The Healthy Aging Index. J. Gerontol. A Biol. Sci. Med. Sci. 69, 479–485.
7. Belsky DW, Caspi A, Houts R, Cohen HJ, Corcoran DL, Danese A, Harrington H, Israel S, Levine ME, Schaefer JD, Sugden K, Williams B, Yashin AI, Poulton R, Moffitt TE (2015) Quantification of biological aging in young adults. Proc. Natl Acad. Sci. USA 112, E4104–E4110.
8. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, Klotzle B, Bibikova M, Fan J-B, Gao Y (2013) Genome-wide methylation profiles reveal quantitative views of human aging rates. Mol. Cell 49, 359–367.
9. Horvath S (2013) DNA methylation age of human tissues and cell types. Genome Biol. 14, R115.
10. Peters MJ, Joehanes R, Pilling LC, Schurmann C, Conneely KN, Powell J, Reinmaa E, Sutphin GL, Zhernakova A, Schramm K (2015) The transcriptional landscape of age in human peripheral blood. Nat. Commun. 6, 8570.
11. ElSharawy A, Keller A, Flachsbart F, Wendschlag A, Jacobs G, Kefer N, Brefort T, Leidinger P, Backes C, Meese E (2012) Genome-wide miRNA signatures of
12. Noren Hooten N, Abdelmohsen K, Gorospe M, Ejiogu N, Zonderman AB, Evans MK (2010) microRNA expression patterns reveal differential expression of target genes with age. PLoS ONE 5, e10724.
13. Zhang H, Yang H, Zhang C, Jing Y, Wang C, Liu C, Zhang R, Wang J, Zhang J, Zen K (2014) Investigation of microRNA expression in human serum during the aging process. J. Gerontol. A Biol. Sci. Med. Sci., 70, 102–109.
14. Huan T, Chen G, Liu C, et al (2017) Age-associated microRNA expression in human peripheral blood is associated with all-cause mortality and age-related traits. Aging Cell. 17, e12687
15. Ikram MA, Brusselle G, Ghanbari M, et al. Objectives, design and main findings until 2020 from the Rotterdam Study. European Journal of Epidemiology. 2020 May;35(5):483-517
16. Friedman J, Hastie T, Tibshirani R (2010) Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. 33, 1.
17. Shah R, Tanriverdi K, Levy D, Larson M, Gerstein M, Mick E, et al. (2016) Discordant Expression of Circulating microRNA from Cellular and Extracellular Sources. PLoS ONE 11(4): e0153691.

18. Ward JA, Esa N, Pidikiti R, Freedman JE, Keaney JF, Tanriverdi K, et al (2013) Circulating Cell and Plasma microRNA Profiles Differ between Non-ST-Segment and ST-Segment-Elevation Myocardial Infarction. Fam Med Med Sci Res. 2(2):108.

19. Rekker K, Saare M, Roost AM, Salumets A, Peters M (2013) Circulating microRNA Profile throughout the menstrual cycle. PLoS One. 8(11):e81166.

20. Duttagupta R, Jiang R, Gollub J, Getts RC, Jones KW. Impact of cellular miRNAs on circulating miRNA biomarker signatures. PLoS One. 2011; 6(6):e20769.

21. Risitano A, Beaulieu LM, Vitseva O, Freedman JE (2012) Platelets and platelet-like particles mediate intercellular RNA transfer. Blood. 119(26):6288–95.

22. Grasedieck S, Sorrentino A, Langer C, Buske C, Dohner H, Mertens D, et al. (2013) Circulating microRNAs in hematological diseases: principles, challenges, and perspectives. Blood. 121(25):4977–84.

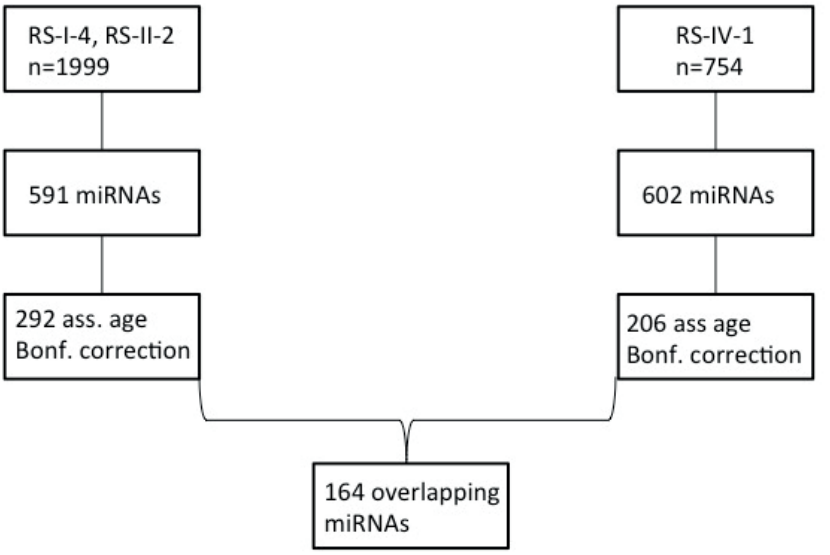
23. Hao, J., Duan, F. F., & Wang, Y. (2017). MicroRNAs and RNA binding protein regulators of microRNAs in the control of pluripotency and reprogramming. Current Opinion in Genetics & Development, 46, 95–103.

24. Ivanovska, I., et al. (2008). MicroRNAs in the miR-106b family regulate p21/CDKN1A and promote cell cycle progression. Molecular and Cellular Biology, 28(7), 2167–2174.

25. Jiménez-Lucena R, Rangel-Zúñiga OA, Alcalá-Díaz JF, et al. Circulating miRNAs as Predictive Biomarkers of Type 2 Diabetes Mellitus Development in Coronary Heart Disease Patients from the CORDIOPREV Study. Mol Ther Nucleic Acids. 2018 Sep 7;12:146-157.

26. Balasubramanyam M, Aravind S, Gokulakrishnan K et al. Impaired miR-146a expression links subclinical inflammation and insulin resistance in Type 2 diabetes. Mol Cell Biochem. 2011 May;351(1-2):197-205.

Supporting information



Supplementary Figure 3.1. Comparing miRNA aging signatures in the old (RS I&II) vs. young (RS IV) group. Around 80% miRNAs that were age- differentially expressed in the young group were overlapped with those identified in the old group

Supplementary Table 3.1. Ascertainment of major morbidities in the Rotterdam Study

Dementia Participants were screened for dementia at baseline and subsequent center visits. Those with a Mini-Mental State Examination score < 26 or Geriatric Mental Schedule score > 0 underwent further investigation and informant interview, including the Cambridge Examination for Mental Disorders of the Elderly. In addition, the entire cohort was under continuous surveillance for dementia through electronic linkage of the study database with medical records from general practitioners and the regional institute for outpatient mental health care. A consensus panel led by a consultant neurologist established the final diagnosis according to standard criteria for dementia (DSM-III-R), Alzheimer's disease (NINCDS–ADRDA), and vascular dementia (NINDS-AIREN).
CHD Prevalent coronary heart disease was defined when the participant suffered a myocardial infarction or underwent a coronary artery bypass grafting or percutaneous coronary revascularization procedure.
Cancer Occurrence of any solid cancer was determined through information obtained by 4-yearly follow-up rounds from the general practitioners (including discharge letters from hospitals) and by linkage with a nationwide registry of histo- and cytopathology in The Netherlands, Pathologisch-Anatomisch Landelijk Geautomatiseerd Archief (PALGA). Two research physicians independently assessed the first date and diagnosis of cancer. All events are pathology based and were classified according to the International Classification of Diseases (ICD) 10th edition. In case of discrepancy, consensus was sought or a cancer epidemiologist decided.
Diabetes Prevalent diabetes mellitus type 2 was identified according to the World Health Organization criteria: fasting glucose levels of ≥7.0 mmol/L, non-fasting glucose levels ≥11.1 mmol/L, or the use of glucose lowering medication. Information regarding the use of glucose lowering medication was obtained from pharmacy records and home interviews.
Stroke Stroke was defined according to the World Health Organization definition as a syndrome of rapidly developing clinical signs of focal or global disturbance of cerebral function, with symptoms lasting 24 hours or longer or leading to death, with no apparent cause other than of vascular origin. We assessed the prevalence of stroke at baseline during interview and verified it using medical records.
COPD The diagnosis and classification of COPD was based on the Global Initiative for Chronic Obstructive Lung Disease (GOLD) criteria (proportion of the forced vital capacity (FVC) exhaled in the first second (forced expiratory volume in 1 s (FEV1)/FVC ratio) <70%).

Supplementary Table 3.2. Top age-related miRNAs identified from whole blood samples in the Framingham Heart Study (14) in comparison to the significant miRNAs identified from plasma samples in the Rotterdam Study: sensitivity analysis

The top 20 miRNAs from whole blood in FHS			Replicated in RS-I-4, RS-II-2		Replicated in RS-IV-1	
miRNA	**Beta**	**P value**	**Beta**	**P value**	**Beta**	**P value**
miR-99b-5p	0.07	1.16E-286	NS		-0.00547758	2.04E-05
miR-130b-5p	0.06	3.04E-227	NS		NS	
miR-505-5p	0.06	3.04E-226	NS		NS	
miR-425-3p	0.08	8.58E-203	-0.0207	4.54331E-28	-0.01337975	6.58E-19
miR-144-5p	0.11	1.55E-165	-0.00894	5.25466E-09	NS	
miR-182-5p	0.1	7.29E-157	NS		NS	
miR-1275	0.08	5.40E-149	0.012146	3.25512E-19	NS	
miR-601	0.09	4.48E-146	NS		NS	
miR-206	0.07	4.48E-139	NS		NS	
miR-30a-5p	0.03	9.07E-132	NS		NS	
miR-218-5p	0.07	4.94E-130	NS		NS	
miR-30d-5p	0.04	9.10E-127	NS		NS	
miR-502-3p	0.05	5.28E-116	-0.02672	5.76452E-24	-0.01743285	1.9E-22
miR-28-3p	-0.08	7.96E-114	NS		NS	
miR-197-3p	0.04	3.69E-111	-0.01459	5.4063E-41	-0.01045392	1.04E-21
miR-320b	0.04	4.77E-110	NS		NS	
miR-576-3p	0.06	2.58E-105	NS		NS	
miR-181a-5p	0.06	2.61E-87	-0.00956	2.47737E-22	-0.01264021	3.79E-18
miR-18a-5p	0.05	2.93E-86	-0.01267	4.81821E-24	-0.00831963	5.67E-11
miR-223-5p	0.05	6.86E-86	NS		NS	

Out of the top twenty age-related whole-blood miRNAs they reported, only seven overlapped with the top age-related plasma miRNAs we identified among the study population (n=1930).

Supplementary Table 3.3. Comparing the plasma miRNA aging signatures between young (RS-IV) vs. advanced (RS-I & II) age groups.

Top 20 (RS-I-4&RS-II-2) n=591 miRNAs

miRNA	Avg.Expr	Lin.base.Age.Est	Lin.base.Age.PVal
miR-19b-3p	11.17401	-0.03947	1.58E-79
miR-3141	10.19511	0.025536	1.16E-66
miR-197-5p	12.1208	0.044259	6.19E-63
miR-19a-3p	9.224243	-0.02287	3.68E-59
miR-93-5p	10.3842	-0.02233	8.28E-59
miR-658	9.854701	0.021527	7.78E-58
miR-17-3p	8.228163	-0.01967	2.49E-57
miR-6887-5p	7.657342	0.028753	2.31E-54
miR-106b-5p	9.754261	-0.01654	6.88E-50
miR-146a-5p	10.87873	-0.03911	7.4E-50
miR-185-5p	10.7233	-0.03174	1.11E-48
miR-92a-3p	11.13629	-0.01846	4.88E-47
miR-4447	8.285225	0.029105	9.77E-46
miR-150-5p	12.19378	-0.01993	1.94E-45
miR-324-3p	8.957732	-0.01812	2.61E-45
miR-425-5p	9.068439	-0.02005	3.06E-45
miR-1287-5p	12.11757	0.030807	7.25E-45
miR-6722-3p	7.645348	0.021346	4.83E-44
miR-25-3p	9.720326	-0.01826	4.65E-43
miR-652-3p	8.472516	-0.01842	3.96E-41

Top 20 (RS-IV-1) n=602 miRNAs

miRNA	Avg.Expr	Lin.base.Age.Est	Lin.base.Age.PVal
miR-185-5p	11.94525	-0.02452	2.83E-46
miR-19b-3p	12.19209	-0.02902	1.55E-42
miR-93-5p	11.17349	-0.01471	2.17E-37
miR-345-5p	8.445861	-0.01124	3.27E-33
miR-425-5p	10.2279	-0.01962	1.09E-32
miR-92a-3p	11.69467	-0.01181	2.65E-31
miR-19a-3p	9.966116	-0.01588	3.25E-31
miR-197-5p	10.98398	0.032256	2.13E-30
miR-17-3p	9.100811	-0.01274	2.25E-30
miR-106b-5p	10.30572	-0.01384	4.2E-30
miR-3141	9.469	0.012579	9.76E-30
miR-324-3p	9.529065	-0.01106	1.49E-27
miR-25-3p	10.08635	-0.01305	1.64E-27
miR-146a-5p	12.28036	-0.02618	1.77E-27
miR-221-3p	11.45481	-0.02226	5.65E-27
miR-4306	8.996389	-0.01222	1.35E-26
miR-501-3p	6.390453	-0.02189	7.87E-26
miR-7107-5p	8.957094	0.014104	1.71E-25
miR-658	9.156562	0.01176	1.07E-23
miR-1255b-2-3p	8.83217	0.015481	4.47E-23

In bold illustrated the miRNAs that were in the top list of age-associated miRNAs in old and young cohort

Around 80% miRNAs that were age- differentially expressed in the young group were overlapped with those identified in the old group

Supplementary Table 3.4a. Comparing the plasma miRNA aging signatures between three normalization cut-offs among the advanced age group (RS-I&II)

RS-I-4, RS-II-2

Avg.Expr	Lin.base.Age.Est	Lin.base.Age.PVal
11.17401	-0.03947	1.58E-79
10.19511	0.025536	1.16E-66
12.1208	0.044259	6.19E-63
9.224243	-0.02287	3.68E-59
10.3842	-0.02233	8.28E-59
9.854701	0.021527	7.78E-58
8.228163	-0.01967	2.49E-57
7.657342	0.028753	2.31E-54
9.754261	-0.01654	6.88E-50
10.87873	-0.03911	7.4E-50
10.7233	-0.03174	1.11E-48
11.13629	-0.01846	4.88E-47
8.285225	0.029105	9.77E-46
12.19378	-0.01993	1.94E-45
8.957732	-0.01812	2.61E-45
9.068439	-0.02005	3.06E-45
12.11757	0.030807	7.25E-45
7.645348	0.021346	4.83E-44
9.720326	-0.01826	4.65E-43
8.472516	-0.01842	3.96E-41

3 miRNAs (0 cutoff)

miRNA	Avg.Expr	Lin.base.Age.Est	Lin.base.Age.PVal
miR.19b.3p	11.17401	-0.03947	1.58E-79
miR.3141	10.19511	0.025536	1.16E-66
miR.197.5p	12.1208	0.044259	6.19E-63
miR.19a.3p	9.224243	-0.02287	3.68E-59
miR.93.5p	10.3842	-0.02233	8.28E-59
miR.658	9.854701	0.021527	7.78E-58
miR.17.3p	8.228163	-0.01967	2.49E-57
miR.6887.5p	7.657342	0.028753	2.31E-54
miR.106b.5p	9.754261	-0.01654	6.88E-50
miR.146a.5p	10.87873	-0.03911	7.4E-50
miR.185.5p	10.7233	-0.03174	1.11E-48
miR.92a.3p	11.13629	-0.01846	4.88E-47
miR.4447	8.285225	0.029105	9.77E-46
miR.150.5p	12.19378	-0.01993	1.94E-45
miR.324.3p	8.957732	-0.01812	2.61E-45
miR.425.5p	9.068439	-0.02005	3.06E-45
miR.1287.5p	12.11757	0.030807	7.25E-45
miR.6722.3p	7.645348	0.021346	4.83E-44
miR.25.3p	9.720326	-0.01826	4.65E-43
miR.652.3p	8.472516	-0.01842	3.96E-41

591 miRNAs (0.5 cutoff)

miRNA	Avg.Expr	Lin.base.Age.Est	Lin.
miR.19b.3p	11.17401	-0.03947	
miR.3141	10.19511	0.025536	
miR.197.5p	12.1208	0.044259	
miR.19a.3p	9.224243	-0.02287	
miR.93.5p	10.3842	-0.02233	
miR.658	9.854701	0.021527	
miR.17.3p	8.228163	-0.01967	
miR.6887.5p	7.657342	0.028753	
miR.106b.5p	9.754261	-0.01654	
miR.146a.5p	10.87873	-0.03911	
miR.185.5p	10.7233	-0.03174	
miR.92a.3p	11.13629	-0.01846	
miR.4447	8.285225	0.029105	
miR.150.5p	12.19378	-0.01993	
miR.324.3p	8.957732	-0.01812	
miR.425.5p	9.068439	-0.02005	
miR.1287.5p	12.11757	0.030807	
miR.6722.3p	7.645348	0.021346	
miR.25.3p	9.720326	-0.01826	
miR.652.3p	8.472516	-0.01842	

687 miRNAs (0.25 cutoff)

Top 20 age-related miRNAs with well-expressed miRNA defined as above 0%, 25% and 50% LLOQ expression values.

Supplementary 3.4b. Comparing the plasma miRNA aging signatures between three normalization cut-offs among the young age group (RS-IV)

RS-IV-1

miRNA	Avg.Expr	Lin.base.Age.Estimate	Lin.base.Age.PValue
miR.185.5p	11.94525	-0.02452	2.83E-46
miR.19b.3p	12.19209	-0.02902	1.55E-42
miR.93.5p	11.17349	-0.01471	2.17E-37
miR.345.5p	8.445861	-0.01124	3.27E-33
miR.425.5p	10.2279	-0.01962	1.09E-32
miR.92a.3p	11.69467	-0.01181	2.65E-31
miR.19a.3p	9.966116	-0.01588	3.25E-31
miR.17.3p	9.100811	-0.01274	2.25E-30
miR.197.5p	10.98398	0.032256	2.13E-30
miR.106b.5p	10.30572	-0.01384	4.2E-30
miR.3141	9.469	0.012579	9.76E-30
miR.324.3p	9.529065	-0.01106	1.49E-27
miR.25.3p	10.08635	-0.01305	1.64E-27
miR.146a.5p	12.28036	-0.02618	1.77E-27
miR.221.3p	11.45481	-0.02226	5.65E-27
miR.4306	8.996389	-0.01222	1.35E-26
miR.501.3p	6.390453	-0.02189	7.87E-26
miR.7107.5p	8.957094	0.014104	1.71E-25
miR.658	9.156562	0.01176	1.07E-23
miR.1255b.2.3p	8.83217	0.015481	4.47E-23

2083 miRNAs (0 cutoff)

miRNA	Avg.Expr	Lin.base.Age.Estimate	Lin.base.Age.PValue
miR1855p	11.94525	-0.02452	2.83E-46
miR19b3p	12.19209	-0.02902	1.55E-42
miR935p	11.17349	-0.01471	2.17E-37
miR3455p	8.445861	-0.01124	3.27E-33
miR4255p	10.2279	-0.01962	1.09E-32
miR92a3p	11.69467	-0.01181	2.65E-31
miR19a3p	9.966116	-0.01588	3.25E-31
miR1975p	10.98398	0.032256	2.13E-30
miR173p	9.100811	-0.01274	2.25E-30
miR106b5p	10.30572	-0.01384	4.2E-30
miR3141	9.469	0.012579	9.76E-30
miR3243p	9.529065	-0.01106	1.49E-27
miR253p	10.08635	-0.01305	1.64E-27
miR146a5p	12.28036	-0.02618	1.77E-27
miR2213p	11.45481	-0.02226	5.65E-27
miR4306	8.996389	-0.01222	1.35E-26
miR5013p	6.390453	-0.02189	7.87E-26
miR71075p	8.957094	0.014104	1.71E-25
miR658	9.156562	0.01176	1.07E-23
miR1255b23p	8.83217	0.015481	4.47E-23

602 miRNAs (0.5 cutoff)

miRNA	Avg.Expr	Lin.base.Age.Estimate	Lin.base.Age.PValue
miR1855p	11.94525	-0.02452	2.83E-46
miR19b3p	12.19209	-0.02902	1.55E-42
miR935p	11.17349	-0.01471	2.17E-37
miR3455p	8.445861	-0.01124	3.27E-33
miR4255p	10.2279	-0.01962	1.09E-32
miR92a3p	11.69467	-0.01181	2.65E-31
miR19a3p	9.966116	-0.01588	3.25E-31
miR1975p	10.98398	0.032256	2.13E-30
miR173p	9.100811	-0.01274	2.25E-30
miR106b5p	10.30572	-0.01384	4.2E-30
miR3141	9.469	0.012579	9.76E-30
miR3243p	9.529065	-0.01106	1.49E-27
miR253p	10.08635	-0.01305	1.64E-27
miR146a5p	12.28036	-0.02618	1.77E-27
miR2213p	11.45481	-0.02226	5.65E-27
miR4306	8.996389	-0.01222	1.35E-26
miR5013p	6.390453	-0.02189	7.87E-26
miR71075p	8.957094	0.014104	1.71E-25
miR658	9.156562	0.01176	1.07E-23
miR1255b23p	8.83217	0.015481	4.47E-23

715 miRNAs (0.25 cutoff)

Top 20 age-related miRNAs with well-expressed miRNA defined as above 0%, 25% and 50% LLOQ expression values

PART II

MicroRNAs in diseases

4

CHAPTER 4

MicroRNA-139-5p is a causal biomarker for type 2 diabetes

Manuscript based on this chapter: Mens MMJ, Mustafa R, Ahmadizar F, Ikram MA, Evangelou M, Kavousi M, Dehghan A, Ghanbari M. MiR-139-5p is a causal biomarker for type 2 diabetes; Results from genome-wide microRNA profiling and Mendelian randomization analysis. *Submitted*

Abstract

Background: MicroRNAs (miRNAs) have emerged as key regulators of gene expression. Differential expression of miRNAs has been linked to diabetes, but underlying pathways remain poorly understood. We performed genome-wide miRNAs profiling and tested the causal associations between miRNAs and type 2 diabetes in the general population. Subsequently, we investigated target genes and metabolites of miRNAs to provide insight into the metabolic disturbances that emerge with diabetes.

Methods: Between 2002 and 2005, plasma levels of 2083 circulatory miRNAs were profiled in 1900 participants (mean age 71.4 years) of the population-based Rotterdam Study cohort. The associations of 591 well-expressed miRNAs with prevalent and incident type 2 diabetes were examined until 2015. Two-sample Mendelian Randomization (MR) was conducted to investigate the causal associations and miRNA-target genes and metabolites were studied in relation to diabetes.

Findings: At baseline, higher plasma levels of miR-139-5p and miR-193a-5p were associated (FDR<0.05) with prevalent type 2 diabetes (n=253 cases). During a follow-up of >9.0 years, 209 participants developed diabetes. Plasma levels of miR-99a-5p, miR-4664-3p, miR-29a-3p, miR-122-5p, and miR-125b-5p were significantly associated with incident diabetes. Two-sample MR confirmed a causal effect for miR-139-5p (MR-IWV-beta=0.10, $p=3.51\times10^{-4}$) on diabetes. We found several target genes and metabolites that could link miR-139-5p to pathways underlying diabetes.

Interpretation: Our study indicates a causal relationship between miR-139-5p and type 2 diabetes and suggests this miRNA as a plasma biomarker of diabetes.

Introduction

MicroRNAs (miRNAs) are small non-coding RNA molecules that have been recognized as the fine-tuners of gene expression, through repressing gene transcription or degradation of messenger (m)RNAs. Previous studies have shown that miRNA expression levels, including levels of miR-126[1] and miR-222[2], vary between type 2 diabetes patients and non-patients. Recent advantages in high-throughput technologies as next-generations sequencing methods have made it possible to investigate a wide variety of miRNAs. This allows observational studies to agnostically test associations between miRNAs and complex diseases, like type 2 diabetes, in the general population. Yet previous studies were often subject to known epidemiologic biases including confounding and reverse causation. Increasing advantages of Mendelian randomization (MR) methods enable to infer causal associations between exposure and outcome using the random inheritance property of genetic variants, therefore, enable us to go beyond observational associations and assess causality.[3]

Large scale research consortia studying genetic and metabolite markers have given us insight into underlying pathways involved in type 2 diabetes development.[4,5] Given the ability of miRNAs to target thousands of genes, miRNAs are likely also to play a regulatory role in the metabolic disturbance that causes type 2 diabetes. However, it remains unclear which specific miRNAs are the major players. In addition, the interaction between miRNAs, genes and metabolites with regard to type 2 diabetes has not yet been well described. In this study, we characterized 2083 miRNAs in an advanced population-based cohort and studied the associations of well-expressed miRNAs with prevalent and incident type 2 diabetes. Further, we studied the potential causal effect of the identified miRNAs on type 2 diabetes using MR approach and elaborate on the possible underlying pathways by investigating putative target genes and metabolites of these miRNAs.

Methods

Study population

This study was embedded within the Rotterdam Study, a prospective population-based cohort among middle-aged and elderly participants in the suburb Ommoord in Rotterdam, the Netherlands. In 1990, 7,983 inhabitants aged 55 years or older were recruited to participate in the first cohort of the Rotterdam Study (RS-I) (78% response rate of 10,215 invitees). In 2000, the Rotterdam Study was extended by 3,011 participants that moved to Ommoord or turned 55 years old (RS-II). A detailed description of the Rotterdam Study can be found elsewhere.[6] In the current study, genome-wide miRNA profiling was performed in a random subset (n=1000) of the fourth visit of Rotterdam Study-I (RS-I-4) and a random subset (n=1000) of the second visit of Rotterdam Study-II (RS-II-2). These visits were performed between 2002 and 2005 with follow-up visits every 4-5 years. From 2000 participants with miRNA data, we excluded participants with missing data on type 2 diabetes at baseline (n=99) and one participant because of missing data for all miRNAs. For the longitudinal study with incident type 2 diabetes, we additionally excluded prevalent type 2 diabetes cases at baseline (n=253).

Procedures

Full details of the Rotterdam Study miRNA expression profiling and quality control have been described previously.[7] In brief, plasma miRNA levels were determined using the HTG EdgeSeq miRNA Whole Transcriptome Assay (WTA), a next-generation sequencing (NGS)-based application that measures the expression of 2083 human miRNAs (HTG Molecular Diagnostics, Tuscon, AZ, USA), by the Illumina NextSeq 500 sequencer (Illumina, San Diego, CA, USA). Quantification of miRNA expression was based on counts per million (CPM). Log2 transformation of CPM was used as standardization and adjusted for total reads within each sample. A lower limit of quantification method was used for the normalization and selection of 591 well-expressed miRNAs that were used in the current analysis.

At baseline and during follow-up, type 2 diabetes cases were ascertained by the use of general practitioners' records, hospital discharge letters, and serum glucose measurements collected from center visits. Glucose measurements were obtained during visits to the research center. Type 2 diabetes was defined according to the World Health Organization definition as fasting glucose levels of ≥7.0 mmol/L, non-fasting glucose levels ≥11.1 mmol/L, or the use of glucose-lowering medications.[8] The follow-up for incident type 2 diabetes is complete until January 1, 2015. Participants were followed from study entry until occurrence of type 2 diabetes, death, last health status update when they were known to be free of type 2 diabetes, or January 1, 2015, whichever came first.

Statistical analysis

We used logistic regression and Cox proportional hazards regression analyses to determine the association between the plasma levels of 591 well-expressed miRNAs with prevalent and incident type 2 diabetes respectively. Odds ratios (OR) and Hazard ratios (HR) with 95% confidence interval (CI) were calculated for each additional unit log2 CPM miRNA expression. The first model was adjusted for age, sex, and cohort (RS-I-4 and RS-II-2). The second model was additionally adjusted for smoking status, alcohol use, body-mass index, high-density lipoprotein, serum total cholesterol, and hypertension. To reduce the possible bias induced by missing values, multiple imputations on confounders was performed. Values were imputed with a maximum iteration number of 10 (N=25 imputations) using the Markov Chain Monte Carlo method, R package "*mice*".[9] The Benjamini-Hochberg method was used to compute the false discovery rate (FDR).[10] Significant associations were considered when FDR<0.05. Furthermore, the identified miRNAs were categorized in 'high or low expressed' based on their median expression values and cumulative hazard graphs were generated. Analyses were performed using R version 3.6.1 (The R Foundation for Statistical Computing, Vienna, Austria).

Mendelian randomization analysis

To assess the causal relationship between the identified miRNAs and type 2 diabetes, we used the two-sample MR approach given the specific conditions illustrated in **Figure 4.1**. Genetic instruments for each identified miRNA were obtained by performing genome-wide association studies (GWAS) in the Rotterdam Study (n=1687) adjusting for age, sex, and population stratification using the first five principal components. Single-nucleotide polymorphisms (SNPs) with minor allele frequency >5% and imputation quality (Rsq) >0.7 were retained. We selected the instruments as SNPs associated with each exposure ($p= <1.0\times10^{-5}$) and F-statistics>10 to avoid the weak instrument bias. SNPs were clumped using a linkage-disequilibrium threshold of $r^2<0.1$ to remove correlated SNPs. Genetic association estimates between the instruments and type 2 diabetes were extracted from GWAS summary results on type 2 diabetes (74124 cases and 824006 controls).[5] Inverse variance weighted method (IVW) was used to combine the effect estimates of the genetic instruments. In the presence of heterogeneity, sensitivity methods such as weighted median (WM), MR-Egger and MR-PRESSO were used.[11-13] MR analysis was performed using the MRCIEU/TwoSampleMR package in R.[14]

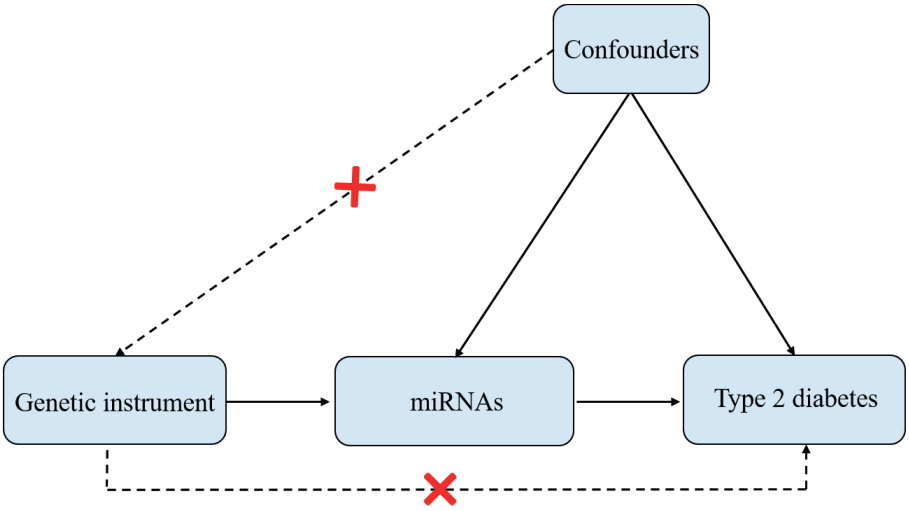


Figure 4.1. Overview of the MR process. Identified miRNAs at the observational approach were used as exposure. Type 2 diabetes was used as outcome.

Post-hoc analysis

To further elucidate the underlying mechanisms involved in the pathology of type 2 diabetes, we tested if the genetic variants were also reported to be associated with metabolite levels using summary statistics of a cross-platform, genome-wide meta-analysis of 174 metabolites across six cohorts with different measurement platforms, consisting of 86507 participants.[4] Subsequently, we performed a lookup to check if the genetic variants associated with the expression levels of miRNA of interest also explain the variation of gene expression.[15]

To dissect the putative target genes of identified miRNAs, three commonly used miRNA target prediction databases, miRTarBase[16], TargetScan (v7.2)[17] and miRDB[18], were used. We chose to include putative target genes available in either two out of three databases or reported as validated target genes by strong validation methods, such as reporter assay, western blot, and qPCR. Then, we extracted SNPs in these target genes and tested the association with type 2 diabetes and fasting glucose using summary statistics data.[5,19]

Results

Baseline characteristics of the study population are illustrated in **Table 4.1**. A total of 253 participants had prevalent type 2 diabetes at baseline and an additional 209 participants were diagnosed with incident diabetes during a mean follow-up period of 9.0 years (standard deviation (SD) 3.1).

Table 4.1. Baseline characteristics of participants of the Rotterdam Study

Variable	Non-cases (n=1438)	Prevalent type 2 diabetes (n=253)	Incident type 2 diabetes (n=209)
Age, years	71.3 ± 7.7	72.4 ± 7.5	70.8 ± 6.7
Female, n (%)	840 (58.4%)	115 (45.5%)	116 (55.5%)
Body mass index, kg/m²	27.1 ± 3.9	29.1 ± 4.5	29.2 ± 4.4
Hypertension, n (%)	1060 (73.7%)	230 (90.9%)	174 (83.3%)
Glucose, mmol/L	5.4 ± 0.5	7.9 ± 2.4	6.2 ± 1.0
Total serum cholesterol, mmol/L	5.7 ± 1.0	5.2 ± 1.0	5.6 ± 1.0
High-density lipoprotein, mmol/L	1.5 ± 0.4	1.3 ± 0.4	1.3 ± 0.3
Smoking			
Current, n (%)	208 (14.5%)	33 (13.0%)	34 (16.3%)
Former, n (%)	787 (54.7%)	156 (61.7%)	122 (58.4%)
Never, n (%)	443 (30.8%)	64 (25.3%)	53 (25.4%)
Alcohol, g/day	12.4 ± 14.9	10.8 ± 14.6	11.0 ± 13.3

Variables are presented as mean ± SD, median (IQR), or number (%)

Across 591 tested miRNAs, miR-139-5p (OR 2.63; 95%CI 1.65-4.26; $p=6.11\times10^{-5}$) and miR-193a-5p (OR 1.91; 95%CI 1.42-2.56 $p=1.86\times10^{-5}$) were significantly associated with prevalent type 2 diabetes (FDR<0.05) after adjustments for covariates in model 2. In total, 51 miRNAs were nominally (p<0.05) associated with prevalent diabetes (**Figure 4.2A; Supplementary Table 4.1**). Furthermore, Cox proportional hazard regression analysis revealed four miRNAs significantly associated with an increased risk for incident diabetes. Among these were miR-99a-5p (HR 2.08; 95%CI 1.40-3.10; $p=3.19\times10^{-4}$), miR-29a-3p (HR 1.96; 95%CI 1.35-2.84; $p=3.58\times10^{-4}$), miR-122-5p (HR 1.29; 95%CI 1.12-1.49; $p=3.65\times10^{-4}$) and miR-125b-5p (HR 1.96; 95%CI 1.35-2.85; $p=4.16\times10^{-4}$). Moreover, we found miR-4664-3p significantly associated with a reduced risk for incident diabetes (HR 0.56; 95%CI 0.41-0.77; $p=3.39\times10^{-4}$). Summary statistics of the 77 nominally associated miRNAs for incident diabetes can be found in **Supplementary Table 4.2**. Notably, miR-139-5p and miR-193a-5p were nominally associated with incident diabetes (**Figure 4.2B**), but none of the five identified miRNAs for incident diabetes were nominally associated with prevalent diabetes. Also, the OR and HR of the seven identified miRNAs in each cohort (RS-I-4 and RS-II-2) are separately illustrated in **Supplementary Figure 4.1**. The observed associations were in a consistent direction between cohorts. Additionally, see **Supplementary Table 4.3** for the baseline characteristics split by median expression values of the seven miRNAs. Notably, glucose was significantly different in low versus high categorized expressed miRNAs of all seven miRNAs.

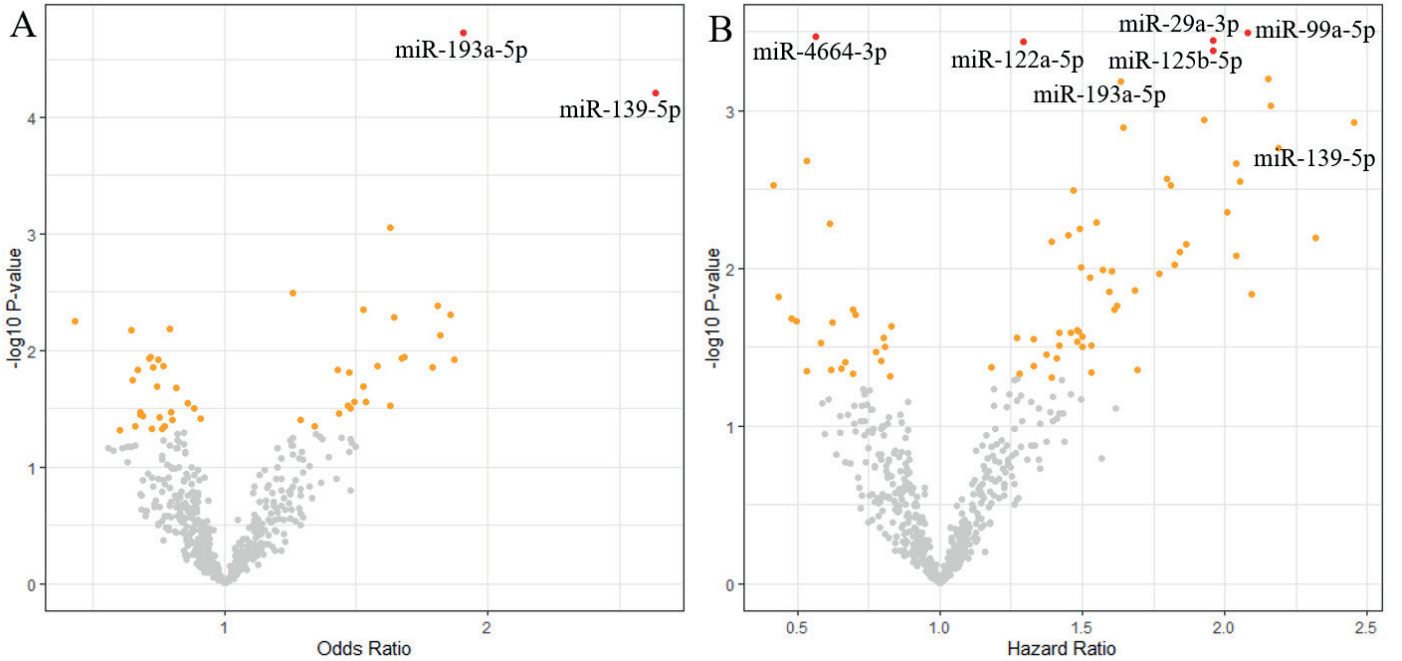


Figure 4.2. Volcano plots of the effect size of the association between plasma miRNA levels and type 2 diabetes. These plots show the odds ratios of plasma miRNA levels and prevalent diabetes (A) and hazard ratios of plasma miRNA levels and incident diabetes (B). The colors of the dots indicate the significance level: gray, non-significant; orange, nominally associated (P-value<0.05); red, significantly associated (FDR<0.05). Identified miRNAs with at least nominal association are labeled in both plots.

To assess the impact of relative expression of the identified miRNAs on diabetes risk, we categorized the expression values of each miRNA into low and high categories based on the median value. We found that high expression levels of miR-99a-5p (median normalized value >8.12), miR-29a-3p (median normalized value >10.21), miR-125b-5p (median normalized value >8.20), miR-139-5p (median normalized value >8.82) and miR-193a-5p (median normalized value >8.12) were also significantly associated with higher cumulative hazard of diabetes (**Figure 4.3**).

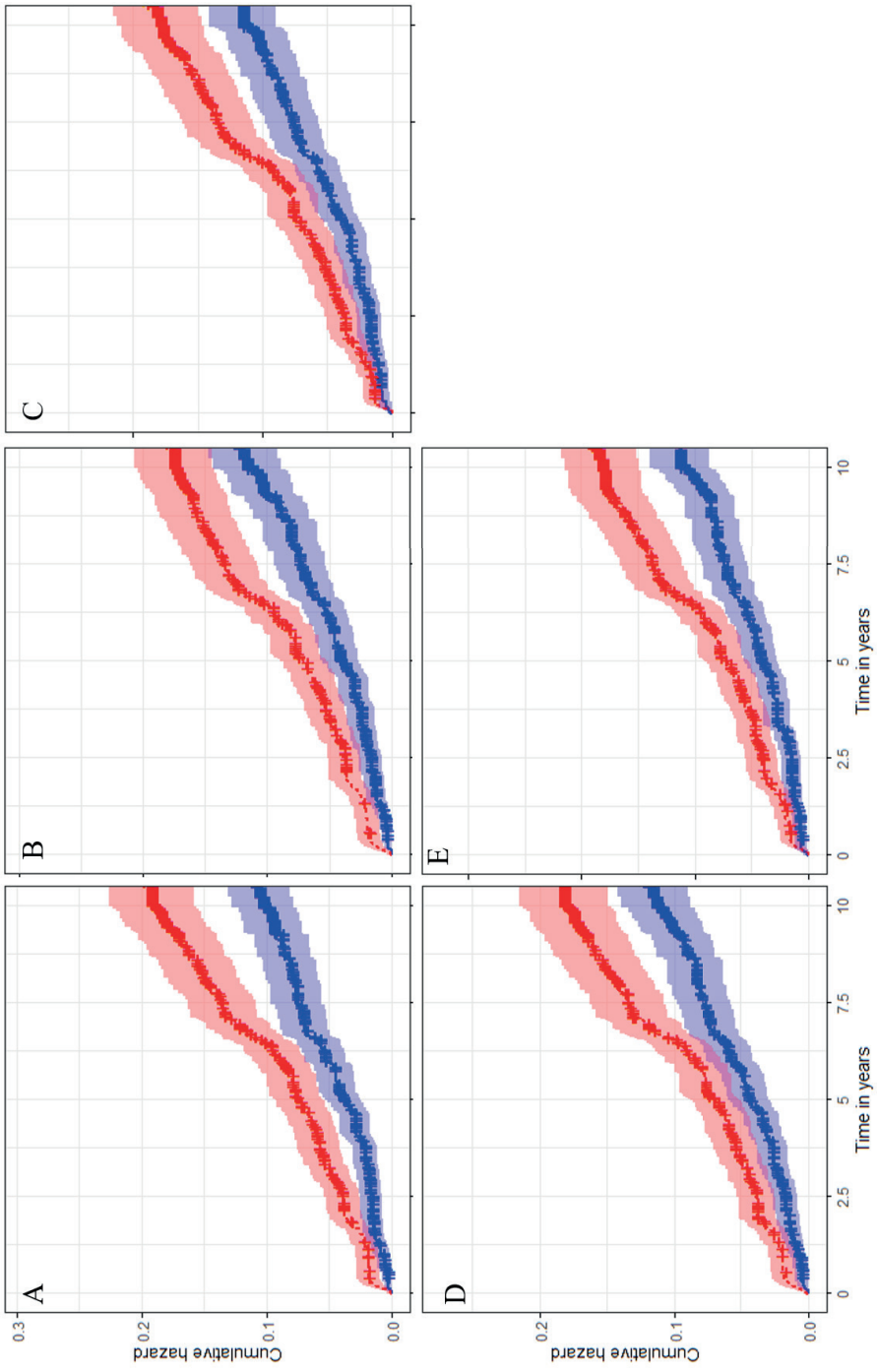

A
B
C
D
E
Cumulative hazard
0.0
0.1
0.2
0.3
Time in years
0
2.5
5
7.5
10

Figure 4.3. Significant cumulative hazard curves of log2 CPM expression of miRNAs and incident type 2 diabetes. The red curves indicate a high expression level and the blue curves indicate a low expression level based on the log2 median values of individual miRNAs. Y-axis indicates the cumulative hazard. The X-axis indicates the time scale in years. **A.** Cumulative hazard on incident diabetes between high and low expression levels of miR-99a-5p. **B.** Cumulative hazard on incident diabetes between high and low expression levels of miR-29a-3p. **C.** Cumulative hazard on incident diabetes between high and low expression levels of miR-125b-5p. **D.** Cumulative hazard on incident diabetes between high and low expression levels of miR-139-5p. **E.** Cumulative hazard on incident diabetes between high and low expression levels of miR-193a-5p.

Next, we assessed the causal relationship between the identified miRNAs and type 2 diabetes. Running GWAS in the Rotterdam Study (n=1687), we obtained genetic instruments for the seven identified miRNAs and performed two-sample MR. The results of the MR studies are presented in **Table 4.2**. MR analysis supported a causal association between miR-139-5p and type 2 diabetes (MR-IVW effect estimate 0.10; IVW $p=3.51\times10^{-4}$). We found significant heterogeneity for miR-139-5p ($p_h=5.00\times10^{-4}$), however, there was no evidence of directional pleiotropy (MR Egger intercept p=0.36). The IVW effect estimate was in the same direction with those from WM and MR-Egger methods (**Table 4.2**; **Supplementary Figures 4.2-4.4**).

Table 4.2. The results of MR analysis between miRNAs associated with type 2 diabetes

Methods	Exposure	Beta	se	P-value	P_h	N SNPs
Inverse-variance weighted	miR-139-5p	0.097	0.027	3.51×10^{-4}	5.00×10^{-4}	22
MR-Egger		0.166	0.079	4.91×10^{-2}		
Weighted Median		0.104	0.031	8.44×10^{-4}		
MR-Egger intercept				3.63×10^{-1}		
MR-PRESSO		0.058	0.025	3.37×10^{-2}		
Inverse-variance weighted	miR-193a-5p	0.067	0.041	1.02×10^{-1}	5.79×10^{-6}	11
MR-Egger		-0.019	0.194	9.24×10^{-1}		
Weighted Median		0.044	0.028	1.19×10^{-1}		
MR-Egger intercept				6.62×10^{-1}		
MR-PRESSO		0.033	0.021	1.56×10^{-1}		
Inverse-variance weighted	miR-99a-5p	0.017	0.038	6.60×10^{-1}	1.25×10^{-2}	11
MR-Egger		-0.191	0.116	1.35×10^{-1}		
Weighted Median		0.033	0.038	5.60×10^{-1}		
MR-Egger intercept				9.46×10^{-2}		
MR-PRESSO		0.041	0.031	2.13×10^{-1}		
Inverse-variance weighted	miR-4664-3p	-0.058	0.054	2.76×10^{-1}	7.43×10^{-6}	11
MR-Egger		0.108	0.224	6.42×10^{-1}		
Weighted Median		0.032	0.041	4.28×10^{-1}		

Methods	Exposure	Beta	se	P-value	P_h	N SNPs
MR-Egger intercept				4.64×10^{-1}		
MR-PRESSO		-0.015	0.041	7.22×10^{-1}		
Inverse-variance weighted	miR-29a-3p	-0.074	0.031	1.65×10^{-2}	4.73×10^{-1}	14
MR-Egger		0.006	0.112	9.56×10^{-1}		
Weighted Median		-0.105	0.043	1.33×10^{-2}		
MR-Egger intercept				4.67×10^{-1}		
MR-PRESSO		NA	NA	NA		
Inverse-variance weighted	miR-122-5p	0.029	0.017	9.51×10^{-02}	2.00×10^{-02}	11
MR-Egger		0.130	0.071	9.93×10^{-02}		
Weighted Median		0.007	0.017	6.63×10^{-01}		
MR-Egger intercept				1.76×10^{-01}		
MR-PRESSO		0.017	0.013	2.25×10^{-01}		
Inverse-variance weighted	miR-125b-5p	-0.001	0.037	9.79×10^{-1}	5.01×10^{-1}	5
MR-Egger		-0.141	0.186	5.04×10^{-1}		
Weighted Median		-0.025	0.047	6.04×10^{-1}		
MR-Egger intercept				4.99×10^{-1}		
MR-PRESSO		NA	NA	NA		

P_h, P-value for heterogeneity; N SNPs, number of genetic instruments used

To address the potential role of miR-139-5p in regulating metabolic pathways, we tested if the 22 genetic variants, that act in *trans* on the expression levels of miR-139-5p, have reported to also change the expression levels of any metabolites using the previous GWAS on plasma metabolites levels.[20] Out of the 22 genetic variants, 21 were associated with at least one metabolite (P<0.05). Over 38% of the SNP-metabolite sets came from phosphatidylcholines, followed by 17% from amino acids (**Figure 4.4**). A large proportion of the SNP-phosphatidylcholines sets is due to the association of five SNPs located on the Chr9q34.2 locus (rs687289, rs558240, rs176694, rs11244083, rs592514). Additionally, we tested if the 22 genetic variants related to miR-139-5p have been associated to change the expression of genes using the previous GWAS.[15] Genes of which their expression levels are strongly regulated by one of the 22 variants are *MED22* (Chr9q34.2), *AP4E1* (Chr15q21.2), *SURF1* (Chr9q34.2), and *SURF6* (Chr9q34.2). The full set of genes can be found in **Supplementary Table 4.4**. Notably, the aforementioned SNPs related to the phosphatidylcholines are also associated with the expression level of *ABO* on the Chr9q34.2 locus.

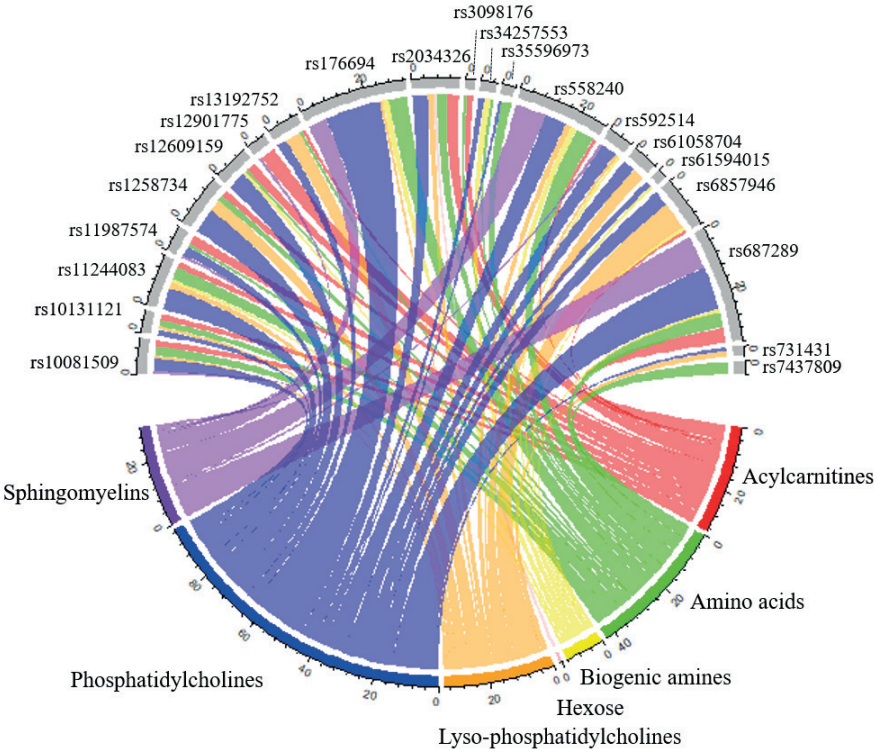


Figure 4.4. Chord diagram illustrating the link between miR-139-5p SNPs and metabolites. Each color presents a class of metabolite (e.g. sphingomyelins, phosphatidylcholines, lyso-phosphatidylcholines, hexose, biogenic amines, amino acids and acylcarnitines). Each line presents an association between SNP and

Given that miRNAs control disease risk through regulating the expression of their target genes, we extracted SNPs in 214 target genes of miR-139-5p and tested their association with type 2 diabetes and fasting glucose using summary statistics from recent GWASs.[5,19] We found 68 putative target genes and 5 validated target genes (incl. *BLC2*, *NR5A2*, *NFKB1*, *OIP5*, *MCL1*) to be associated with type 2 diabetes (FDR<0.05). In addition, we found 13 putative target genes and 2 validated target genes (incl. *IGF1R*, *NOTCH1*) to be associated with fasting glucose (**Supplementary Table 4.5**). According to GTEx v8[21], the associated target genes of miR-139-5p are most specific to be downregulated in pancreatic, kidney, heart and liver tissue. The expression of these target genes across different tissues was obtained using FUMA[21] and can be found in **Supplementary Figures 4.5-4.6**.

Discussion

In this study, we characterized 591 well-expressed miRNAs to identify associations with prevalent and incident type 2 diabetes in a population-based cohort. Using an agnostic approach, we found higher plasma levels of miR-139-5p and miR-193a-5p in diabetes patients at the baseline. Moreover, four miRNAs (miR-99a-5p, miR-29a-3p, miR-122-5p and miR-125b-5p) were positively associated with incident diabetes during follow-up. Notable, we found miR-4664-3p to be associated with incident diabetes in an opposite direction, suggesting a protective effect of this miRNA. MR analysis confirmed a causal relationship between the plasma level of miR-139-5p and the risk of type 2 diabetes. Our post-hoc analysis further provided insight into the underlying molecular pathways that could link miR-139-5p to the pathophysiology of type 2 diabetes.

The link between miRNAs and type 2 diabetes has gained increasing attention in recent years.[22-24] In particular, population-based prospective cohort studies with long follow-up time are needed to detect differentially expressed miRNAs at different stages of the disease. The observed differences in miRNA profiles between a disease versus a healthy state are a result of tissue-specific processes occurring at different stages of the disease. Therefore, miRNAs in the associations with prevalent and incident type 2 diabetes do not necessarily need to be consistent to serve as disease biomarkers. Out of the seven identified miRNAs in our study, miR-122-5p has been well-studied miRNA concerning diabetes and liver diseases. For example, a previous study[25] reported an association between miR-122-5p with type 2 diabetes, impaired fasting glucose, and HbA1c despite a small number of cases (n=24). Another study[24] found a higher risk for metabolic syndrome and diabetes over 15 years and additionally reported a strong correlation between miR-122, and lipid levels (triglycerides, LDL and HDL). Furthermore, we recently demonstrated that miR-122-5p is linked to fatty liver disease, which is a common problem in patients with diabetes.[7]

In this study, we used a genome-wide sequencing-based miRNA profiling method to investigate the causal association between plasma-derived miRNA expression and type 2 diabetes. The consistency of findings between regression analysis in the observational data and MR analysis supports a causal role for miR-139-5p in developing type 2 diabetes. MR exploits the random allocation of genetic variants during conception that is unlikely to be subject to reverse causation and confounding, therefore is comparable to randomly allocated intervention in a clinical trial.[3] Different sensitivity analyses for MR has also supported a robust estimate of causal effects. Collectively, our findings indicate that higher levels of miR-139-5p in plasma increase the risk of diabetes, which in addition to its potential as biomarker, might have therapeutic potential that warrant further investigation.

The increasing interest in discovering miRNAs for complex traits such as diabetes, has led to better miRNA quantification platforms. In particular, the EdgeSeq array, which has been used in the current study, is a high-throughput technology that is reported to be more accurate, sensitive and specific when compared to a traditional RT-qPCR method.[26] Although this EdgeSeq method is still little used in epidemiological studies, it has been shown to detect novel miRNAs with the least bias detection.[26] A recent study[27] using the same EdgeSeq array has also researched genetic variants associated with expression levels of miR-139-5p. In the latter study, they focused on different cardio-metabolic phenotypes, but not on diabetes, including lipid and glycemic traits and coronary artery disease. Also, a second study[28] that measured the expression of 750 whole-blood derived miRNAs by RT-qPCR, had identified *cis*-regulatory variants that regulate miR-139-5p. They tested the causal role of miR-139-5p on diabetes but did not find any significant effect. A possible explanation for this includes the differences in miRNA source. Given that whole blood reflects platelet and electrolyte cellular miRNAs, among other things, it may show conflicting expression from extracellular sources, such as plasma.[29] Previously identified SNPs for miR-139-5p were all regulating in *cis*[27,28], some of which were successfully replicated in our GWAS ($p<0.05$). We were able to identify *trans*-regulatory variants for miR-139-5p due to the relatively larger sample size in our study compared to the previously mentioned study[27] which used the same miRNA profiling EdgeSeq array. Our study, therefore, has a bigger power to identify genetic instruments for miRNAs for MR analysis.

MiRNA expression is often tissue specific, however, little is known about the tissue-specificity of miR-139-5p in humans. Since miR-139-5p was reported to be embedded within the intron region of *PDE2A* in a sense direction and likely to be co-expressed, we tested whether *PDE2A* was expressed in tissue relevant to diabetes, including pancreas and liver. According to the Human Protein Atlas (https://www.proteinatlas.org/), *PDE2A* is among others expressed in endocrine tissue and mainly in the adrenal gland.[30] Moreover, *PDE2A* is reported to be expressed in endothelial cells in liver and pancreas.

In order to test whether miR-139-5p is expressed in tissue relevant to diabetes, such as the pancreas and liver, the host gene (PDE2A) that share the same promoter was used as proxy. Finally, according to FANTOM5[31], miR-139-5p is similar to *PDE2A*[30] enriched in endothelial cells, of which dysfunction is a major mediator of diabetes.[32]
As a secondary aim of this study, we investigated downstream effects, including metabolites and target genes to gain better insights into the potential mechanisms miR-139-5p is involved in the pathogenesis of diabetes. We described the association between miR-139-5p and metabolites, of which phosphatidylcholines were the most abundant. Phosphatidylcholines are phospholipids attached to a choline particle and can be found in foods, such as eggs and liver. A recent prospective study[33] has shown a protective effect of phosphatidylcholines intake and diabetes. It is plausible that miR-139-5p plays role in the metabolism of phosphatidylcholines and thus influences

the risk of diabetes. Furthermore, we found multiple putative and validated target genes of miR-139-5p that are known for their role in pathways underlying diabetes. For example, *BLC2* is involved in regulating both pro-inflammatory cytokines, glucose metabolism, and pancreatic β-cell homeostasis.[34] Another target gene *IGF1R* is a tyrosine kinase receptor that is activated by *IGF-1* and is related to risk for diabetes. This is in line with a previous study[35] that found a positive relation between miR-139-5p expression and the suppression of *IRS1* gene in diabetic rats. *IRS1* plays a major role in transmitting signals from *IGF-1* to downstream cellular pathways such as the MAPK pathway. Moreover, a depressed function of *IRS1* is associated with insulin resistance and pancreatic β-cell function.[35]

Strengths of our study include, first, the investigation for well-expressed miRNAs in an agnostic manner in the population level, which could identify novel miRNAs that were not previously recognized for type 2 diabetes. Second, the use of the EdgeSeq array that allowed us to measure the levels of cell-free miRNAs in plasma in a highly specific, sensitive, and reproducible manner. Additionally, we triangulated the evidence integrating cross-sectional, and longitudinal analysis of observational data with MR analysis to examine the causal association of circulatory miRNAs and type 2 diabetes. However, some limitations should be acknowledged when interpreting the results of this study. First, during the baseline examination of this study, traditional parameters such as insulin, HbA1c, HOMA-β, and HOMA-IR were not measured. Therefore, we could not calculate the differences in disease prediction capacity between the identified miRNAs and the abovementioned parameters. Second, glucose and insulin metabolism take place in specific tissues, such as the pancreas and liver, but the analyzed miRNAs in the current study are cell-free in plasma, meaning that the tissue of origin is unknown. As miRNAs are tissue-specific, it would be interesting yet to study the identified miRNAs in relevant tissues and cell types. Moreover, further experimental validation of the identified miRNAs is warranted to confirm their regulatory roles in the pathophysiology of diabetes.

In summary, this study indicates that plasma levels of two miRNAs linked to prevalent diabetes and five miRNAs associated with incident type 2 diabetes. The identified miRNAs are possibly involved in pathways underlying the metabolic disturbance of type 2 diabetes. In particular, elevated levels of miR-139-5p have a causal effect on type 2 diabetes potentially through affecting several important genes and metabolites involved in the disease pathology.

References

1. Zampetaki A, Kiechl S, Drozdov I, et al. Plasma microRNA profiling reveals loss of endothelial miR-126 and other microRNAs in type 2 diabetes. *Circ Res* 2010; **107**(6): 810-7.
2. Ortega FJ, Mercader JM, Moreno-Navarrete JM, et al. Profiling of circulating microRNAs reveals common microRNAs linked to type 2 diabetes that change with insulin sensitization. *Diabetes Care* 2014; **37**(5): 1375-83.
3. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003; **32**(1): 1-22.
4. Lotta LA, Pietzner M, Stewart ID, et al. A cross-platform approach identifies genetic regulators of human metabolism and health. *Nature Genetics* 2021; **53**(1): 54-64.
5. Mahajan A, Taliun D, Thurner M, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet* 2018; **50**(11): 1505-13.
6. Ikram MA, Brusselle G, Ghanbari M, et al. Objectives, design and main findings until 2020 from the Rotterdam Study. *European Journal of Epidemiology* 2020.
7. Zhang X, Mens MMJ, Abozaid YJ, et al. Circulatory microRNAs as potential biomarkers for fatty liver disease: the Rotterdam study. *Aliment Pharmacol Ther* 2020.
8. World Health O, International Diabetes F. Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia : report of a WHO/IDF consultation. Geneva: World Health Organization; 2006.
9. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *2011* 2011; **45**(3): 67.
10. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J R Stat Soc B* 1995; **57**(1): 289-300.
11. Bowden J, Del Greco MF, Minelli C, Davey Smith G, Sheehan N, Thompson J. A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Stat Med* 2017; **36**(11): 1783-802.
12. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* 2015; **44**(2): 512-25.
13. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet Epidemiol* 2016; **40**(4): 304-14.
14. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife* 2018; **7**.
15. Võsa U, Claringbould A, Westra H-J, et al. Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. *bioRxiv* 2018: 447367.
16. Chou C-H, Shrestha S, Yang C-D, et al. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Research* 2017; **46**(D1): D296-D302.
17. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. *Elife* 2015; **4**.
18. Chen Y, Wang X. miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res* 2020; **48**(D1): D127-D31.

19. Lagou V, Mägi R, Hottenga J-J, et al. Sex-dimorphic genetic effects and novel loci for fasting glucose and insulin variability. *Nature Communications* 2021; **12**(1): 24.

20. Lotta LA, Pietzner M, Stewart ID, et al. Cross-platform genetic discovery of small molecule products of metabolism and application to clinical outcomes. *bioRxiv* 2020: 2020.02.03.932541.

21. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nature Communications* 2017; **8**(1): 1826.

22. Jaeger A, Zollinger L, Saely CH, et al. Circulating microRNAs -192 and -194 are associated with the presence and incidence of diabetes mellitus. *Sci Rep* 2018; **8**(1): 14274.

23. Jiménez-Lucena R, Camargo A, Alcalá-Diaz JF, et al. A plasma circulating miRNAs profile predicts type 2 diabetes mellitus and prediabetes: from the CORDIOPREV study. *Experimental & Molecular Medicine* 2018; **50**(12): 1-12.

24. Willeit P, Skroblin P, Moschen AR, et al. Circulating MicroRNA-122 Is Associated With the Risk of New-Onset Metabolic Syndrome and Type 2 Diabetes. *Diabetes* 2017; **66**(2): 347-57.

25. Mononen N, Lyytikäinen L-P, Seppälä I, et al. Whole blood microRNA levels associate with glycemic status and correlate with target mRNAs in pathways important to type 2 diabetes. *Scientific Reports* 2019; **9**(1): 8887.

26. Godoy PM, Barczak AJ, DeHoff P, et al. Comparison of Reproducibility, Accuracy, Sensitivity, and Specificity of miRNA Quantification Platforms. *Cell Reports* 2019; **29**(12): 4212-22.e5.

27. Nikpay M, Beehler K, Valsesia A, et al. Genome-wide identification of circulating-miRNA expression quantitative trait loci reveals the role of several miRNAs in the regulation of cardiometabolic phenotypes. *Cardiovasc Res* 2019; **115**(11): 1629-45.

28. Huan T, Rong J, Liu C, et al. Genome-wide identification of microRNA expression quantitative trait loci. *Nat Commun* 2015; **6**: 6601.

29. Shah R, Tanriverdi K, Levy D, et al. Discordant Expression of Circulating microRNA from Cellular and Extracellular Sources. *PLoS One* 2016; **11**(4): e0153691.

30. Uhlen M, Fagerberg L, Hallstrom BM, et al. Proteomics. Tissue-based map of the human proteome. *Science* 2015; **347**(6220): 1260419.

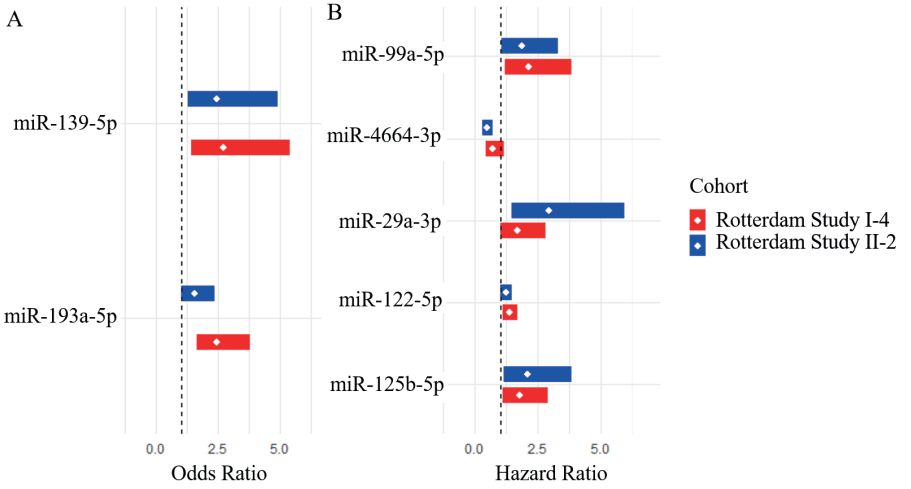
31. de Rie D, Abugessaisa I, Alam T, et al. An integrated expression atlas of miRNAs and their promoters in human and mouse. *Nature Biotechnology* 2017; **35**(9): 872-8.

32. Kaur R, Kaur M, Singh J. Endothelial dysfunction and platelet hyperactivity in type 2 diabetes mellitus: molecular insights and therapeutic strategies. *Cardiovasc Diabetol* 2018; **17**(1): 121.

33. Virtanen JK, Tuomainen T-P, Voutilainen S. Dietary intake of choline and phosphatidylcholine and risk of type 2 diabetes in men: The Kuopio Ischaemic Heart Disease Risk Factor Study. *European Journal of Nutrition* 2020; **59**(8): 3857-61.

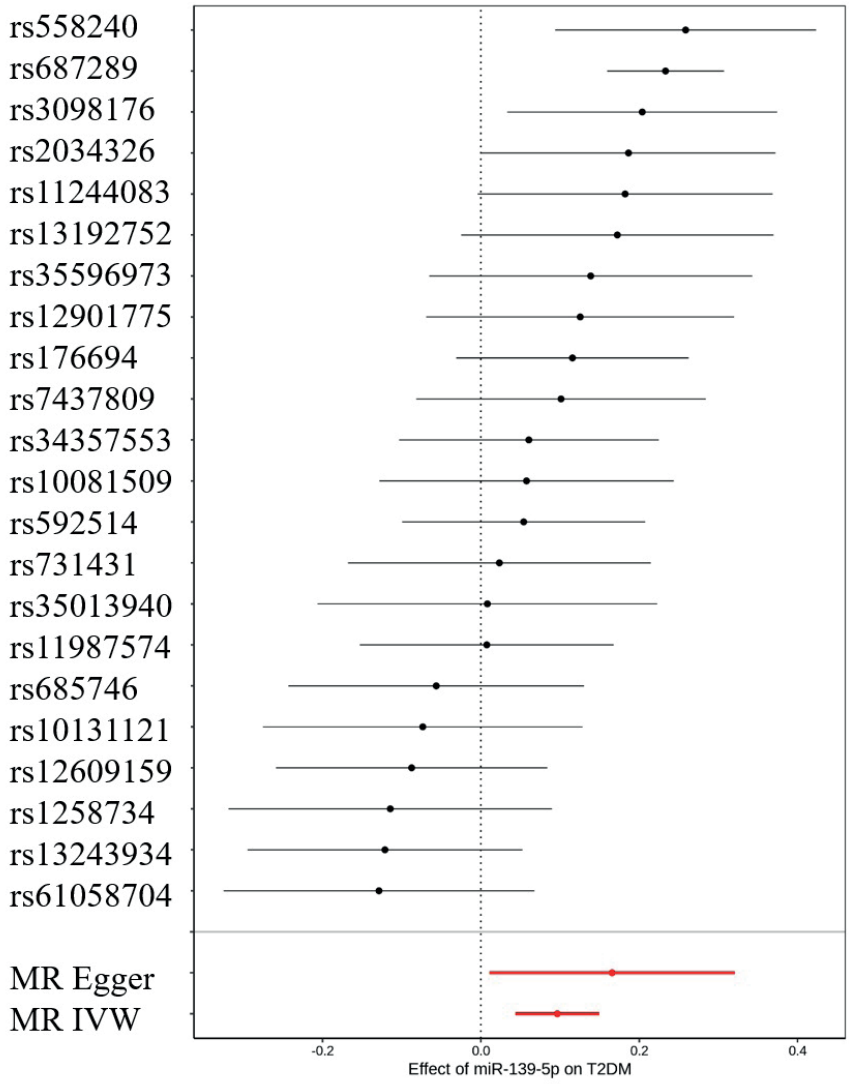
34. Gurzov EN, Eizirik DL. Bcl-2 proteins in diabetes: mitochondrial pathways of beta-cell death and dysfunction. *Trends Cell Biol* 2011; **21**(7): 424-31.

35. Li J, Su L, Gong YY, et al. Downregulation of miR-139-5p contributes to the antiapoptotic effect of liraglutide on the diabetic rat pancreas and INS-1 cells by targeting IRS1. *PLoS One* 2017; **12**(3): e0173576.

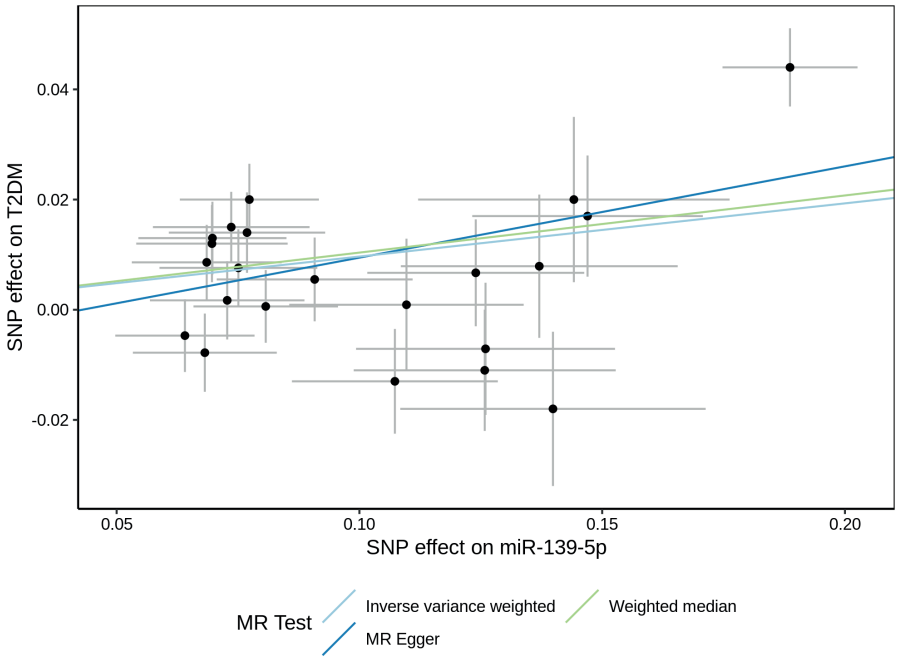


Rotterdam Study-I-4 (n=912)				Rotterdam Study-II-2 (n=988)			
Prevalent diabetes (n=126)				**Prevalent diabetes (n=127)**			
miRNA	OR	95%CI	P value	miRNA	OR	95%CI	P value
miR-139-5p	2.71	1.40-5.41	3.89×10^{-3}	miR-139-5p	2.47	1.28-4.91	8.25×10^{-3}
miR-193a-5p	2.47	1.62-3.79	2.82×10^{-5}	miR-193a-5p	1.54	1.01-2.35	4.59×10^{-2}
Incident diabetes (n=103)				**Incident diabetes (n=106)**			
miRNA	HR	95%CI	P value	miRNA	HR	95%CI	P value
miR-99a-5p	2.15	1.20-3.85	9.64×10^{-3}	miR-99a-5p	1.88	1.08-3.29	2.62×10^{-2}
miR-4664-3p	0.69	0.43-1.13	1.43×10^{-1}	miR-4664-3p	0.47	0.31-0.72	6.04×10^{-4}
miR-29a-3p	1.69	1.02-2.80	4.04×10^{-2}	miR-29a-3p	2.94	1.45-5.94	2.71×10^{-3}
miR-122-5p	1.36	1.09-1.69	6.23×10^{-3}	miR-122-5p	1.22	1.01-1.47	4.39×10^{-2}
miR-125b-5p	1.78	1.09-2.90	2.20×10^{-2}	miR-125b-5p	2.08	1.13-3.82	1.82×10^{-2}

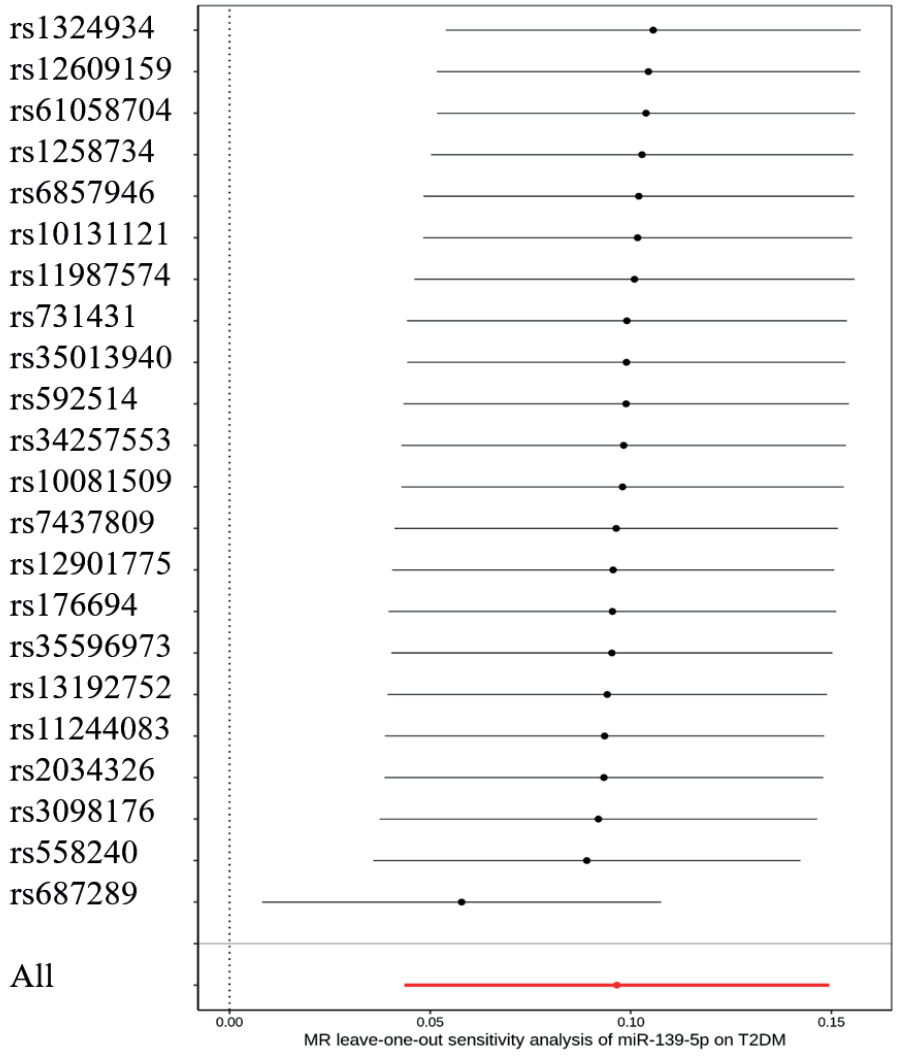
Supplementary Figure 4.1. Forest plot of the association between miRNA expression and prevalent and incident diabetes stratified by cohort. Results are displayed as the odds ratio for prevalent diabetes (A), and hazard ratio for incident diabetes (B) with 95% confidence interval. In red displayed the results of Rotterdam Study I-4. In blue displayed the results of Rotterdam Study II-2. The table presents the odds ratio or hazard ratio with 95%CI and P-value.



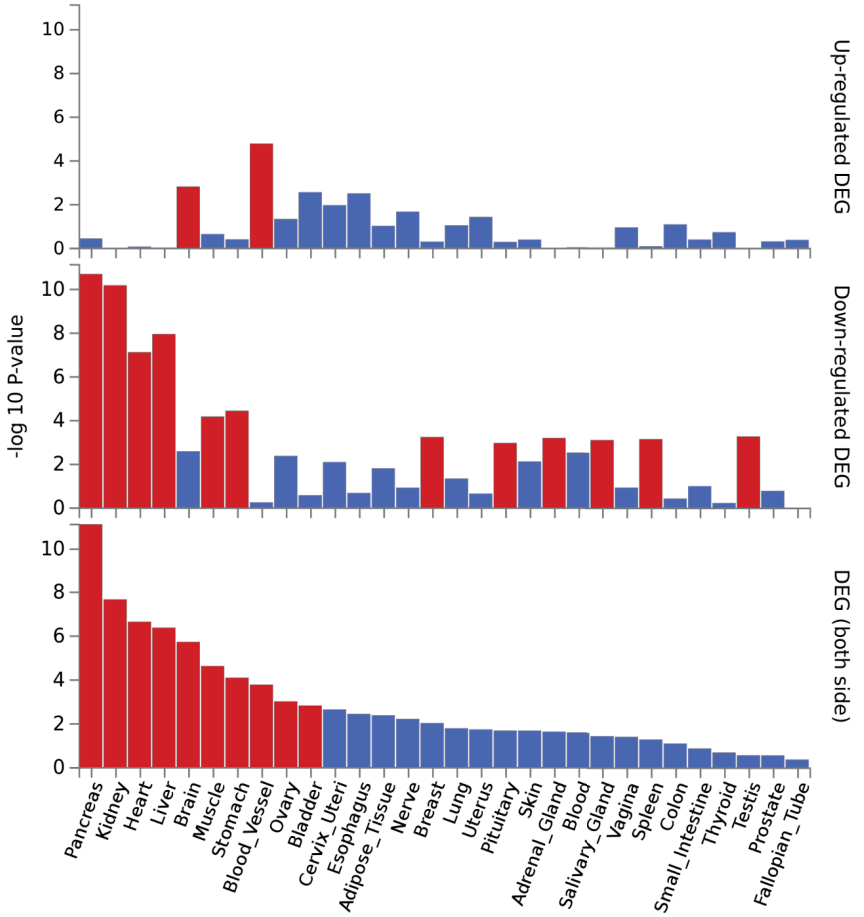
Supplementary Figure 4.2. The effect for individual genetic instruments for miR-139-5p and type 2 diabetes. Results are displayed as Beta estimates with 95% CIs. The x-axis presents the effect for each of the individual genetic instrumental variables and type 2 diabetes. In red are displayed the results of MR Egger and MR IVW.



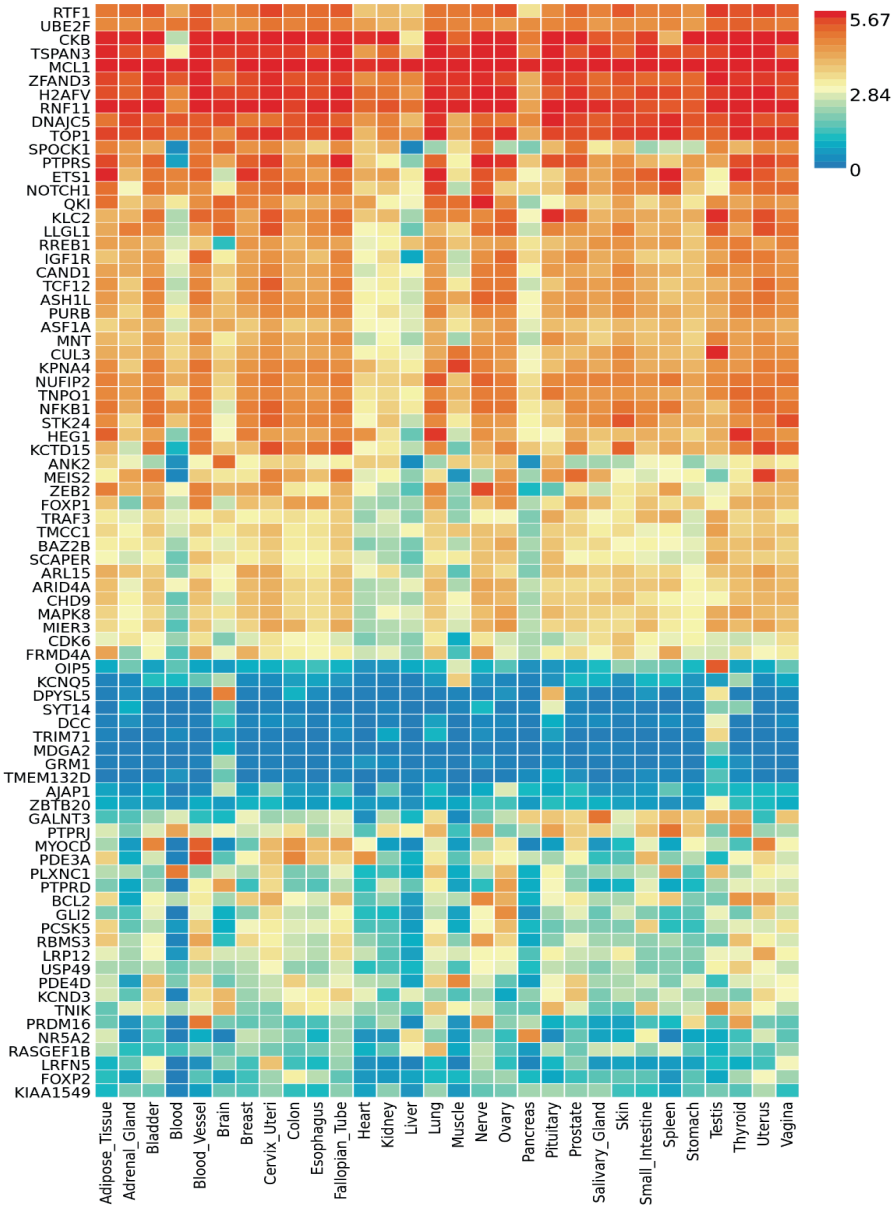
Supplementary Figure 4.3. Scatter plot of different MR analysis. Each dot represents an individual genetic instrument with 95%CI. The x-axis presents the individual effect of each genetic variant on miR-139-5p and the y-axis presents the effect on type 2 diabetes. Each color presents a different MR test; light blue MR Inverse variance weighted; dark blue MR Egger; green MR Weighted median.



Supplementary Figure 4.4. Leave-one-out plot of MR analysis. This sensitivity analysis of MR analysis (using IVW) excludes a particular genetic instrument. Red presents the overall effect.



Supplementary Figure 4.5. Differential expression of miR-139-5p target genes across tissues. The top graph presents upregulated genes per tissue, the middle graph down-regulated genes, and the bottom graph overall differential expressed genes. Red color presents significant differences.



Supplementary Figure 4.6. Heatmap of expressed miR-139-5p target genes across tissues. Red indicates a relatively high expression whereas blue indicates a relatively low expression.

Supplementary Table 4.1. MiRNAs nominally associated with prevalent type 2 diabetes

miRNA	Beta	Std. Error	Z value	P value
miR-193a-5p	0.64	0.15	4.28	1.86E-05
miR-139-5p	0.97	0.24	4.01	6.11E-05
miR-3687	0.49	0.15	3.33	8.73E-04
miR-375	0.23	0.08	2.94	3.24E-03
miR-99b-5p	0.59	0.21	2.87	4.11E-03
miR-6799-5p	0.42	0.15	2.84	4.50E-03
miR-125a-5p	0.62	0.22	2.81	4.98E-03
miR-193a-3p	0.50	0.18	2.79	5.22E-03
miR-6729-3p	-0.84	0.30	-2.77	5.58E-03
miR-3667-3p	-0.23	0.09	-2.72	6.49E-03
miR-6728-3p	-0.44	0.16	-2.71	6.73E-03
miR-378i	0.60	0.22	2.68	7.45E-03
miR-1247-5p	-0.33	0.13	-2.53	1.13E-02
miR-10b-5p	0.52	0.21	2.53	1.13E-02
miR-6792-3p	-0.34	0.13	-2.52	1.16E-02
miR-342-3p	0.52	0.20	2.52	1.17E-02
miR-6825-3p	-0.29	0.11	-2.52	1.19E-02
let-7c-5p	0.63	0.25	2.51	1.20E-02
let-7e-5p	0.46	0.19	2.47	1.36E-02
miR-4721	-0.27	0.11	-2.47	1.36E-02
miR-6870-3p	-0.31	0.13	-2.45	1.41E-02
miR-30c-5p	0.58	0.24	2.46	1.41E-02
miR-920	-0.40	0.16	-2.44	1.45E-02
miR-193b-3p	0.36	0.15	2.44	1.49E-02
miR-150-5p	0.39	0.16	2.42	1.55E-02
miR-6731-3p	-0.43	0.18	-2.36	1.82E-02
let-7a-5p	0.42	0.18	2.32	2.03E-02
miR-5196-3p	-0.30	0.13	-2.32	2.06E-02
miR-1229-3p	-0.20	0.09	-2.31	2.11E-02
miR-27b-3p	0.43	0.20	2.20	2.77E-02
miR-23b-3p	0.40	0.18	2.20	2.78E-02
miR-205-5p	-0.15	0.07	-2.19	2.86E-02
miR-422a	0.39	0.18	2.17	2.99E-02
miR-30b-5p	0.49	0.22	2.17	3.00E-02
miR-6797-3p	-0.12	0.06	-2.16	3.11E-02

miRNA	Beta	Std. Error	Z value	P value
miR-378e	0.39	0.18	2.15	3.13E-02
miR-4723-3p	-0.38	0.18	-2.13	3.35E-02
miR-299-3p	-0.23	0.11	-2.12	3.40E-02
miR-126-3p	0.36	0.17	2.11	3.46E-02
miR-548j-3p	-0.38	0.18	-2.10	3.54E-02
miR-4726-5p	-0.37	0.18	-2.09	3.66E-02
miR-4306	-0.28	0.14	-2.08	3.78E-02
miR-449b-3p	-0.10	0.05	-2.07	3.84E-02
miR-4713-3p	-0.22	0.11	-2.06	3.90E-02
miR-192-5p	0.25	0.12	2.06	3.95E-02
miR-6127	0.29	0.15	2.01	4.48E-02
miR-877-3p	-0.41	0.21	-2.01	4.48E-02
miR-548ay-5p	-0.26	0.13	-2.00	4.49E-02
miR-1233-3p	-0.27	0.14	-1.98	4.74E-02
miR-25-3p	-0.32	0.16	-1.98	4.76E-02
miR-6795-3p	-0.51	0.26	-1.97	4.83E-02

Supplementary Table 4.2. MiRNAs nominally associated with incident type 2 diabetes

miRNA	Beta	HR	Std.Error	Z value	P value
miR-99a-5p	0.73	2.08	0.20	3.60	3.19E-04
miR-4664-3p	-0.57	0.56	0.16	-3.58	3.39E-04
miR-29a-3p	0.67	1.96	0.19	3.57	3.58E-04
miR-122-5p	0.26	1.29	0.07	3.56	3.65E-04
miR-125b-5p	0.67	1.96	0.19	3.53	4.16E-04
miR-99b-5p	0.77	2.15	0.22	3.42	6.27E-04
miR-193a-5p	0.49	1.64	0.14	3.41	6.53E-04
miR-125a-5p	0.77	2.16	0.23	3.31	9.32E-04
miR-342-3p	0.66	1.93	0.20	3.25	1.14E-03
let-7c-5p	0.90	2.45	0.28	3.24	1.20E-03
miR-193b-3p	0.50	1.64	0.15	3.22	1.29E-03
miR-139-5p	0.78	2.19	0.25	3.13	1.72E-03
miR-34b-3p	-0.63	0.54	0.20	-3.08	2.07E-03
miR-27b-3p	0.71	2.04	0.23	3.07	2.15E-03
miR-10a-5p	0.59	1.80	0.20	3.00	2.73E-03
miR-181a-5p	0.72	2.05	0.24	2.99	2.82E-03
miR-4655-5p	-0.88	0.41	0.30	-2.97	2.96E-03
miR-181b-5p	0.59	1.81	0.20	2.97	2.99E-03
miR-194-5p	0.38	1.47	0.13	2.94	3.24E-03
miR-30b-5p	0.70	2.01	0.24	2.85	4.45E-03
miR-155-5p	0.44	1.55	0.16	2.80	5.18E-03
miR-1207-5p	-0.49	0.61	0.17	-2.79	5.20E-03
miR-1303	0.40	1.49	0.14	2.77	5.66E-03
miR-1244	0.37	1.45	0.14	2.74	6.16E-03
miR-30a-5p	0.84	2.32	0.31	2.73	6.41E-03
miR-424-3p	0.33	1.39	0.12	2.71	6.79E-03
miR-378i	0.62	1.86	0.23	2.69	7.08E-03
miR-222-3p	0.61	1.84	0.23	2.65	7.93E-03
let-7g-5p	0.71	2.04	0.27	2.64	8.38E-03
miR-29b-3p	0.60	1.82	0.23	2.60	9.44E-03
miR-1273c	0.40	1.50	0.16	2.58	9.93E-03
miR-28-5p	0.45	1.57	0.18	2.57	1.03E-02
miR-140-5p	0.47	1.61	0.18	2.56	1.04E-02
miR-10b-5p	0.57	1.77	0.22	2.54	1.09E-02
miR-23a-3p	0.42	1.53	0.17	2.53	1.15E-02
let-7b-5p	0.52	1.69	0.21	2.46	1.37E-02

miRNA	Beta	HR	Std.Error	Z value	P value
miR-23b-3p	0.47	1.59	0.19	2.46	1.41E-02
miR-146b-5p	0.74	2.09	0.30	2.44	1.47E-02
miR-6769b-3p	-0.84	0.43	0.34	-2.43	1.52E-02
miR-378f	0.48	1.62	0.20	2.38	1.74E-02
let-7e-5p	0.48	1.61	0.20	2.36	1.84E-02
miR-1908-5p	-0.37	0.69	0.16	-2.36	1.85E-02
miR-6789-5p	-0.35	0.70	0.15	-2.33	1.99E-02
miR-6785-3p	-0.74	0.48	0.32	-2.31	2.11E-02
miR-6729-3p	-0.70	0.50	0.30	-2.29	2.17E-02
miR-877-3p	-0.47	0.62	0.21	-2.29	2.21E-02
miR-4697-5p	-0.19	0.83	0.08	-2.27	2.35E-02
miR-378d	0.39	1.48	0.18	2.25	2.48E-02
miR-100-5p	0.40	1.49	0.18	2.24	2.53E-02
miR-1322	0.35	1.42	0.16	2.23	2.56E-02
miR-27a-3p	0.38	1.46	0.17	2.23	2.56E-02
miR-378e	0.40	1.50	0.18	2.21	2.73E-02
miR-4421	0.24	1.27	0.11	2.20	2.76E-02
miR-4728-5p	-0.22	0.80	0.10	-2.20	2.79E-02
miR-4539	0.28	1.33	0.13	2.20	2.80E-02
miR-24-3p	0.39	1.48	0.18	2.18	2.95E-02
miR-648	-0.54	0.58	0.25	-2.17	2.98E-02
let-7a-5p	0.42	1.53	0.20	2.16	3.08E-02
miR-660-5p	0.35	1.42	0.16	2.16	3.10E-02
miR-422a	0.40	1.50	0.19	2.15	3.15E-02
miR-4417	-0.21	0.81	0.10	-2.15	3.15E-02
miR-6821-3p	-0.26	0.77	0.12	-2.12	3.41E-02
miR-152-3p	0.32	1.38	0.15	2.10	3.56E-02
miR-150-5p	0.34	1.41	0.17	2.08	3.74E-02
miR-6165	-0.23	0.80	0.11	-2.07	3.87E-02
miR-4795-5p	-0.40	0.67	0.20	-2.06	3.94E-02
miR-3674	0.28	1.33	0.14	2.03	4.22E-02
miR-649	0.17	1.18	0.08	2.03	4.24E-02
miR-374b-3p	-0.42	0.65	0.21	-2.02	4.33E-02
miR-6742-5p	-0.48	0.62	0.24	-2.01	4.41E-02
miR-30c-5p	0.53	1.69	0.26	2.01	4.43E-02
miR-1234-3p	-0.63	0.53	0.31	-2.00	4.52E-02

miRNA	Beta	HR	Std.Error	Z value	P value
miR-378a-3p	0.43	1.53	0.21	2.00	4.60E-02
miR-7114-3p	-0.36	0.70	0.18	-1.99	4.65E-02
miR-192-5p	0.25	1.28	0.12	1.98	4.72E-02
miR-6819-5p	-0.19	0.82	0.10	-1.98	4.83E-02
miR-30a-3p	0.33	1.39	0.17	1.96	4.99E-02

5

CHAPTER 5

Circulatory microRNAs as potential biomarkers for fatty liver disease: The Rotterdam Study

Manuscript based on this chapter: Zhang X, Mens MMJ, Abozaid YJ, Bos D, Darwish Murad S, de Knegt RJ, Ikram MA, Pan Q, Ghanbari M. Circulatory microRNAs as potential biomarkers for fatty liver disease: the Rotterdam study. Aliment Pharmacol Ther. 2021 Feb;53(3):432-442. doi: 10.1111/apt.16177.

Abstract

Background: Fatty liver disease (FLD) is the most common cause of liver dysfunction in developed countries. There is a great interest in developing clinically valid and minimally-invasive biomarkers to enhance early diagnosis of FLD. The aim was to investigate the potential of circulatory microRNAs (miRNAs) as biomarkers of FLD at the population level.

Methods: Plasma levels of 2083 miRNAs were measured by RNA-sequencing in 1,999 participants from the prospective population-based Rotterdam Study cohort. The Hounsfield Unit (HU) attenuation of liver was measured using non-enhanced computed tomography (CT-scan). Logistic and linear regression models adjusting for potential confounders were used to examine the association of circulatory miRNAs with liver enzymes (n=1,991) and CT-based FLD (n=954). Moreover, the association of miRNAs with hepatic steatosis and liver fibrosis were assessed longitudinally in individuals who underwent abdominal ultrasound (n=1,211) and transient elastography (n=777) after a median follow-up of >6 years.

Findings: Cross-sectional analysis showed 61 miRNAs significantly associated with serum Gamma-glutamyl transferase and/or Alkaline phosphatase levels (Bonferroni-corrected p-value$<8.46\times10^{-5}$). Moreover, 17 miRNAs were significantly associated with CT-based FLD (p-value$<8.46\times10^{-5}$), 14 of them were associated with liver enzymes. Longitudinal analysis showed that four of the 14 miRNAs (miR-193a-5p, miR-122-5p, miR-378d, and miR-187-3p) were significantly associated with hepatic steatosis (p-value$<3.57\times10^{-3}$) and three (miR-193a-5p, miR-122-5p and miR-193b-3p) were nominally associated with liver fibrosis (p-value<0.05). Nine of the 14 identified miRNAs were involved in pathways underlying liver diseases.

Interpretation: This study indicates that plasma levels of several miRNAs can be used as biomarkers of FLD, laying the groundwork for future clinical applications.

Introduction

Fatty liver disease (FLD), is the most common cause of liver dysfunction in developed countries, that is also increasing in developing countries [1], is defined as an excess accumulation of fat in hepatocytes [2]. Specifically, non-alcoholic fatty liver disease (NAFLD) is characterized by fat accumulation in hepatocytes not due to excess alcohol consumption [3]. The disorder covers a broad spectrum of underlying conditions, ranging from simple fatty liver to inflammation, which can progress to fibrosis, cirrhosis and even liver cancer [4]. FLD is strongly associated with obesity, hypertension, dyslipidemia and insulin resistance, regarded as hepatic manifestation of the metabolic syndrome [5]. Currently, liver biopsy is the gold standard for diagnosing and staging of FLD, but its application is limited by the invasive nature, risk of complications and high cost [6]. Various imaging modalities, such as computed tomography (CT) scan and ultrasound, have also been used for detecting the presence or quantifying the severity of liver fat noninvasively [7]. However, the limited diagnostic accuracy of detecting mild degree hepatic steatosis with CT and ultrasound is an issue that should be taken into consideration [7]. Transient elastography is non-invasive technique that uses both ultrasound and low-frequency elastic waves to qualify liver fibrosis. However, recent research suggests that steatosis may influence its diagnostic performance [8]. Controlled attenuation parameter (CAP) is an ultrasound-based diagnostic method and added to transient elastography enables simultaneous assessment of steatosis and fibrosis [9], but the clinical application of CAP is limited by influences of covariates [10]. Therefore, the development of clinically valid and minimally-invasive methods are required to enhance early diagnosis of FLD.

MicroRNAs (miRNAs) are small non-coding RNA molecules of 20–25 nucleotides in length that regulate gene expression at the post-transcriptional level [11]. Recently, the interest in miRNAs has increased tremendously because they offer new insights into disease mechanisms and have a great potential to be used in the clinic as diagnostic biomarkers and/or even therapeutic targets [12, 13]. In line with this, numerous studies have reported increased levels of circulating miR-122 in liver diseases with different etiologies and suggested this miRNA as a potential biomarker and target of therapy in liver dysfunction [14, 15]. Although extensive research has explored the role of miRNAs in the pathophysiology of liver diseases, little is known about the potential of circulatory miRNAs as FLD biomarker in the population level. Moreover, the available studies have mainly used qPCR-based methods and limited to small number of miRNAs and sample sizes [16, 17].

The aim of this study was to systematically investigate the association of circulating miRNAs in plasma with FLD in a population-based setting. To achieve this aim, we conducted regression models to identify miRNAs that are associated with FLD and liver enzymes at the baseline in the Rotterdam Study cohort. Moreover, we performed

subsequent analyses to check whether the identified miRNAs are linked to the risk of hepatic steatosis or liver fibrosis after follow-up and are involved in the known pathways underlying liver diseases.

Materials and methods

Study population

This study was embedded within the framework of the Rotterdam Study (RS), a prospective cohort study of individuals aged ≥45 years living in the Ommoord district of Rotterdam, the Netherlands. The objectives and design of the Rotterdam Study have been described in detail elsewhere [18]. In 1989, the first cohort of study participants (RS-I) comprised 7983 persons aged 55 years or over. In 2000, the second cohort (RS-II) was extended to include an additional 3011 participants who moved into the study district or had become 55 years of age. A further extension of the Rotterdam Study cohort (RS-III) formed in 2006 and include 3932 participants living in the research area and aged 45 years and older. Follow-up examinations were scheduled periodically, approximately every 3-5 years. All participants in the study provided written inform consent to participate and to obtain information from their treating physicians.

For this study, we used the expression profiles of circulating miRNA in plasma, collected between 2002 and 2005, from a random subset (n=1,000) of the fourth visit of the first cohort (RS-I-4) and a random subset (n=999) of the second visit of the second cohort (RS-II-2). Among them, 1,991 participants had serum Gamma-glutamyl transferase (GGT) and Alkaline phosphatase (ALP) levels available at baseline that were included to investigate the association of miRNAs with liver enzymes. Moreover, 954 participants who underwent CT-scan from June 2003 to February 2006 were included for investigating the associations of miRNAs with FLD in a cross-sectional setting (**Figure 5.1**).

For the longitudinal analysis, 1,999 participants at the baseline were followed-up >6 years, until the fifth visit of the first cohort (RS-I-5) and the third visit of the second cohort (RS-II-3). Among these, 1,211 participants who underwent abdominal ultrasound between January 2009 and June 2014 were included to investigate the association of miRNAs with hepatic steatosis (424 cases). Of these, 1,147 participants were included to investigate the association of miRNAs with NAFLD (321 cases) and alcoholic FLD (76 cases). Moreover, out of the 1,999 individuals, 777 participants who underwent transient elatography were included to investigate the association of miRNAs with liver fibrosis (33 cases). A more detailed flow chart for the selection of study participants is shown in **Figure 5.1**.

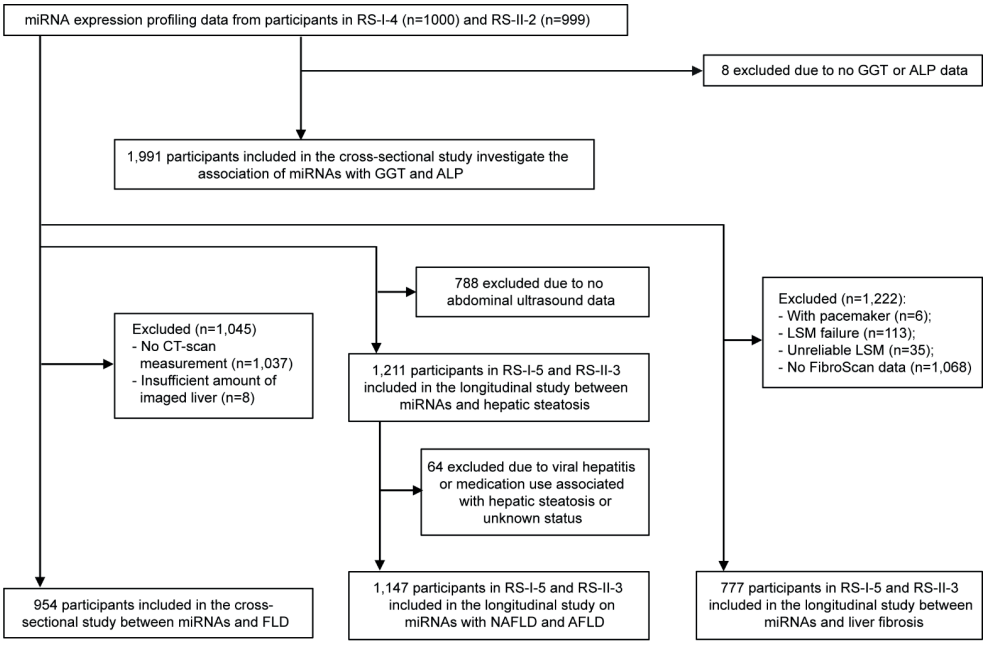


Figure 5.1. An overview of the study to identify circulatory miRNAs associated with FLD. Cross-sectional studies at baseline were performed in participants from RS-I-4 and RS-II-2 with liver enzymes and CT-scan data available. Longitudinal studies were performed in participants from RS-I-5 and RS-II-3 who underwent abdominal ultrasound or transient elastography. **Abbreviations**: miRNAs, microRNAs; GGT, Gamma-glutamyl transferase; ALP, Alkaline phosphatase; FLD, fatty liver disease; RS, Rotterdam Study; RS-I-4, the fourth visit of the first cohort; RS-II-2, the second visit of the second cohort; RS-I-5, the fifth visit of the first cohort; RS-II-3, the third visit of the second cohort; CT, computed tomography; NAFLD, non-alcoholic fatty liver disease; AFLD, alcoholic fatty liver disease; LSM, liver stiffness measurement.

MiRNA expression profiling

Plasma levels of cell-free miRNAs were determined using the HTG EdgeSeq miRNA Whole Transcriptome Assay (WTA), which measures the expression of 2083 human mature miRNAs. The WTA characterizes miRNA expression patterns, and measures the expression of 13 housekeeping genes, that allows flexibility in data normalization and analysis. Plasma samples, for two re-measurements that generally is sufficient to obtain a valid result for all samples, were sent to HTG Molecular Diagnostics (AZ, USA) for sequencing. Each sample was tagged individually with molecular barcodes, tagged samples were pooled and sequenced on an Illumina NextSeq 500 sequencer (Illumina, San Diego, CA, USA). Quantification of miRNA expression was based on counts per million (CPM). The log2 transformation of CPM was used as standardization and adjustment for total reads within each sample. The miRNAs with log2 CPM< 1.0 were considered as not

expressed in the samples. Of the 2083 miRNAs, 591 miRNAs were expressed at good levels in plasma. These 591 well-expressed miRNAs are those with >50% values above Lower Limit of Quantification (LLOQ). The LLOQ level is based on a monotonic decreasing spline curve fit between the means and standard deviations of all miRNAs.

Assessment of liver fat with CT scan

As part of a larger project on the assessment of vascular calcification, ECG-gated, cardiac, non-enhanced CT scanning on a 16-slice (n=251) or 64-slice (n=703) multi-detector CT scanner (Somatom Sensation 16 or 64, Siemens, Forchheim, Germany). Imaging parameters of the scans are described in detail elsewhere [19].

Using this cardiac scans, we evaluated the liver density (attenuation) using a standardized procedure. First, we placed three circular regions of interest (ROIs) in the liver and calculated the mean liver attenuation (LA) within these regions [20]. These ROIs are delineated throughout the imaged liver tissue (including both the left and right liver lobes) are carefully chosen to include only liver tissue, and avoiding the large blood vessels, cysts, or focal lesions. Next, we calculated the mean Hounsfield unit (HU) value from these three measurements as a marker of total amount of liver fat, which is a reliable proxy for the mean LA value of the whole liver [20]. All measurements were done using Philips iSite Enterprise software (Royal Philips Electronics N.V. 2006) and described in detail elsewhere [21].

The CT diagnosis of liver fat is made by measuring mean LA in HU or the difference between the liver and spleen [20]. As the amount of liver fat increases, the measured LA decreases, that means low LA was equal to high risk of fatty liver. However, in the present study fatty liver disease was defined as mean LA< 40 HU [22].

Assessment of hepatic steatosis and liver fibrosis

Hepatic steatosis was assessed by using abdominal ultrasound, which was carried out by a certified and skilled technician (Pavel Taimr) on Hitachi HI VISION 900 [23]. Images were stored digitally and re-assessed by a single hepatologist with more than ten years of experience in ultrasonography. Diagnosis of steatosis was determined dichotomously as presence of a hyperechogenic liver parenchyma according to the protocol by Hamaguchi et al [24].

Moreover, liver fibrosis was assessed using transient elastography (FibroScan®, EchoSens, Paris, France). Applied implementation of this examination has been described in detail previously [25]. Liver stiffness measurement (LSM) was performed by a single certified and experienced operator, who obtained 10 serial measurements using either the M or XL-probe dependent on the thickness of the subcutaneous fat

layer [23]. Moreover, LSM interquartile range/median LSM >0.3 kilopascals (kPa) and LSM ≥7.1 kPa were regarded as poorly reliable [26]. In the present study, LSM ≥9.0 kPa was used as a cutoff suggesting clinically relevant liver fibrosis [8].

Assessment of covariates and liver enzymes

Information on smoking behavior, medication use and blood sampling, was obtained during home interviews [18]. Height and weight were measured, and the body mass index (BMI) [(weight in kg)/(height in m)^2] was calculated. Waist circumferences was measured at the level midway between the lower rib margin and the iliac crest with the participant in a standing position. Smoking status was categorized into never, current or former and were classified (yes/no), for ever-smokers were regarded as current and former smokers combined. Alcohol consumption was assessed in grams of ethanol per day and were classified (yes/no). Excessive alcohol consumption was defined as alcohol intake >30 g/day for men and >20 g/day for women [23]. Hypertension was defined as a systolic blood pressure (BP) ≥140mmHg or a diastolic BP ≥90 mmHg or the use of BP-lowering drugs prescribed for hypertension [27]. Diabetes mellitus was defined according to recent WHO guidelines [28] as fasting blood glucose ≥7.0 mmol/L or non-fasting blood glucose between ≥11.1 mmol/L or the use of antidiabetic medication. From the blood samples, concentrations high-density lipoprotein (HDL) cholesterol were determined using enzymatic procedures [29]. Serum GGT and ALP levels were determined within 2 weeks using a Merck Diagnostica kit on an Elan Autoanalyzer (Merc, Darmstadt, Germany). According to local cutoffs, elevation of GGT was defined as >34 U/L for women and >49 U/L for men, and elevation of ALP was defined as >97 U/L for women and >114 U/L for men [30].

Statistical analysis

Continuous variables are reported as mean ± standard deviation (SD) unless stated otherwise and categorical variables were presented as sample sizes and percentages. To obtain a normal distribution, skewed variables (serum HDL cholesterol, GGT and ALP) were log transformed. In addition, the amount of liver fat (A) had a left skewed and we used exponential transformed values (B) using the formula [$B=A^{3.5}/10000$] [21].

Multivariable linear and logistic regression models were used to investigate the association between miRNA levels and CT-based FLD (with the HU continues and dichotomous data). The multivariable linear regression models were used to check the association of miRNA levels with serum GGT and ALP levels. Beta, standard error (SE), p-value were reported. The Bonferroni-corrected p-value threshold was calculated based on the number of tested miRNAs ($0.05/591=8.46\times10^{-5}$). In basic model (model 1), we adjusted the analysis for age and sex. The multivariable model (model 2), was additionally adjusted for waist circumference, ever smoking, alcohol consumption, hypertension, diabetes mellitus and serum HDL cholesterol. Because

the missing values were likely to be missing at random and for avoidance of loss in efficiency, missing values on covariates (ranging from 0.1% to 1.7%) were imputed using a multiple imputation technique (N=5 imputations). All analyses were done using SPSS statistical software (SPSS, version 25; IBM Corp, Armonk) and R software version 3.5.2 (The R Foundation for Statistical Computing, Vienna, Austria).

Sensitivity analyses were performed by adjusting for more variables. Model 3 was built by adding GGT and ALP to model 2. In model 4, we further adjusted for potential intermediator factors including in the model 2, use of lipid-lowering medication, use of bile and liver medications. In model 5, we adjusted for all potential intermediator factors (including model 2, GGT, ALP, use of lipid-lowering medication, use of bile and liver medications).

Furthermore, multivariable logistic regression models were used to investigate longitudinally the association of the plasma levels of the identified miRNAs with prevalence of hepatic steatosis and liver fibrosis after a median follow-up of 6.4 years [interquartile range (IQR): 5.9-7.0 years]. The Bonferroni correction was used to set the significance threshold.

Two databases, the Human miRNA tissue atlas (https://ccb-web.cs.uni-saarland.de/tissueatlas)[31] and Human miRNA expression profiles (https://guanfiles.dcmb.med.umich.edu/mirmine/ index.html), were used to check whether the identified miRNAs are expressed in the liver. We also searched the literature [32-35] and several web tools (e.g., miR2Diease and GWAS catalog) to see whether the identified miRNAs are associated with liver function and diseases.

Results

At baseline, 954 participants who had miRNA expression data and CT-based liver fat measurement were included to test the association of miRNAs and FLD. The mean age of the study population was 68.8±6.7 years, and 46.6% were male. The mean LA in the population was 61.6 HU (IQR: 55.4-65.6 HU). Among the study participants, 14.8% were diagnosed with cancer, but none of them was diagnosed with liver cancer. At follow-up, 1,211 participants who had underwent abdominal ultrasound were included to test the association of miRNAs and hepatic steatosis. The man age of the study population was 76.3±6.5 years, and 42.4% were male. Lifestyle, clinical and biochemical characteristics of all study participants are presented in **Table 5.1**. Comparison of characteristics between healthy controls and FLD patients based on CT-scan and ultrasound data (in baseline and follow-up study) are shown in Supplementary **Table 5..1**. The participants with FLD have significantly higher BMI, waist circumference and alcohol consumption than healthy controls. In addition, compare to healthy controls, individuals with FLD have significantly lower serum HDL cholesterol.

Table 5.1. Characteristics of the study population

Characteristic	Baseline (n=954) with CT-scan data	Follow-up (n=1,211) with ultrasound data
Age, years	68.8±6.7	76.3±6.5
Male, n (%)	445 (46.6)	513 (42.4)
Body mass index , kg/m²	27.9±4.0	27.5±4.1
Waist circumference, cm	94.4±11.7	93.2±12.0
Hypertension, n (%)	709 (74.3)	1,047 (86.5)
Blood-pressure-lowering medication, n (%)	385 (40.4)	644 (53.2)
Smoking status, n (%)		
Ever	677 (71.0)	791 (65.3)
Current	130 (13.6)	116 (9.6)
Former	547 (57.4)	675 (55.7)
Diabetes mellitus, n (%)	123 (12.9)	156 (12.9)
Use of lipid-lowering medication, n (%)	246 (25.8)	372 (30.7)
Alcohol intake, grams/day	8.6 (1.4-20.0)	8.6 (1.6-8.6)
Mean liver attenuation, HU	61.6 (55.4-65.6)	–
Serum HDL cholesterol, mmol/L	1.4 (1.2-1.7)	1.4 (1.2-1.7)
GGT level, U/L	26.0 (18.0-39.0)	24.0 (17.0-34.2)
ALP level, U/L	77.0 (66.0-91.0)	68.0 (57.0-80.0)
Cancer, n (%)	141 (14.8)	–
Liver cancer, n (%)	0 (0)	–
Fatty liver, n (%)	47 (4.93)	424 (35)

The table shows characteristics of 954 participants with CT-scan data at baseline and 1,211 participants with ultrasound data at follow-up. Values are represented as mean (±standard deviation), sample sizes (percentage), or median (inter-quartile range) for characteristics with skewed distributions. Abbreviations: BP: blood pressure; HU, Hounsfield unit; HDL, high-density lipoprotein; GGT, Gamma glutamyl transferase; ALP, Alkaline phosphatase.

In the linear regression analysis with liver enzymes, 37 miRNAs were significantly associated with serum GGT levels and 29 miRNAs with serum ALP levels, at the Bonferroni-corrected p-value< 8.46×10^{-5} (0.05/591 well-expressed miRNAs) (**Supplementary Table 5.2** and **Supplementary Table 5.3**, respectively). Volcano plot showing differently expressed miRNAs in relation to GGT and ALP levels is depicted in **Figure 5.2A-B**.

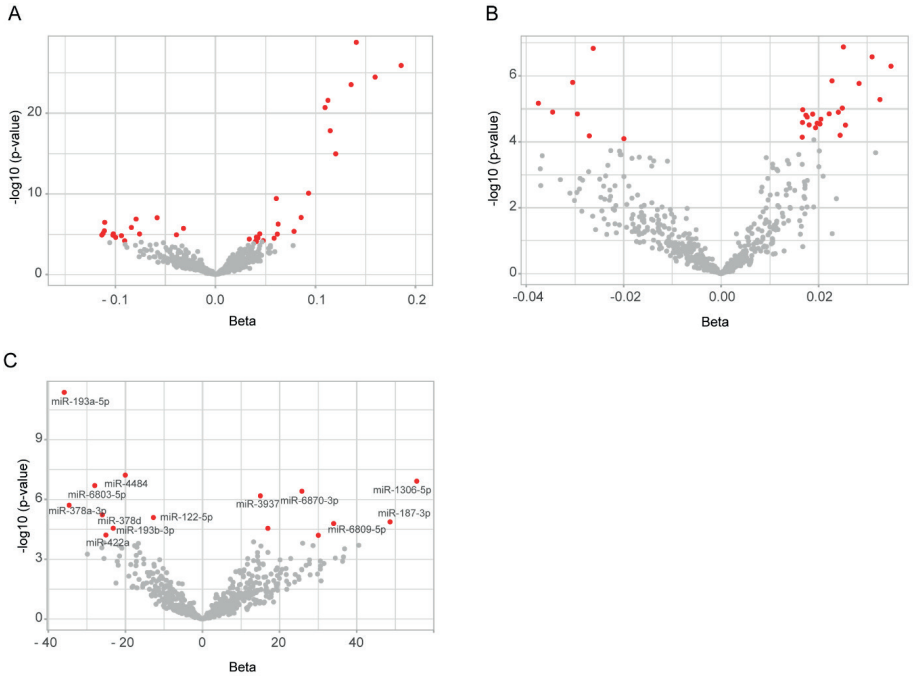


Figure 5.2. Volcano plots showing correlation between plasma levels of miRNAs and GGT (A), ALP (B), and continuous Hounsfield Unit values (C). The red dots indicate miRNAs significantly associated at Bonferroni-corrected p-value< 8.46×10^{-5}. The gray dots indicate miRNAs with no significant association. The name of miRNAs that were significantly associated with liver enzymes and the continuous Hounsfield Unit values are mentioned in the Figure 2C. **Abbreviations**: miRNAs, microRNAs; GGT, Gamma-glutamyl transferase; ALP, Alkaline phosphatase.

Using a linear regression analysis of the continuous HU values, in the multivariable model 2, we found 15 miRNAs to be significantly associated at Bonferroni-corrected p-value< 8.46×10^{-5} (**Supplementary Table 5.4**). Volcano plot showing differently expressed miRNAs in relation to the continuous HU values is depicted in **Figure 5.2C**. In addition, in a logistic regression model testing the association of miRNA levels with the dichotomous HU values assessing FLD (mean LA ≤ or >40), 6 miRNAs were significantly associated (p-value< 8.46×10^{-5}) (**Figure 5.3**). In model 3, further adjusting

for GGT and ALP changed slightly the associations and with less miRNAs significant associated, but 2 out of the 17 and 6 out of 17 miRNAs remained significant using dichotomous and continuous Hounsfield Unit values, respectively (p-value< 8.46×10^{-5}) (**Supplementary Table 5.5** and **Supplementary Table 5.6**). Additional adjustment for use of lipid-lowering medication, use of bile and liver medications (model 4) did not change the observed associations between miRNAs and FLD (**Supplementary Table 5.5** and **Supplementary Table 5.6**). In model 5, we added GGT, ALP, use of serum lipid reducing agents, use of bile and liver medications to model 2, the associations of miRNAs with FLD were similar to model 3, the same miRNAs remained significant (**Supplementary Table 5.5** and **Supplementary Table 5.6**).

Collectively, our cross-sectional studies with the baseline data revealed 61 unique miRNAs associated with liver enzymes (GGT and/or ALP) and 17 unique miRNAs associated with CT-based FLD (continuous or dichotomous HU values). Of these, 14 miRNAs were common in both lists that were selected for further analyses (**Table 5.2**).

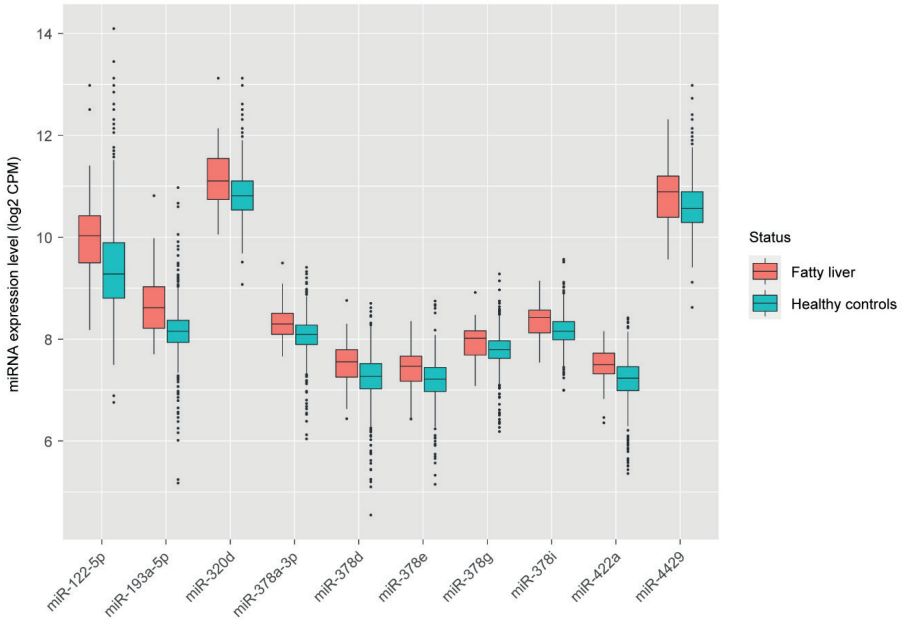


Figure 5.3. Comparison between expression levels of the top 10 miRNAs in patients with fatty liver disease and healthy controls. Of these 10 miRNAs, 6 of them (miR-193a-5p, miR-378a-3p, miR-422a, miR-378d, miR-320d and miR-378e) were significantly associated with fatty liver disease at Bonferroni-corrected p-value< 8.46×10^{-5}. **Abbreviations**: miRNAs, microRNAs; CPM, counts per million.

Table 5.2. Circulatory miRNAs significantly associated with CT-scan based fatty liver disease and serum liver enzymes

miRNA ID	CT-based Dichotomous HU values		FLD Continues HU values		Liver GGT		enzymes ALP	
	Beta	p-value	Beta	p-value	Beta	p-value	Beta	p-value
miR-193a-5p	1.86	**2.02×10^{-10}**	-35.88	**4.26×10^{-12}**	0.14	**6.15×10^{-29}**	0.001	7.99×10^{-01}
miR-4484	0.71	8.33×10^{-4}	-20.04	**6.02×10^{-08}**	0.03	**4.02×10^{-05}**	0.01	3.68×10^{-04}
miR-1306-5p	-0.60	3.15×10^{-01}	55.59	**1.21×10^{-07}**	-0.09	**6.45×10^{-05}**	-0.02	1.12×10^{-01}
miR-378a-3p	2.36	**1.90×10^{-07}**	-34.61	**1.95×10^{-06}**	0.09	**8.45×10^{-08}**	-0.01	3.98×10^{-01}
miR-6803-5p	0.70	2.68×10^{-02}	-27.96	**2.01×10^{-07}**	0.05	**6.31×10^{-05}**	0.01	1.03×10^{-02}
miR-6870-3p	-0.59	1.60×10^{-02}	25.79	**3.87×10^{-07}**	-0.06	**9.04×10^{-08}**	-0.02	**8.00×10^{-05}**
miR-3937	-0.54	1.83×10^{-03}	15.02	**6.59×10^{-07}**	-0.03	**1.90×10^{-06}**	-0.01	3.84×10^{-04}
miR-122-5p	0.54	1.72×10^{-04}	-12.74	**8.05×10^{-06}**	0.14	**2.39×10^{-116}**	0.01	4.73×10^{-04}
miR-422a	2.01	**9.87×10^{-06}**	-25.07	**6.09×10^{-05}**	0.06	**9.71×10^{-06}**	0.002	7.37×10^{-01}
miR-378d	1.75	**1.11×10^{-05}**	-26.00	**5.87×10^{-06}**	0.06	**5.55×10^{-07}**	-0.004	4.92×10^{-01}
miR-187-3p	-1.31	2.52×10^{-02}	48.67	**1.33×10^{-05}**	-0.11	**3.78×10^{-06}**	-0.03	1.80×10^{-02}
miR-6809-5p	-0.73	5.51×10^{-02}	34.03	**1.61×10^{-05}**	-0.08	**1.45×10^{-06}**	-0.01	1.11×10^{-01}
miR-193b-3p	1.01	2.70×10^{-03}	-23.17	**2.77×10^{-05}**	0.11	**2.01×10^{-21}**	-0.001	7.85×10^{-01}
miR-4713-3p	-0.55	1.02×10^{-02}	16.96	**2.80×10^{-05}**	-0.03	8.53×10^{-04}	-0.01	1.13×10^{-02}
miR-320d	1.16	**2.87×10^{-05}**	-14.30	4.54×10^{-03}	0.02	1.72×10^{-01}	0.004	4.10×10^{-01}
miR-34b-3p	-0.76	8.88×10^{-02}	30.06	**6.27×10^{-05}**	-0.06	1.79×10^{-04}	-0.01	3.41×10^{-01}
miR-378e	1.64	**6.88×10^{-05}**	-21.54	6.86×10^{-04}	0.06	**3.13×10^{-05}**	-0.01	4.42×10^{-01}

Model 2: adjusted for age, sex, waist circumference, ever smoking, alcohol consumption, hypertension, diabetes mellitus and serum HDL cholesterol. The table is sorted based on Bonferroni-corrected p-value association of miRNAs with dichotomous or continues Hounsfield Unit values in model 2. The Bonferroni-corrected significance threshold is p-value< 8.46×10^{-5} (0.05/591 miRNAs). The p-values surpassing the significance threshold are bold. Abbreviations: miRNA, microRNA; FLD, fatty liver disease; GGT, Gamma-glutamyl transferase; ALP, Alkaline phosphatase; HU, Hounsfield Unit.

The longitudinal analysis was performed for the 14 miRNAs by using ultrasound and FibroScan data. Using a logistic regression in the multivariable model 2, we found significant association between four of the 14 miRNAs (miR-193a-5p, miR-122-5p, miR-378d, and miR-187-3p) with hepatic steatosis at Bonferroni-corrected p-value< 3.57×10^{-3} (0.05/14 miRNAs) (**Table 5.3**). Moreover, we found significant association of miR-122-5p and miR-187-3p with NAFLD (p-value< 3.57×10^{-3}), and miR-3937 was nominally (p-value< 0.05) associated with alcoholic FLD (**Supplementary 5.7**). Using FibroScan data and in a multivariable logistic regression model, we found miR-193a-5p (p-value= 5.58×10^{-3}, β= 1.11), miR-122-5p (p-value=0.0147, β= 0.45) and miR-193b-3p (p-value= 0.0102, β= 1.19) were to be nominally associated with liver fibrosis (**Table 5.3**).

Table 5.3. Longitudinal study of the 14 identified miRNAs with hepatic steatosis and liver fibrosis

miRNA ID	Hepatic Steatosis			Liver Fibrosis		
	Beta	SE	p-value	Beta	SE	p-value
miR-193a-5p	0.54	0.16	**2.02×10^{-10}**	1.11	0.40	5.58×10^{-03}
miR-4484	0.21	0.10	4.65×10^{-02}	0.41	0.28	1.37×10^{-01}
miR-1306-5p	-0.33	0.28	2.36×10^{-01}	-1.16	-0.73	1.14×10^{-01}
miR-378a-3p	0.37	0.21	8.19×10^{-02}	1.23	0.64	5.67×10^{-02}
miR-6803-5p	0.13	0.15	3.79×10^{-01}	0.70	0.41	8.42×10^{-02}
miR-6870-3p	-0.22	0.14	1.18×10^{-01}	0.07	0.45	8.80×10^{-01}
miR-3937	-0.13	0.09	1.44×10^{-01}	-0.18	0.26	4.87×10^{-01}
miR-122-5p	0.33	0.08	**6.06×10^{-05}**	0.45	0.18	1.47×10^{-02}
miR-422a	0.37	0.18	4.12×10^{-02}	0.96	0.54	7.61×10^{-02}
miR-378d	0.54	0.17	**1.65×10^{-03}**	0.73	0.53	1.69×10^{-01}
miR-187-3p	-1.01	0.34	**2.76×10^{-03}**	-0.42	0.98	6.68×10^{-01}
miR-6809-5p	-0.55	0.24	2.49×10^{-02}	-0.66	-0.76	3.83×10^{-01}
miR-193b-3p	0.22	0.15	1.42×10^{-01}	1.19	0.46	1.02×10^{-02}
miR-378e	0.50	0.18	7.25×10^{-03}	0.94	0.60	1.14×10^{-01}

Model 2: adjusted for age, sex, waist circumference, ever smoking, alcohol consumption, hypertension, diabetes mellitus and serum HDL cholesterol. The table is sorted based on the association of 14 miRNAs with the continues Hounsfield Unit values in the cross-sectional study. The p-values surpassing the Bonferroni-corrected threshold of p-value< 3.57×10^{-3} (0.05/14 miRNAs) are bold and nominal associations with p-value <0.05 are underlined. Abbreviations: miRNA, microRNA; SE, standard error.

Additionally, we searched the Human miRNA tissue atlas and the miRmine database to see whether the 14 miRNAs associated with FLD in plasma are also expressed in the liver that are shown in **Supplementary Table 5.8**. Among them, miR-122-5p is a specifically expressed miRNA with the tissue specificity index (TSI) of 0.97 and highly expressed in the liver. Then, we sought to find whether the 14 identified miRNAs are reported in previous studies to be associated with liver function or/and diseases. A summary of evidence for associations between 9 of these miRNAs and liver diseases are shown in **Supplementary Table 5.9**. Finally, we extracted SNPs annotated to the 16 identified miRNAs and checked their associations with FLD and liver enzymes using summary statistics data from pervious GWAS [34, 35]. There were 63 SNPs related to miR-193a-5p, miR-378d and miR-193b-3p, none of them showed significant association after correcting the p-value for multiple testing.

Discussion

In this study, we investigated the association between circulating miRNAs and liver enzymes in a population-based setting and found 61 unique miRNAs to be associated with serum GGT or ALP levels. Moreover, we found plasma levels of 17 miRNAs to be associated with CT-based FLD, 14 of these were also associated with the liver enzymes. Higher plasma levels of three and lower plasma level of one of these 14 miRNAs were significantly associated with hepatic steatosis after >6 years follow-up. These findings indicate that plasma levels of miRNAs can be considered as potential biomarkers of FLD and hepatic steatosis in the general population.

Several studies have demonstrated the potential of miRNAs to be used as biomarkers for liver diseases [36-38]. However, previous studies have conducted for subset of miRNAs, using qPCR-based methods, or on the modest sample sizes. While our study is embedded within the Rotterdam Study with much larger sample size, based on RNA-sequencing method, conducted genome-wide profiling of almost all important cell-free miRNAs, and adjusted for a broad range of potential confounders, such as waist circumference, smoking status, alcohol consumption, hypertension and diabetes mellitus, which have been overlooked in most of previous studies [36, 39]. Such a large-scale population-based study with long term follow-up data provided a more statistical power to detect multiple significant associations. Compared to microarray or qPCR-based profiling techniques, the cell-free RNA-seq analysis can provide higher sensitivity to measure miRNAs expression levels over a wide dynamic range and with ability to identify novel miRNAs [40]. Additionally, due to the high stability of cell-free miRNAs in body fluids and accessibility of plasma compared with the target tissue, the identified miRNAs can be considered as potential easy-to-use biomarkers in clinical routine [41].

Previous studies on human or mouse model have demonstrated particularly miR-122-5p as potential biomarkers and therapeutic target for liver diseases [39, 42, 43]. In line with previous studies, we found that the higher plasma miR-122-5p level is significantly associated with FLD and liver enzymes also in a population-based setting. In addition to the well-established liver-associated miR-122, we found evidence in previous studies for 8 of the other identified miRNAs in our study to be associated with liver diseases, indicating the importance of these miRNAs in pathways underlying liver function and diseases. In particular, the expression of miR-193a-5p, which is one of our top miRNAs associated with FLD and hepatic steatosis, is reported to be upregulated in HCC tissues [44], whereas miR-193a-5p can distinguish HCC from other non-HCC individuals [43] and inhibited HCC development through targeting *SPOCK1* [45]. Similarly, miR-422a was related to NAFLD [46], miR-378d [47] miR-187-3p [48], miR-6809-5p [49] and miR-4484 [50] were associated with HCC. Moreover, miR-193b-3p which were significantly associated with FLD and nominally associated with liver fibrosis in our study, have been verified previously to be involved in the pathogenesis of liver fibrosis in vitro [51]. Finally, the members of miR-378 family, in particular miR-378a-3p that has been also proposed to have a therapeutic potential for liver fibrosis [52].

Our results showed a minimal of the use of serum lipid reducing agents, use of bile and liver medications on the observed associations between miRNAs and FLD. We observed slightly change in the associations between miRNAs and FLD by adding GGT and ALP to the model. This difference may indicate that liver enzymes have more stronger links to FLD compared to medication use. Also, we did not find significant association between SNPs related to the identified miRNAs and FLD in the summary statistics from previous GWAS, but we need to take into consideration the sample sizes of available GWAS of liver diseases. To date, the GWAS on liver diseases are mainly from two studies [34, 35], Nakamura et al. conducted a study to identify susceptibility loci for primary biliary cirrhosis, a GWAS in 963 Japanese individuals and in a subsequent replication study including 1,402 other Japanese individuals. The sample sizes of this study is limited and the analysis conducted in Asia population, while our study was conducted in European population, and miRNAs expression might exhibits population differences [53]. In addition, Namjou et al. conducted a GWAS using both adult and pediatric participants (1,106 NAFLD cases and 8,571 controls) from electronic medical records to identify genetic contributions to NAFLD. As the cohorts in that GWAS study represent many geographic area in USA, other ancestry groups are under-represented in the electronic medical records. Thus, it is possible that future trans-ethnic GWAS with larger samples sizes find association between some of the identified miRNAs and FLD.

Our study has some limitations that should be considered. First, in the cross-sectional observational study the ability to assess causality or temporality is limited. We therefore assessed additionally the associations of the identified miRNAs as biomarkers for diagnosis hepatic steatosis after follow-up. Future studies are still needed to confirm our findings in longitudinal settings considering the incidence date, longer follow-up time and in different age groups. Second, we defined the mean LA< 40 HU as FLD and found 47 cases out of 954 individuals (5%), which is lower than the excepted prevalence of FLD in the general population. A liver-to-spleen ratio< 1.0 is comparable to using a mean LA cut off ≤ 51 HU for diagnosis of mild liver fat [54]. However, previous studies have demonstrated different cut-offs, mainly a cut-off value of 40 HU on non-enhanced CT as the most clinically indicator for moderate-to severe steatosis [20, 22]. In our study, the cut-off value is 40 HU for FLD, it is relatively strict than using the mean LA≤ 51 HU, which increases the certainty of identifying participants with true FLD and also results in a lower prevalence. Therefore, we performed cross-sectional analysis at baseline with the continuous HU values and liver enzymes as well. The majority (14 out of 17) of the identified miRNAs with CT-scan data showed significant association with liver enzyme, indicating the robustness of our results. Yet, compare to the liver biopsy as the gold standard, but invasive method, to measure FLD, CT-based liver fat has limited diagnostic accuracy of detecting mild degree hepatic steatosis. Therefore, there might be some known or unknown causes for low density of the liver on CT scan. Also, there might be some inconsistency between CT-scans and ultrasound data for diagnosing FLD and hepatic steatosis. In an optimal setting, one should use the repeated measurement of liver fat by a similar diagnostic method for longitudinal analysis.

In conclusion, we found that plasma levels of several miRNAs were significantly associated with FLD and hepatic steatosis that can be considered as plasma disease biomarkers in this population-based study. Future research need to be conducted even with more sample sizes and longer follow-up times in order to confirm the potential of the identified miRNAs as biomarkers for early diagnosis and progression of FLD and also to uncover underlying molecular mechanisms by which these miRNAs may control liver fat.

References

1. Younossi Z, Tacke F, Arrese M, et al. Global Perspectives on Nonalcoholic Fatty Liver Disease and Nonalcoholic Steatohepatitis. *Hepatology* 2019;**69**(6):2672-2682.
2. Fatiha Nassir, R. Scott Rector, Ghassan M. Hammoud, Ibdah JA. Pathogenesis and Prevention of Hepatic Steatosis. *Gastroenterology & Hepatology* 2015;**11**(3):167-175.
3. VanWagner LB, Ning H, Allen NB, et al. Alcohol Use and Cardiovascular Disease Risk in Patients With Nonalcoholic Fatty Liver Disease. *Gastroenterology* 2017;**153**(5):1260-1272 e3.
4. DiStefano JK, Gerhard GS. Circulating microRNAs in nonalcoholic fatty liver disease. *Expert Rev Gastroenterol Hepatol* 2016;**10**(2):161-3.
5. Eslam M, Sanyal AJ, George J, International Consensus P. MAFLD: A Consensus-Driven Proposed Nomenclature for Metabolic Associated Fatty Liver Disease. *Gastroenterology* 2020;**158**(7):1999-2014 e1.
6. Musaddaq G, Shahzad N, Ashraf MA, Arshad MI. Circulating liver-specific microRNAs as noninvasive diagnostic biomarkers of hepatic diseases in human. *Biomarkers* 2019;**24**(2):103-109.
7. Lee DH. Imaging evaluation of non-alcoholic fatty liver disease: focused on quantification. *Clin Mol Hepatol* 2017;**23**(4):290-301.
8. Eddowes PJ, Sasso M, Allison M, et al. Accuracy of FibroScan Controlled Attenuation Parameter and Liver Stiffness Measurement in Assessing Steatosis and Fibrosis in Patients With Nonalcoholic Fatty Liver Disease. *Gastroenterology* 2019;**156**(6):1717-1730.
9. Karlas T, Petroff D, Sasso M, et al. Impact of controlled attenuation parameter on detecting fibrosis using liver stiffness measurement. *Aliment Pharmacol Ther* 2018;**47**(7):989-1000.
10. Oeda S, Takahashi H, Imajo K, et al. Accuracy of liver stiffness measurement and controlled attenuation parameter using FibroScan((R)) M/XL probes to diagnose liver fibrosis and steatosis in patients with nonalcoholic fatty liver disease: a multicenter prospective study. *J Gastroenterol* 2020;**55**(4):428-440.
11. Otsuka M, Kishikawa T, Yoshikawa T, et al. MicroRNAs and liver disease. *Journal Of Human Genetics* 2016;**62**:75.
12. Schueller F, Roy S, Vucur M, Trautwein C, Luedde T, Roderburg C. The Role of miRNAs in the Pathophysiology of Liver Diseases and Toxicity. *Int J Mol Sci* 2018;**19**(1).
13. Pirola CJ, Fernandez Gianotti T, Castano GO, et al. Circulating microRNA signature in non-alcoholic fatty liver disease: from serum non-coding RNAs to liver histology and disease pathogenesis. *Gut* 2015;**64**(5):800-12.
14. Roderburg C, Luedde T. Circulating microRNAs as markers of liver inflammation, fibrosis and cancer. *J Hepatol* 2014;**61**(6):1434-7.
15. Liu CH, Ampuero J, Gil-Gomez A, et al. miRNAs in patients with non-alcoholic fatty liver disease: A systematic review and meta-analysis. *J Hepatol* 2018;**69**(6):1335-1348.
16. Mehta R, Otgonsuren M, Younoszai Z, Allawi H, Raybuck B, Younossi Z. Circulating miRNA in patients with non-alcoholic fatty liver disease and coronary artery disease. *BMJ Open Gastroenterol* 2016;**3**(1):e000096.
17. Zarrinpar A, Gupta S, Maurya MR, Subramaniam S, Loomba R. Serum microRNAs explain discordance of non-alcoholic fatty liver disease in monozygotic and dizygotic twins: a prospective study. *Gut* 2016;**65**(9):1546-54.

18. Ikram MA, Brusselle G, Ghanbari M, et al. Objectives, design and main findings until 2020 from the Rotterdam Study. *Eur J Epidemiol* 2020.

19. Odink AE, van der Lugt A, Hofman A, et al. Association between calcification in the coronary arteries, aortic arch and carotid arteries: the Rotterdam study. *Atherosclerosis* 2007;**193**(2):408-13.

20. Kodama Y, Ng CS, Wu TT, et al. Comparison of CT methods for determining the fat content of the liver. *AJR Am J Roentgenol* 2007;**188**(5):1307-12.

21. Wolff L, Bos D, Murad SD, et al. Liver fat is related to cardiovascular risk factors and subclinical vascular disease: the Rotterdam Study. *Eur Heart J Cardiovasc Imaging* 2016;**17**(12):1361-1367.

22. Seong Ho P, Pyo Nyun K, Kyoung Won K, et al. Macrovesicular Hepatic Steatosis in Living Liver Donors: Use of CT for Quantitative and Qualitative Assessment. *Radiology* 2006;**239**(1):105-112.

23. Alferink LJM, Trajanoska K, Erler NS, et al. Nonalcoholic Fatty Liver Disease in The Rotterdam Study: About Muscle Mass, Sarcopenia, Fat Mass, and Fat Distribution. *J Bone Miner Res* 2019;**34**(7):1254-1263.

24. Hamaguchi M, Kojima T, Itoh Y, et al. The severity of ultrasonographic findings in nonalcoholic fatty liver disease reflects the metabolic syndrome and visceral fat accumulation. *Am J Gastroenterol* 2007;**102**(12):2708-15.

25. Koehler EM, Plompen EP, Schouten JN, et al. Presence of diabetes mellitus and steatosis is associated with liver stiffness in a general population: The Rotterdam study. *Hepatology* 2016;**63**(1):138-47.

26. Boursier J, Zarski JP, de Ledinghen V, et al. Determination of reliability criteria for liver stiffness evaluation by transient elastography. *Hepatology* 2013;**57**(3):1182-91.

27. Williams B, Mancia G, Spiering W, et al. 2018 ESC/ESH Guidelines for the management of arterial hypertension. *Eur Heart J* 2018;**39**(33):3021-3104.

28. Origanization.WH. guidelines - World Health Organization. 2016.

29. Braun KVE, Dhana K, de Vries PS, et al. Epigenome-wide association study (EWAS) on lipids: the Rotterdam Study. *Clin Epigenetics* 2017;**9**:15.

30. Koehler EM, Sanna D, Hansen BE, et al. Serum liver enzymes are associated with all-cause mortality in an elderly population. *Liver Int* 2014;**34**(2):296-304.

31. Ludwig N, Leidinger P, Becker K, et al. Distribution of miRNA expression across human tissues. *Nucleic Acids Res* 2016;**44**(8):3865-77.

32. Huan T, Rong J, Liu C, et al. Genome-wide identification of microRNA expression quantitative trait loci. *Nat Commun* 2015;**6**:6601.

33. Majid Nikpay, Kaitlyn Beehler, Armand Valsesia, et al. Genome-wide identification of circulating-miRNA expression quantitative trait loci reveals the role of several miRNAs in the regulation ofcardiometabolic phenotypes. *European Society of Cardiology* 2019;**115**(11):1629-1645.

34. Nakamura M, Nishida N, Kawashima M, et al. Genome-wide association study identifies TNFSF15 and POU2AF1 as susceptibility loci for primary biliary cirrhosis in the Japanese population. *Am J Hum Genet* 2012;**91**(4):721-8.

35. Namjou B, Lingren T, Huang Y, et al. GWAS and enrichment analyses of non-alcoholic fatty liver disease identify new trait-associated genes and pathways across eMERGE Network. *BMC Med* 2019;**17**(1):135.

36. Salvoza NC, Klinzing DC, Gopez-Cervantes J, Baclig MO. Association of Circulating Serum miR-34a and miR-122 with Dyslipidemia among Patients with Non-Alcoholic Fatty Liver Disease. *PLoS One* 2016;**11**(4):e0153497.

37. Auguet T, Aragones G, Berlanga A, et al. miR33a/miR33b* and miR122 as Possible Contributors to Hepatic Lipid Metabolism in Obese Women with Nonalcoholic Fatty Liver Disease. *Int J Mol Sci* 2016;**17**(10).
38. Raitoharju E, Seppala I, Lyytikainen LP, et al. Blood hsa-miR-122-5p and hsa-miR-885-5p levels associate with fatty liver and related lipoprotein metabolism-The Young Finns Study. *Sci Rep* 2016;**6**:38262.
39. Brandt S, Roos J, Inzaghi E, et al. Circulating levels of miR-122 and nonalcoholic fatty liver disease in pre-pubertal obese children. *Pediatr Obes* 2018;**13**(3):175-182.
40. Eminaga S, Christodoulou DC, Vigneault F, Church GM, Seidman JG. Quantification of microRNA expression with next-generation sequencing. *Curr Protoc Mol Biol* 2013;**Chapter 4**:Unit 4 17.
41. Szelenberger R, Kacprzak M, Saluk-Bijak J, Zielinska M, Bijak M. Plasma MicroRNA as a novel diagnostic. *Clin Chim Acta* 2019;**499**:98-107.
42. Cheng JL, Zhao H, Yang SG, Chen EM, Chen WQ, Li LJ. Plasma miRNA-122-5p and miRNA-151a-3p identified as potential biomarkers for liver injury among CHB patients with PNALT. *Hepatol Int* 2018;**12**(3):277-287.
43. Jin Y, Wong YS, Goh BKP, et al. Circulating microRNAs as Potential Diagnostic and Prognostic Biomarkers in Hepatocellular Carcinoma. *Sci Rep* 2019;**9**(1):10464.
44. Wang JT, Wang ZH. Role of miR-193a-5p in the proliferation and apoptosis of hepatocellular carcinoma. *European Review for Medical and Pharmacological Sciences* 2018;**22**:7233-7239.
45. Li P, Xiao Z, Luo J, Zhang Y, Lin L. MiR-139-5p, miR-940 and miR-193a-5p inhibit the growth of hepatocellular carcinoma by targeting SPOCK1. *J Cell Mol Med* 2019;**23**(4):2475-2488.
46. Latorre J, Moreno-Navarrete JM, Mercader JM, et al. Decreased lipid metabolism but increased FA biosynthesis are coupled with changes in liver microRNAs in obese subjects with NAFLD. *Int J Obes (Lond)* 2017;**41**(4):620-630.
47. Gao L, Xiong D-D, He R-Q, et al. Identifying TF-miRNA-mRNA regulatory modules in nitidine chloride treated HCC xenograft of nude mice. *American journal of translational research* 2019;**11**:7503-7522.
48. Dou C, Liu Z, Xu M, et al. miR-187-3p inhibits the metastasis and epithelial-mesenchymal transition of hepatocellular carcinoma by targeting S100A4. *Cancer Lett* 2016;**381**(2):380-90.
49. Yang PW, Lu ZY, Pan Q, et al. MicroRNA-6809-5p mediates luteolin-induced anticancer effects against hepatoma by targeting flotillin 1. *Phytomedicine* 2019;**57**:18-29.
50. Pascut D, Cavalletto L, Pratama MY, et al. Serum miRNA Are Promising Biomarkers for the Detection of Early Hepatocellular Carcinoma after Treatment with Direct-Acting Antivirals. *Cancers (Basel)* 2019;**11**(11).
51. Ju B, Nie Y, Yang X, et al. miR-193a/b-3p relieves hepatic fibrosis and restrains proliferation and activation of hepatic stellate cells. *J Cell Mol Med* 2019;**23**(6):3824-3832.
52. Hyun J, Wang S, Kim J, et al. MicroRNA-378 limits activation of hepatic stellate cells and liver fibrosis by suppressing Gli3 expression. *Nat Commun* 2016;**7**:10993.
53. Huang RS, Gamazon ER, Ziliak D, et al. Population differences in microRNA expression and biological implications. *RNA Biol* 2011;**8**(4):692-701.
54. Zeb I, Li D, Nasir K, Katz R, Larijani VN, Budoff MJ. Computed tomography scans in the evaluation of fatty liver disease in a population based study: the multi-ethnic study of atherosclerosis. *Acad Radiol* 2012;**19**(7):811-8.

Supplementary Table 5.1. Comparison of the characteristics between healthy controls and fatty liver disease patients in this study

Characteristic	Baseline CT-scan data			Follow-up ultrasound data		
	Controls (n=907)	Cases (n=47)	p-value	Controls (n=787)	Cases (n=424)	p-value
Age, years	68.8±6.7	67.7±5.5	0.18	76.6±6.7	75.8±5.9	**0.02**
Male, n (%)	419 (46.2)	26 (55.3)	0.23	320 (40.7)	193 (45.5)	0.10
Body mass index, kg/m2	27.4±4.0	30.6±3.4	**1.10E-06**	26.2±3.5	29.9±4.2	**6.53E-47**
Waist circumference, cm	93.9±11.5	104.6±10.6	**1.09E-06**	89.4±10.7	100.4±11.0	**2.02E-54**
Hypertension, n (%)	671 (74.0)	38 (80.9)	0.26	658 (83.6)	389 (91.7)	**1.64E-05**
BP lowering medication, n (%)	360 (40.0)	25 (53.2)	0.41	378 (48.0)	266 (62.7)	**7.45E-07**
Smoking status, n (%)						
Ever	640 (70.6)	37 (78.7)	0.2	490 (62.5)	301 (71.0)	**0.007**
Current	123 (13.6)	7 (14.9)	0.77	83 (10.5)	33 (7.8)	0.10
Former	517 (57.0)	30 (63.8)	0.28	407 (51.7)	268 (63.2)	**0.002**
Diabetes mellitus, n (%)	113 (12.5)	10 (21.3)	0.16	63 (8.0)	92 (21.7)	**9.37E-10**
Alcohol intake, grams/day	8.6 (1.1-20.0)	17.4 (2.1-28.9)	**0.02**	8.57 (1.6-8.6)	8.6 (1.6-15.0)	**3.43E-05**
Mean liver attenuation, HU	62.0 (56.6-65.8)	31.0 (27.0-36.4)	**3.27E-29**	–	–	–
Serum HDL cholesterol, mmol/L	1.4 (1.2-1.7)	1.2 (1.0-1.5)	**5.10E-03**	1.5 (1.3-1.8)	1.3 (1.1-1.6)	**1.90E-16**
Use of lipid-lowering medication, n (%)	235 (25.9)	11 (23.4)	0.7	213 (27.1)	159 (37.5)	**2.51E-04**
GGT level, U/L	25.0 (18.0-38.0)	43.0 (32.0-67.0)	0.27	21.0 (16.0-30.0)	30.0 (20.0-42.0)	**3.84E-06**
ALP level, U/L	77.0 (66.0-91.0)	78.0 (66.0-95.0)	0.9	68.0 (58.0-78.8)	69.5 (56.0-82.0)	0.38
Cancer, n (%)	135 (14.9)	6 (12.8)	0.99	–	–	–
Liver cancer, n (%)	0 (0)	0 (0)	1	–	–	–

Values are represented as mean (±standard deviation), sample sizes (percentage), or median (interquartile range) for characteristics with skewed distributions. The characteristics between healthy controls and fatty liver patients were compared using Student's t-test. A p-value of < 0.05 was considered significant and bold. Abbreviations: BP: blood pressure; HU, Hounsfield unit; HDL: high-density lipoprotein; GGT, Gamma glutamyl transferase; ALP, Alkaline phosphatase.

Supplementary Table 5.2. 37 miRNAs significantly associated with GGT in model 2

miRNA ID	Beta	SE	p-value
miR-122-5p	0.14	0.01	2.39E-116
miR-194-5p	0.19	0.01	8.77E-79
miR-192-5p	0.16	0.01	6.25E-65
miR-193a-5p	0.14	0.01	6.15E-29
miR-148a-3p	0.11	0.01	2.56E-22
miR-193b-3p	0.11	0.01	2.01E-21
miR-100-5p	0.11	0.01	1.50E-18
miR-99a-5p	0.12	0.01	1.09E-15
miR-125b-5p	0.09	0.01	8.31E-11
miR-34a-5p	0.06	0.01	3.78E-10
miR-378a-3p	0.09	0.02	8.45E-08
miR-6870-3p	-0.06	0.01	9.04E-08
miR-4433b-5p	-0.08	0.01	1.33E-07
miR-4655-3p	-0.11	0.02	3.39E-07
miR-378d	0.06	0.01	5.55E-07
miR-6809-5p	-0.08	0.02	1.45E-06
miR-3937	-0.03	0.01	1.90E-06
miR-187-3p	-0.11	0.02	3.78E-06
miR-378g	0.08	0.02	4.42E-06
miR-6775-3p	-0.11	0.03	7.80E-06
miR-3140-5p	-0.10	0.02	9.03E-06
miR-6780b-5p	0.04	0.01	9.09E-06
miR-4257	-0.08	0.02	9.28E-06
miR-422a	0.06	0.01	9.71E-06
miR-6165	-0.04	0.01	1.18E-05
miR-1225-5p	-0.11	0.03	1.26E-05
miR-2115-5p	-0.10	0.02	1.29E-05
miR-212-3p	-0.09	0.02	1.52E-05
miR-6870-5p	0.04	0.01	2.35E-05
miR-6794-3p	-0.10	0.02	2.45E-05
miR-7851-3p	0.04	0.01	2.90E-05
miR-378e	0.06	0.01	3.13E-05
miR-4484	0.03	0.01	4.02E-05
miR-6803-5p	0.05	0.01	6.31E-05
miR-1306-5p	-0.09	0.02	6.45E-05
miR-548d-5p	0.04	0.01	7.64E-05
miR-1299	0.04	0.01	7.83E-05

The table shows 37 miRNAs that are significantly associated with GGT in Model 2. The significant threshold is the Bonferroni-corrected p-value of 8.46E-05 (0.05/591 miRNAs). Model 2: adjusted for age, sex, waist circumference, ever smoking, alcohol consumption, hypertension, diabetes mellitus and serum HDL cholesterol). Abbreviations: miRNA, microRNA; GGT, Gamma-glutamyl transferase; SE, standard error; HDL, high-density lipoprotein.

Supplementary Table 5.3. 29 miRNAs significantly associated with ALP in model 2

miRNA ID	Beta	SE	p-value
miR-3674	0.03	0.005	1.32E-07
miR-126-5p	-0.03	0.005	1.46E-07
miR-4463	0.03	0.006	2.64E-07
miR-4459	0.03	0.007	5.07E-07
miR-566	0.02	0.005	1.41E-06
miR-126-3p	-0.03	0.006	1.57E-06
miR-6799-5p	0.03	0.006	1.69E-06
miR-1304-3p	0.03	0.007	5.21E-06
miR-139-5p	-0.04	0.008	6.75E-06
miR-3135a	0.02	0.006	9.46E-06
miR-1273a	0.02	0.004	1.06E-05
miR-4257	-0.03	0.008	1.25E-05
miR-1273c	0.02	0.005	1.26E-05
miR-1254	0.02	0.005	1.41E-05
miR-195-5p	-0.03	0.007	1.42E-05
miR-1255b-2-3p	0.02	0.004	1.44E-05
miR-1273e	0.02	0.004	1.55E-05
miR-5585-3p	0.02	0.004	1.77E-05
miR-1299	0.02	0.005	2.05E-05
miR-1285-5p	0.02	0.004	2.58E-05
miR-7851-3p	0.02	0.005	2.72E-05
miR-1303	0.02	0.005	2.86E-05
miR-1273g-5p	0.02	0.004	3.06E-05
miR-1914-5p	0.03	0.006	3.10E-05
miR-548d-5p	0.02	0.005	3.72E-05
miR-4478	0.02	0.006	6.30E-05
miR-10b-5p	-0.03	0.007	6.57E-05
miR-4512	0.02	0.004	7.21E-05
miR-6870-3p	-0.02	0.005	8.00E-05

The table shows 29 miRNAs that are significantly associated with ALP in model 2. The significant threshold is the Bonferroni-corrected p-value of 8.46E-05 (0.05/591 miRNAs). Model 2: adjusted for age, sex, waist circumference, ever smoking, alcohol consumption, hypertension, diabetes mellitus and serum HDL cholesterol).Abbreviations: miRNA, microRNA; ALP, Alkaline phosphatase ; SE, standard error; HDL, high-density lipoprotein.

Supplementary Table 5.4. 15 miRNAs significantly associated with fatty liver disease (continuous Hounsfield Unit values)

MiRNA ID	Beta	SE	p-value
miR-193a-5p	-35.88	5.11	4.26E-12
miR-4484	-20.04	3.67	6.02E-08
miR-1306-5p	55.59	10.42	1.21E-07
miR-6803-5p	-27.96	5.34	2.01E-07
miR-6870-3p	25.79	5.04	3.87E-07
miR-3937	15.02	3.00	6.59E-07
miR-378a-3p	-34.61	7.23	1.95E-06
miR-378d	-26.00	5.70	5.87E-06
miR-122-5p	-12.74	2.84	8.05E-06
miR-187-3p	48.67	11.12	1.33E-05
miR-6809-5p	34.03	7.85	1.61E-05
miR-193b-3p	-23.17	5.50	2.77E-05
miR-4713-3p	16.96	4.03	2.80E-05
miR-34b-3p	30.06	7.48	6.27E-05
miR-422a	-25.07	6.22	6.09E-05

Model 2: adjusted for age, sex, waist circumference, ever smoking, alcohol consumption, hypertension, diabetes mellitus and HDL cholesterol. The significant threshold is the Bonferroni-corrected p-value of 8.46E-05 (0.05/591 miRNAs). Abbreviations: miRNA, microRNA; FLD, fatty liver disease; SE, standard error; HDL,high-density lipoprotein.

Supplementary Table 5.5. Additional adjustments for liver enzymes and medication use in the association between 17 identified miRNAs and FLD (using dichotomous Hounsfield Unit values)

miRNA ID	Model 3			Model 4			Model 5		
	Beta	SE	p-value	Beta	SE	p-value	Beta	SE	p-value
miR-193a-5p	1.58	0.30	**1.43E-07**	1.86	0.29	**1.77E-10**	1.59	0.29	**5.59E-08**
miR-4484	0.58	0.22	9.98E-03	0.72	0.22	8.80E-04	0.57	0.22	1.05E-02
miR-1306-5p	-0.11	0.63	8.64E-01	-0.63	0.60	2.90E-01	-0.01	0.63	9.82E-01
miR-378a-3p	2.05	0.47	**1.17E-05**	2.35	0.45	**2.06E-07**	2.10	0.46	**5.77E-06**
miR-6803-5p	0.51	0.34	1.29E-01	0.72	0.32	2.43E-02	0.47	0.33	1.61E-01
miR-6870-3p	-0.46	0.26	7.90E-02	-0.59	0.24	1.64E-02	-0.43	0.26	9.90E-02
miR-3937	-0.43	0.18	1.87E-02	-0.57	0.17	1.14E-03	-0.41	0.18	2.50E-02
miR-122-5p	0.24	0.17	1.54E-01	0.54	0.14	1.49E-04	0.27	0.17	1.07E-01
miR-422a	1.73	0.46	1.94E-04	2.02	0.46	**9.31E-06**	1.76	0.46	1.34E-04
miR-378d	1.50	0.41	2.84E-04	1.75	0.40	**1.21E-05**	1.54	0.41	1.60E-04
miR-187-3p	-1.25	0.64	4.93E-02	-1.36	0.59	1.98E-02	-1.15	0.62	6.54E-02
miR-6809-5p	-0.63	0.43	1.38E-01	-0.76	0.38	4.47E-02	-0.60	0.42	1.56E-01
miR-193b-3p	0.56	0.35	1.08E-01	1.02	0.34	2.47E-03	0.63	0.34	6.72E-02
miR-4713-3p	-0.46	0.24	5.28E-02	-0.60	0.22	6.11E-03	-0.42	0.23	7.41E-02
miR-320d	1.11	0.29	1.55E-04	1.16	0.28	**3.20E-05**	1.13	0.29	1.16E-04
miR-34b-3p	-0.45	0.46	3.25E-01	-0.80	0.45	7.67E-02	-0.45	0.46	3.30E-01
miR-378e	1.41	0.43	1.02E-03	1.64	0.41	**7.43E-05**	1.45	0.43	6.63E-04

The table is sorted based on Bonferroni-corrected p-value association of miRNAs with FLD based on Hounsfield Unit values in model 2. The significant threshold is the Bonferroni-corrected p-value of 8.46E-05 (0.05/591 miRNAs). MiRNAs passing the thresholds are shown in bold. Model 2: adjusted for age, sex, waist circumference, ever smoking, alcohol consumption, hypertension, diabetes mellitus and HDL cholesterol. Model 3: model 2, + additionally adjusted for GGT and ALP. Model 4: model 2, + additionally adjusted for use of lipid-lowering medication, use of bile and liver medications. Model 5: model 2, + additionally adjusted for GGT, ALP, use of lipid-lowering medication, use of bile and liver medications. Abbreviations: miRNA, microRNA; FLD, fatty liver disease; GGT, Gamma-glutamyl transferase; ALP, Alkaline phosphatase; SE, standard error.

Supplementary Table 5.6. Additional adjustments for liver enzymes and medication use in the association between 17 identified miRNAs and FLD (using continuous Hounsfield Unit values)

miRNA ID	Model 3			Model 4			Model 5		
	Beta	**SE**	**p-value**	**Beta**	**SE**	**p-value**	**Beta**	**SE**	**p-value**
miR-193a-5p	-25.28	5.21	**1.45E-06**	-36.09	5.11	**3.26E-12**	-25.34	5.21	**1.35E-06**
miR-4484	-18.01	3.56	**4.90E-07**	-20.12	3.67	**5.34E-08**	-18.08	3.55	**4.36E-07**
miR-1306-5p	45.15	10.20	**1.07E-05**	56.74	10.45	**7.26E-08**	46.65	10.21	**5.54E-06**
miR-378a-3p	-25.69	7.10	3.16E-04	-34.64	7.23	**1.92E-06**	-25.51	7.10	3.45E-04
miR-6803-5p	-21.59	5.24	**4.17E-05**	-28.04	5.35	**1.97E-07**	-21.73	5.24	**3.70E-05**
miR-6870-3p	20.80	4.94	**2.83E-05**	26.01	5.05	**3.12E-07**	20.94	4.94	**2.49E-05**
miR-3937	11.80	2.95	**6.81E-05**	15.32	3.03	**5.13E-07**	12.25	2.97	**4.04E-05**
miR-122-5p	-1.62	3.12	6.05E-01	-12.82	2.84	**7.05E-06**	-1.58	3.12	6.13E-01
miR-422a	-18.16	6.07	2.85E-03	-25.22	6.22	**5.50E-05**	-18.23	6.06	2.72E-03
miR-378d	-19.32	5.60	5.92E-04	-26.00	5.71	**6.03E-06**	-18.96	5.61	7.64E-04
miR-187-3p	40.87	10.83	1.71E-04	49.31	11.14	**1.07E-05**	41.84	10.83	1.19E-04
miR-6809-5p	28.14	7.63	2.38E-04	34.45	7.87	**1.34E-05**	28.82	7.63	1.70E-04
miR-193b-3p	-12.91	5.50	1.92E-02	-23.26	5.50	**2.59E-05**	-12.92	5.49	1.89E-02
miR-4713-3p	13.56	3.93	5.81E-04	17.40	4.06	**2.02E-05**	14.28	3.95	3.14E-04
miR-320d	-12.38	4.89	1.15E-02	-14.57	5.03	3.87E-03	-12.60	4.89	1.01E-02
miR-34b-3p	23.99	7.25	9.83E-04	30.15	7.49	**6.11E-05**	24.21	7.25	8.78E-04
miR-378e	-15.58	6.17	1.17E-02	-21.72	6.33	6.29E-04	-15.49	6.18	1.23E-02

The table is sorted based on Bonferroni-corrected p-value association of miRNAs with FLD based on Hounsfield Unit values in model 2. The significant threshold is the Bonferroni-corrected p-value of 8.46E-05 (0.05/591 miRNAs). MiRNAs passing the thresholds are shown in bold. Model 2: adjusted for age, sex, waist circumference, ever smoking, alcohol consumption, hypertension, diabetes mellitus and HDL cholesterol. Model 3: model 2, + additionally adjusted for GGT and ALP. Model 4: model 2, + additionally adjusted for use of lipid-lowering medication, use of bile and liver medications. Model 5: model 2, + additionally adjusted for GGT, ALP, use of lipid-lowering medication, use of bile and liver medications. Abbreviations: miRNA, microRNA; FLD, fatty liver disease; GGT, Gamma-glutamyl transferase; ALP, Alkaline phosphatase; SE, standard error.

Supplementary Table 5.7. The expression of 14 identified miRNAs in the liver and relevant tissues

miRNA ID	Database	Tissue specifity index	Liver
miR-193a-5p	TissueAtlas	0.61	42.94
miR-4484	TissueAtlas	0.57	159.15
miR-1306-5p	miRmine	–	3.30
miR-378a-3p	TissueAtlas	0.97	444.41
miR-6803-5p	miRmine	–	0.00
miR-6870-3p	miRmine	–	0.00
miR-3937	TissueAtlas	0.58	316.39
miR-122-5p	TissueAtlas	0.97	4468.79
miR-422a	TissueAtlas	0.62	52.80
miR-378d	TissueAtlas	0.98	68.44
miR-187-3p	miRmine	–	7.40
miR-6809-5p	miRmine	–	0.00
miR-193b-3p	TissueAtlas	0.86	62.19
miR-378e	TissueAtlas	0.92	22.18

TissueAtlas database: Human miRNA tissue atlas database [1] (https://ccb-web.cs.uni-saarland.de/tissueatlas). miRmine database: Human miRNA expression database (https://guanfiles.dcmb.med.umich.edu/mirmine/index.html). The table shows the expression level of 14 identified miRNAs in the liver . Tissue specifity index range from 0 to 1 with the score of 0 corresponding to ubiquitously expressed miRNAs and a score of 1 for miRNAs that are expressed in a single tissue. The threshold originally proposed for defining housekeeping and specifically expressed miRNAs of tissue specifity index <0.15 and >0.85. Abbreviations: miRNA, microRNA; FLD, fatty liver disease; HDL: high-density lipoprotein.

References

[1] Ludwig N, Leidinger P, Becker K, et al. Distribution of miRNA expression across human tissues. Nucleic Acids Res 2016;44(8):3865-77

Supplementary Table 5.8. miRNA associated with NAFLD and AFLD based on logistic regression models using ultrasound data in model 2

MiRNA ID	NAFLD Model 2			AFLD Model 2		
	Beta	SE	p-value	Beta	SE	p-value
miR-193a-5p	0.30	0.17	7.10E-02	0.55	0.40	1.69E-01
miR-4484	0.15	0.11	1.86E-01	0.45	0.28	1.08E-01
miR-1306-5p	-0.30	0.33	3.53E-01	-1.41	0.76	6.21E-02
miR-378a-3p	0.17	0.23	4.59E-01	0.13	0.58	8.28E-01
miR-6803-5p	0.10	0.16	5.63E-01	0.64	0.41	1.12E-01
miR-6870-3p	-0.33	0.15	2.89E-02	0.08	0.43	8.49E-01
miR-3937	-0.07	0.10	4.69E-01	-0.59	0.23	**9.75E-03**
miR-122-5p	0.30	0.09	**8.56E-04**	0.27	0.19	1.52E-01
miR-422a	0.39	0.20	4.98E-02	-0.25	0.43	5.63E-01
miR-378d	0.34	0.19	6.74E-02	0.21	0.45	6.38E-01
miR-187-3p	-1.14	0.37	**1.86E-03**	-1.18	0.97	2.24E-01
miR-6809-5p	-0.46	0.27	8.31E-02	-1.21	0.66	6.72E-02
miR-193b-3p	0.16	0.17	3.48E-01	0.08	0.33	8.03E-01
miR-378e	0.51	0.21	1.43E-02	-0.24	0.41	5.62E-01

Model 2: Additionally adjusted for waist circumference, ever smoking, hypertension, diabetes mellitus and HDL cholesterol. The significant threshold is the Bonferroni-corrected p-value of 3.57E-03 (0.05/14 miRNAs). The miRNAs significantly associated with NAFLD in model 2 are bold. The miRNAs nominally (p-value< 0.05) associated with AFLD in model 2 are blod. Abbreviations: miRNA, microRNA; NAFLD, non-alcoholic fatty liver disease; AFLD, alcoholic fatty liver disease;SE, standard error; HDL, high-density lipoprotein.

Additional supplementary material for this chapter can be found online:
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7839694/bin/APT-53-432-s001.xlsx

6

CHAPTER 6

Multi-omics analysis reveals microRNAs associated with cardiometabolic traits

Manuscript based on this chapter: Mens MMJ, Maas SCE, Klap J, Weverling GJ, Klatser P, Brakenhoff JPJ, van Meurs JBJ, Uitterlinden AG, Ikram MA, Kavousi M, Ghanbari M. Multi-Omics Analysis Reveals MicroRNAs Associated With Cardiometabolic Traits. Front Genet. 2020 Feb 27;11:110. doi: 10.3389/fgene.2020.00110.

Abstract

Background: MicroRNAs (miRNAs) are non-coding RNA molecules that regulate gene expression. Extensive research has explored the role of miRNAs in the risk for type 2 diabetes (T2D) and coronary heart disease (CHD) using single-omics data, but much less by leveraging population-based omics data. Here we aimed to conduct a multi-omics analysis to identify miRNAs associated with cardiometabolic risk factors and diseases.

Methods: First, we used publicly available summary statistics from large-scale genome-wide association studies to find genetic variants in miRNA-related sequences associated with various cardiometabolic traits, including lipid and obesity-related traits, glycemic indices, blood pressure, and disease prevalence of T2D and CHD. Then, we used DNA methylation and miRNA expression data from participants of the Rotterdam Study to further investigate the link between associated miRNAs and cardiometabolic traits.

Findings: After correcting for multiple testing, 180 genetic variants annotated to 67 independent miRNAs were associated with the studied traits. Alterations in DNA methylation levels of CpG sites annotated to 38 of these miRNAs were associated with the same trait(s). Moreover, we found that plasma expression levels of 8 of the 67 identified miRNAs were also associated with the same trait. Integrating the results of different omics data showed miR-10b-5p, miR-148a-3p, miR-125b-5p and miR-100-5p to be strongly linked to lipid traits.

Interpretation: Collectively, our multi-omics analysis revealed multiple miRNAs that could be considered as potential biomarkers for early diagnosis and progression of cardiometabolic diseases.

Introduction

Type 2 diabetes mellitus (T2D) is a complex metabolic disease that is characterized by insulin resistance and impairment of insulin secretion, which leads to hyperglycemia. The presence of T2D leads to a two- to four-fold increase risk of developing coronary heart disease (CHD) [1], which is among the leading causes of morbidity and mortality worldwide [2]. Many risk factors are identified as mediators of these diseases, including hypertension, dyslipidemia, central adiposity and elevated blood glucose, which are together known as cardiometabolic traits [3]. Despite substantial advances in diagnosis and widely prescribed drugs for these diseases, their rate continue to increase worldwide, emphasizing the need for deeper insights into underlying mechanisms and innovative therapeutic strategies. Cardiometabolic traits and diseases have underlying genetic components and many loci have been discovered through large-scale genome- and epigenome-wide association studies [4, 5]. However, most of the identified genetic variants do not affect protein sequences, but are thought to affect gene regulation. One of the potential regulatory mechanisms involved might be microRNAs (miRNAs).

MiRNAs represent a class of small non-coding RNAs, which function as post-transcriptional regulators of gene expression via targeting the 3' untranslated region of target transcripts [6]. Over the past years, miRNAs have emerged as key regulators of biological processes underlying T2D and CHD. In this context, aberrant expression and function of miRNAs, such as miR-33, miR-208, miR-133 and miR-124, have been shown to be associated with lipid metabolism, insulin secretion, myocardial infarction and T2D [7-9]. Most of the disease-associated miRNAs have been discovered in cells originated from tissue of interest in small number of samples or animal studies. But advances in high-throughput technologies make it possible to study miRNAs in a population-based manner. In particular cell-derived vesicles, known as exosomes, release miRNAs in the blood stream that are very stable and can be used as biomarkers for disease [10].

Similar to other regulatory RNA molecules, the function and expression of miRNAs can be affected by genetic variants. Single-nucleotide polymorphisms (SNPs) can occur at various stages of the miRNA biogenesis including precursor- and mature miRNA sequences [11] as well as within regulatory elements, such as promoter regions [12]. Also, DNA methylation can control transcription, which have been reported to be associated with the expression level of miRNAs [13]. In this context, epigenome-wide association studies (EWAS) have shown that altered DNA methylation within miRNA promoters is associated with miRNAs expression levels and therewith modify disease risk [14]. However, previous studies are mainly based on single omics data or small sample size [15, 16]. As each type of omics data provides associations that can be useful for detecting development or progression of disease, integrating different omics layers can limit passive correlations and provide a more comprehensive view of the disease biology.

In this study, we applied a multi-omics approach to identify miRNAs associated with cardiometabolic traits. First, we identified genetic variants in miRNA sequences and their potential regulatory regions associated with different cardiometabolic risk factors and diseases using genetic association data from the available genome-wide association studies (GWAS). We then integrated population-based DNA methylation and miRNA expression data from the Rotterdam Study to link omics layers, strengthening the association of the identified miRNAs with cardiometabolic traits. We envision that the identified miRNAs could be considered as potential biomarkers for early diagnosis of cardiometabolic diseases.

Material and Methods

A graphical overview of the multi-omics approach used in this study is illustrated in **Figure 6.1.**

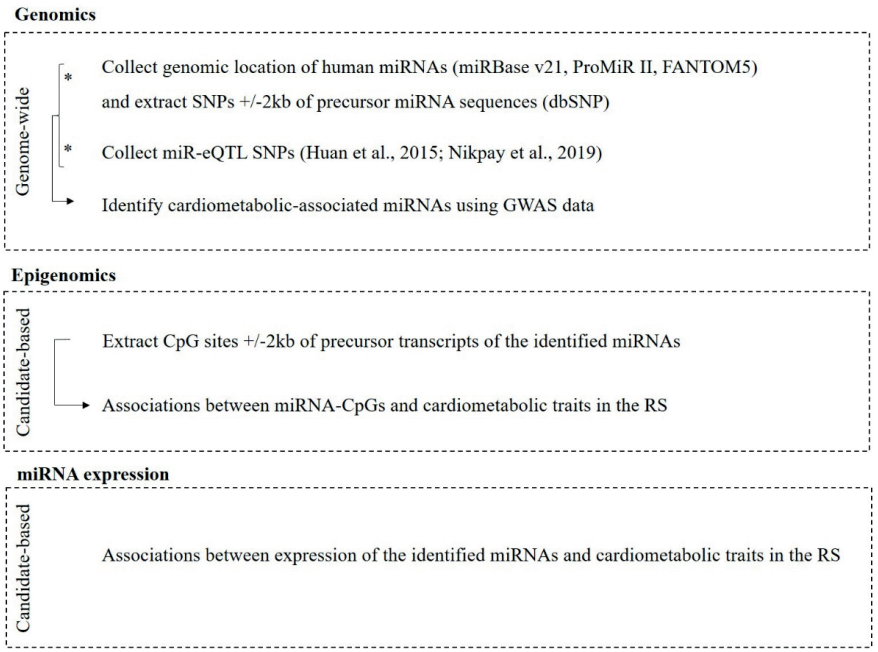


Figure 6.1. Overview of the multi-omics layers used in this study.

Genetic association study

Retrieval of SNPs in miRNA-related regions

The primary transcripts of miRNAs for the processing to mature miRNAs are approximately 3-4kb in length [17]. We collected the genomic position of all human miRNAs employing the miRBase database (v21) [18], ProMiR II [19] and FANTOM5 [12]. Using dbSNP database [20], we extracted 18,545 SNPs located in +/-2kb of the precursor miRNA sequences (pre-miRNA) of 1,554 known miRNAs. Of these, 2,420 SNPs are located in pre- and mature sequences of miRNAs. Genetic variants have been found to alter miRNA expression and are known as miRNA expression quantitative trait loci (miR-eQTLs). To this end, we included 5,528 miR-eQTLs that change the expression of 221 mature miRNA using data from the Framingham Heart Study (FHS) [21] and from the Ottawa Hospital Bariatric Centre [22]. The FHS focused on *cis*-miR-eQTLs, of which the majority was located 300-500kb away from their target miRNA. Nikpay et al. (2019) investigated both *cis*-miR-eQTLs and *trans*-miR-eQTLs, however, they reported similar to the FHS that most cis-miR-eQTLs were distal regulators of the miRNAs. There were 83 miR-eQTLs overlapping with the SNPs in +/-2kb of the precursor miRNA sequence. Altogether, 23,990 unique SNPs were included in our analysis.

The genomic location of miRNAs can be discriminated among intergenic and intragenic. Roughly half of the known miRNAs are found to be transcribed from intergenic regions of the genome, suggesting that these miRNAs are transcribed under independent control of regulatory elements [23]. The intragenic miRNAs are embedded within sequences of protein-coding genes, including intronic and exonic regions. If the intragenic miRNA and its host gene share the same promoter, the miRNA is likely to be co-expressed with the host gene [24]. Here, the genomic location of the identified miRNAs was obtained using miRIAD [25].

Genome-wide association studies of cardiometabolic traits

Cardiometabolic risk factors and diseases in this study were classified into four specific trait groups based on their shared pathophysiology and underlying pathways. These include i) Anthropometric traits: body mass index (BMI), waist to hip ratio (WHR) and waist circumference (WC); ii) Glycemic traits: fasting glucose (FG), glucose 2 hours (G2H), fasting insulin (FI), proinsulin (Pro-Ins), hemoglobin A1c (HbA1c), homeostatic model assessment of insulin resistance (HOMA-IR), β-cell function (HOMA-β) and type 2 diabetes mellitus (T2D); iii) Lipid traits: low-density lipoprotein (LDL), high-density lipoprotein (HDL), total serum cholesterol (TC) and triglycerides (TG); and iv) Cardiovascular traits: coronary artery disease (CAD), diastolic (DBP) and systolic blood pressure (SBP). To test the association of miRNA-related SNPs with cardiometabolic traits we used publicly available GWAS summary statistics. A description of GWAS

meta-analysis data and corresponding consortia used in this study is provided in **Supplementary Table 6.1.** 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/), in which we excluded the SNPs with $R^2 > 0.7$. Bonferroni correction was used to adjust for multiple testing based on the number of independent SNPs available in the GWAS data (HapMap or 1000G project imputed data).

Prioritization of miRNA-related SNPs associated with cardiometabolic traits

For miRNA-related SNPs significantly associated with cardiometabolic traits, we performed *in silico* analysis to prioritize the SNPs that are more likely to be functional in their corresponding loci based on the following criteria: i) association between the miRNA-related SNP and the cardiometabolic trait, ii) association between the miRNA-related SNP and the expression level of miRNA/miRNA hosting gene, and iii) expression of the miRNA in tissues relevant to cardiometabolic traits. In this regard, regional association plots were generated (using LocusZoom web tool, Version 1.1) to visualize the physical position and evaluate the association of the cardiometabolic traits with the miRNA-related SNP and its proxy SNPs ($R^2 > 0.8$) in the corresponding locus: (i) To explore whether the SNP is associated with the expression of related miRNA or miRNA hosting genes in relevant tissues (e.g., adipose tissue, liver, pancreas, muscle and blood), we used eQTL data from GTEx Portal (https://www. gtexportal.org/home/), (ii) We used two online databases; miRmine and Human miRNA tissue atlas [26, 27] to test where a miRNA is expressed in tissues relevant to cardiometabolic traits (e.g., adipose tissue, liver, pancreas, muscle, and blood), (iii) The Vienna RNAfold algorithm was used to check miRNA secondary structure and free energy changes with wild-type and mutant alleles of SNPs located in miRNA sequences [28].

Determination of methylation quantitative trait loci (me-QTLs)

To determine if the identified SNPs have an effect on the methylation levels of CpG sites (me-QTLs), we used data of a recent me-QTL study performed in five cohorts, including the RS, with a total of 3,841 individuals [29]. We incorporated both *cis*-me-QTLs and *trans*-me-QTLs. Where *cis*-me-QTLs were defined as the effect of SNPs on the methylation levels of a CpG sites no further than 250kb apart, *trans*-me-QTLs were defined as the effect of distal SNPs on the CpG methylation levels. Details on the me-QTL mapping are described elsewhere [29]. We tested if the cardiometabolic-associated SNPs found in the current study were identified as me-QTLs.

DNA methylation analysis in the Rotterdam Study

The Rotterdam Study (RS) is a large prospective population-based cohort study conducted among middle-aged and elderly people in the suburb Ommoord in Rotterdam, the Netherlands. In 1989, 7,983 inhabitants aged 55 and older were recruited in the first cohort (RS-I) (78% of 10,215 invitees). In 2000, the RS was extended with a second cohort of 3,011 participants that moved to Ommoord or turned 55 years old (RS-II). In 2006, the third cohort (RS-III) was initiated in which inhabitants aged 45-54 years were invited and included 3,932 participants. A detailed description of RS can be found elsewhere [30]. In the current study, we used DNA methylation data from a random subset (n=717) of the third visit of RS-II (RS-II-3) and second visit of RS-III (RS-III-2) and a random subset (n=721) of the first visit of RS-III (RS-III-1). There was no overlap in participants. The RS has been approved by the institutional review board (Medical Ethics Committee) of the Erasmus Medical Center and by the review board of The Netherlands Ministry of Health, Welfare and Sports. All participants gave written consent before participation in the study. Participant characteristics are presented in **Supplementary Table 6.2.**

DNA was extracted from whole peripheral blood using standardized salting out methods, of which 500ng was bisulfite treated using the Zymo EZ-96 DNA methylation kit (Zymo Research, Irvine, CA, USA). Bisulfite converted DNA was hybridized to the Illumina Human 450K array (Illumina, San Diego, CA, USA), according to manufacturer's protocol. Data preprocessing was performed using an R programming pipeline based on the pipeline developed by Touleimat and Tost [31]. The genome coordinates provided by Illumina (GRCh37/hg19) were used to identify independent loci. We extracted 12,939 unique CpGs located in +/-2kb of the pre-miRNA sequences using the Illumina450K array annotation file as provided by Illumina [32]. Among these, 12,617 CpGs were located in the regulatory region of 1,269 miRNAs and 450 CpGs were located in the pre- and mature sequence of 391 miRNAs. We tested the association of these CpGs with different cardiometabolic traits using linear mixed models. Data collection on these traits in the RS is described in **Supplementary Methods**. The models were adjusted for age, gender, current smoking, blood cell counts (monocytes, granulocytes, lymphocytes) as fixed effects and technical covariates as random effects. Models were further adjusted for covariates per group as follows: i) for Anthropometric traits we adjusted WC and WHR for BMI, ii) for Glycemic traits we adjusted for BMI and diabetic medication, iii) for Lipid traits we adjusted for BMI and lipid medication, and iv) for Cardiovascular traits we adjusted for BMI, blood pressure lowering medication and lipid medication. A candidate-based approach was used to sought overlap between identified miRNAs. A nominal p-value of <0.05 was found to be significant.

Determination of miR-eQTMs

To identify association between the methylation level of CpGs and the expression of miRNAs (miR-eQTMs), we used miR-eQTM data from a recent study [13]. The latter study analyzed associations of expression levels of 283 miRNAs with methylation of CpGs from 3,565 individuals, in which they identified 227 miR-eQTMs at FDR<0.01. We tested if any of the cardiometabolic-associated CpGs in the current study was among the identified miR-eQTM [13].

MiRNA expression profiling in the Rotterdam Study

We performed miRNA expression analysis in 2,000 RS participants, including a random subset (n=1,000) of the fourth visit of RS-I (RS-I-4) and a random subset (n=1,000) of the second visit of RS-II (RS-II-2). Plasma miRNA levels were determined using the HTG EdgeSeq miRNA Whole Transcriptome Assay (WTA), which measures the expression of 2,083 mature human miRNAs (HTG Molecular Diagnostics, Tuscon, AZ, USA) and using the Illumina NextSeq 500 sequencer (Illumina, San Diego, CA, USA). The WTA characterizes miRNA expression patterns, and measures the expression of 13 housekeeping genes, that allows flexibility in data normalization and analysis. Quantification of miRNA expression was based on counts per million (CPM). Log2 transformation of CPM was used as standardization and adjustment for total reads within each sample. MiRNAs with Log2 CPM <1.0 were indicated as not expressed in the samples. The lower limit of quantification (LLOQ) was used to select well-expressed miRNAs. The LLOQ level was based on a monotonic decreasing spline curve fit between the means and standard deviations of all miRNAs. In our definition well-expressed miRNA levels in plasma were those with >50% values above LLOQ. Out of the 2,083 measured miRNAs, 591 miRNAs were expressed at good levels in plasma.

The miRNAs significantly associated with cardiometabolic traits, in the genetic association studies, were tested for the association of their plasma expression levels with the same cardiometabolic trait(s). Linear models were used to test the association between available continuous traits in the RS (incl. BMI, WC, WHR, FG, HDL, TC, SBD and DBP) and miRNA expression. Additionally, we used binomial models to test the association between disease prevalence (incl. T2D and CHD) and miRNA expression. We used the cardiometabolic traits as dependent variable and plasma miRNAs level as explanatory variable, adjusting for age, gender and current smoking. Models were further adjusted for covariates per group as follows: i) for Anthropometric traits we adjusted WC and WHR for BMI, ii) for Glycemic traits we adjusted for BMI and diabetic medication, iii) for Lipid traits we adjusted for BMI and lipid medication, and iv) for Cardiovascular traits we adjusted for BMI, blood pressure lowering medication and lipid medication. A candidate-based approach was used to sought overlap between identified miRNAs. A nominal p-value of <0.05 was found to be significant.

In addition, we extracted strongly validated target genes, defined as being validated by western blot and/or luciferase reporter assay, of the identified miRNAs from the miRTarBase database [33]. Next, we extracted SNPs in these target genes and tested their associations with cardiometabolic traits using summary statistics of previously mentioned GWAS data.

Results

Association of miRNA-SNPs with cardiometabolic traits and diseases

Out of 23,990 miRNA-related SNPs, 2,358 independent SNPs were present in the GWAS data based on HapMap and 8,652 independent SNPs were present in the 1000G project. Bonferroni correction was used to set the significance threshold, at p-value $<2.12\times10^{-5}$ (0.05/2,358) for GWAS with HapMap imputed data and p-value $<5.78\times10^{-6}$ (0.05/8,652) for GWAS with 1000 Genomes project imputed data. Of these, 180 SNPs annotated to 67 miRNAs passed the significance threshold to be associated with at least one cardiometabolic trait (**Table 6.1**). Out of the 180 identified SNPs, 89 SNPs were located in +/-2kb of 57 primary miRNA transcripts (**Supplementary Table 6.3**) and 92 SNPs were among the previously reported miR-eQTLs of 15 mature miRNAs (**Supplementary Table 6.4**). Manhattan plots illustrated in **Figure 6.2** present the miRNA-annotated genetic variants associated with lipid traits and the prevalence of T2D and CHD. **Table 6.2** shows the top miRNA-related SNPs associated with cardiometabolic traits, which were annotated to 20 miRNAs.

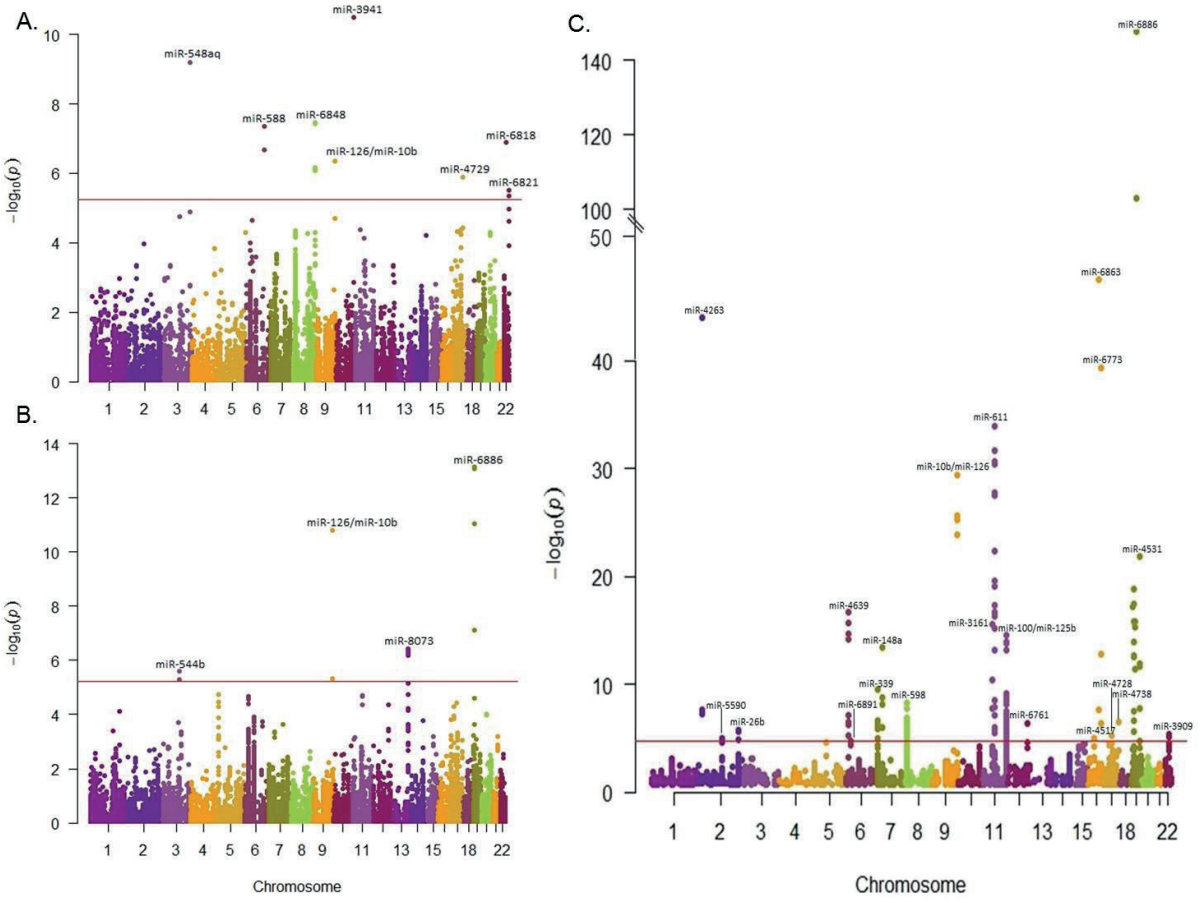


Figure 6.2. Manhattan plots showing the association of miRNA-SNPs with T2D, CAD and lipid traits. The association miRNA-related SNPs and cardiometabolic traits were examined using the publicly available GWAS data. We reported the most significantly associated miRNA of each SNP loci. The horizontal red line indicates the study significance threshold. (**A)** Manhattan plot showing the association of miRNA-SNPs with T2D in which 12 SNPs in 8 miRNAs passed the significant threshold. (**B)** Manhattan plot showing the association of miRNA-SNPs with CAD in which 13 SNPs in 9 miRNAs passed the significance threshold. (**C)** Manhattan plot showing the association of miRNA-SNPs with lipid traits in which 107 SNPs in 36 miRNAs passed the significant threshold. When SNPs were present in more traits, the most associated SNP was plotted.

Table 6.1. Description of GWAS of cardiometabolic traits and associated miRNA SNPs

Phenotype	Consortium	SNPs in +/- 2kb miR*	SNPs in miR-seq*	SNPs in miR-QTL*	Associated miR loci†
Anthropometric traits					
Body-mass index	GIANT[48]	9	0	9	7
Waist to hip ratio	GIANT[49]	2	1	1	4
Waist circumference	GIANT[49]	10	0	1	8
Glycemic traits					
Glucose fasting	MAGIC[50]	3	0	1	4
Glucose after 2h	MAGIC[51]	0	0	0	0
Insulin fasting	MAGIC[50]	1	0	2	2
Proinsulin	MAGIC[52]	3	0	4	3
HbA1c	MAGIC[53]	1	0	15	4
HOMA-IR	MAGIC[54]	0	0	0	0
HOMA-β	MAGIC[54]	0	0	0	0
Type 2 diabetes	DIAGRAM[55]	12	0	1	8
Lipid traits					
LDL	GLGC[56]	22	1	20	11
HDL	GLGC[56]	12	1	23	9
Total cholesterol	GLGC[56]	26	1	40	13
Triglyceride	GLGC[56]	8	1	27	7
Cardiovascular traits					
CAD	CARDIoGRM plusC4D[57]	10	0	2	4
DBP	ICBP[58]	3	0	-	2
SBP	ICBP[58]	2	0	-	2

Shown are SNPs located within +/-2kb of primary miRNA transcripts, pre- and mature miRNA sequences, miRNA-eQTL SNPs).

* Number of SNPs that passed the significance threshold (p-value $<2.12\times10^{-5}$ for SNPs imputed with HapMap and p-value $<5.78\times10^{-6}$ for SNPs imputed with 1000G)

† Number of independent loci.

In order to prioritize miRNA-related SNPs based on potential functionality in relation to the associated cardiometabolic traits, we created regional association plots to visualize the LD of miRNA SNP with the top SNP in the corresponding locus (**Figure 6.3**). We found three top SNPs in their loci, including rs7117842 associated with TC ($p=2.48\times10^{-15}$, $\beta=0.029$) and located ~512kb upstream of miR-100-5p/miR-125b-5p (**Figure 6.3A**), rs1997243 associated with TC ($p=2.72\times10^{-10}$, $\beta=0.033$) and located ~21kB upstream of miR-339-3p (**Figure 6.3B**), and rs7607369 associated with BMI ($p=1.10\times10^{-7}$, $\beta=-0.016$) and located ~11.7kb upstream of miR-26b-5p (**Figure 6.3C**). These three SNPs were previously identified as miR-eQTLs that change the expression levels of related miRNAs in blood [21]. In addition, rs4722551 located ~2kb upstream of miR-148a shows the strongest association with LDL ($p=3.95\times10^{-14}$, $\beta=0.039$) on the Chr7p15.2 locus (**Figure 6.3D**).

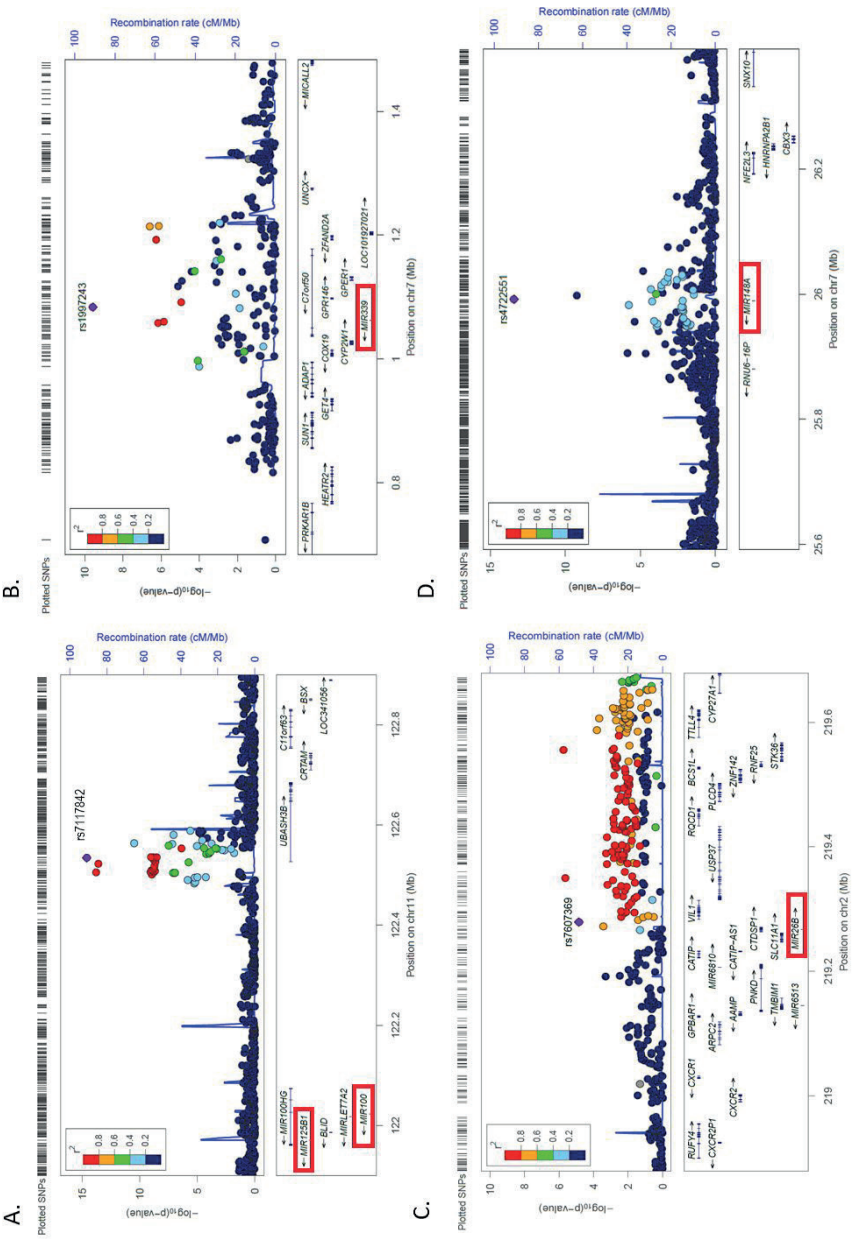


Figure 6.3. Regional plots showing the association of four top miRNA-SNPs with cardiometabolic traits. The most significant SNP in the region, according to P-value, is represented by a purple diamond, and the degree of linkage disequilibrium of other SNPs in the region to the lead SNP is representative by the color scale

shown in the legend. Genes are illustrated below. The associated miRNA is illustrated with a red box. (**A)** Regional plot showing the association of rs7117842 located ~512kb upstream of miR-100-5p/125b-5p with TC, LDL and HDL on the Chr11q24.1 locus. (**B)** Regional plot showing the association of rs1997243 located ~21kB upstream of the primary transcript of miR-339-3p with TC and HDL on the Chr7p22.3 locus. (**C)** Regional plot showing the association of rs7607369 located ~11.7kb upstream of the primary transcript of miR-26b-5p with BMI and TG on the Chr2q35 locus. (**D)** Regional plot showing the association of rs472551 located ~2kb upstream of the primary transcript of miR-148a with LDL, TG and TC on the Chr7p15.2 locus.

Moreover, rs174561 has previously been reported by [22] to change the expression of miR-1908-5p. We found this SNP, located in the coding sequence of miR-1908-5p, to be associated with lipid traits (LDL, HDL,TC and TG), and rs11614913, located in the coding sequence of miR-196a2-3p, to be associated with WHR. These two variants have previously been reported to be associated with lipid traits and WHR and have been suggested to change the miRNA structure and expression [34]. We also found a suggestive association between rs58834075, located in the pre-miR-656 sequence (T>C, Chr14:101066756) and T2D ($p=6.30\times10^{-5}$, $\beta= -0.170$). The miRNA secondary structure and free energy changes of both wild-type and mutant alleles of these three SNPs (rs174561, rs11614913 and rs58834075) are illustrated in **Supplementary Figure 6.1**.

Table 6.2. The top twenty miRNAs with SNPs in related regions association with cardiometabolic traits

miRNA	SNPID	Chr.	Position	Alleles (A/R)	Annotated gene	Genomic location miRNA	Associated trait	β-coefficient	P value
miR-6886†	rs17248720	19	11198187	C/T	*LDLR*	Intronic	LDL	0.226	$2.40x10^{-148}$
miR-6863†	rs13306673	16	56900931	C/T	*SLC12A3*	Intronic	HDL	0.098	$2.76x10^{-48}$
miR-4263†	rs2305929	2	28113911	G/A	*BRE*	Intronic	TG	0.064	$1.13x10^{-44}$
miR-6773†	rs8057119	16	68268836	T/C	*ESRP2*	Intronic	HDL	0.072	$5.21x10^{-40}$
miR-611†	rs174538	11	61560081	G/A	*THEM258*	Exonic	LDL	0.050	$1.07x10^{-34}$
miR-1908-5p†	rs174548	11	61571348	C/G	*FADS1*	Exonic	LDL	0.047	$2.29x10^{-31}$
miR-10b-5p/126-5p‡	rs532436	9	136149830	A/G	*ABO*	Exonic/Intronic	LDL	0.079	$4.02x10^{-30}$
miR-4721†	rs4788099	16	28763228	G/A	*TUMF*	Exonic	BMI	0.031	$1.09x10^{-24}$
miR-4531†	rs6509170	19	45159636	C/A	*LOC107985305*	Intronic	LDL	0.127	$1.54x10^{-22}$
miR-199a-1†	rs11085748	19	10927540	T/C	*DNM2*	Intronic	LDL	0.055	$1.46x10^{-19}$
miR-4999†	rs7254882	19	8359822	C/T	*MIR4999*	Intergenic	HDL	0.033	$6.66x10^{-18}$
miR-4639†	rs3757354	6	16127407	C/T	*MYLIP*	Intronic	LDL	0.038	$2.09x10^{-17}$
miR-640†	rs1000237	19	19518316	T/A	*GATAD2A*	Intronic	TG	0.033	$1.61x10^{-16}$
miR-3161†	rs79837139	11	48000780	C/T	*PTPRJ*	Intronic	HDL	0.062	$2.99x10^{-16}$
miR-100-5p/125b-5p‡	rs7117842	11	122663796	C/T	*UBASH3B*	Intergenic/Intergenic	TC	0.029	$2.48x10^{-15}$
miR-148a†	rs4722551	7	25991826	C/T	*MIR148A*	Intergenic	LDL	0.039	$3.95x10^{-14}$
miR-139†	rs11605042	11	72700619	A/G	*ARAP1*	Intronic	Pro-Ins	-0.053	$5.24x10^{-13}$
miR-3941†	rs71486610	10	124134803	C/G	*PLEKHA1*	Intronic	T2D	-0.081	$3.30x10^{-11}$
miR-6745†	rs901750	11	47209472	A/G	*PACSIN3*	Intronic	HDL	0.024	$3.95x10^{-11}$
miR-196a-2-3p*	rs11614913	12	53991815	C/T	*MIR196A2*	Intergenic	WHR	0.029	$6.90x10^{-11}$

* SNP located in pre- and mature miRNA sequence

† SNP located within +/- 2kb of primary miRNA transcript

‡ miR-eQTL SNPs

Identification of methylation quantitative trait loci (me-QTLs)

We identified 29 *cis*-me-QTL effects for 47 independent CpGs at FDR< 0.05 (49 SNP-CpG pairs). Among these, we found 14 *cis*-me-QTLs that were associated with both the expression level of 8 miRNAs and the methylation level of 26 CpGs (**Supplementary Table 6.5**). In total there were 7 *cis*-me-QTLs (for 8 CpGs) that were associated with a cardiometabolic trait in the current study (**Table 6.3).** Furthermore, 4 *trans*-me-QTL effects for 21 independent CpGs were found at FDR<0.05 (27 SNP-CpG pairs) (**Supplementary Table 6.5).** Two out of the four *trans*-me-QTL were miR-eQTL SNPs (rs174548 for miR-1908-5p and rs1997243 for miR-339-3p). None of the associated CpGs *in trans* were found in the current study to be associated with cardiometabolic traits.

Table 6.3. Identified me-QTLs with cardiometabolic-associated CpGs

miRNA	SNPID	CpG	Cis[†]/ Trans	miR-eQTL SNP[*]	SNP associated with cardiometabolic trait	CpG associated with cardiometabolic trait
miR-611	rs174538	cg16150798	Cis	-	FG, LDL, HDL, TG, TC	WC
miR-588	rs9388486	cg20229609	Cis	-	T2D	SBP, DBP
miR-1908-5p	rs174548	cg03921599	Cis	√	FG, HbA1c, LDL, HDL, TG, TC	LDL, TC
miR-199a-1	rs3786719	cg02907064	Cis	-	LDL, TC	LDL
miR-6745	rs901750	cg00724111	Cis	-	HDL	FI, SBP, DBP
miR-8073	rs3809346	cg22382805	Cis	-	CAD	FI
miR-653, miR-489	rs2528521	cg06934092	Cis	-	BMI	FG, FI, TC
miR-8073	rs3809346	cg19700260	Cis	-	CAD	DBP

Shown are 7 me-QTLs associated with methylation levels of 8 cardiometabolic-associated CpG sites.

[*] miR-eQTL SNP is associated to change the expression of miRNA level

[†] SNP and CpG are located not further away than 250kb

Testing DNA methylation and expression of miRNAs associated with cardiometabolic traits

To access the relationship between miRNAs and cardiometabolic traits in other omics layers, we performed a candidate-based test to check whether the 67 identified miRNAs, with SNPs associated with cardiometabolic traits, show also association between DNA methylation and miRNA expression with cardiometabolic traits. Using DNA methylation data from 1,438 RS participants, we found 278 CpG sites annotated to 64 out of the 67 miRNAs, to be associated with any cardiometabolic trait (**Supplementary Table 6.6**). By integrating our DNA methylation results with the GWAS data, we observed an overlap of 38 miRNAs (79 CpGs) that had both a SNP and a CpG associated with the same trait (**Supplementary Table 6.7**). The CpG site showing the most significant association was cg15616915 which is located in the regulatory region of miR-26b and is positively associated with TG ($p=1.59\times10^{-4}$, $\beta=0.009$). We found 16 cardiometabolic-associated CpGs that are annotated to more than one miRNA. For example, cg03722243 associated with BMI ($p=1.55\times10^{-3}$, $\beta=0.001$) is annotated to miR-489 and miR-653, which are clustered on chromosome 7. In addition, cg15334028 associated with WC, HDL, LDL and TG is annotated to three miRNAs (miR-638, miR-6793 and miR-4748) on chromosome 19.

We identified two CpGs that are associated with the expression level of miRNAs (miR-eQTM) at FDR<0.01. The most significant cis-miR-eQTM, cg26363555 has been reported to be negatively associated with both miR-125b-5p (~2kb downstream) and miR-100-5p (~50kb upstream) expression levels [13]. The CpG cg26363555 was positively associated with FG ($\beta=0.012$) and DBP ($\beta=2.00x10-4$) and negatively associated with HDL ($\beta=-0.004$) in the RS. In addition, cg03891346 has been reported to be negatively associated with the expression level of miR-100-5p (~53kb downstream) [13]. This CpG, which is also annotated to MIR125B1, was positively associated with WC ($\beta=5.00x10^{-4}$) in the RS.

Next, we tested whether the 67 identified miRNAs show differential expression in plasma in relation to the associated cardiometabolic trait(s). Of the 67 miRNAs, we could only test the association of 28 mature miRNAs that were well-expressed in plasma and of which the phenotype of interest was available in the RS. Of these, plasma levels of 22 miRNAs were nominally associated with at least one cardiometabolic traits (**Supplementary Table 6.8**). Furthermore, out of the 67 miRNAs, we found 12 differently expressed mature miRNAs to be associated with the same trait (**Table 6.4**). Plasma levels of miR-126-3p, miR-126-5p, miR-10b-5p, miR-148a-3p, miR-199a-1-3p, miR-199a-1-5p, miR-125b-5p and miR-100-5p were positively associated with serum TC levels. In contrast, miR-6886 was negatively associated with serum TC levels. A negative association between miR-126-5p and miR-126-3p and CHD was found. Furthermore, we observed a negative association between miR-4681 levels and WC. An overview of the number of associated miRNAs using different omics data is illustrated in **Figure 6.4**.

Table 6.4. Plasma expression levels of miRNAs associated with cardiometabolic traits

miRNA	β-coefficient	P value	Associated trait
miR-126-3p	0.379	$1.09x10^{-14}$	TC†
miR-10b-5p	0.352	$3.30x10^{-11}$	TC†
miR-126-5p	0.258	$3.75x10^{-11}$	TC†
miR-148a-3p	0.189	$8.01x10^{-06}$	TC†
miR-199a-1-3p	0.171	$3.38x10^{-05}$	TC†
miR-125b-5p	0.159	$2.43x10^{-03}$	TC†
miR-100-5p	0.141	$3.15x10^{-03}$	TC†
miR-6886-3p	-0.083	$9.49x10^{-03}$	TC†
miR-126-5p	-0.365	$1.24x10^{-02}$	CHD‡
miR-4681	-0.440	$2.13x10^{-02}$	WC*
miR-199a-1-5p	0.074	$3.38x10^{-02}$	TC†
miR-126-3p	-0.385	$3.54x10^{-02}$	CHD‡

Model 1: adjusted for: age, gender, current smoking

* Model 1 + BMI

† Model 1 + BMI, lipid medication

‡ Model 1 + BMI, blood pressure lowering medication, lipid medication

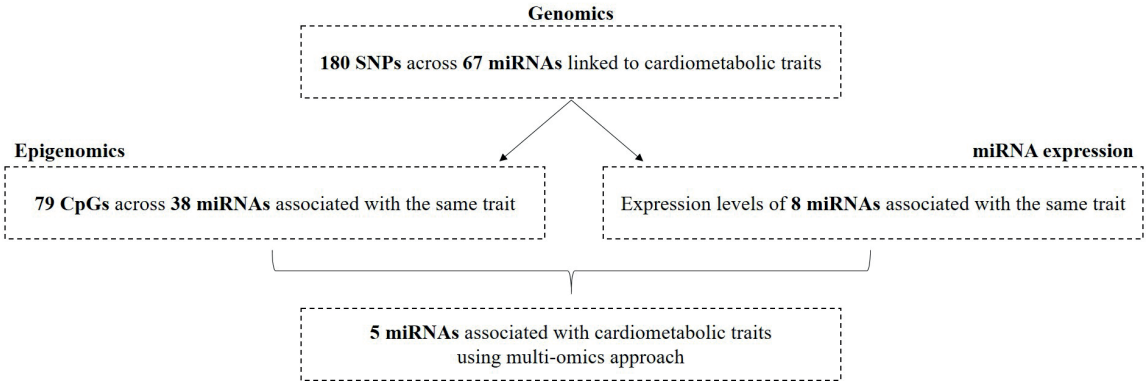


Figure 6.4. Overview of miRNAs associated with cardiometabolic traits by integrating three omics layers.

Furthermore, out of 22 miRNAs that were associated with at least one cardiometabolic trait, we found validated target genes for 14 miRNAs. We tested the association between these target genes and cardiometabolic traits using summary statistics GWAS data. After correcting for multiple testing, based on the number of tested SNPs in the target genes of a miRNA, we found 24 unique target genes for 9 of the 14 miRNAs to be associated with cardiometabolic traits (**Supplementary Table 6.9**).

Finally, we sought overlapping miRNAs that were associated with the same cardiometabolic trait in the three different omics analyses (**Supplementary Table 6.10**). Since not all related phenotypes were available within the RS and not all miRNAs were expressed, we tested 64 miRNAs that had DNA methylation sites and 22 mature miRNAs that were available for miRNA expression analyses using the RS. We found five miRNAs, including miR-10b-5p, miR-148a-3p, miR-100-5p, miR-125b-5p and miR-6886 that had at least one CpG and of which the expression was also associated with the same cardiometabolic trait. After prioritization based on the suggested criteria for potential functionality, miR-10b-5p, miR-148a-3p, miR-125b-5p and miR-100-5p were highlighted as the most likely miRNAs involved in the pathogenesis of risk factors for T2D and CHD (**Table 6.5**).

Table 6.5. Cardiometabolic-associated miRNAs after integrative multi-omics data and in silico prioritization analysis

miRNA	Associated traits	SNPID	P value	Proxy SNPs (Non-syn.)	GWAS	miRNA CpG	P value	Associated trait	miR-eQTM	Associated trait	miRNA Exp.	Associated trait
miR-10b-5p*	LDL†, TC, CAD, T2D	rs532436(‡)	$4.02x10^{-30}$	9 (0)	√	cg25820279	$4.37x10^{-02}$	TC	-	-	$3.30x10^{-11}$	TC
miR-148a-3p*	LDL†, TC, TG	rs4722551	$3.95x10^{-14}$	1 (0)	√	cg18188200	$1.94x10^{-03}$	TC	-	-	$8.01x10^{-06}$	TC
miR-125b-5p*	TC†, LDL, HDL	rs7117842(‡)	$2.48x10^{-15}$	52 (0)	√	cg06749053	$2.39x10^{-02}$	LDL	cg26363555	HDL	$2.43x10^{-03}$	TC
miR-100-5p*	TC†, LDL, HDL	rs7117842(‡)	$2.48x10^{-15}$	52 (0)	√	cg14724899	$4.02x10^{-02}$	HDL	-	-	$3.15x10^{-03}$	TC
miR-6886-3p	LDL†, TC, CAD	rs17248720	$2.40x10^{-148}$	49 (0)	-	cg19751789	$8.03x10^{-04}$	TC	-	-	$9.49x10^{-03}$	TC

* MiRNAs which were associated with cardiometabolic traits using multi-omics approach and were according to the prioritization analysis more likely to be involved in the development of cardiometabolic traits. † Trait to which the SNP is associated. ‡ miR-eQTL SNP

Shown here are 5 miRNAs that were association with cardiometabolic traits using multi-omics approach and being tested with additional in silico analysis. Proxy: Number of proxy SNPs (R^2>0.8) in strong linkage disequilibrium (LD) with the given variant. Non-synonymous: Number of non-synonymous variants that are in high LD with the given variant. GWAS: (√ SNP represented to be top SNP with the strongest association with the related trait on the certain genomic position). miR-eQTM: CpG is associated with expression level of miRNA.

Discussion

In this study, we integrated different population-based omics data (including genetics, epigenetics and miRNA expression) to identify miRNAs associated with cardiometabolic traits. Genetic variants related to 67 miRNAs were associated with the studied traits. Alterations in DNA methylation of CpG sites annotated to 38 of these miRNAs and plasma expression levels of 8 of them were also associated with the same trait. In principle, the association between a miRNA and trait of interest in more than two layers of omics may strengthen its potential to play a role in the disease underlying mechanisms. In this context, we sought to identify overlap between miRNAs that were associated with the same cardiometabolic trait across different approaches. This integration analysis revealed the correlation between four miRNAs (miR-10b-5p, miR-148a-3p, miR-125b-5p and miR-100-5p) and lipid traits.

MiR-10b-5p is a highly conserved miRNA across multiple species and is located inside the homeobox D cluster on chromosome 2. A recent study showed a mediating role for miR-10b between obesity and primary breast cancer [35]. Moreover, previous research in mice found a negative regulatory role of miR-10b on cholesterol efflux via targeting the ATP binding cassette transporter gene (ABCA1) [36]. MiR-10b has been also shown to be involved in the progression of atherosclerosis, which is a major cause of cardiovascular disease [37]. We found a genetic variant (rs532436;A>G) annotated to the Alpha 1-3-N-acetylgalactosaminyltransferase (*ABO*) gene to be positively associated with LDL, TC, CAD and T2D. The *ABO* gene has been linked to cholesterol absorption and cardiovascular disease [38]. Rs532436, located on chromosome 9, has been reported as *trans*-miR-eQTL for miR-10b-5p [22]. In this study, we further showed that a CpG site (cg25820279) annotated to Homeobox D3 (*HOXD3*), is located in the regulatory region of miR-10b and is associated with total cholesterol levels in serum. In addition, the expression level of miR-10b-5p in plasma showed a positive association with total cholesterol levels, which further support the crucial role of miR-10b-5p in lipid metabolism.

MiR-148a-3p has been shown to control the LDL uptake and cholesterol efflux through affecting the expression of low-density lipoprotein receptor (LDLR) [39]. Moreover, *in vivo* studies in mouse models have confirmed that miR-148a-3p is upregulated in adipogenesis and highly expressed in liver tissue [40]. We found rs4722551, located ~2kb upstream of miR-148a, associated with LDL, TC and TG. It has been suggested previously that a large part of regulatory elements such as promoter regions are located within +/-2kb of pre-miRNAs [17]. Rs4722551 has previously been reported to be positively associated with serum lipid levels via *cis*-miR-eQTL in liver tissue [41]. Our findings may shed light on the mechanism that associates the rs4722551 risk allele (T>C) with an increased miR-148a-3p expression, which is subsequently associated with higher serum cholesterol levels. Furthermore, our results showed a CpG site (cg18188200) in the regulatory region of miR-148a to be associated with LDL, TC

and TG and demonstrated that the plasma expression level of miR-148a-3p is also associated with total serum cholesterol levels. These data are in line with the findings from previous studies reporting a functional role for miR-148a-3p in lipid metabolism confirmed by various *in vivo* and *in vitro* validation experiments [39, 41].

We found strong associations of rs7117842, located ~512kb upstream of miR-100-5p/125b-5p, with TC, LDL and HDL, suggesting these two miRNAs to play a role in lipid metabolism. The SNP has been previously shown to be negatively associated with the expression levels of miR-100-5p and miR-125b-5p in blood [21]. In our analysis, plasma expression levels of miR-100-5p and miR-125b-5p are positively associated with TC. These findings could be interpreted in a way that carrying the risk allele of rs7117842 (T>C) is associated with decreased expression of miR-100-5p/125b-5p, which is associated with a reduced increase of total serum cholesterol levels. In addition, cg26363555, located in the promoter region of miR-125b-5p, was previously reported to act as miR-eQTM by changing the expression levels of both miR-100-5p and miR-125b-5p [13]. We found cg26363555 associated with HDL in the RS. In addition, cg03891346 annotated to MIR125B1 was reported to be associated with the expression level of miR-100-5p [13]. Our DNA methylation analysis results showed the association between cg03891346 and waist circumference in the RS. Our findings are partly in line with previous research investigating the role of miR-125b-5p on adipogenesis where it is observed that miR-125b-5p downregulates the anti-adipogenic gene *MMP11* in human, indicating that miR-125b-5p via *MMP11* positively regulate adipogenesis [42]. Conversely, the same study demonstrated a direct effect of reduction in fat accumulation through overexpression of miR-125b-5p [42]. In addition to the role of miR-125b-5p on lipid metabolism in human, its regulatory role has been investigated in other organisms including zebrafish and mice. Over-expression of miR-125b in zebrafish is linked to lipid metabolism in brain, heart and liver tissue [43]. This study observed that overexpression of miR-125b inhibits osteoblastic differentiation and promotes fat synthesis. Moreover, the expression of miR-125b is activated by estrogen via ERα *in vitro* and *in vivo* in mice, in which they demonstrated that miR-125b can limit fat accumulation in liver tissue [44]. These contradictory findings may implicate that miR-125b-5p plays an important role in lipid metabolism via a complex molecular cascade. However, the role of miR-100-5p in regard to lipid metabolism and cardiovascular disease yet to be further investigated. Since miR-100-5p and miR-125b-5p are located in the same locus on chromosome 11, it could be possible that miR-125b-5p is the driving miRNA in relation to the associated lipid traits. Future research is warranted to confirm the regulatory role of miR-100-5p in lipid metabolism.

The main strengths of this study include the use of robust data from the large-scale GWAS studies and multi-omics implementation of a large sample size in the Rotterdam Study, which indicates with more confidence that miRNAs are involved in the pathophysiology of cardiometabolic diseases. Our study, however, does not

come without limitations. First, our study design is based on associations rather than causations, therefore this approach does not prove that the identified miRNAs play a causal role in the studied traits. To test for causal inferences between miRNAs and disease risk, future studies should test mediating effects and incorporate functional follow-up experiments. Furthermore, our study design was based on a cross-sectional approach, which means that individuals included in this study were not free of CHD or T2D. In regard to test whether the identified miRNAs are associated with the risk of developing disease, future longitudinal studies are warranted. Another limited factor is that we were unable to link all identified miRNAs with epigenetic and expression analyses in the RS, since not all phenotypic data were available for each trait of interest nor were all miRNAs well-expressed in plasma. In addition, different sub cohorts of the RS were used for DNA methylation and miRNA expression analysis due to the availability of data. DNA methylation and miRNA signatures are dynamic over time and could have yield in confounding results. The challenge of this multi-omics approach includes the intra-individual variation and thereby lack of generalizability between datasets. However, the sub cohorts of RS-II and RS-III are extensions of the RS-I cohort. Previous epigenetic (DNA methylation) studies using the RS data showed that the results are replicated after additional adjustment for sub cohort [45-47]. This may indicate that the intra-individual differences between variables in these RS sub cohorts have not significantly affected by exposing to different environmental factors. Yet in an optimal setting one should apply the multi-omics analysis in the same individuals and the same timeframe. Furthermore, we used whole blood to determine DNA methylation and plasma to check expression levels of miRNAs, which are not the most relevant tissue for cardiometabolic traits. This could have resulted in overlooking some of the miRNAs, but the found associations are comparable because both analyses were performed in the same tissue. In an optimal setting one should examine the observed associations using next-generation sequencing covering all miRNAs in target tissues (e.g. adipose tissue, heart, pancreas and liver). Such infrastructure is not yet available in large epidemiologic studies with validated clinical data. However, for the use of miRNAs as targets for early diagnosis or progression of T2D and CHD, blood might be a very good test tissue since it is a non-invasive method for biomarker measurements in clinical diagnosis. In addition, regarding potential missed cardiometabolic-associated SNPs, our study could have benefited from denser genotyping methods including 1000 Genomes project or the Haplotype Reference Consortium (HRC).

Conclusion

In this study, we systematically examined the association of miRNAs with cardiometabolic risk factors and diseases using population-based genetic, DNA methylation and miRNA expression data. By integrating these omics data we found several cardiometabolic- associated miRNAs, such as miR-10b-5p, miR-148a-3p, miR-125b-5p and miR-100-5p involved in lipid metabolism, that can be viewed as potential biomarkers for early diagnosis or progression of T2D and CHD. Future experimental studies are warranted to elucidate pathways underlying the link between these miRNAs and cardiometabolic risk factors such as dyslipidemia, central adiposity and elevated blood glucose levels.

References

1. Kannel WB, McGee DL. Diabetes and cardiovascular disease. The Framingham study. JAMA 1979;241(19):2035-8.
2. Roth GA, Abate D, Abate KH, Abay SM, Abbafati C, Abbassi N. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018;392(10159):1736-1788.
3. Wilson PW, D'Agostino RB, Parise H, Sullivan L, Meigs JB. Metabolic syndrome as a precursor of cardiovascular disease and type 2 diabetes mellitus. Circulation 2005;112(20):3066-72.
4. Peloso GM, Auer PL, Bis JC, Voorman A, Morrison AC, Stitziel NO, et al. Association of low-frequency and rare coding-sequence variants with blood lipids and coronary heart disease in 56,000 whites and blacks. Am J Hum Genet 2014;94(2):223-32.
5. De Rosa S, Arcidiacono B, Chiefari E, Brunetti A, Indolfi C, Foti DP. Type 2 Diabetes Mellitus and Cardiovascular Disease: Genetic and Epigenetic Links. Front Endocrinol (Lausanne) 2018;9:2.
6. Friedman RC, Farh KK, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. Genome Res 2009;19(1):92-105.
7. Stauffer BL, Russell G, Nunley K, Miyamoto SD, Sucharov CC. miRNA expression in pediatric failing human heart. J Mol Cell Cardiol 2013;57:43-6.
8. Wang ZH, Sun XY, Li CL, Sun YM, Li J, Wang LF, et al. miRNA-21 Expression in the Serum of Elderly Patients with Acute Myocardial Infarction. Med Sci Monit 2017;23:5728-5734.
9. Rotllan N, Price N, Pati P, Goedeke L, Fernandez-Hernando C. microRNAs in lipoprotein metabolism and cardiometabolic disorders. Atherosclerosis 2016;246:352-60.
10. Grasedieck S, Scholer N, Bommer M, Niess JH, Tumani H, Rouhi A, et al. Impact of serum storage conditions on microRNA stability. Leukemia 2012;26(11):2414-6.
11. Gong J, Tong Y, Zhang HM, Wang K, Hu T, Shan G, et al. Genome-wide identification of SNPs in microRNA genes and the SNP effects on microRNA target binding and biogenesis. Hum Mutat 2012;33(1):254-63.
12. de Rie D, Abugessaisa I, Alam T, Arner E, Arner P, Ashoor H, et al. An integrated expression atlas of miRNAs and their promoters in human and mouse. Nat Biotechnol 2017;35(9):872-878.
13. Huan T, Mendelson M, Joehanes R, Yao C, Liu C, Song C, et al. Epigenome-wide association study of DNA methylation and microRNA expression highlights novel pathways for human complex traits. Epigenetics 2019:1-16.
14. Aure MR, Leivonen SK, Fleischer T, Zhu Q, Overgaard J, Alsner J, et al. Individual and combined effects of DNA methylation and copy number alterations on miRNA expression in breast tumors. Genome Biol 2013;14(11):R126.
15. Hijmans JG, Diehl KJ, Bammert TD, Kavlich PJ, Lincenberg GM, Greiner JJ, et al. Association between hypertension and circulating vascular-related microRNAs. J Hum Hypertens 2018;32(6):440-447.
16. de Candia P, Spinetti G, Specchia C, Sangalli E, La Sala L, Uccellatore A, et al. A unique plasma microRNA profile defines type 2 diabetes progression. PLoS One 2017;12(12):e0188980.
17. Saini HK, Griffiths-Jones S, Enright AJ. Genomic analysis of human microRNA transcripts. Proc Natl Acad Sci U S A 2007;104(45):17719-24.

18. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res 2014;42(Database issue):D68-73.

19. Nam JW, Kim J, Kim SK, Zhang BT. ProMiR II: a web server for the probabilistic prediction of clustered, nonclustered, conserved and nonconserved microRNAs. Nucleic Acids Res 2006;34(Web Server issue):W455-8.

20. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 2001;29(1):308-11.

21. Huan T, Rong J, Liu C, Zhang X, Tanriverdi K, Joehanes R, et al. Genome-wide identification of microRNA expression quantitative trait loci. Nat Commun 2015;6:6601.

22. Nikpay M, Beehler K, Valsesia A, Hager J, Harper ME, Dent R, et al. Genome-wide identification of circulating-miRNA expression quantitative trait loci reveals the role of several miRNAs in the regulation of Cardiometabolic phenotypes. Cardiovasc Res 2019.

23. Liu B, Shyr Y, Cai J, Liu Q. Interplay between miRNAs and host genes and their role in cancer. Brief Funct Genomics 2018;18(4):255-266.

24. Lutter D, Marr C, Krumsiek J, Lang EW, Theis FJ. Intronic microRNAs support their host genes by mediating synergistic and antagonistic regulatory effects. BMC Genomics 2010;11:224.

25. Hinske LC, França GS, Torres HA, Ohara DT, Lopes-Ramos CM, Heyn J, et al. miRIAD-integrating microRNA inter- and intragenic data. Database (Oxford) 2014;2014.

26. Ludwig N, Leidinger P, Becker K, Backes C, Fehlmann T, Pallasch C, et al. Distribution of miRNA expression across human tissues. Nucleic Acids Res 2016;44(8):3865-77.

27. Panwar B, Omenn GS, Guan Y. miRmine: a database of human miRNA expression profiles. Bioinformatics 2017;33(10):1554-1560.

28. Lorenz R, Bernhart SH, Honer Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, et al. ViennaRNA Package 2.0. Algorithms Mol Biol 2011;6:26.

29. Bonder MJ, Luijk R, Zhernakova DV, Moed M, Deelen P, Vermaat M, et al. Disease variants alter transcription factor levels and methylation of their binding sites. Nat Genet 2017;49(1):131-138.

30. Ikram MA, Brusselle GGO, Murad SD, van Duijn CM, Franco OH, Goedegebure A, et al. The Rotterdam Study: 2018 update on objectives, design and main results. Eur J Epidemiol 2017;32(9):807-850.

31. Touleimat N, Tost J. Complete pipeline for Infinium((R)) Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. Epigenomics 2012;4(3):325-41.

32. Sandoval J, Heyn H, Moran S, Serra-Musach J, Pujana MA, Bibikova M, et al. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. Epigenetics 2011;6(6): 692-702.

33. Chou CH, Shrestha S, Yang CD, Chang NW, Lin YL, Liao KW, et al. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. Nucleic Acids Res 2018;46(D1):D296-D302.

34. Ghanbari M, Sedaghat S, de Looper HW, Hofman A, Erkeland SJ, Franco OH, et al. 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(2):495-503.

35. Meerson A, Eliraz Y, Yehuda H, Knight B, Crundwell M, Ferguson D, et al. Obesity impacts the regulation of miR-10b and its targets in primary breast tumors. BMC Cancer 2019;19(1):86.

36. Wang D, Xia M, Yan X, Li D, Wang L, Xu Y, et al. Gut microbiota metabolism of anthocyanin promotes reverse cholesterol transport in mice via repressing miRNA-10b. Circ Res 2012;111(8):967-81.

37. Wang D, Wang W, Lin W, Yang W, Zhang P, Chen M, et al. Apoptotic cell induction of miR-10b in macrophages contributes to advanced atherosclerosis progression in ApoE-/- mice. Cardiovasc Res 2018;114(13):1794-1805.

38. Silbernagel G, Chapman MJ, Genser B, Kleber ME, Fauler G, Scharnagl H, et al. High intestinal cholesterol absorption is associated with cardiovascular disease and risk alleles in ABCG8 and ABO: evidence from the LURIC and YFS cohorts and from a meta-analysis. J Am Coll Cardiol 2013;62(4):291-9.

39. Goedeke L, Rotllan N, Canfran-Duque A, Aranda JF, Ramirez CM, Araldi E, et al. MicroRNA-148a regulates LDL receptor and ABCA1 expression to control circulating lipoprotein levels. Nat Med 2015;21(11):1280-9.

40. Gailhouste L, Gomez-Santos L, Hagiwara K, Hatada I, Kitagawa N, Kawaharada K, et al. miR-148a plays a pivotal role in the liver by promoting the hepatospecific phenotype and suppressing the invasiveness of transformed cells. Hepatology 2013;58(3):1153-65.

41. Wagschal A, Najafi-Shoushtari SH, Wang L, Goedeke L, Sinha S, deLemos AS, et al. Genome-wide identification of microRNAs regulating cholesterol and triglyceride homeostasis. Nat Med 2015;21(11):1290-7.

42. Rockstroh D, Loffler D, Kiess W, Landgraf K, Korner A. Regulation of human adipogenesis by miR125b-5p. Adipocyte 2016;5(3):283-97.

43. Wang X, Zheng Y, Ma Y, Du L, Chu F, Gu H, et al. Lipid metabolism disorder induced by up-regulation of miR-125b and miR-144 following beta-diketone antibiotic exposure to F0-zebrafish (Danio rerio). Ecotoxicol Environ Saf 2018;164:243-252.

44. Zhang ZC, Liu Y, Xiao LL, Li SF, Jiang JH, Zhao Y, et al. Upregulation of miR-125b by estrogen protects against non-alcoholic fatty liver in female mice. J Hepatol 2015;63(6):1466-75.

45. Braun KVE, Dhana K, de Vries PS, Voortman T, van Meurs JBJ, Uitterlinden AG, et al. Epigenome-wide association study (EWAS) on lipids: the Rotterdam Study. Clin Epigenetics 2017;9:15.

46. Ligthart S, Steenaard RV, Peters MJ, van Meurs JB, Sijbrands EJ, Uitterlinden AG, et al. Tobacco smoking is associated with DNA methylation of diabetes susceptibility genes. Diabetologia 2016;59(5):998-1006.

47. Nano J, Ghanbari M, Wang W, de Vries PS, Dhana K, Muka T, et al. Epigenome-Wide Association Study Identifies Methylation Sites Associated With Liver Enzymes and Hepatic Steatosis. Gastroenterology 2017;153(4):1096-1106 e2.

48. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature 2015;518(7538):197-206.

49. Shungin D, Winkler TW, Croteau-Chonka DC, Ferreira T, Locke AE, Magi R, et al. New genetic loci link adipose and insulin biology to body fat distribution. Nature 2015;518(7538):187-196.

50. 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(6):659-69.

51. 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(2):142-8.

52. 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(10):2624-34.

53. Wheeler E, Leong A, Liu CT, Hivert MF, Strawbridge RJ, Podmore C, et al. Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. PLoS Med 2017;14(9):e1002383.

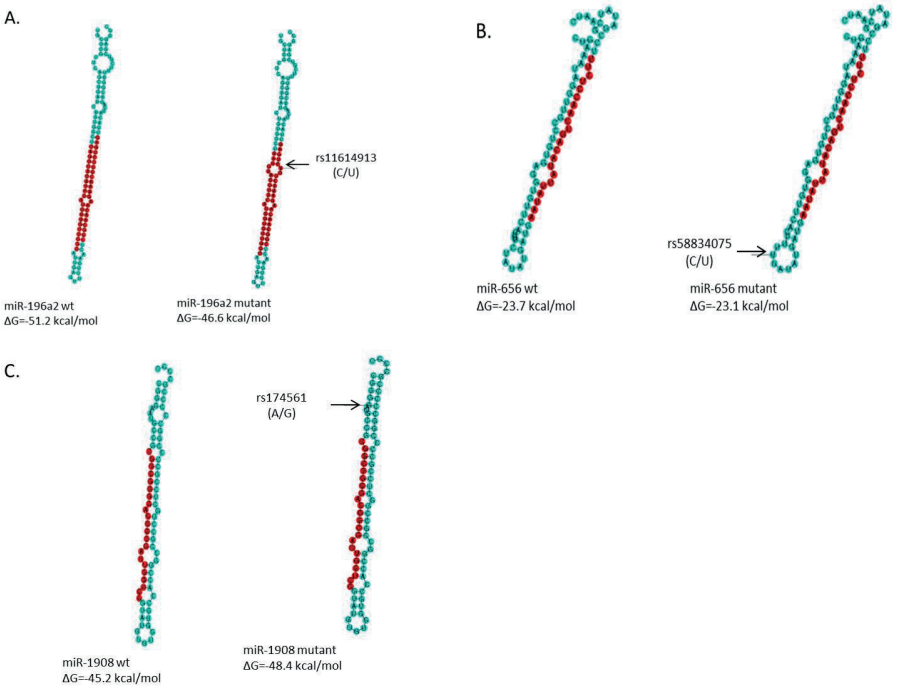
54. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet 2010;42(2):105-16.

55. Scott RA, Scott LJ, Magi R, Marullo L, Gaulton KJ, Kaakinen M, et al. An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. Diabetes 2017;66(11):2888-2902.

56. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, et al. Discovery and refinement of loci associated with lipid levels. Nat Genet 2013;45(11):1274-1283.

57. Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat Genet 2015;47(10):1121-1130.

58. 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(7367):103-9.



Supplementary Figure 6.1. Predicted secondary structure of miRNA wild type and variant. Location of the SNP is demonstrated by an arrow. The red part shows the mature sequence and the blue part shows the rest of the pre-miR. The corresponding minimum free energy (MFE) is illustrated with the thermodynamic ensemble ΔG. **A**, Secondary structure of miR- 196a2-3p wildtype and variant (rs11614913) located in mature miRNA sequence. MFE changes by -4.6kcal/mol. **B,** Secondary structure of miR-656 wildtype and variant (rs58834075) located in precursor miRNA sequence. MFE changes by -0.6kcal/mol. **C,** Secondary structure of miR-1908-5p wildtype and variant (rs174561) located in precursor miRNA sequence. MFE changes by +3.2kcal/mol.

Supplementary Table 6.2. Participant characteristics of the Rotterdam Study for DNA methylation analysis and miRNA expression profiling

	DNA methylation (RS-II-3 & RS-III-2)	DNA methylation (RS-III-1)	miRNA expression (RS-I-4 & RS-II-2)	P value*
N	717	721	1999	
Female	413 (57.6%)	391 (54.2%)	1141 (57.1%)	<0.001
Age (years)	67.5 (5.93)	59.8 (8.16)	71.6 (7.58)	<0.001
BMI (kg/m^2)	27.7 (4.12)	27.6 (4.63)	27.7 (4.11)	
Waist circumference	94.4 (12.00)	93.75 (12.92)	93.6 (11.98)	
WHR	0.9 (0.09)	0.9 (0.08)	0.9 (0.09)	
Current smoking (yes)	76 (10.6%)	193 (26.8%)	288 (14.4%)	
Triglycerides (mmol/L)	1.5 (0.79)	1.5 (0.88)	NA	
HDL-cholesterol (mmol/L)	1.5 (0.44)	1.4 (0.41)	1.4 (0.39)	<0.001
LDL-cholesterol (mmol/L)	3.7 (0.94)	3.9 (1.00)	NA	<0.001
Total cholesterol (mmol/L)	5.5 (1.02)	5.6 (1.07)	5.6 (0.99)	0.004
Lipid lowering medication (yes)	225 (31.4%)	191 (26.5%)	450 (22.5%)	<0.001
Systolic blood pressure	144.8 (21.91)	134.2 (19.76)	148.2 (20.82)	
Diastolic blood pressure	84.4 (11.66)	82.8 (11.38)	79.6 (10.84)	
Coronary heart disease	28 (3.9%)	46 (6.4%)	214 (10.7%)	<0.001
Anti-hypertensive medication (yes)	310 (43.2%)	217 (30.1%)	880 (44.0%)	<0.001
Glucose (mmol/L)	5.7 (1.11)	5.6 (1.04)	5.8 (1.09)	0.001
Insulin (pmol/L)	82.6 (48.26)	96.0 (63.04)	NA	<0.001
Prevalence type 2 diabetes	96 (13.4%)	74 (10.3%)	278 (13.9%)	0.04
Anti-diabetic medication	59 (8.2%)	39 (5.4%)	132 (6.6%)	

Values are presented as mean ±(SD) or N (%).

*Differences between groups were addressed using ANOVA in the case variables were available among three groups, Student's T-tests in the case variables were available in two groups.

NA: Not Available

Additional supplementary material for this chapter can be found online:
https://www.frontiersin.org/articles/10.3389/fgene.2020.00110/full#supplementary-material

7

CHAPTER 7

Circulatory microRNAs as potential biomarkers for stroke risk: the Rotterdam Study

Manuscript based on this chapter: Mens MMJ, Heshmatollah A, Fani L, Ikram MA, Ikram MK, Ghanbari M. Circulatory MicroRNAs as Potential Biomarkers for Stroke Risk: The Rotterdam Study. Stroke. 2021 Mar;52(3):945-953. doi: 10.1161/STROKEAHA.120.031543.

Abstract

Background: MicroRNAs (miRNAs) are post-transcriptionally regulators of gene expression that can be released extracellularly upon pathophysiological processes. By complementary binding of target transcripts, miRNAs can modulate the expression of an abundance of genes. Increasing evidence recognize miRNAs as promising biomarkers for complex traits, including cardiovascular disease and stroke. We conducted a longitudinal study to determine the association between circulatory miRNAs and incident stroke in a population-based setting.

Methods: Next-generation sequencing was used to measure expression levels of 2,083 miRNAs in plasma samples, collected between 2002 and 2005, from 1,914 stroke-free participants of the Rotterdam Study. Participants were assessed for incident stroke through continuous monitoring of medical records until January 1st 2016. Cox proportional hazards regression models adjusted for age, sex and vascular risk factors were used to investigate the association between the levels of 591 miRNAs well-expressed in plasma and incident stroke. Furthermore, stroke-subtype analysis was performed to assess the link between identified miRNAs and ischemic, hemorrhagic and unspecified stroke. Subsequently, post-hoc analyses were conducted to gain insight into the association between putative target genes of miRNAs and stroke.

Findings: Of 1,914 participants (mean age 71.5 years±7.6; 57.7% women), 138 were diagnosed with incident stroke during a mean follow-up of 9.7±3.2 years. After adjusting for potential confounders, we found plasma levels of three miRNAs to be associated with incident stroke (false discovery rate-adjusted P-value<0.05). These include miR-6124 (hazard ratio (HR) 1.66, 95%CI 1.31-2.09), miR-5196-5p (HR 1.90, 95%CI 1.39-2.61) and miR-4292 (HR 2.65, 95%CI 1.62-4.34). *In-silico* analysis of the miRNA putative target genes showed associations of variants in several target genes with stroke.

Interpretation: This study indicates that plasma levels of three miRNAs are associated with the risk of stroke, proposing them as potential biomarkers for early detecting of the disease.

Introduction

Although environmental and vascular risk factors are well-known to be involved in the pathology of stroke, genetic factors also contribute to the risk of developing ischemic and hemorrhagic stroke [1-5]. It has been previously shown that not only coding genes, in particular loci such as 12q24.2 and *ABO*, account for a proportion of the heritable risk, but also non-coding regions of the genome, such as microRNAs (miRNAs), can play a role in the pathophysiology of stroke [6]. A better understanding of the role of miRNAs in the development of stroke and their potential as biomarkers in the diagnosis of disease could lead to early detection of individuals with high risk profiles.

MiRNAs are small RNA molecules of approximately 22 nucleotides that regulate post-transcriptionally gene expression through complementary binding of target transcripts. They can potentially target hundreds-to thousands genes and are involved in various molecular pathways [7]. In addition to their ability to modulate cellular processes in both disease and non-disease states, miRNAs are released from cells into body fluids, such as plasma [8]. These circulatory miRNAs are very stable in plasma, due to their packaging into membranous vesicles including exosomes, and are thereby potentially clinically relevant as biomarkers for diseases [9]. Previous studies have linked changes in plasma miRNA levels with cardiovascular disease and stroke [10-16]. Most of the stroke-associated miRNAs have been identified cross-sectionally in patient cohorts with acute-stroke, leaving it uncertain whether the identified miRNAs are related to the risk of developing stroke. Prospective cohorts could overcome this limitation. Indeed, a recent longitudinal study with a modest sample size (n=51 cases) found the association between plasma levels of miR-656-3p and miR-941 and incident stroke with a mean follow-up time of 2.5±1.6 years [17]. However, longitudinal studies with a relatively long follow-up period are limited. In the current study, we determined the association between circulatory miRNAs and the risk of all stroke and its subtypes in a prospective-population based study with almost 10 years follow-up. Subsequently, we performed post-hoc analyses of the miRNA putative target genes to gain insight into the potential pathways by which the identified miRNAs may play a role in stroke.

Methods

Study population

This study was conducted within the Rotterdam Study, a large prospective population-based cohort among participants aged ≥45 years in the suburb Ommoord in Rotterdam, the Netherlands. In 1990, 7,983 inhabitants aged 55 years or older were recruited to participate in the first cohort of the Rotterdam Study (RS-I) (78% response rate of 10,215 invitees). In 2000, the Rotterdam Study was extended by 3,011 participants that moved to Ommoord or turned 55 years old (RS-II). A detailed description of the Rotterdam Study can be found elsewhere [18]. In the current study, miRNA expression profiling (n=2,000) was performed in a random subset (n=1,000) of the fourth visit of RS-I (RS-I-4) and a random subset (n=1,000) of the second visit of RS-II (RS-II-2). These visits of the Rotterdam Study were performed between 2002 and 2005, and with follow-up visits every 4-5 years. From 2,000 participants with miRNA data, we excluded participants with prevalent stroke (n=83) and individuals with no informed consent for follow-up (n=2). Furthermore, one participant was excluded because of missing data for all miRNAs. For more information regarding the inclusion criteria, see **Figure 7.1**.

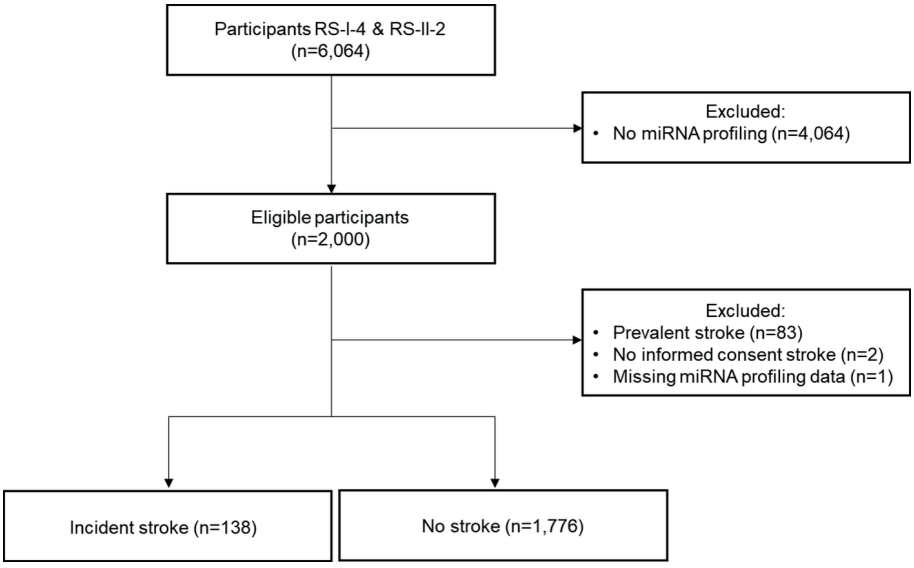


Figure 7.1. Flowchart of the study participants. RS, Rotterdam Study.

Informed Consent and Ethics Approval

The Rotterdam Study has been approved by the medical ethics committee at the Erasmus University of Rotterdam and the Ministry of Health, Welfare and Sport of the Netherlands. The study is implemented in the Population Studies Act: Rotterdam Study (Wet Bevolkingsonderzoek ERGO). All participants included in the current study provided written informed consent for participation and for researchers to access medical information from their personal physicians.

Measurements on circulatory miRNAs

Blood samples were collected in EDTA treated containers and centrifuged. Plasma was then aliquoted and frozen at -80°C according to standard procedures. Subsequently, plasma miRNA levels were determined using the HTG EdgeSeq miRNA Whole Transcriptome Assay, which measures the expression of 2,083 mature human miRNAs (HTG Molecular Diagnostics, Tuscon, AZ, USA) and using the Illumina NextSeq 500 sequencer (Illumina, San Diego, CA, USA). The Whole Transcriptome Assay characterizes miRNA expression patterns and measures the expression of 13 housekeeping genes, allowing flexibility in data normalization and analysis. Quantification of miRNA expression was based on counts per million (CPM). Log2 transformation of CPM was used as standardization and adjusted for total reads within each sample. The lower limit of quantification (LLOQ) was used to select well-expressed miRNAs. The LLOQ level was based on a monotonic decreasing spline curve fit between the means and standard deviations of all miRNAs. In our definition well-expressed miRNA levels in plasma were those with >50% values above LLOQ. This includes a set of 591 miRNAs that were used for association analysis.

Assessment of stroke

Stroke was defined according to the World Health Organization definition as a syndrome of rapidly developing clinical signs of focal or global disturbance of cerebral function, with symptoms lasting 24 hours or longer or leading to death, with no apparent cause other than of vascular origin [19, 20]. We assessed the prevalence of stroke at baseline during interview and verified it using medical records. After enrollment, participants were continuously monitored for incident stroke by linking the study database to files of general practitioners. Records of nursing homes and records from general practitioners of the participants who moved out of the study district were also checked. Additional information, such as clinical notes and neuroimaging reports, were obtained from hospital records.

Stroke patients were further classified as ischemic or hemorrhagic based on neuroimaging reports. If neuroimaging was lacking, a stroke was classified as unspecified. This classification corresponds with ICD-10 codes I61, I63 and I64. Potential stroke patients were reviewed by research physicians and verified in a consensus panel led by an experienced stroke neurologist.

The follow-up for incident stroke was conducted until January 1st, 2016. Participants were followed from study entry until stroke, death, last health status update when they were known to be stroke-free, or January 1st, 2016, whichever came first. Follow-up was complete for 92.9% of potential person-years.

Assessment on covariates

Information on age, sex and smoking status (current/former/never) was obtained from questionnaires. Body-mass index (BMI) was calculated based on weight in kilograms divided by the height in meters squared. Blood samples of participants were obtained during the visit to the research center. Using automatic enzymatic method, levels of high-density lipoprotein (HDL) and total cholesterol were measured in serum (mmol/L). Systolic and diastolic blood pressure were measured (mm Hg) in a seating position after a 5 minute rest period on the right upper arm of the participant using a random-zero sphygmomanometer. Prevalent diabetes mellitus type 2 was identified according to the World Health Organization criteria: fasting glucose levels of ≥7.0 mmol/L, non-fasting glucose levels ≥11.1 mmol/L, or the use of glucose lowering medication. Prevalent coronary heart disease was defined when the participant suffered a myocardial infarction or underwent a coronary artery bypass grafting or percutaneous coronary revascularization procedure. Prevalent atrial fibrillation was assessed using a 10 seconds 12-lead electrocardiogram and medical records. Information regarding the use of lipid lowering medication, blood pressure lowering medication and glucose lowering medication was obtained from pharmacy records and home interviews. All covariates included for statistical analysis were obtained at baseline examination.

Statistical analysis

We used Cox proportional hazards regression to determine hazard ratios (HR) with 95% confidence intervals for the association between miRNA expression and stroke risk. Hazard ratios were reported per each additional unit of log2 CPM of miRNA expression. Models were adjusted for age, sex and cohort. Additionally, we adjusted for smoking, BMI, HDL, total cholesterol, systolic blood pressure, diastolic blood pressure, prevalent diabetes mellitus type 2, prevalent coronary heart disease, prevalent atrial fibrillation, lipid lowering medication and blood pressure lowering medication. To reduce the false positive results, the Benjamini-Hochberg (BH) procedure was used to adjust the p-values into false discovery rate (FDR) [21]. In this study, a FDR-corrected

p-value <0.05 (5%) was set as significance threshold. Subsequently, we analyzed the association between the expression of the identified miRNAs and risk of ischemic, hemorrhagic and unspecified stroke separately. To avoid the risk of overfitting, we adjusted models for the association between miRNA expression and hemorrhagic and unspecified stroke only for age, sex and cohort. To reduce the possible bias induced by missing values, multiple imputation on confounders was performed based on outcome and included covariates for predictors of missing data. Values were imputed with a maximum iteration number of 10 (n=25 imputations) using the Markov Chain Monte Carlo method, R package "*mice*". Areas under the curve (AUC) for time-dependent receiver operating characteristics (ROC) at 10-years were calculated for the vascular risk factor model (covariates included in model 2) and the vascular risk factor model complemented by the identified miRNAs. The AUC were calculated using the Kaplan-Meier method, R package "*survivalROC*". Furthermore, the identified miRNAs were categorized in 'high- or low expressed' based on their median expression values and cumulative hazard graphs were generated. Additionally, we performed a permutation test to check the enrichment of the identified miRNAs in the association with stroke compared to sets of randomly selected miRNAs. Analyses were performed using R version 3.6.1 (The R foundation for Statistical Computing, Vienna, Austria).

Post-hoc genetic analyses

To test whether the stroke-associated miRNAs are potentially involved in the pathways underlying disease, we performed *in silico* analysis on their putative target genes. Three commonly used miRNA target prediction databases, TargetScan (v7.2) [22], miRTarBase [23] and miRDB [24], were used to retrieve putative target genes of the stroke-associated miRNAs. Then, we extracted single-nucleotide polymorphisms (SNPs) in these target genes. Moreover, we extracted SNPs located in +/-2kb of the precursor sequences of the associated miRNAs [25]. Subsequently, we performed a look-up to test whether the SNPs located in target genes and in +/-2kb of the miRNA precursor sequences are associated with stroke by using the summary statistics of a previous genome-wide association study (GWAS) on stroke [3]. Target genes were considered significantly associated with stroke based on FDR<0.1 (10%).

Furthermore, we sought to explore whether the identified miRNAs are expressed in the brain, a relevant tissue for stroke. We also retrieved the miRNA host genes as proxy for the identified miRNAs to check their expression in the brain using the Human Protein Atlas (https://www.proteinatlas.org/) [26]. The rationale for this analysis is that the genomic location of miRNAs can be discriminated among intergenic and intragenic. The intragenic miRNAs and their host genes may share the same promoter, and these miRNA are likely to be co-expressed with their host genes [27]. The genomic location of the identified miRNAs was obtained using miRIAD [28].

Results

The baseline characteristics of the 1,914 Rotterdam Study participants used for this study are presented in **Table 7.1**. The individuals that were diagnosed with incident stroke were on average slightly older compared to non-cases (74.6 years vs. 71.2 years).

Table 7.1. Baseline characteristics of the Rotterdam Study participants of this study

Characteristic	Incident stroke N=138	No incident stroke N=1776	P value
Age, years	74.6 ± 7.5	71.2 ± 7.5	<0.001
Female, n(%)	76 (55.1%)	1029 (57.9%)	0.51
Body mass index, kg/m2	27.6 ± 3.4	27.6 ± 4.2	0.87
Systolic blood pressure, mmHg	156.5 ± 23.0	147.5 ± 20.6	<0.001
Diastolic blood pressure, mmHg	80.6 ± 12.2	79.4 ± 10.8	0.25
Total serum cholesterol, mmol/L	5.5 ± 1.1	5.7 ± 1.0	0.13
High-density lipoprotein, mmol/L	1.4 ± 0.4	1.5 ± 0.4	0.5
C-reactive protein, mg/L	3.9 ± 6.5	3.0 ± 6.3	0.11
Smoking			0.35
current, n(%)	20 (14.5%)	260 (14.6%)	
former, n(%)	83 (60.1%)	966 (54.4%)	
never, n(%)	35 (25.4%)	550 (31.0%)	
Prevalent diabetes mellitus type 2, n(%)	26 (18.8%)	217 (12.2%)	<0.05
Prevalent coronary heart disease, n(%)	22 (15.9%)	172 (9.7%)	<0.05
Prevalent atrial fibrillation, n(%)	8 (5.8%)	79 (4.4%)	0.46
Prevalent heart failure, n(%)	13 (9.4%)	77 (4.3%)	<0.01
Chronic kidney disease, n(%)	3 (2.2%)	22 (1.2%)	0.34
Any Internal carotid artery plaque, n(%)	89 (64.5%)	913 (51.4%)	<0.01
Any Internal carotid artery stenosis, n(%)	16 (11.6%)	89 (5.0%)	<0.01
Blood pressure lowering medication, n(%)	77 (55.8%)	737 (41.5%)	<0.01
Lipid lowering medication, n(%)	29 (21.0%)	385 (21.7%)	0.86
Anti-coagulant medication, n(%)	50 (36.3%)	348 (19.6%)	<0.001
Anti-platelet medication, n(%)	7 (5.1%)	93 (5.2%)	0.93

Variables are presented as mean ± SD, or number (%). Missing values were imputed. Number of missing values for final study population: 8 (0.4%) for systolic blood pressure, 8 (0.4%) for diastolic blood pressure, 12 (0.6%) for blood pressure lowering medication, 12 (0.6%) for lipid lowering medication, 26 (1.4%) for coronary heart disease, 28 (1.5%) for smoking, 37 (1.9%) for body-mass index and 75 (3.9%) for diabetes mellitus type 2. Differences between incident stroke cases and non-cases were examined by T-test for continuous variables and Chi-square test for categorical variables.

During a mean follow-up of 9.7 years (± 3.2), 138 individuals were diagnosed with incident stroke. Of these, 96 were ischemic, 19 were hemorrhagic and 23 cases were unspecified. The results of the association between miRNA levels in plasma and the incidence of stroke are presented in **Table 7.2**. We found significant associations between the expression levels of three miRNAs and risk of stroke, miR-6124 (HR 1.66, 95%CI 1.31-2.09, P-value= 2.00×10^{-5}), miR-5196-5p (HR 1.90, 95%CI 1.39-2.61, P-value= 6.07×10^{-5}), and miR-4292 (HR 2.65, 95%CI 1.62-4.34, P-value= 1.09×10^{-4}). In total, 39 miRNAs were at least nominally associated (P-value<0.05) with incident stroke (**Supplemental Table 7.1**). The distribution of plasma expression levels of the three identified miRNAs in the study population are illustrated in **Figure 7.2**. The AUC for the time-dependent ROC at 10 years was 0.735 for the vascular risk factor model and 0.761 for the model complemented by miR-6124, miR-5196-5p and miR-4292 (**Figure 7.3**). Analyses by stroke subtype showed associations of miR-6124, miR-5196-5p and miR-4292 with both ischemic and hemorrhagic stroke. However, the number of hemorrhagic cases was small (n=19), which resulted in a wide 95% confidence interval.

Table 7.2. Association between miRNA expression and stroke incidence

	Any stroke (n=138)				Ischemic stroke (n=96)				Hemorrhagic stroke (n=19)		Unspecified stroke (n=23)	
	Model 1		Model 2		Model 1		Model 2		Model 1		Model 1	
	HR (95% CI)	P value	HR (95% CI)	P value	HR (95% CI)	P value	HR (95% CI)	P value	HR (95% CI)	P value	HR (95% CI)	P value
miR-6124	1.68 (1.35-2.09)	**2.49×10^{-6}**	1.66 (1.31-2.09)	**2.00×10^{-5}**	1.62 (1.22-2.15)	8.82×10^{-4}	1.61 (1.17-2.20)	3.03×10^{-3}	2.50 (1.70-3.68)	3.61×10^{-6}	1.38 (0.71-2.68)	3.49×10^{-1}
miR-5196-5p	1.94 (1.44-2.63)	**1.71×10^{-5}**	1.90 (1.39-2.61)	**6.07×10^{-5}**	1.85 (1.23-2.78)	2.95×10^{-3}	1.88 (1.21-2.92)	4.98×10^{-3}	3.13 (1.88-5.21)	1.15×10^{-5}	1.38 (0.48-3.91)	5.48×10^{-1}
miR-4292	2.68 (1.63-4.42)	**1.03×10^{-4}**	2.65 (1.62-4.34)	**1.09×10^{-4}**	2.61 (1.38-4.97)	3.34×10^{-3}	2.79 (1.43-5.45)	2.64×10^{-3}	4.87 (1.66-14.33)	4.03×10^{-3}	2.31 (0.57-9.33)	2.39×10^{-1}

[1] Model 1: adjusted for age, sex and cohort

[2] Model 2: Model 1 + smoking, BMI, HDL, total cholesterol, systolic blood pressure, diastolic blood pressure, prevalent diabetes mellitus type 2, prevalent coronary heart disease, prevalent atrial fibrillation, lipid lowering medication and blood pressure lowering medication

In bold: miRNAs survived correction for multiplicity (FDR<0.05)

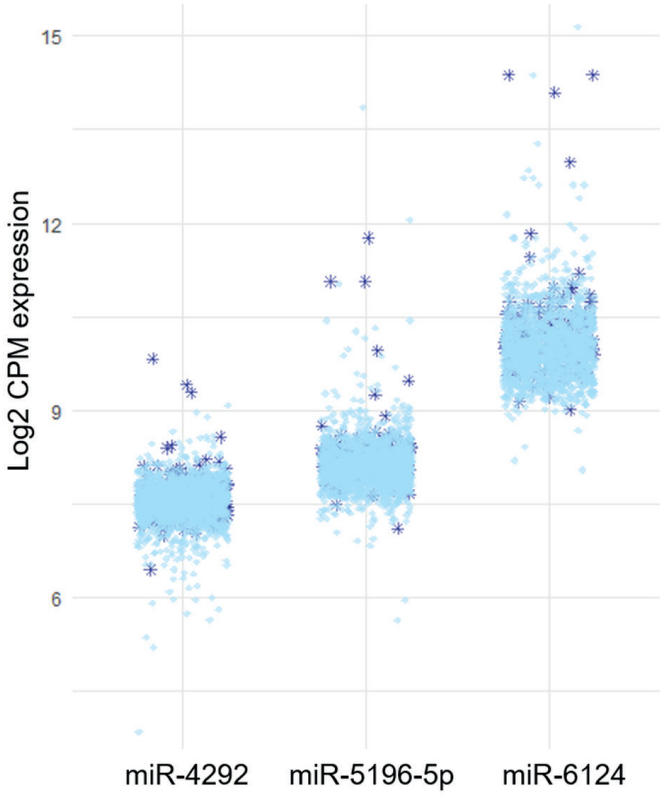


Figure 7.2. Scatter plots of the distribution of log2 expression values of the three identified miRNAs in the study participants. Stars indicate cases (incident stroke) and dots indicate non-cases.

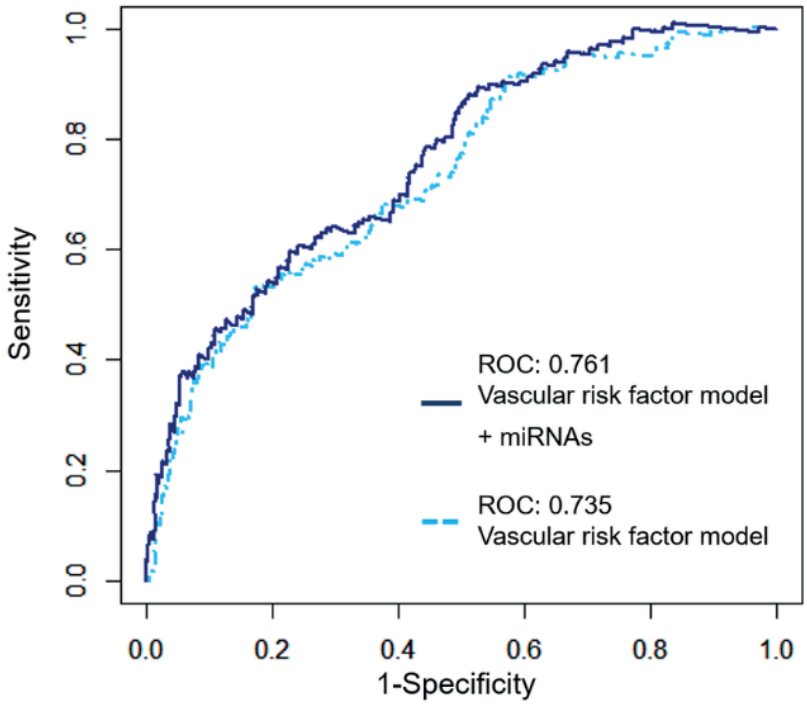


Figure 7.3. Time-dependent ROC curve analysis at 10 years. The lower line indicate a ROC curve of the vascular risk factor model with an AUC of 0.735. The upper line indicate a ROC of the vascular risk factor model combined with the plasma values of miR-6214, miR-5196-5p and miR-4292, with an AUC of 0.761.

To assess the impact of relative expression of the identified miRNAs on 10-year stroke risk, we categorized the expression values based on the median value of each miRNA into two categories, including low and high. See **Supplemental Table 7.2** for the baseline characteristics split by the median expression values of the three miRNAs. We found that high expression levels of miR-6124 (median normalized value >10.01) and miR-5196-5p (median normalized value >8.09) were also significantly associated with higher cumulative hazard of stroke (**Figure 7.4**). A combination of the high values of miR-6124, miR-5196-5p and miR-4292 is represented in 23.7% of the study participants and was associated with a higher cumulative hazard for the risk of stroke (P-value= 4.20×10^{-4}).

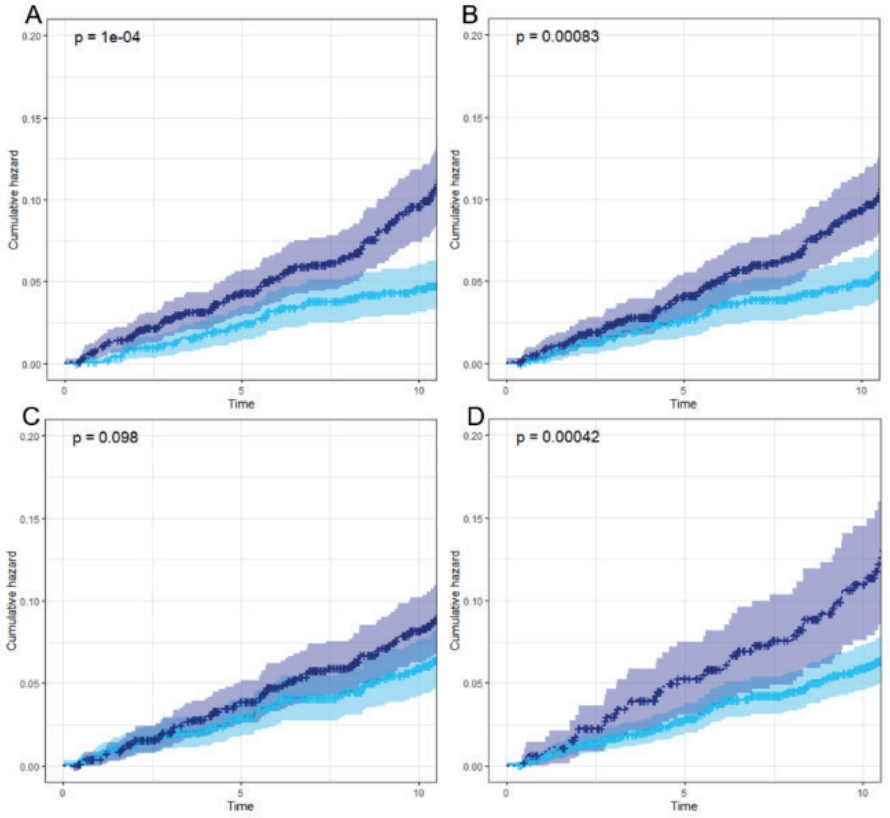


Figure 7.4. Cumulative hazard curves of log2 CPM expression of the three identified miRNAs and the risk of incident stroke. The upper curves indicate a high expression level and the lower curves indicate a low expression level based on the log2 median values of individual miRNAs. P value indicates difference between the two curves. Y-axis indicates cumulative hazard. X-axis indicates time scale in years. **A.** Cumulative hazard on incident stroke between high and low expression levels of miR-6124. **B.** Cumulative hazard on incident stroke between high and low expression levels of miR-5196-5p. **C.** Cumulative hazard on incident stroke between high and low expression levels of miR-4292. **D.** Cumulative hazard on incident stroke between high and low expression levels of miR-6124, miR-5196-5p and miR-4292.

Next, we retrieved the predicted target genes of miR-6124, miR-5196-5p and miR-4929 from the online miRNA target prediction databases. We focused on the target genes overlapping in at least two out of the three databases. This resulted in 1,633 target genes for miR-6124, 821 target genes for miR-5196-5p and 384 target genes for miR-4292. We extracted SNPs in these target genes and tested their associations with risk of stroke using summary statistic from a recent GWAS on stroke [3]. Based on FDR<0.1, we found variants in 10 unique target genes to be associated with stroke (**Supplemental Table 7.3**). Among these, nine were putative target genes of miR-6124 (*CASZ1*, *COL15A1*, *HDAC9*, *NRP2*, *RERE*, *SH3PXD2A*, *SLC4A8*, *STXBP5*, *ZFHX3*), one was target gene of miR-5196-5p (*SH3PXD2A*), and three were target genes of miR-4292 (*CASZ1*, *FURIN*, *STXBP5*). Furthermore, we found that rs79684932 (T>C), located on chromosome 19, ~15.9kb upstream of the transcription start site of miR-5196 is nominally associated with stroke in the GWAS data (P-value= 1.20×10^{-2}, Beta= -0.065) [3]. This SNP was also reported to be associated with the expression of *CD22*, the host gene of miR-5196, in whole blood [29].

Finally, we sought to explore whether the three identified miRNAs are expressed in the brain. We did not find any reports regarding the brain expression levels of the identified miRNAs in literature. Alternatively, we explored the expression in brain of the miRNA host genes as proxy for the identified miRNAs. Two of the three identified miRNAs (miR-6124 and miR-4292) were reported to be embedded within an intron of the protein-coding genes (*MICAL2* and *RABL6*) in a sense orientation [28]. Moreover, miR-5196-5p is located in the exonic region of *CD22*. All three host genes are expressed in brain tissue, of which *MICAL2* has been reported to be expressed at relatively high levels in the brain according to the Human Protein Atlas (https://www.proteinatlas.org/) [26]. Out of the 10 stroke-associated target genes, expression levels of 8 genes were detected in the brain (*HDAC9*, *FURIN*, *NRP2*, *RERE*, *SH3PXD2A*, *SLC4A8*, *STXBP5*, *ZFHX3*), of which *SLC4A8* has been reported to be expressed at relatively high levels in the brain [26].

Discussion

In this longitudinal study, we found that higher plasma levels of three miRNAs (miR-6124, miR-5196-5p and miR-4292) were significantly associated with stroke risk at the population level, suggesting them as potential plasma biomarkers for the disease. In addition, we identified several putative target genes of these miRNAs to be associated with stroke using GWAS data. These observations may suggest that the identified miRNAs are also involved in the pathophysiology of stroke that warrant further investigations.

The findings of this study that circulatory miRNAs are associated with incident stroke is consistent with a previous longitudinal study [17]. In their longitudinal study on ~2,700 participants from the Framingham Heart Study (FHS), Mick *et al.* found plasma levels of miR-656-3p and miR-941 to be associated with incident stroke (n=51 cases) [17]. This may confirm that miRNAs are relevant as biomarkers for stroke risk in the general population. But the major differences between Mick *et al.* and the current study include the methodology and data normalization. Therefore, the number of available miRNAs measured by qPCR in the FHS (301 miRNAs) and next-generation sequencing in the current study (591 miRNAs) do not have to overlap completely between studies. Notably, the expression levels of miR-6124, miR-5196-5p and miR-4292 were not included in the study of Mick *et al.*, and similarly, miR-656-3p and miR-941 were not among the 591 well-expressed miRNAs included in the current study. An independent study from the FHS observed no significant differences of 257 blood-derived miRNA levels between incident stroke cases (n=80) and non-stroke participants [30]. In a similar study Salinas *et al.* linked the levels of miR-574-3p to prevalent stroke [30]. However, the follow-up time of previous studies was considerably shorter than the current study (3.2 and 2.5 years vs. 9.7 years). Given the relatively small effect of a single miRNA, changes in expression levels are likely to be more impactful over a longer period of time. This may explain why Salinas *et al.* found no association between miRNA expression levels and incident stroke [30]. In addition, a different study population is likely to reflect a different miRNA abundance. The FHS participants were 5.2 years younger than the participants of the RS, and age is shown to be highly associated with both miRNA expression and stroke risk [1, 31].

A notable observation of previous studies regards the difference in miRNA detection between different types of extracellular biofluids such as, serum, plasma, urine and saliva, but also between extracellular and cellular fluids, like whole blood [8, 32]. Most previous studies have measured miRNA expression in whole blood [11, 15, 30]. Although, blood is an easy accessible tissue, blood biomarkers are more likely to reflect blood-based features than phenomena caused by disease. [33]. In particular, cell-derived vesicles, known as exosomes release miRNAs into extracellular fluids, which can be linked to different pathological conditions. We measured cell free miRNA levels in plasma, which is therefore unbiased compared to miRNAs levels in whole blood.

MiRNAs can affect the expression of protein-coding genes by complementary binding of their target sequences. Alteration in expression of stroke-associated genes can influence the disease pathophysiology. We found that the three identified miRNAs can potentially target 10 genes that are associated with stroke using the GWAS data [3]. Among these, 8 genes are expressed in the brain [26]. Furthermore, the three identified-miRNAs are located in intergenic regions of coding genes and are therefore likely to be expressed together with their host genes in the brain [27]. For instance, miR-6124 is located in the intronic region of *MICAL2*, which has been reported to be expressed at relatively high levels in brain tissue [26]. To our knowledge, *MICAL2* has not been previously linked to stroke in human. However, Hou *et al.* demonstrated that Mical2 is involved in the endothelial and vascular mechanisms that are disturbed with an ischemic stroke in mice [34]. In addition, according Malik *et al.* some of the identified putative target genes are related to vascular traits, such as blood pressure (*CASZ1, FURIN*), coronary artery disease (*HDAC9*) and atrial fibrillation (*ZFHX3*) [3]. The latter gene, *ZFHX3*, is associated with cardioembolic stroke, which accounts for a significant proportion of ischemic strokes [3]. However, it is beyond the scope of this study to elaborate on the pathophysiologic implications of putative target genes. Future studies are needed to experimentally confirm the regulatory interaction between the identified miRNAs and target genes and their relation to stroke subtypes.

Furthermore, we observed differences in the association between the median expression levels of the three identified miRNAs and various stroke risk factors (**Supplemental Table 7.2**). While miR-6124 and miR-5196-5p are associated with nearly all stroke risk factors, miR-4292 shows only a nominal association with HDL and BMI. This may indicate that the two former miRNAs play a role in the causal pathways of stroke. MiR-4292, unlike miR-6124 and miR-5196-5p, might be involved differently in stroke, but is still useful as a predictive biomarker of the disease. Future studies are needed to identify the molecular mechanisms by which these miRNAs promote stroke.

Our study has limitations that need to be considered for interpretation of the results. The pathology of ischemic and hemorrhagic stroke is different, and for risk stratification purposes, it is important to know whether a biomarker is specific for a subtype. With our study design, we were able to elaborate on the associations between some miRNAs and stroke subtypes. However, the number of cases of hemorrhagic stroke was small (n=19) and with a wide 95% confidence interval; future studies with larger samples are needed to provide more certainty. In addition, we used the brain-tissue expression levels of the host genes as proxy for the identified miRNAs. Since most tissue-specific expression levels are obtained via qPCR methods, no information was available on the RNA sequencing-measured expression levels of the identified miRNAs in brain tissue. Strengths of our study include the prospective design, almost 10 years of follow-up and the ability to measure virtually all known miRNAs to date with a highly specific and sensitive method. The ideal biomarker is easy assessable, non-invasive, cost-

effective, disease-specific and can be detected before the onset of the disease. To this end, plasma-derived miRNAs are favorable for biomarker discovery and unbiased compared to blood-derived miRNAs that might reflect blood-specific features.

Summary

This study indicates that higher plasma levels of miR-6124, miR-5196-5p and miR-4292 are associated with the risk of stroke in the general population. The elevated levels of these miRNAs may reflect the risk of stroke and therefore could serve as plasma biomarkers for early diagnosis of the disease. Future studies with even more samples and experimental validation are warranted to replicate and verify the potential of the identified miRNAs as biomarker of stroke and their roles in the onset of disease.

References

1. Virani SS, Alonso A, Benjamin EJ, Bittencourt MS, Callaway CW, Carson AP, Chamberlain AM, Chang AR, Cheng S, Delling FN, et al. Heart disease and stroke statistics-2020 update: A report from the american heart association. *Circulation*. 2020;141:e139-e596
2. Network NSG, International Stroke Genetics C. Loci associated with ischaemic stroke and its subtypes (sign): A genome-wide association study. *Lancet Neurol*. 2016;15:174-184
3. Malik R, Chauhan G, Traylor M, Sargurupremraj M, Okada Y, Mishra A, Rutten-Jacobs L, Giese AK, van der Laan SW, Gretarsdottir S, et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet*. 2018;50:524-537
4. Malik R, Traylor M, Pulit SL, Bevan S, Hopewell JC, Holliday EG, Zhao W, Abrantes P, Amouyel P, Attia JR, et al. Low-frequency and common genetic variation in ischemic stroke: The metastroke collaboration. *Neurology*. 2016;86:1217-1226
5. Kilarski LL, Achterberg S, Devan WJ, Traylor M, Malik R, Lindgren A, Pare G, Sharma P, Slowik A, Thijs V, et al. Meta-analysis in more than 17,900 cases of ischemic stroke reveals a novel association at 12q24.12. *Neurology*. 2014;83:678-685
6. Tiedt S, Dichgans M. Role of non-coding rnas in stroke. *Stroke*. 2018;49:3098-3106
7. Friedman RC, Farh KK, Burge CB, Bartel DP. Most mammalian mrnas are conserved targets of microrna. *Genome Res*. 2009;19:92-105
8. Weber JA, Baxter DH, Zhang S, Huang DY, Huang KH, Lee MJ, Galas DJ, Wang K. The microrna spectrum in 12 body fluids. *Clin Chem*. 2010;56:1733-1741
9. Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, Peterson A, Noteboom J, O'Briant KC, Allen A, et al. Circulating microrna as stable blood-based markers for cancer detection. *Proc Natl Acad Sci U S A*. 2008;105:10513-10518
10. Li P, Teng F, Gao F, Zhang M, Wu J, Zhang C. Identification of circulating microrna as potential biomarkers for detecting acute ischemic stroke. *Cell Mol Neurobiol*. 2015;35:433-447
11. Sepramaniam S, Tan JR, Tan KS, DeSilva DA, Tavintharan S, Woon FP, Wang CW, Yong FL, Karolina DS, Kaur P, et al. Circulating microrna as biomarkers of acute stroke. *Int J Mol Sci*. 2014;15:1418-1432
12. 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
13. Sonoda T, Matsuzaki J, Yamamoto Y, Sakurai T, Aoki Y, Takizawa S, Niida S, Ochiya T. Serum microrna-based risk prediction for stroke. *Stroke*. 2019;50:1510-1518
14. Tiedt S, Prestel M, Malik R, Schieferdecker N, Duering M, Kautzky V, Stoycheva I, Bock J, Northoff BH, Klein M, et al. Rna-seq identifies circulating mir-125a-5p, mir-125b-5p, and mir-143-3p as potential biomarkers for acute ischemic stroke. *Circ Res*. 2017;121:970-980
15. Huang S, Lv Z, Guo Y, Li L, Zhang Y, Zhou L, Yang B, Wu S, Zhang Y, Xie C, et al. Identification of blood let-7e-5p as a biomarker for ischemic stroke. *PLoS One*. 2016;11:e0163951
16. Tan KS, Armugam A, Sepramaniam S, Lim KY, Setyowati KD, Wang CW, Jeyaseelan K. Expression profile of microrna in young stroke patients. *PLoS One*. 2009;4:e7689
17. Shah R, Murthy V, Pacold M, Danielson K, Tanriverdi K, Larson MG, Hanspers K, Pico A, Mick E, Reis J et al. Extracellular rnas are associated with insulin resistance and metabolic phenotypes. *Diabetes Care*. 2017;40:546-553
18. Ikram MA, Brusselle G, Ghanbari M, Goedegebure A, Ikram MK, Kavousi M, Kieboom BCT, Klaver CCW, de Knegt RJ, Luik AI, et al. Objectives, design and main findings until 2020 from the

rotterdam study. *European Journal of Epidemiology*. 2020

19. Wieberdink RG, Ikram MA, Hofman A, Koudstaal PJ, Breteler MM. Trends in stroke incidence rates and stroke risk factors in rotterdam, the netherlands from 1990 to 2008. *Eur J Epidemiol*. 2012;27:287-295
20. Sacco RL, Kasner SE, Broderick JP, Caplan LR, Connors JJ, Culebras A, Elkind MS, George MG, Hamdan AD, Higashida RT, et al. An updated definition of stroke for the 21st century: A statement for healthcare professionals from the american heart association/american stroke association. *Stroke*. 2013;44:2064-2089
21. Benjamini Y, Hochberg Y. Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J R Stat Soc B*. 1995;57:289-300
22. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microrna target sites in mammalian mrnas. *Elife*. 2015;4
23. Huang HY, Lin YC, Li J, Huang KY, Shrestha S, Hong HC, Tang Y, Chen YG, Jin CN, Yu Y, et al. Mirtarbase 2020: Updates to the experimentally validated microrna-target interaction database. *Nucleic Acids Res*. 2020;48:D148-D154
24. Chen Y, Wang X. Mirdb: An online database for prediction of functional microrna targets. *Nucleic Acids Res*. 2020;48:D127-D131
25. Mens MMJ, Maas SCE, Klap J, Weverling GJ, Klatser P, Brakenhoff JPJ, van Meurs JBJ, Uitterlinden AG, Ikram MA, Kavousi M, et al. Multi-omics analysis reveals micrornas associated with cardiometabolic traits. *Front Genet*. 2020;11:110
26. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A, Kampf C, Sjostedt E, Asplund A, et al. Tissue-based map of the human proteome. *Science*. 2015;347:1260419
27. Lutter D, Marr C, Krumsiek J, Lang EW, Theis FJ. Intronic micrornas support their host genes by mediating synergistic and antagonistic regulatory effects. *BMC Genomics*. 2010;11:224
28. Hinske LC, Franca GS, Torres HA, Ohara DT, Lopes-Ramos CM, Heyn J, Reis LF, Ohno-Machado L, Kreth S, Galante PA. Miriad-integrating microrna inter- and intragenic data. *Database (Oxford)*. 2014;2014
29. Võsa U, Claringbould A, Weststra HJ, Bonder MJ, Deelen P, Zeng B, Kirsten H, Saha A, Kreuzhuber R, Kasela S, et al. Unraveling the polygenic architecture of complex traits using blood eqtl meta-analysis. bioRxiv 447367, 2018
30. Salinas J, Lin H, Aparico HJ, Huan T, Liu C, Rong J, Beiser A, Himali JJ, Freedman JE, Larson MG, et al. Whole blood microrna expression associated with stroke: Results from the framingham heart study. *PLoS One*. 2019;14:e0219261
31. Huan T, Chen G, Liu C, Bhattacharya A, Rong J, Chen BH, Seshadri S, Tanriverdi K, Freedman JE, Larson MG, et al. Age-associated microrna expression in human peripheral blood is associated with all-cause mortality and age-related traits. *Aging Cell*. 2018;17
32. Shah R, Tanriverdi K, Levy D, Larson M, Gerstein M, Mick E, Rozowsky J, Kitchen R, Murthy V, Mikalev E, et al. Discordant expression of circulating microrna from cellular and extracellular sources. *PLoS One*. 2016;11:e0153691
33. Pritchard CC, Kroh E, Wood B, Arroyo JD, Dougherty KJ, Miyaji MM, Tait JF, Tewari M. Blood cell origin of circulating micrornas: A cautionary note for cancer biomarker studies. *Cancer Prev Res (Phila)*. 2012;5:492-497
34. Hou ST, Nilchi L, Li X, Gangaraju S, Jiang SX, Aylsworth A, Monette R, Slinn J. Semaphorin3a elevates vascular permeability and contributes to cerebral ischemia-induced brain damage. *Sci Rep*. 2015;5:7890

Supplemental Table 7.1. MiRNAs nominally associated with incident stroke

miRNA	HR	Std.Error	P value	P value adjusted
miR-6124	1,66	0,12	2,00E-05	1,18E-02
miR-5196-5p	1,90	0,16	6,07E-05	1,79E-02
miR-4292	2,65	0,25	1,09E-04	2,15E-02
miR-4534	1,97	0,19	4,50E-04	6,65E-02
miR-6870-5p	1,47	0,12	1,15E-03	1,36E-01
miR-7111-5p	1,54	0,14	1,85E-03	1,68E-01
miR-6728-3p	0,57	0,19	2,31E-03	1,68E-01
miR-3940-5p	1,48	0,13	2,39E-03	1,68E-01
miR-4676-3p	1,78	0,19	2,56E-03	1,68E-01
miR-3141	1,64	0,17	3,07E-03	1,78E-01
miR-7150	1,70	0,18	3,31E-03	1,78E-01
miR-3648	1,55	0,16	5,25E-03	2,59E-01
miR-92b-3p	0,60	0,18	6,05E-03	2,75E-01
miR-6798-5p	1,36	0,12	8,45E-03	3,48E-01
miR-4478	1,59	0,18	9,42E-03	3,48E-01
miR-1915-3p	1,52	0,16	1,03E-02	3,48E-01
miR-4449	1,39	0,13	1,03E-02	3,48E-01
miR-374b-5p	0,56	0,22	1,06E-02	3,48E-01
miR-126-3p	0,60	0,21	1,21E-02	3,67E-01
miR-5703	1,50	0,16	1,24E-02	3,67E-01
miR-30a-3p	0,70	0,15	1,35E-02	3,80E-01
miR-16-5p	0,60	0,21	1,43E-02	3,84E-01
miR-345-5p	0,55	0,25	1,70E-02	4,36E-01
miR-765	1,52	0,18	1,79E-02	4,42E-01
miR-6873-3p	0,76	0,12	2,00E-02	4,63E-01
miR-6127	1,45	0,16	2,04E-02	4,63E-01
miR-374b-3p	0,55	0,26	2,12E-02	4,63E-01
miR-6778-5p	1,70	0,23	2,24E-02	4,67E-01
miR-4459	1,64	0,22	2,29E-02	4,67E-01
miR-4784	1,88	0,28	2,57E-02	5,05E-01
miR-30e-3p	0,50	0,33	3,09E-02	5,89E-01
miR-133a-3p	0,79	0,11	3,50E-02	6,47E-01
miR-4430	1,39	0,16	3,64E-02	6,47E-01
miR-532-3p	0,82	0,10	3,72E-02	6,47E-01
miR-3620-3p	0,74	0,14	3,84E-02	6,48E-01

miRNA	HR	Std.Error	P value	P value adjusted
miR-205-5p	0,85	0,08	4,29E-02	7,04E-01
miR-1275	1,47	0,19	4,53E-02	7,06E-01
miR-3937	0,82	0,10	4,61E-02	7,06E-01
miR-4745-3p	0,65	0,22	4,66E-02	7,06E-01

Association between plasma miRNA levels and incident stroke. Models are adjusted for: age, sex, cohort, smoking, BMI, HDL, total cholesterol, systolic blood pressure, diastolic blood pressure, prevalent diabetes mellitus type 2, prevalent coronary heart disease, prevalent atrial fibrillation, lipid lowering medication and blood pressure lowering medication

Additional supplementary material for this chapter can be found online: https://www.ahajournals.org/action/downloadSupplement?doi=10.1161%2FSTROKEAHA.120.031543&file=STR_STROKE-2020-031543_supp1.pdf

8

CHAPTER 8

Circulatory MicroRNAs in Plasma and Atrial Fibrillation among Men and Women in the General Population: the Rotterdam Study

Manuscript based on this chapter: Geurts S, Mens MMJ, Bos MM, Ikram MA, Ghanbari M, Kavousi M. Association between Circulatory microRNAs in Plasma and Atrial Fibrillation among Men and Women from the General Population: the Rotterdam Study. *Submitted*

Abstract

Objectives: MicroRNAs (miRNAs), small non-coding RNAs regulating gene expression, have shown to play an important role in cardiovascular disease. Limited population-based data regarding the relationship between circulatory miRNAs in plasma and atrial fibrillation (AF) exist. Moreover, it remains unclear if the relationship differs by sex. We therefore aimed to determine the association between circulatory miRNAs in plasma and AF.

Methods: Plasma levels of miRNAs were measured in 1999 participants (858 men and 1141 women) from the population-based Rotterdam Study. Logistic regression and Cox proportional hazards models were used to assess the associations of 591 well-expressed miRNAs with the prevalence and incidence of AF. Models were adjusted for cardiovascular risk factors. We further examined the link between predicted target genes of identified miRNAs.

Findings: The mean age was 71.7 years (57.1% women), 98 participants (58 men and 40 women) had prevalent AF at baseline. Moreover, 196 participants (96 men and 100 women) developed AF during a median follow-up of 9.0 years. After adjusting for multiple testing, miR-4798-3p was significantly associated with prevalent AF among men; odds ratio: 0.39 (95% confidence interval, 0.24-0.66). No miRNAs were significantly associated with incident AF. MiR-4798-3p could potentially regulate the expression of AF-related genes, including genes involved in calcium and potassium handling in myocytes, protection of cells against oxidative stress, and cardiac fibrosis.

Interpretation: Plasma levels of miR-4798-3p were significantly associated with prevalent AF among men. Various target genes in relation to AF pathophysiology could potentially be regulated by miR-4798-3p that warrant further investigations in future experimental studies.

Introduction

Atrial fibrillation (AF) is the most common cardiac arrhythmia worldwide.[1, 2] The prevalence of AF is expected to increase steeply in the coming decades due to aging of the population.[1-3] Despite the identification of risk factors for AF(4-6) and improvement in its management, AF still confers a high morbidity and mortality risk.[1, 2, 7] Furthermore, recent evidence suggests that sex differences in AF pathophysiology and prognosis exist.[8] Women with AF are older at diagnosis, have a higher prevalence of hypertension, valvular heart disease and have an increased risk of stroke, myocardial infarction and mortality in comparison to men.[8]

MicroRNAs (miRNAs) are a class of small non-coding RNAs that post-transcriptionally regulate gene expression by complementary binding to target transcripts. Dysregulation of miRNA function could affect pathology of diseases.[9] Extensive studies have also shown the potential of miRNAs to be used as disease biomarkers, since their expression remains stable after drawing blood and they are easy accessible in different types of body fluid.[10] Over the past years, the role of miRNAs in various cardiovascular diseases have received a major interest.[11] MiRNAs have been suggested, among others, as key regulators of electrical remodeling,[12] structural remodeling,[13] autonomic nerve remodeling,[14] calcium handling abnormalities,[15] and inflammation[16] of the heart. These functions suggest a role for miRNAs in AF pathophysiology.

Previous studies have identified plasma levels of several miRNAs to be associated with AF.[17-26] However, most of these studies were limited to cross-sectional analysis, a subgroup of AF patients, or hypothesis-driven by studying only subsets of specific miRNAs.[24] To date, limited data exists on the association between circulatory miRNAs in plasma and AF in the general population.[25, 26] Furthermore, research regarding sex differences in the associations of miRNAs with AF is sparse.

In this study, we aimed to investigate the association between circulatory miRNAs in plasma with prevalent and incident AF among men and women in the general population using data from the prospective population-based Rotterdam Study to gain more insight into AF pathophysiology. We further retrieved the predicted target genes of identified miRNAs and examined if any of these target genes have been associated with AF pathophysiology by previous literature. Moreover, we provided an extensive literature overview of the previously-reported circulatory miRNAs in blood/plasma in association with AF and we provided a detailed overview per type of miRNA and the corresponding study characteristics. Subsequently, we tested the association of these previously-reported circulatory miRNAs with prevalent and incident AF in our study.

Methods

Patient Involvement

Patients were involved in the design and conduct of this research. Participants will be informed of the results through a dedicated website and/or will be sent details of the results in a study newsletter suitable for a non-specialist audience.

Study Population

This study was embedded in the Rotterdam Study.[27, 28] In short, the Rotterdam Study is a prospective population-based cohort study that investigates the occurrence and progression of risk factors for chronic diseases in middle-age and elderly persons. The design of the Rotterdam Study is explained in detail in the **Supplementary Methods.**

For the present study, we included 1000 participants from the fourth visit of Rotterdam Study-I and 1000 participants from the second visit of Rotterdam Study-II for whom miRNA expression data was obtained (n=2000). These visits took place between 2002 and 2005 and we considered this as the baseline of our study. From these 2000 randomly chosen participants, one participant was excluded, because of insufficient baseline data on AF for the cross-sectional study with prevalent AF. For the longitudinal study with incident AF, we additionally excluded prevalent AF cases (n=98). A total of 1999 participants, 858 men and 1141 women, were included in the current study.

The Rotterdam Study complies with the Declaration of Helsinki and has been approved by the Medical Ethics Committee of the Erasmus Medical Center (registration number MEC 02.1015) and by the Dutch Ministry of Health, Welfare and Sport (Population Screening Act WBO, license number 1071272-159521-PG). The Rotterdam Study has been entered into the Netherlands National Trial Register (NTR; www.trialregister.nl) and into the WHO International Clinical Trials Registry Platform (ICTRP; www.who.int/ictrp/network/primary/en/) under shared catalogue number NTR6831. All participants provided written informed consent to participate, prior to inclusion, in the study and to have their information obtained from treating physicians.

Assessment of Circulatory miRNAs in Plasma

Methods on plasma miRNA level measurement have been described previously.[29] In short, the expression levels of 2083 mature human miRNAs (HTG Molecular Diagnostics, Tuscon, AZ, USA) and using the Illumina NextSeq 500 sequencer (Illumina, San Diego, CA, USA) were measured. Out of 2083 measured miRNAs, 591 miRNAs were well-expressed in plasma.[29] The assessment of miRNAs is further explained in the **Supplementary Methods.**

Assessment of Atrial Fibrillation

AF was defined in accordance with the European Society of Cardiology (ESC) guidelines.[7] Methods on event adjudication for prevalent and incident AF have been described previously.[3, 27] In short, a 10 seconds 12-lead electrocardiogram (ECG) was used to assess AF at baseline and additional follow-up information was obtained from medical files and follow-up examinations at the research center. All participants were followed from the date of enrolment in the Rotterdam Study until the date of onset of AF, date of death, loss to follow-up, or to January 1st, 2014, whichever occurred first. The assessment of AF is further explained in detail in the **Supplementary Methods**.

Assessment of Cardiovascular Risk Factors

The cardiovascular risk factors included in the study were body mass index (BMI), serum total and high-density lipoprotein (HDL) cholesterol, hypertension, history of diabetes mellitus (DM), history of coronary heart disease (CHD), history of heart failure (HF), left ventricular hypertrophy (LVH) on the ECG, smoking status, use of lipid-lowering and of cardiac medication. Methods for measurements of cardiovascular risk factors are explained in details in the **Supplementary Methods**.[27, 30]

Assessment of Predictive Target Genes

We retrieved the list of predicted targets genes of identified miRNAs associated with AF using the two commonly used target prediction databases: miRDB[31, 32] and TargetScan.[33] Furthermore, we assessed if any of these predicted target genes have been associated previously with AF by a systematic review and a genome-wide association study[24, 34] and we assessed if any of these genes are potentially involved in AF pathophysiology by electrical and/or structural remodeling of the heart.

Statistical Analysis

Participant characteristics are presented as mean with standard deviation or numbers with percentages as appropriate. Group differences between men and women were examined by Student T test for continuous variables and chi-square test for categorical variables.

Logistic regression and Cox proportional hazards models were used to assess the association between plasma miRNAs at baseline with prevalent and incident AF, respectively. Odds Ratios (ORs) and hazard ratios (HRs) with 95% confidence intervals (CIs) were calculated to quantify the associations. An examination of the shape of relation with AF was performed using natural cubic splines for continuous variables and no deviation from linearity was found. No influential values were observed when using Cook's distance and no multicollinearity among the variables was observed using

a variance inflation factors threshold of <5. The proportional hazards assumptions were tested by Schoenfeld tests and were found to be satisfied. Additionally, we examined the interaction of miRNAs and sex before subsequently stratifying our analyses.

Analyses were performed in the total study population and for men and women separately. All models were adjusted for age, sex (if applicable) and cohort (model 1) and additionally for cardiovascular risk factors including BMI, total and HDL cholesterol, hypertension, history of DM, history of CHD, history of HF, LVH on the ECG, smoking status, use of lipid lowering and of cardiac medication (model 2). Missing values of variables were imputed under the assumption of missing at random using multiple imputation. For multiple imputation all available data were used to generate 10 imputed data sets. The results from each complete data set were combined to present single estimates.

The p-value threshold was corrected for multiple testing based on the eigenvalues of the correlation matrix from all the miRNAs. This adapted method was proposed by Li[35] and is based on a method which was introduced by Cheverud[36] to adjust correlated tests as if they were independent, according to an 'effective number' of independent tests, since there is evidence that miRNAs are clustered together or may be co-expressed.[37] This means that the miRNAs are thereby correlated with each other and by adopting this p-value correction proposed by Li[35] we take this correlation into account when adjusting. Based on the aforementioned method from Li, the significance p-value cutoff was set at 0.000352 based on 142 identified independent tests (0.05/142).

The data management and analyses were done using IBM SPSS Statistics version 25.0 for Windows (IBM Corp, Armonk, New York) and R software (R 3.6.3; R Foundation for Statistical Computing, Vienna, Austria).

Literature Overview

We searched the literature (PubMed) to identify studies that reported on circulatory miRNAs in blood/plasma in association with AF. Subsequently, we tested the association of these previously-reported circulatory miRNAs with prevalent and incident AF in our study in attempt to replicate previous findings.

Genetic Analyses

We also sought to investigate whether the identified miRNAs are expressed in the heart. In addition, we retrieved the miRNA host genes as proxy for the identified miRNAs to evaluate their expression in the heart using the Human Protein Atlas.[38] The idea behind this is that intragenic miRNAs and their host genes are likely to be co-expressed.[39] Furthermore, the genomic location of the identified miRNAs was obtained using miRIAD.(40)

Results

A total of 1999 participants (858 men and 1141 women) were eligible for the analyses of miRNAs associated with prevalent AF. The baseline characteristics for the study sample are depicted in **Table 8.1**. For the longitudinal analysis of incident AF, after exclusion of prevalent AF cases, 1901 individuals (800 men and 1101 women) were included. Characteristics of this study population are presented in **Supplementary Table 8.1**. Compared to men, women were slightly older, more often hypertensive and never smokers. DM, CHD, HF and LVH on the ECG were less prevalent among women. Women used lipid lowering and cardiac medication less frequently than men.

Table 8.1. Baseline characteristics of the study population

Baseline characteristics *	Total population n=1999	Men n=858	Women n=1141	P-value ‡
Age, years	71.7 ± 7.6	71.4 ± 7.3	71.9 ± 7.8	0.116
Women, n (%)	1141 (57.1)	NA	1141 (100)	
Body mass index, kg/m²	27.7 ± 4.1	27.6 ± 3.4	27.7 ± 4.6	0.382
Hypertension, n (%)	1558 (77.9)	654 (76.2)	904 (79.2)	0.109
Total serum cholesterol, mmol/L †	5.6 ± 1.0	5.3 ± 1.0	5.9 ± 1.0	<0.001
High-density lipoprotein, mmol/L †	1.4 ± 0.4	1.0 ± 0.3	1.6 ± 0.4	<0.001
Smoking, n (%)				<0.001
Current	306 (15.3)	149 (17.4)	157 (13.8)	
Former	1094 (54.7)	592 (69.0)	502 (44.0)	
Never	599 (30.0)	117 (13.6)	482 (42.2)	
History of diabetes mellitus, n (%)	268 (13.4)	145 (16.9)	123 (10.8)	<0.001
History of coronary heart disease, n (%)	213 (10.7)	145 (16.9)	68 (6.0)	<0.001
History of heart failure, n (%)	101 (5.1)	50 (5.8)	51 (4.5)	0.170
Left ventricular hypertrophy, n (%)	108 (5.4)	62 (7.2)	46 (4.0)	0.002
Lipid lowering medication, n (%)	450 (22.5)	208 (24.2)	242 (21.2)	0.080
Cardiac medication, n (%)	210 (10.5)	102 (11.9)	108 (9.5)	0.108

* Values are mean (standard deviation) or numbers (percentages).

† SI conversion factors: To convert cholesterol to mg/dL divide by 0.0259.

‡ Statistical significance between men and women for continuous data was tested using the Student T test and categorical data was tested using the Chi Square Test.

Abbreviations: NA, not applicable.

At baseline, 98 cases (4.9%) of prevalent AF were identified from which 58 cases (6.8%) in men and 40 cases (3.5%) in women. Logistic regression showed that 47 miRNAs in the total study population, 45 miRNAs in men and 31 miRNAs in women were nominally significantly (p-value <0.05) associated with prevalent AF after adjustment for age and cardiovascular risk factors (model 2). See **Supplementary Table 8.2** for an overview of the nominally significantly associated miRNAs. For one unit increase in miR-4798-3p plasma levels at baseline, the odds for prevalent AF in the total population was OR 0.64 (95% CI: 0.44-0.97) (model 1). After adjusting for cardiovascular risk factors the odds did not attenuate; OR 0.63 (95% CI: 0.42-0.99) (model 2). The odds for prevalent AF were lower in men than in women. After adjustment for cardiovascular risk factors, ORs (95% CIs) were 0.39 (0.24-0.66) in men and 1.84 (95% CI: 0.76-4.97) in women (model 2). However, after adjusting for multiple testing (0.05/142=0.000352), only one miRNA, miR-4798-3p, remained statistically significantly associated with prevalent AF among men. See **Table 8.2** for more details. The interaction term between miRNA-4798-3p and sex in relation to the odds of prevalent AF in the total study population using logistic regression

was significant (p-value= 0.004730). This significant sex interaction further highlights our observed sex differences for miR-4798-3p. **Figure 8.1** illustrates the nominally significant miRNAs associated with prevalent AF among men described by a Volcano plot.

Table 8.2. Association between miR-4798-3p with prevalent atrial fibrillation

	OR (95% CI)			
	Model 1 *	**P-value**	**Model 2 †**	**P-value**
Total population				
miR-4798-3p	0.64 (0.44-0.97)	0.028033	0.63 (0.42-0.99)	0.034433
Men				
miR-4798-3p	0.42 (0.27-0.69)	0.000254	0.39 (0.24-0.66)	0.000248
Women				
miR-4798-3p	1.53 (0.71-3.70)	0.311964	1.84 (0.76-4.97)	0.203587

* Adjusted for age and cohort (model 1).

† Adjusted for age, cohort, body mass index, total and high-density lipoprotein cholesterol, hypertension, history of diabetes, history of coronary heart disease, history of heart failure, left ventricular hypertrophy on the electrocardiogram, smoking status, use of lipid lowering and of cardiac medication (model 2). Abbreviations: OR, Odds ratio; CI: confidence interval.

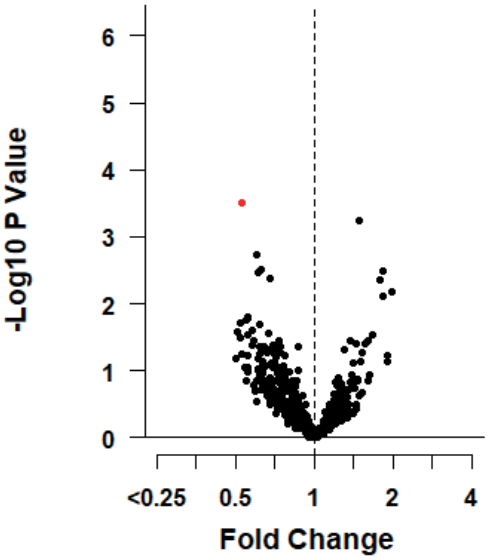


Figure 8.1. Volcano plot of nominally significant miRNAs in association with prevalent atrial fibrillation among men. The y-axis represents the negative log of the p-value on the y-axis and the x-axis represents the log of the fold change for prevalent AF. The red dot indicates the significant association after correction for multiple testing for miR-4798-3p.

During a median follow-up of 9.0 years, 196 (10.3%) cases of incident AF were identified from which 96 cases in men and 100 cases in women. Incidence rates were 15.2 per 1000 person-years in men and 10.7 per 1000 person-years in women. Cox proportional hazards models showed that a total of 17 miRNAs in the total study population, 26 miRNAs in men and 13 miRNAs in women were nominally significant in association with incident AF (model 2), but none of them remained statistically significant after adjustment for multiple testing. **Supplementary Table 8.3** shows a complete list of the nominally significantly associated miRNAs with incident AF.

There was little overlap in similarity between the effect estimates of the miRNAs among the prevalent AF cases when we compared them to the effect estimates in the incident AF sample and vice versa. This was also the case when we compared the effect estimates of miR-4798-3p for the association with prevalent AF among men with the effect estimates of miR-4798-3p for incident AF among men (OR 0.39 (95% CI: 0.24-0.66) vs. OR 1.02 (95% CI: 0.74-1.39)) (model 2).

We additionally examined the predicted target genes of miR-4798-3p using the two miRNA target prediction databases: miRDB[31, 32] and TargetScan.[33] Among predicted target genes of miR-4798-3p are *CACNB2*,[41] *KCNN3*,[42] *SIRT1*[43] and *STAT3*[44, 45] that are suggested to be involved in electrical and structural remodeling of the heart. **Supplementary Table 8.4** depicts the genes, that were among the predicted target genes of miR-4798-3p, and that have been previously associated with AF by a systematic review and a genome-wide association study.[24, 34] In addition, the potential remodeling mechanisms of the heart for these genes is also provided.[46]

Additionally, we provided an extensive literature overview of circulatory miRNAs in blood/plasma in association with AF that have been reported in the literature before **(Supplementary Table 8.5)**. In this overview we provided detailed information per type of miRNA and the corresponding study characteristics including study design, study population, baseline characteristics, reported effect estimates, the statistical model and adjustments. Furthermore, we did a look-up for these AF-associated miRNAs in our results. The effect estimates and p-values for the association of these previously-reported miRNAs with prevalent and incident AF in our data are reported in **Supplementary Table 8.6** and **Supplementary Table 8.7**, respectively. For the prevalent AF analyses, we were able to compare 39 miRNAs. Among these 39 miRNAs, the direction of the effect estimates reported in the literature were in line with our results for 18 miRNAs. The reported effect estimate in the literature that was most similar to our findings was the effect estimate for miR-20a-5p (literature-reported OR 1.36 (95% CI: 1.14-1.61)[25] while we found an OR 1.30 (95% CI: 0.68-2.58)). Moreover, for the incident AF analyses, we were able to compare 10 miRNAs from the literature with our results. The direction of the effect was similar for 4 miRNAs and the miRNA with the most similar effect estimate was miR-193a-5p with a literature-reported OR 0.87 (95% CI: 0.77-0.98)[24] while we found an OR 0.93 (0.67-1.28).

Finally, we explored whether miR-4798-3p was expressed in the heart and we were unable to find any information regarding its expression levels within the heart.(38) Alternatively, we evaluated the expression of its host gene as a proxy for miR-4798-3p.[39] MiR-4798-3p is located within an intron of the protein-coding gene *SORCS2*.[40] *SORCS2* is especially found within the central nervous system and it is well-expressed within the brain and to a lesser degree within the heart.[38] Moreover, the expression levels of AF-associated target genes of miR-4798-3p (**Supplementary Table 8.4)** were also detected in varies degrees within the heart.[38]

Discussion

In this prospective population-based study, we performed a systematic analysis of 591 circulatory miRNAs well-expressed in plasma with risk of AF in the general population and men and women separately. We found that plasma levels of miR-4798-3p were significantly associated with prevalent AF among men after extensive adjustment for potential confounders and correcting for multiple testing. Several predicted target genes of miR-4798-3p have been associated previously with AF in a systematic review[24] and data from a recent genome-wide association study on AF.[34] These target genes are namely potentially involved in electrical and structural remodeling of the heart and thereby may mediate the effect of miR-4798-3p in AF pathophysiology. Future experimental studies are warranted to investigate the potential (sex-specific) role of this miRNA in molecular pathways underlying AF. We also provided an extensive and detailed literature review of the previously-reported miRNAs linked to AF and compared these literature-reported associations with the associations observed in our study.

Plasma levels of miR-4798-3p were significantly associated with prevalent AF in our study. The exact pathology behind the associations of many miRNAs with cardiovascular diseases is not completely understood. In general, miRNAs are involved in every biological pathway through regulating expression of target genes/transcripts. The host gene *SORCS2* is profoundly expressed within the central nervous system and may thereby potentially exert an effect on AF vulnerability, since an effect on the extensive network of vagal ganglionated plexi are known to affect AF risk.[47] MiR-4798-3p has a predicted number of more than 50 target genes that it may regulate.[31-33] Various genes that are potentially regulated by this miRNA are involved in calcium and potassium handling in myocytes (*CACNB2*, *KCNN3*), in protection of cells against oxidative stress (*SIRT1*), and in regulating cardiac fibrosis (*STAT3*). These aforementioned mechanisms are linked to electrical and structural remodeling of the heart which are associated with AF pathophysiology.[8, 12, 13, 24, 34] However, if these circulating levels of miRNA in plasma by themselves cause AF or that the circulating levels of miRNAs are merely a reflection of an underlying pathology that may lead to the pathogenesis of AF is not clear. In addition, it is beyond the scope of this investigation to elucidate on the

pathophysiologic implications of putative target genes. Future experimental studies are warranted to investigate the interaction between miR-4798-3p, its target genes and their relation to AF. The little overlap between miRNAs associated with prevalent and incident AF might suggest that miRNAs may indeed be a reflection of underlying pathology that is associated with prevalent AF instead of that miRNAs may cause incident AF over time. The potential discrepancy between cell-specific expression of miRNAs and circulatory (cell-free) miRNAs in plasma makes it more difficult to disentangle this pathophysiology. However, it has been shown that circulatory miRNAs constitute a way of cell-to-cell communication and miRNAs are released to extracellular matrix and blood by exosome from the diseased tissue/cells. Moreover, the duration of time that is involved in the release of miRNA in plasma (by a pathological event) and the effect that it may have is still elusive. Nevertheless, miR-4798-3p could still be a potential plasma biomarker for AF prediction or prognosis.

Sex differences in AF pathophysiology are increasingly gaining interest.[8] The association of miR-4798-3p with prevalent AF in our study was only significant among men. This difference could be explained by the different target genes of miR-4798-3p and their potential sex specific effects. For example, *KCNN3* and *CACNB2* regulate L-type calcium channels and may thereby influence QT intervals of the heart.[48,44] Women have different and longer QT-intervals than men[8] and a long QT-interval has been associated with AF initiation.[49] *SIRT1* is known to be upregulated in patients with CHD possibly as a potential compensatory mechanism to counteract the adverse effects of oxidative stress caused by CHD.[50, 51] CHD is more prevalent among men than in women[8] and is implicated in AF pathophysiology.[4-8] *STAT3* is involved in cardiac fibrosis and previous research has shown that women with AF have more atrial fibrosis than men.[8] Although these effects could be sex specific, further exploration is warranted to examine the exact underlying molecular mechanisms that might explain these sex differences.

To the best of our knowledge, the Framingham Heart Study is the only population-based cohort that has previously investigated the association between miRNAs and AF at the population level. McManus et al[25] identified one miRNA that was significantly associated with prevalent AF (miR-328) in the Framingham Heart Study, while they did not find any significant miRNAs that were associated with incident AF. Vaze et al[26] identified 6 miRNAs that were significantly associated with incident AF in the Framingham Heart Study, including 4 also significantly associated with prevalent AF (miR-106b, miR-26a-5p, miR-484 and miR-20a-5p). We could not replicate the findings from McManus et al[25] and Vaze et al[26] or the results from the studies assessed during the literature review. These differences may be due to the fact that we measured circulatory miRNA levels in plasma instead of whole blood[52] as in the Framingham Heart Study, differences in miRNA expression profiling,[53] differences in adjusting for confounders, differences in correcting for multiple testing, and differential expression

of miRNAs related to the type and phase of AF. It is worth noting that an internationally adopted standardized method to evaluate miRNA expression could potentially improve any future miRNA studies (for example plasma vs. blood, or circulatory vs. tissue). Since such a standardization could then improve the comparability between future miRNA studies and this would also benefit any potential clinical applications of miRNA-based therapies in the future. Our findings, however, extend the previous mentioned studies by examining 591 (instead of 253-339) miRNAs, a longer follow-up time, and more extensive adjustment for potential confounding. We also examined potential sex differences in the associations between miRNAs and AF in our study population and we thereby also add to the emerging evidence that circulating miRNAs play a critical role in the pathophysiology of AF that may potentially be sex specific.

Major strengths of our study include its population-based nature, large sample size, detailed information on cardiovascular risk factors, precise adjudication of prevalent and incident AF, a long follow-up time, including a well-expressed set of 591 miRNAs, extensive adjustment for potential confounders and the examination of potential sex differences between miRNAs and AF. Nonetheless, there are some limitations. We could not distinguish between paroxysmal and permanent AF as Holter monitoring is not available in this cohort. Although we extensively adjusted for confounders, residual confounding cannot be entirely ruled out. Furthermore, our study population included mainly elderly participants from European descent and our results may therefore not be generalizable to younger populations or other ethnicities.

References

1. Chugh SS, Havmoeller R, Narayanan K, Singh D, Rienstra M, Benjamin EJ, et al. Worldwide epidemiology of atrial fibrillation: a Global Burden of Disease 2010 Study. Circulation. 2014;129(8):837-47.
2. Krijthe BP, Kunst A, Benjamin EJ, Lip GY, Franco OH, Hofman A, et al. Projections on the number of individuals with atrial fibrillation in the European Union, from 2000 to 2060. Eur Heart J. 2013;34(35):2746-51.
3. Heeringa J, van der Kuip DA, Hofman A, Kors JA, van Herpen G, Stricker BH, et al. Prevalence, incidence and lifetime risk of atrial fibrillation: the Rotterdam study. Eur Heart J. 2006;27(8):949-53.
4. Benjamin EJ, Levy D, Vaziri SM, D'Agostino RB, Belanger AJ, Wolf PA. Independent risk factors for atrial fibrillation in a population-based cohort. The Framingham Heart Study. JAMA. 1994;271(11):840-4.
5. Schnabel RB, Yin X, Gona P, Larson MG, Beiser AS, McManus DD, et al. 50 year trends in atrial fibrillation prevalence, incidence, risk factors, and mortality in the Framingham Heart Study: a cohort study. Lancet. 2015;386(9989):154-62.
6. Brandes A SM, Nguyen BO, Rienstra M, Van Gelder IC. Risk Factor Management in Atrial Fibrillation. Arrhythm Electrophysiol Rev. 2018;7(2):118-27.
7. Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomstrom-Lundqvist C, et al. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS). Eur Heart J. 2021;42(5):373-498.
8. Ko D, Rahman F, Schnabel RB, Yin X, Benjamin EJ, Christophersen IE. Atrial fibrillation in women: epidemiology, pathophysiology, presentation, and prognosis. Nat Rev Cardiol. 2016;13(6):321-32.
9. Friedman RC FK, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. 2009;19(1):92-105.
10. Grasedieck S, Scholer N, Bommer M, Niess JH, Tumani H, Rouhi A, et al. Impact of serum storage conditions on microRNA stability. Leukemia. 2012;26(11):2414-6.
11. Small EM, Olson EN. Pervasive roles of microRNAs in cardiovascular biology. Nature. 2011;469(7330):336-42.
12. Chaldoupi SM, Loh P, Hauer RN, de Bakker JM, van Rijen HV. The role of connexin40 in atrial fibrillation. Cardiovasc Res. 2009;84(1):15-23.
13. Xu J, Cui G, Esmailian F, Plunkett M, Marelli D, Ardehali A, et al. Atrial extracellular matrix remodeling and the maintenance of atrial fibrillation. Circulation. 2004;109(3):363-8.
14. Wei J, Zhang Y, Li Z, Wang X, Chen L, Du J, et al. GCH1 attenuates cardiac autonomic nervous remodeling in canines with atrial-tachypacing via tetrahydrobiopterin pathway regulated by microRNA-206. Pacing Clin Electrophysiol. 2018;41(5):459-71.
15. Bingen BO, Askar SF, Neshati Z, Feola I, Panfilov AV, de Vries AA, et al. Constitutively Active Acetylcholine-Dependent Potassium Current Increases Atrial Defibrillation Threshold by Favoring Post-Shock Re-Initiation. Sci Rep. 2015;5:15187.
16. Zhang H, Liu L, Hu J, Song L. MicroRNA Regulatory Network Revealing the Mechanism of Inflammation in Atrial Fibrillation. Med Sci Monit. 2015;21:3505-13.
17. Tsoporis JN FA, Rizos IK, Izhar S, Proteau G, Salpeas V, Rigopoulos A, Sakadakis E, Toumpoulis IK, Parker TG. Increased right atrial appendage apoptosis is associated with differential regulation of

candidate MicroRNAs 1 and 133A in patients who developed atrial fibrillation after cardiac surgery. J Mol Cell Cardiol. 2018;121:25-32.

18. Dawson K, Wakili R, Ordog B, Clauss S, Chen Y, Iwasaki Y, et al. MicroRNA29: a mechanistic contributor and potential biomarker in atrial fibrillation. Circulation. 2013;127(14):1466-75, 75e1-28.
19. McManus DD, Tanriverdi K, Lin H, Esa N, Kinno M, Mandapati D, et al. Plasma microRNAs are associated with atrial fibrillation and change after catheter ablation (the miRhythm study). Heart Rhythm. 2015;12(1):3-10.
20. Wei XJ, Han M, Yang FY, Wei GC, Liang ZG, Yao H, et al. Biological significance of miR-126 expression in atrial fibrillation and heart failure. Braz J Med Biol Res. 2015;48(11):983-9.
21. Soeki T MT, Bando S, Tobiume T, Uematsu E, Ise T, Kusunose K, Yamaguchi K, Yagi S, Fukuda D, Yamada H, Wakatsuki T, Shimabukuro M, Sata M. Relationship between local production of microRNA-328 and atrial substrate remodeling in atrial fibrillation. J Cardiol. 2016;68(6):472-7.
22. Liu T, Zhong S, Rao F, Xue Y, Qi Z, Wu S. Catheter ablation restores decreased plasma miR-409-3p and miR-432 in atrial fibrillation patients. Europace. 2016;18(1):92-9.
23. Harling L LJ, Ashrafian H, Darzi A, Gooderham NJ, Athanasiou T. Elevated serum microRNA 483-5p levels may predict patients at risk of post-operative atrial fibrillation. Eur J Cardiothorac Surg. 2017;51(1):73-8.
24. Komal S, Yin JJ, Wang SH, Huang CZ, Tao HL, Dong JZ, et al. MicroRNAs: Emerging biomarkers for atrial fibrillation. J Cardiol. 2019;74(6):475-82.
25. McManus DD LH, Tanriverdi K, Quercio M, Yin X, Larson MG, Ellinor PT, Levy D, Freedman JE, Benjamin EJ. Relations between circulating microRNAs and atrial fibrillation: data from the Framingham Offspring Study. Heart Rhythm. 2014;11(4):663-9.
26. Vaze A TK, Tanriverdi K, Sardana M, Lessard D, Donahue JK, Barton B, Aurigemma G, Lubitz SA, Lin H, Nasr GH, Mandapati A, Benjamin EJ, Vasan RS, Freedman JE, McManus DD. Relations between plasma microRNAs, echocardiographic markers of atrial remodeling, and atrial fibrillation: Data from the Framingham Offspring study. PLoS One. 2020;15(8):e0236960.
27. Leening MJ, Kavousi M, Heeringa J, van Rooij FJ, Verkroost-van Heemst J, Deckers JW, et al. Methods of data collection and definitions of cardiac outcomes in the Rotterdam Study. Eur J Epidemiol. 2012;27(3):173-85.
28. Ikram MA, Brusselle G, Ghanbari M, Goedegebure A, Ikram MK, Kavousi M, et al. Objectives, design and main findings until 2020 from the Rotterdam Study. Eur J Epidemiol. 2020;35(5):483-517.
29. Zhang X MM, Abozaid YJ, Bos D, Darwish Murad S, de Knegt RJ, Ikram MA, Pan Q, Ghanbari M. Circulatory microRNAs as potential biomarkers for fatty liver disease: the Rotterdam study. Aliment Pharmacol Ther. 2021;53(3):432-42.
30. Heeringa J, van der Kuip DA, Hofman A, Kors JA, van Rooij FJ, Lip GY, et al. Subclinical atherosclerosis and risk of atrial fibrillation: the rotterdam study. Arch Intern Med. 2007;167(4):382-7.
31. Liu W, Wang X. Prediction of functional microRNA targets by integrative modeling of microRNA binding and target expression data. Genome Biol. 2019;20(1):18.
32. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. Elife. 2015;4.
33. Chen Y, Wang X. miRDB: an online database for prediction of functional microRNA targets. Nucleic Acids Res. 2020;48(D1):D127-D31.

34. Nielsen JB TR, Fritsche LG, Zhou W, Skov MW, Graham SE, Herron TJ, McCarthy S, Schmidt EM, Sveinbjornsson G, Surakka I, Mathis MR, Yamazaki M, Crawford RD, Gabrielsen ME, Skogholt AH, Holmen OL, Lin M, Wolford BN, Dey R, Dalen H, Sulem P, Chung JH, Backman JD, Arnar DO, Thorsteinsdottir U, Baras A, O'Dushlaine C, Holst AG, Wen X, Hornsby W, Dewey FE, Boehnke M, Kheterpal S, Mukherjee B, Lee S, Kang HM, Holm H, Kitzman J, Shavit JA, Jalife J, Brummett CM, Teslovich TM, Carey DJ, Gudbjartsson DF, Stefansson K, Abecasis GR, Hveem K, Willer CJ. Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. Nat Genet. 2018;50(9):1234-9.

35. Li J JL. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity (Edinb). 2005;95(3):221-7.

36. Cheverud JM. A simple correction for multiple comparisons in interval mapping genome scans. Heredity (Edinb). 2001;87(Pt 1):52-8.

37. Wang Y, Luo J, Zhang H, Lu J. microRNAs in the Same Clusters Evolve to Coordinately Regulate Functionally Related Genes. Mol Biol Evol. 2016;33(9):2232-47.

38. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. Science. 2015;347(6220):1260419.

39. Lutter D MC, Krumsiek J, Lang EW, Theis FJ. Intronic microRNAs support their host genes by mediating synergistic and antagonistic regulatory effects. BMC Genomics. 2010;11:224.

40. Hinske LC, Franca GS, Torres HA, Ohara DT, Lopes-Ramos CM, Heyn J, et al. miRIAD-integrating microRNA inter- and intragenic data. Database (Oxford). 2014;2014.

41. Ling TY, Wang XL, Chai Q, Lu T, Stulak JM, Joyce LD, et al. Regulation of cardiac CACNB2 by microRNA-499: Potential role in atrial fibrillation. BBA Clin. 2017;7:78-84.

42. Ling TY, Wang XL, Chai Q, Lau TW, Koestler CM, Park SJ, et al. Regulation of the SK3 channel by microRNA-499--potential role in atrial fibrillation. Heart Rhythm. 2013;10(7):1001-9.

43. Yamac AH, Kucukbuzcu S, Ozansoy M, Gok O, Oz K, Erturk M, et al. Altered expression of microRNA 199a and increased levels of cardiac SIRT1 protein are associated with the occurrence of atrial fibrillation after coronary artery bypass graft surgery. Cardiovasc Pathol. 2016;25(3):232-6.

44. He X, Zhang K, Gao X, Li L, Tan H, Chen J, et al. Rapid atrial pacing induces myocardial fibrosis by down-regulating Smad7 via microRNA-21 in rabbit. Heart Vessels. 2016;31(10):1696-708.

45. Huang Z, Chen XJ, Qian C, Dong Q, Ding D, Wu QF, et al. Signal Transducer and Activator of Transcription 3/MicroRNA-21 Feedback Loop Contributes to Atrial Fibrillation by Promoting Atrial Fibrosis in a Rat Sterile Pericarditis Model. Circ Arrhythm Electrophysiol. 2016;9(7).

46. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. Curr Protoc Bioinformatics. 2016;54:1 30 1-1 3.

47. Choi EK ZY, Everett TH 4th, Chen PS. Ganglionated plexi as neuromodulation targets for atrial fibrillation. J Cardiovasc Electrophysiol. 2017;28(12):1485-91.

48. Antzelevitch C, Pollevick GD, Cordeiro JM, Casis O, Sanguinetti MC, Aizawa Y, et al. Loss-of-function mutations in the cardiac calcium channel underlie a new clinical entity characterized by ST-segment elevation, short QT intervals, and sudden cardiac death. Circulation. 2007;115(4):442-9.

49. Mandyam MC, Soliman EZ, Alonso A, Dewland TA, Heckbert SR, Vittinghoff E, et al. The QT interval and risk of incident atrial fibrillation. Heart Rhythm. 2013;10(10):1562-8.

50. Saxena A DD, Smith JA, Shardey GC, Reid CM, Newcomb AE. Usefulness of postoperative atrial fibrillation as an independent predictor for worse early and late outcomes after isolated

coronary artery bypass grafting (multicenter Australian study of 19,497 patients). Am J Cardiol. 2012;109(2):219-25.

51. Mostafa A, El-Haddad MA, Shenoy M, Tuliani T. Atrial fibrillation post cardiac bypass surgery. Avicenna J Med. 2012;2(3):65-70.

52. Shah R, Tanriverdi K, Levy D, Larson M, Gerstein M, Mick E, et al. Discordant Expression of Circulating microRNA from Cellular and Extracellular Sources. PLoS One. 2016;11(4):e0153691.

53. Godoy PM, Barczak AJ, DeHoff P, Srinivasan S, Etheridge A, Galas D, et al. Comparison of Reproducibility, Accuracy, Sensitivity, and Specificity of miRNA Quantification Platforms. Cell Rep. 2019;29(12):4212-22 e5.

Supplementary Table 8.1. Baseline characteristics of the study population for incident atrial fibrillation analyses

Baseline characteristics *	Total population n=1901	Men n=800	Women n=1101	P-value ‡
Age, years	71.4 ± 7.5	71.0 ± 7.1	71.7 ± 7.7	0.022
Women, n (%)	1101 (57.9)	NA	1101 (100)	
Body mass index, kg/m²	27.7 ± 4.1	27.6 ± 3.4	27.7 ± 4.6	0.582
Hypertension, n (%)	1469 (77.3)	604 (75.5)	865 (78.6)	0.115
Total serum cholesterol, mmol/L †	5.7 ± 1.0	5.4 ± 1.0	5.9 ± 1.0	<0.001
High-density lipoprotein, mmol/L †	1.4 ± 0.4	1.3 ± 0.3	1.6 ± 0.4	<0.001
Smoking, n (%)				<0.001
Current	298 (15.7)	143 (17.9)	155 (14.1)	
Former	1031 (54.2)	545 (68.1)	486 (44.1)	
Never	572 (30.1)	112 (14.0)	460 (41.7)	
History of diabetes mellitus, n (%)	248 (13.0)	129 (16.1)	119 (10.8)	0.001
History of coronary heart disease, n (%)	188 (9.9)	121 (15.1)	67 (6.1)	<0.001
History of heart failure, n (%)	83 (4.4)	37 (4.6)	46 (4.2)	0.638
Left ventricular hypertrophy, n (%)	95 (5.0)	50 (6.3)	45 (4.1)	0.033
Lipid lowering medication, n (%)	419 (22.0)	191 (23.9)	228 (20.7)	0.386
Cardiac medication, n (%)	168 (8.8)	76 (9.5)	92 (8.4)	0.100

* Values are mean (standard deviation) or numbers (percentages).

† SI conversion factors: To convert cholesterol to mg/dL divide by 0.0259.

‡ Statistical significance between men and women for categorical data was tested using the Chi Square Test and continuous data was tested using the Student T test.

Abbreviations: NA, not applicable.

Supplementary Table 8.2. MicroRNAs nominally associated with prevalent atrial fibrillation in total population and among men and women

Total population			Men			Women		
miRNA	**OR (95% CI) ***	**P-value**	**miRNA**	**OR (95% CI) ***	**P-value**	**miRNA**	**OR (95% CI) ***	**P-value**
miR-122-5p	1.42 (1.12-1.78)	0.002967	miR-4798-3p	0.39 (0.24-0.66)	0.000248	miR-1273h-5p	2.55 (1.36-4.89)	0.004270
miR-8078	1.54 (1.16-2.06)	0.003384	miR-122-5p	1.77 (1.28-2.46)	0.000539	miR-6788-3p	0.55 (0.37-0.86)	0.004616
miR-194-5p	1.82 (1.21-2.74)	0.004300	miR-4798-5p	0.47 (0.30-0.77)	0.001531	miR-1273d	3.06 (1.45-6.82)	0.004675
miR-1322	2.01 (1.24-3.32)	0.005358	miR-605-3p	0.50 (0.32-0.80)	0.002291	miR-1304-3p	5.56 (1.74-18.97)	0.004847
miR-378e	0.49 (0.30-0.83)	0.006478	miR-4784	0.48 (0.30-0.82)	0.003185	miR-548ay-5p	4.81 (1.65-15.27)	0.005846
miR-1254	1.77 (1.17-2.69)	0.007442	miR-192-5p	2.38 (1.35-4.29)	0.003233	miR-8078	1.97 (1.21-3.32)	0.008302
miR-574-3p	1.99 (1.19-3.31)	0.008025	miR-4721	0.57 (0.39-0.86)	0.003958	miR-1322	3.04 (1.36-7.10)	0.008354
miR-4721	0.66 (0.50-0.92)	0.008546	miR-194-5p	2.29 (1.29-4.11)	0.005029	miR-1231	0.53 (0.33-0.87)	0.009599
miR-6852-5p	0.54 (0.34-0.88)	0.009831	miR-4512	2.74 (1.37-5.70)	0.005538	miR-574-3p	2.94 (1.24-6.98)	0.014333
miR-4798-5p	0.60 (0.42-0.91)	0.010017	miR-6747-3p	2.41 (1.30-4.68)	0.007422	miR-5585-3p	2.36 (1.22-4.84)	0.014432
miR-1909-3p	1.51 (1.10-2.07)	0.010265	miR-6794-5p	0.15 (0.03-0.67)	0.012026	miR-6797-5p	0.64 (0.45-0.95)	0.017631
miR-204-3p	0.36 (0.16-0.80)	0.011161	miR-2116-5p	0.42 (0.22-0.89)	0.012516	miR-6500-3p	1.89 (1.15-3.37)	0.020219
miR-1285-5p	1.58 (1.11-2.27)	0.012415	miR-125a-5p	0.41 (0.21-0.87)	0.015047	miR-1285-5p	2.06 (1.14-3.89)	0.020352
miR-1273d	1.71 (1.13-2.65)	0.013537	miR-3124-3p	0.43 (0.22-0.88)	0.016097	miR-1254	2.30 (1.13-4.71)	0.021772
miR-4512	1.84 (1.16-3.03)	0.013550	miR-3667-5p	0.48 (0.28-0.92)	0.016510	miR-1273e	2.06 (1.13-3.88)	0.022274
miR-5585-3p	1.63 (1.11-2.43)	0.014015	miR-4723-3p	0.39 (0.18-0.89)	0.017808	miR-378e	0.43 (0.21-0.94)	0.024289
miR-192-5p	1.64 (1.11-2.45)	0.014723	miR-335-5p	0.54 (0.32-0.91)	0.020447	miR-566	2.73 (1.19-6.91)	0.025762
miR-1304-3p	2.22 (1.17-4.30)	0.016146	miR-10b-5p	0.45 (0.23-0.93)	0.021481	miR-4421	2.09 (1.14-4.16)	0.026404
miR-1255b2-3p	1.54 (1.08-2.18)	0.017186	miR-3912-5p	0.36 (0.16-0.91)	0.022020	miR-1255b2-3p	1.95 (1.08-3.51)	0.026521
miR-4784	0.61 (0.41-0.95)	0.019058	miR-4734	2.16 (1.08-4.07)	0.022058	miR-1273a	2.23 (1.13-4.68)	0.026830
miR-616-3p	1.62 (1.10-2.48)	0.019789	miR-6715b-3p	0.42 (0.20-0.91)	0.023262	miR-213p	0.47 (0.25-0.94)	0.027541

Total population			Men			Women		
miRNA	**OR (95% CI) ***	**P-value**	**miRNA**	**OR (95% CI) ***	**P-value**	**miRNA**	**OR (95% CI) ***	**P-value**
miR-4421	1.54 (1.08-2.23)	0.019957	miR-30a-5p	0.34 (0.14-0.89)	0.023364	miR-1273c	2.49 (1.10-5.65)	0.028319
miR-1273h-5p	1.54 (1.07-2.25)	0.020961	miR-3687	1.98 (1.05-3.54)	0.027593	miR-1269b	2.10 (1.13-4.25)	0.028337
miR-29c-5p	0.69 (0.51-0.96)	0.021375	miR-4713-3p	0.63 (0.42-0.98)	0.030730	miR-548d-5p	2.18 (1.08-4.50)	0.031964
miR-3687	1.67 (1.06-2.58)	0.022307	miR-1909-3p	1.58 (1.05-2.41)	0.030829	miR-3674	2.31 (1.09-5.07)	0.032957
miR-141-3p	0.51 (0.29-0.93)	0.022368	miR-100-5p	1.99 (1.09-4.00)	0.031450	miR-3135a	2.26 (1.06-4.83)	0.035470
miR-6794-5p	0.28 (0.09-0.85)	0.023037	miR-670-3p	0.38 (0.16-0.95)	0.032241	miR-4539	2.03 (1.05-3.98)	0.037016
miR-6788-3p	0.70 (0.52-0.97)	0.023570	miR-920	0.46 (0.23-0.96)	0.032879	miR-616-3p	2.19 (1.10-4.89)	0.039789
miR-3674	1.63 (1.07-2.51)	0.024082	miR-4695-5p	1.70 (1.01-2.74)	0.034056	miR-765	0.57 (0.34-1.02)	0.041683
miR-566	1.74 (1.08-2.87)	0.026028	miR-133a-5p	0.45 (0.22-0.99)	0.036036	miR-1539	2.02 (1.12-4.30)	0.041994
miR-765	0.64 (0.43-0.96)	0.027041	miR-6761-5p	0.52 (0.29-0.99)	0.037263	miR-4459	2.55 (0.98-6.29)	0.048123
miR-4753-5p	0.49 (0.27-0.95)	0.027462	miR-4543p	0.59 (0.36-0.99)	0.039182			
miR-5684	1.41 (1.05-1.97)	0.029441	miR-4319	0.55 (0.32-1.01)	0.039455			
miR-1273e	1.49 (1.04-2.15)	0.030004	miR-1365p	0.81 (0.66-1.00)	0.040095			
miR-7106-5p	0.71 (0.52-0.98)	0.032349	miR-193a-5p	0.49 (0.25-0.96)	0.040877			
miR-4798-3p	0.63 (0.42-0.99)	0.034433	miR-362-5p	0.64 (0.43-1.00)	0.041871			
miR-6877-3p	2.12 (1.06-4.31)	0.034941	miR-6791-5p	1.48 (1.01-2.15)	0.043547			
miR-6716-3p	0.57 (0.34-0.99)	0.035978	miR-155-5p	0.59 (0.36-1.00)	0.043632			
miR-4426	0.76 (0.59-0.99)	0.036714	miR-548ay-5p	0.58 (0.36-1.04)	0.045788			
miR-3667-5p	0.63 (0.42-1.01)	0.038123	miR-2116-3p	0.39 (0.16-1.05)	0.045877			
miR-1228-3p	1.93 (1.04-3.66)	0.041108	miR-196b-3p	0.51 (0.26-1.00)	0.047484			
miR-1303	1.58 (1.02-2.47)	0.041760	miR-4722-3p	1.86 (1.00-3.39)	0.047545			
miR-1273c	1.62 (1.01-2.57)	0.042953	miR-217	0.49 (0.25-1.03)	0.048351			

Total population			Men			Women		
miRNA	**OR (95% CI) ***	**P-value**	**miRNA**	**OR (95% CI) ***	**P-value**	**miRNA**	**OR (95% CI) ***	**P-value**
miR-6887-5p	0.71 (0.51-0.99)	0.043152	let-7d-3p	0.49 (0.24-1.00)	0.048457			
miR-1269b	1.44 (1.03-2.08)	0.043577	miR-6887-5p	0.64 (0.41-1.00)	0.049277			
miR-6747-3p	1.53 (1.02-2.36)	0.047232						
miR-30a-5p	0.48 (0.23-1.02)	0.049055						

* Adjusted for age, cohort, body mass index, total and high-density lipoprotein cholesterol, hypertension, history of diabetes, history of coronary heart disease, history of heart failure, left ventricular hypertrophy on the electrocardiogram, smoking status, use of lipid lowering and of cardiac medication (model 2).

Supplementary Table 8.3. MicroRNAs nominally associated with incident atrial fibrillation in total population and among men and women

Total population			Men			Women		
miRNA	**HR (95% CI) ***	**P-value**	**miRNA**	**HR (95% CI) ***	**P-value**	**miRNA**	**HR (95% CI) ***	**P-value**
miR-197-3p	0.58 (0.40-0.84)	0.003531	miR-378e	0.43 (0.26-0.71)	0.001137	miR-197-3p	0.47 (0.29-0.79)	0.003940
miR-548w	1.38 (1.10-1.74)	0.006042	miR-6769b-3p	4.22 (1.51-11.74)	0.005900	miR-181b-5p	0.56 (0.37-0.84)	0.005111
miR-6765-5p	1.54 (1.08-2.20)	0.017530	miR-1290	0.67 (0.49-0.90)	0.008000	miR-185-3p	0.78 (0.64-0.94)	0.010826
miR-6894-5p	0.75 (0.59-0.97)	0.025632	miR-7111-3p	0.69 (0.52-0.91)	0.008684	miR-181d-5p	0.81 (0.68-0.96)	0.017475
miR-548b-5p	1.24 (1.02-1.51)	0.033402	miR-5006-5p	2.61 (1.24-5.49)	0.011750	miR-548w	1.53 (1.08-2.18)	0.018272
miR-6126	1.17 (1.01-1.36)	0.033909	miR-4270	0.58 (0.38-0.89)	0.013263	miR-561-5p	1.98 (1.11-3.53)	0.019960
miR-3197	0.89 (0.79-0.99)	0.036662	miR-4758-5p	0.61 (0.40-0.92)	0.019000	miR-6765-5p	1.82 (1.06-3.12)	0.028780
miR-324-3p	0.70 (0.49-0.98)	0.038061	miR-532-5p	0.65 (0.46-0.93)	0.019790	miR-3689a-3p	0.81 (0.67-0.98)	0.031271
miR-1307-5p	1.32 (1.01-1.73)	0.039834	miR-378a-3p	0.53 (0.31-0.91)	0.021570	miR-204-3p	2.50 (1.08-5.77)	0.032405
miR-17-3p	0.69 (0.48-0.99)	0.041655	miR-6500-3p	0.80 (0.66-0.97)	0.023000	miR-125a-5p	0.55 (0.31-0.97)	0.038775
miR-6715b-3p	1.57 (1.02-2.43)	0.041864	miR-7978	1.74 (1.07-2.81)	0.024209	miR-4676-3p	1.54 (1.02-2.33)	0.041641
miR-1287-5p	1.23 (1.01-1.51)	0.042095	miR-4279	2.20 (1.09-4.44)	0.027113	miR-3169	1.41 (1.01-1.96)	0.042371
miR-4800-5p	0.73 (0.54-0.99)	0.044090	miR-6892-3p	1.82 (1.07-3.10)	0.027160	miR-324-3p	0.60 (0.36-0.99)	0.043760
miR-4522	1.32 (1.01-1.72)	0.044800	miR-3157-3p	2.39 (1.08-5.29)	0.031678			
miR-658	1.40 (1.01-1.94)	0.044996	miR-2115-3p	3.15 (1.10-9.02)	0.032931			
miR-4449	0.72 (0.52-0.99)	0.045036	miR-299-5p	1.54 (1.03-2.29)	0.034224			
miR-93-5p	0.71 (0.51-0.99)	0.046484	miR-4646-3p	0.75 (0.57-0.98)	0.035213			
			miR-149-3p	1.42 (1.02-1.98)	0.035383			
			miR-122-5p	0.75 (0.57-0.98)	0.036170			
			miR-2276-3p	1.62 (1.03-2.55)	0.037520			
			miR-4644	1.30 (1.02-1.66)	0.037586			
			miR-658	1.64 (1.03-2.62)	0.038120			

Total population			Men			Women		
miRNA	HR (95% CI) *	P-value	miRNA	HR (95% CI) *	P-value	miRNA	HR (95% CI) *	P-value
			miR-6742-5p	2.05 (1.04-4.06)	0.039417			
			miR-4309	2.05 (1.01-4.16)	0.046510			
			miR-1287-5p	1.34 (1.00-1.79)	0.046800			
			miR-1229-3p	0.81 (0.66-1.00)	0.047938			

* Adjusted for age, cohort, body mass index, total and high-density lipoprotein cholesterol, hypertension, history of diabetes, history of coronary heart disease, history of heart failure, left ventricular hypertrophy on the electrocardiogram, smoking status, use of lipid lowering and of cardiac medication (model 2).

Supplementary Table 8.4. Predicted target genes of miR-4798-3p that have been previously associated with atrial fibrillation and their potential remodeling mechanism

Predicted target gene	Target gene name	Remodeling mechanism
CACNB2	Calcium channel, voltage-dependent, beta 2 subunit (41)	Electrical
KCNN3	Potassium intermediate/small conductance calcium-activated channel, subfamily N, member 3 (42)	Electrical
SIRT1	Sirtuin 1 (43)	Structural
STAT3	Signal transducer and activator of transcription 3 (44,45)	Structural
SLIT3	Slit homolog 3 (46)	Structural
NAV2	Neuron navigator 2 (46)	Structural
MYOCD	Myocardin (46)	Structural
PHLDA1	Pleckstrin homology-like domain, family A, member 1 (34)	Structural
REEP1	Receptor accessory protein 1 (34)	Structural
FAM13B	Family with sequence similarity 13, member B (34)	Structural
ATXN1	Ataxin 1-like (34)	Structural
UST	Uronyl-2-sulfotransferase (34)	Structural
CDK6	Cyclin-dependent kinase 6 (34)	Structural
CAV2	Caveolin 2 (34)	Structural
XPO7	Exportin 7 (34)	Structural
FBXO32	F-box protein 32 (34)	Structural
FUT11	Fucosyltransferase 11 (34)	Structural
ASAH1	N-acylsphingosine amidohydrolase 1 (34)	Structural
ORMDL3	ORM1-like 3 (34)	Structural
WNT3	Wingless-type MMTV integration site family, member 3 (34)	Structural
USP36	Ubiquitin specific peptidase 36 (34)	Structural

PART III

9

CHAPTER 9

General Discussion

Aim

In this thesis, I investigated the potential role of miRNAs as regulators and biomarkers of aging and age-related diseases. I have made efforts in improving the understanding about the biological functions of miRNAs in both cellular developmental and pathological mechanisms of these complex diseases. This epidemiological approach made it possible to characterize circulatory miRNAs at the population level within the Rotterdam Study. To this end, I have identified differential miRNA expression patterns for individuals with underlying conditions towards metabolic and vascular diseases. This thesis makes a small contribution to unleashing the potential of miRNAs as diagnostic and prognostic biomarkers, as well as to better understanding of their role in functional mechanisms in complex systems. Although I have not fully been able to investigate the molecular mechanisms and complex interplay of genetics, DNA methylation, metabolites and miRNA expression, the work presented in this thesis is a preview towards the important contribution of miRNAs in human biology. In this chapter, the main findings, research challenges, clinical relevance and future directions are discussed.

Main findings

Aging is a change in an organism that leads to an increased risk of tissue dysfunction, diseases and death[1,2]. As we tend to define aging in a chronological manner, this may not be true. Individuals with the same chronological age could differ in terms of age-associated functional decline. A proposed cause of biological aging involves epigenetics, in which the established epigenetic role in embryonic development is increasingly seen as relevant to later life stages[3]. To this end, epigenetic markers such as DNA methylation are associated with cellular dysfunction that lead to accelerative aging[4]. A common phenomenon of cellular aging is the increase in size and the reduced ability to divide[5]. In **Part I**, I described the role of miRNAs in aging. I highlighted several miRNAs in **chapter 2** that regulate cell cycle initiation and progression of stem cells and have impact on the grow and divide phases (G1 and S-phase in respective) of various cell types, in particular miR-290-295, miR-302, miR-17-92, miR-106b-25 and miR-106a-363. The reported findings of **chapter 3** indicated that miR-19b, a member of the miR-17-92 cluster, is the leading identified age-associated miRNA. Additionally members of the miR-17-92 cluster, such as miR-19a (number 4), miR-17 (number 8) and miR-92a (number 12) were also presented among the top age-associated miRNAs, suggesting that the miR-17-92 cluster plays an important role in aging. Through a systematically assessment of plasma-derived miRNA age signature I noted a correlation with chronological age (r=0.60). Although this correlation is very moderate compared to the DNA methylation age (r=0.97) described by Horvath in 2013[6], a relationship exists. I found that miRNA age is predictive of all-cause mortality, with each additional miRNA-age year increasing the

risk of all-cause mortality with 9%. This finding is consistent with a previous study that found whole-blood derived miRNA-delta age to be associated with all-cause mortality with 10% per each year increase[7].

Increased age is a major risk factor for complex diseases including, neurodegeneration and cardiovascular diseases. In **chapters 4**, **5** and **8**, I discussed several miRNAs that were differently expressed between participants diagnosed with disease compared to healthy participants. The expression levels of identified miRNAs, such as miR-4798-3p for atrial fibrillation, miR-193a-5p and miR-122-5p for fatty liver disease, and miR-139-5p for type 2 diabetes, explain some of the susceptibility to the disease. In addition, these miRNAs may explain some etiological aspects of the disease. For example, the identified association of miR-4798-3p with prevalent atrial fibrillation was only significant among men, suggesting that miR-4798-3p may play a role in the not yet fully understood sex difference in atrial fibrillation[8]. In the post-hoc target gene analyzes, I shed light on the underlying disease pathways of the identified miRNAs. For example, in **chapter 4**, I highlighted the link between miR-139-5p and *BLC2* gene, which has been validated previously by *in vitro* experiments[9]. Previous studies have linked *BLC2* to glucose metabolism, suggesting that there may be a mediating role for miR-139-5p in glucose metabolism[10]. Another example is described in **chapter 7**, where I found miR-6124, miR-5196-5p and miR-4292 to be associated with the risk of stroke. Of the 10 target genes of the stroke-associated miRNAs, 8 genes have been reported to be expressed in the brain, which is a predominant tissue for stroke[11]. In addition, a previous stroke GWAS reported that some of the identified target genes of the stroke-associated miRNAs are related to vascular features, such as blood pressure and coronary artery disease[12]. While future research is needed to elucidate the link between identified miRNAs and stroke, this indicates a promising pathophysiological role for miRNAs.

In **chapter 4** and **7**, I found that expression levels of certain miRNAs were associated with an increased risk of developing type 2 diabetes and stroke in respective. Because the miRNAs were quantified before the onset of disease, the associated miRNAs may give important information about the pathophysiology of diseases. During a follow-up period of almost 10 years, I identified positive relationships between four miRNAs (miR-99a-5p, miR-29a-3p, miR-122-5p and miR-125b-5p) and incident diabetes. These findings can contribute to an early detection of type 2 diabetes. On the other hand, I found miR-4664-3p associated with incident diabetes in a negative direction, suggesting a protective effect of miRNA in disease onset. Although functional studies are needed to confirm this protective effect *in vivo*, over-expression of miR-4664-3p could be beneficial to prevent type 2 diabetes. In **chapter 7**, I identified three miRNAs (miR-6124, miR-5196-5p and miR-4292) associated with stroke risk. A similar longitudinal study on indecent stroke found plasma levels of miR-656-3p and miR-941 associated with stroke, confirming that miRNAs are useful as indicators for stroke risk in the general population[13]. Stroke is classified into ischemic stroke and hemorrhagic

stroke. Since the pathophysiology of ischemic strokes differs from hemorrhagic stroke, it is relevant to gain knowledge about the role of miRNAs in underlying processes. Interestingly, I showed that, unlike miR-4292, miR-6124 and miR-5196-5p were associated with coronary artery disease, heart failure, and carotid plaques in a cross-sectional analysis. This could indicate that miR-6124 and miR-5196-5p are involved in pathophysiological processes underlying stroke that are different from miR-4292.

The realization that complex diseases, such as cardiometabolic disorders, are better not investigated by means of one exposure and one outcome, has led us to integrate multiple data in this thesis. The information that can be gathered in each of the omics layers (e.g., genomics, epigenomics, metabolomics) improves our understanding of the disease[14]. To this end, I found four miRNAs (miR-10b-5p, miR-148a-3p, miR-125b-5p and miR-100-5p) associated with lipid traits in three omics layers (genetic variation, DNA methylation, miRNA expression) in **chapter 6**. While an observed association in only one omics layer should not rule out a possible causal effect for a miRNA, significant associations in all layers make this more likely. Moreover, in **chapter 4**, I show that some genetic variants associated with the miRNA expression, miR-eQTLs, are also associated with diabetes-related metabolites[15]. This suggests an interaction between miRNAs and metabolites in which they together affect type 2 diabetes.

The battle between etiology and prediction, how can microRNAs cross this line?

Epidemiology makes a distinction between prediction research and etiological research. In terms of prediction research, the probability of having (diagnostic) or developing (prognostic) an event can be estimated. The main purpose here is risk assessment irrespective of underlying causes. On the other hand, etiological research provides insight into the cause and origin of observations, and does not predict the outcome. Based on their nature, miRNAs can be linked to the underlying mechanisms of diseases. The ability to bind complementarily to gene transcripts and thereby affect gene expression is one aspect that makes miRNAs unmistakably related to pathophysiological conditions. Nevertheless, the observed effects of miRNAs on repressing translation of gene transcripts are discovered inside a cell. Considering miRNAs are released from the cells into the bloodstream upon pathophysiological processes, quantifying extracellular miRNAs may reflect certain states of the body, suggesting the potential of miRNAs as disease biomarkers[16,17]. With these motivations, miRNAs hold promise for both prediction and etiological research. In contrast to etiological research, where one has to correct for possible confounding, in prediction research all that matters is the accuracy of the predictive biomarker. From that perspective, it does not matter whether the identified miRNAs are causal or non-causal for the disease in question. Circulatory miRNAs derived from plasma that reflects a pathophysiological process for

disease A can play a (causal) role in *disease B* by suppressing the translation of target transcripts in tissues relevant to disease B. While miRNAs can cross the line between etiology and prediction based on their nature, I think researchers should be careful about including both prediction and etiological research areas. Most importantly, the research question must be leading in the choice of modeling.

In this thesis, I investigated the prediction ability of the miRNAs by answering the following: are certain miRNAs differentially expressed in plasma of patients versus non-patients and can therefore serve as diagnostic or even prognostic biomarkers of the disease? Then, I investigated whether the identified miRNAs are potentially causally related to the pathophysiology of disease of interest by studying the genomic region of the miRNAs and their potential downstream target genes and pathways using *in silico* analysis.

Is a microRNA a good biomarker?

A biomarker has been defined as biological marker that is objectively measured and evaluated as indicator of normal biological processes, pathological processes, or pharmacological responses to therapeutic interventions[18,19]. Since the main purpose of using a biomarker is to evaluate status ('how') and not to collect information about the cause of status ('why'), biomarker research belongs to prediction research. Some criteria describing an ideal biomarker include: (I) non-invasively detectable from accessible sources such as blood or urine, (II) high sensitivity, and (III) high specificity for the disease.

While this concept is still in its early stages, miRNAs have characteristics that make these small non-coding RNAs as potential promising biomarkers for complex diseases. **Figure 9.1** illustrates a steady increasing number of publications on miRNAs as disease biomarkers in the last two decades; from 2002 till 2020.

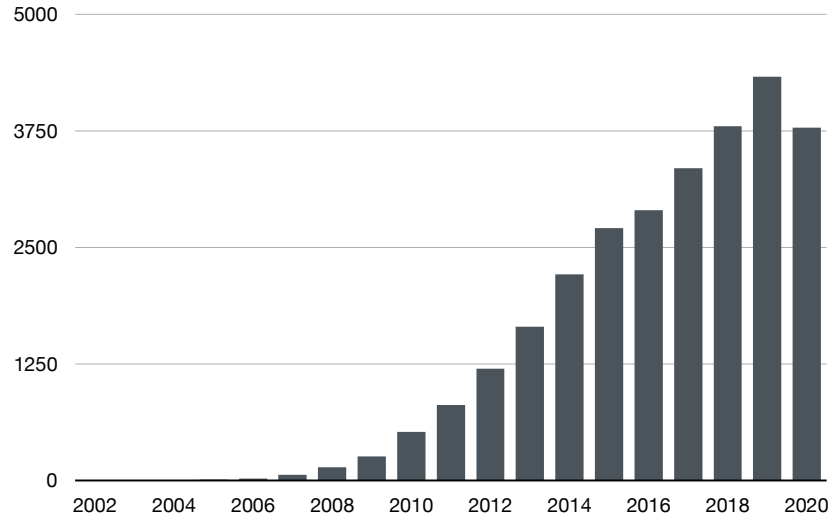


Figure 9.1. Number of publications on miRNAs biomarkers from 2002 till 2020 (PubMed).

Especially, some aspects that makes circulatory miRNAs in body fluids interesting for biomarker discovery is their stability even under extreme conditions (e.g. long term storage and freezing)[20]. This is due to the protective effect of extracellular vesicles that protect the cargo against degradation of, for example, RNAse. These lipid bilayer-delimited particles, such as exosomes or microvesicles, are released from a cell[21,22]. They carry among others proteins, lipids and miRNAs and facilitate the intercellular communication[23]. These extracellular vesicles, and with them miRNAs, are easy to detect in body fluids such as blood, urine and saliva, allowing us to tick the first criteria as ideal biomarker. But the main challenge is to what extent can circulatory miRNAs be specific disease indicators?

It has been established that miRNAs can function as pleiotropic modulators[24]. An example in this thesis is the observed effects of miR-122-5p and miR-193a-5p that were found to be associated with both type 2 diabetes and fatty liver disease. Given that type 2 diabetes and fatty liver disease are both metabolic diseases, these miRNAs may be specific for shared characteristics and common obesity pathways. Another possibility is that the same miRNA has a different function in different phenotypes. For example, previous studies have shown that miR-181b was highly expressed in bladder cancer[25] whereas it was found to be down-regulated in prostate cancer[26], suggesting that the function of miR-181b is different across different tissues. Notably, the use of circulatory miRNAs does not provide information about the origin of tissue. A previous study has demonstrated that the vast majority (~83%) of the mature miRNAs in human were neither tissue specific nor expressed in all tissues (house-keeping miRNAs). A small

proportion (~14%) was found to be tissue specific, including miR-1-3p which was highly expressed in muscle and myocardium tissue[27]. This latter study has determined the abundance of 1997 mature miRNAs in 61 tissues of different organs from two subjects collected post mortem. Unfortunately, little information is currently available about the miRNA expression in different tissues for relatively novel miRNAs (e.g. four digits), such as miR-6124 and miR-5196-5p I identified in **chapter 7**. To circumvent this issue, one approach is to test if the host gene was expressed in the tissue of interest. However, using the host gene expression as proxy for the miRNA is only informative if the miRNA is located in the exonic- or intronic region and likely to be co-expressed with the host gene[28]. For instance, I found that the stroke-associated miR-5195-5p in **chapter 7** was embedded in the exonic region *CD22* gene, which was reported to be expressed in brain tissue[11,29]. Another example is the diabetes-associated miR-139-5p in **chapter 4** that was embedded in the intronic region of *PDE2A*[29]. Previous research has demonstrated that *PDE2A* was expressed in endothelial cells in liver and pancreas tissue[11]. Moreover, a dysfunction of endothelial cells is known to be related to type 2 diabetes[30]. This might suggest that higher levels of miR-139-5p in the circulation are a result of the cellular damage to endothelial cells caused by type 2 diabetes. But this scenario should be still validated by testing the levels of miRNAs in the target tissues.

In this thesis, I aimed to shed light on miRNAs as prognostic biomarkers. Using cross-validation in **chapter 3**, I tested the ability of a miRNA-age signature to predict mortality. This miRNA-age signature was built by 291 miRNAs. In this study I found that miRNA-age could predict mortality in the validation set. Given that the magnitude of the effect estimates of individual miRNAs on age was small, ranging from 0.003-0.04, a large sample size of a multiple-miRNA signature is needed for use as an age-related biomarker. This is in line with the receiver operator characteristics analysis in **chapter 7**, in which I demonstrated a moderate improvement in stroke prediction by adding identified miRNAs to a general vascular risk factor model (ROC vascular= 0.735 versus ROC vascular+miRNAs= 0.761).

Novel statistical methods help to identify miRNAs involved in disease etiology

Causal relationships are important in etiological research, but extremely difficult to prove in observational studies, using cohort data such as the Rotterdam Study[31], as well as in clinical trials. The miRNAs that I have found to be associated with disease prevalence in **chapter 4, 5** and **8** can either reflect the cause or effect of the disease and therefore excluded from cause- and effect relationship. Although the longitudinal analyzes performed in **chapters 4** and **7** provide a temporal framework indicating that exposure (miRNA expression) precedes outcome (type 2 diabetes and stroke), these studies are still subject to (unmeasured) confounding and systemic errors. In recent years, a popular method used in epidemiological research to examine the causal

effect of an exposure on the outcome is Mendelian randomization (MR)[32,33]. This approach is based on the second law of Mendel; the law of independent assortment, which states the allele of one gene separates independently of an allele of other gene. This method uses genetic variants as instrumental variable for the exposure. Since genetic variants are randomly inherited, Mendelian randomization is an approach that is free from confounding and reverse causality and can therefore differentiate between a true causal and non-causal exposure. Yet, valid instrumental variables must meet three assumptions (**Figure 9.2**) [34]:

I. The genetic instrument must associate with the exposure
II. The genetic instrument must influence the outcome only through the exposure
III. The genetic instrument must not associate with measured or unmeasured confounders

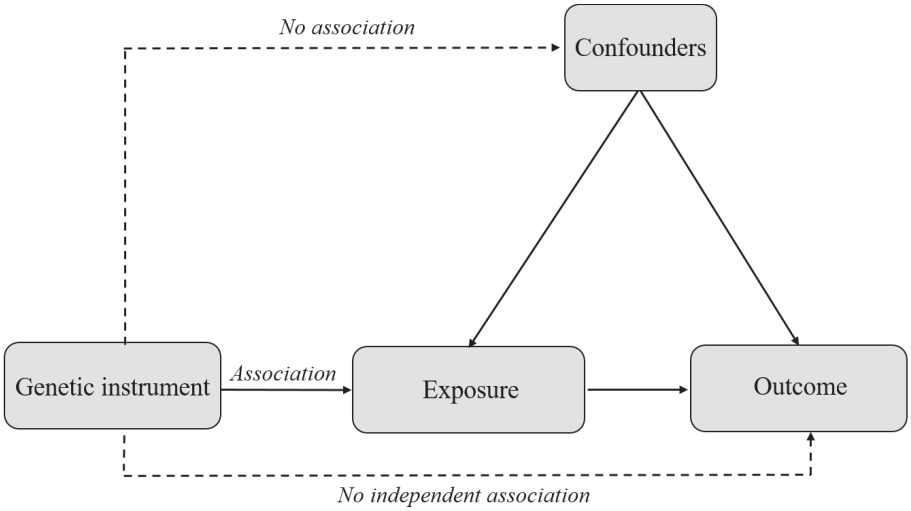


Figure 9.2. Schemetic presentation of the MR method.

In **chapter 4**, I identified a causal effect of miR-139-5p on type 2 diabetes using Mendelian randomization. I used 22 genetic variants for miR-139-5p that were valid according to the MR three assumptions. However, I was not always able to identify valid genetic tools for miRNAs and therefore unable to perform Mendelian randomization testing the causal inference between the identified miRNAs and several traits. Nevertheless, this does not exclude the potential causal role of the identified miRNAs in the pathophysiology of disease. In this context, an important component is the miRNA target gene analysis I performed in **chapters 4, 6** and **7**. By testing

the association between genetic variants in miRNA target genes and the disease of interest, I was able to provide some evidence for a possible mediating effect between the identified miRNAs and disease.

The lack of generalization

In this thesis I provide an overview of miRNAs, that emerged as small emperors of physiological conditions. These findings form the basis of an epidemiological approach to the utility of miRNAs that can serve as accessible biomarkers as well as potential drug targets. While I have marked several miRNAs as biomarkers for metabolic disorders and cardiovascular disease, the clinical implications of miRNAs still have a long way to go. In particular, the lack of external validity is a big challenge that has led to miRNAs remaining in the research phase as biomarkers. The non-standardized quantification and normalization of miRNA expression data are important factors why the thus far identified miRNA biomarkers do not take off the ground.

Quantification of miRNAs

The increasing interest in discovering miRNAs for complex traits has led to better miRNA quantification platforms. In particular, the EdgeSeq array, which has been used in several **chapters** of this thesis, is a next-generation high-throughput technology that is able to measure 2083 mature miRNAs and is reported to quantify with a high accuracy, sensitivity and specificity. Although this EdgeSeq method is still little used in epidemiological studies, it has been shown to detect novel miRNAs with the least bias detection.[35] Aside from the above benefits, the EdgeSeq method is relatively expensive and time consuming, and therefore not often used in similar studies of comparable sample size. In addition, one of the reasons why the relatively novel miRNAs (e.g. miR-6124, miR-5196-5p, miR-4292, miR-4798-3p) were not identified in previous miRNA studies using a RT-qPCR tool is probably due to limited available assays. Future studies that have the ability to use a high-throughput technology like EdgeSeq, should overcome this limitation.

Normalization of miRNAs

In the studies included in this thesis, I used a set of miRNAs that was according the normalization method defined as well-expressed miRNAs. The benefit of using well-expressed miRNAs instead of all measured miRNAs is that I could minimize false-positive findings (type 1 error). This type of error is likely to increase when there is a relative large contribution of accidental outliers compared to the overall effect. The normalization method used in this thesis is based on a monotonic decreasing spline curve fit between the means and standard deviations of all miRNAs. This excludes miRNAs with standard deviations that exceed the means and have low expression levels. Using a lower limit of quantification (LLOQ), which is based on the lowest

concentration of the measured miRNAs that can be accurately quantified, a threshold for defining well-expressed miRNAs was established. In the definition well-expressed miRNA levels in plasma, were those with >50% values above LLOQ. Out of the 2083 measured miRNAs, a set of 591 miRNAs was qualified as well-expressed and used as exposure variables in the studies conducted in this thesis. As this normalization approach is not standardized across studies, I performed a series of sensitivity analyses in **chapter 3** to test if normalization is likely to be a source of discrepancy between studies. I reported that the top hits of associated miRNAs were nearly identical between different cut-offs, from 50% (n=591 miRNAs) to 25% (n=687 miRNAs) and 0% (n=2083 miRNAs) expression values above LLOQ, indicating that normalization is probably not the main reason for the inability to replicate the findings.

The independency of miRNAs and how to correct for that

In the association analyses conducted in this thesis, I tested more than one hypothesis simultaneously. If I do not take into account this multitude of tests, there is a good chance that some of the true null hypothesis will be rejected by chance alone. For example, if I test 50 hypotheses simultaneously, the probability that at least one true null hypothesis will be rejected is 92% ($1-0.95^{50} = 0.92$). By focusing on a particular hypothesis like I did partly in **chapter 6**, the nominal cutoff of P value <0.05 is well-accepted. In the event that 591 miRNAs are tested simultaneously, as I have done in the majority of this thesis, I have to take into account this multitude of tests. There are multiple ways to correct for multiple testing, each with their pros and cons. The Bonferroni correction sets the significant cutoff at 0.05/number of independent tests. This means that for testing 591 miRNAs, the significant threshold will be 8.46×10^{-5}.

Chapters 3, 5 and 6 are examples where I have used the Bonferroni correction. While this method is well established in epidemiological research, the correction is usually a bit too conservative, especially for miRNAs. For example, in **chapter 2**, I discussed several miRNAs that are clustered in the genome. It is very likely that these miRNAs share the same promoter and therefore transcribed at the same time, which is often reflected in miRNA expression profiles[26]. While this is not always the case as with families, such as the miR-499 family, whose members have different expression patterns in different tissues[27], it may raise the question of how independent miRNAs are from each other? As a counter-argument to the dependence on miRNAs, I look at the biological function of the miRNAs. The ability of highly “correlated" miRNAs, including -3p/-5p miRNAs to repress gene transcripts could be different. Because of having different seed sequences, the miRNAs could bind to different target transcripts. For example, according to TargetScan7.2[36], there are 365 putative target genes for miR-150-5p, while for miR-150-3p there are 2570, leading to them targeting different protein-coding transcripts and being involved in different pathways.

Another widely used method to correct for multiple testing is the false discovery rate (FDR). This approach is based on the proportion of false positives among all significant results. By allowing a false positive threshold of 5%, the adjusted P value must be less than 0.05 (FDR <0.05). I used this method in **chapters 4** and **7** of this thesis. The main difference between the Bonferroni and FDR is that the Bonferroni penalizes all P values equally while the FDR penalties the P values according to their rank. Nevertheless, both Bonferroni and FDR correction are accepted methods in epidemiological studies. In addition, in **chapter 8**, I found that among the 591 miRNAs, 142 were independent based on the methodology described by Li & Ji in 2005[37]. By correcting for the independent miRNAs, I used $3.52x10^{-4}$ (0.05/142) as significant threshold. Although this is significantly less stringent compared to the previously mentioned $8.46x10^{-5}$ threshold, this method is widely used in genetic research. Elastic net regularization is a method used to deal with correlation of high-dimensional data (high number of rows and columns) that I used in **chapter 3**. High-dimensional datasets are prone to noise and uncertainties. The elastic net regularization regression combines two penalties methods, lasso (L1) and ridge (L2)[38]. Lasso regression takes into account the magnitudes of the coefficient, resulting in that the effect of some miRNAs can be completely neglected. An advantage is that lasso regression helps in miRNA selection. In the ridge regression, the coefficient of each miRNA is equally penalized, which is useful when there is collinearity between miRNAs, but not for miRNA selection. In the elastic net the penalty value (alpha value) can range between 0 and 1, in which an alpha of 0 corresponds to ridge and alpha of 1 to lasso. In **chapter 3**, I reported an alpha parameter of 0.90, indicating that the L1 (lasso) penalization is predominant.

What does the future hold for miRNAs?

From bench to bedside: microRNA-based therapy

As this chapter is being written, a global pandemic is underway that is having an extreme impact on society. The coronavirus outbreak has infected millions of people and the severity of the COVID-19 infection caused is linked to older and obese individuals, which are major risk factors for cardiovascular diseases[39,40]. This once again shows the burden of this vulnerable population and the importance of risk management. The fight against the coronavirus has taken RNA-based vaccinology into gear. Biotech companies Pfizer / BioNTech and Moderna have developed RNA-based vaccines with over 90% efficiency[41]. Continuously evolving RNA therapies and vaccines underscore the exciting future for this RNA-based drug. Similarly, during the last 20 years of miRNA research in humans, multiple pharmaceutical and biotech companies have addressed the therapeutic potential of miRNAs[42-46]. In this context, miRNA mimics can be used to increase the expression of miRNAs that exhibit a protective effect on the disease[45,47]. For example, in theory miRNA mimicry of miR-4664-3p could be beneficial for type 2 diabetes. In addition, miRNA inhibitors, known as anti-miRs, may suppress the expression or function of miRNAs

associated with a deleterious effect on the disease[48]. In the field of cardiovascular disease, several anti-miRs have been developed, including anti-miR-21 which reverses cardiac remodeling and anti-miR-92a which promotes blood vessel growth[49,50]. One of the major challenges in miRNA-based therapeutics is the off-target effect in tissues other than the tissue of interest[51]. For instance, when miRNAs are expressed in a non-tissue specific manner, manipulation of a specific miRNA can affect target genes in different tissues, which can have both beneficial and deleterious consequences. In addition, most studies will stall in preclinical stages due to inefficient delivery, toxicity and pharmacokinetics[52]. Nonetheless, a success story of anti-miR therapies is that of miR-122 for the treatment of hepatitis C virus[46]. Clearly, we still have a long way to go to manifest miRNA therapies in clinical practice. Nevertheless, the biological function of miRNAs during disease and the success of anti-miR-122 showing evidence for safe and efficient miRNA therapy hold great promise for miRNA as drug targets in the future.

Your personal microRNA makeup

In the coming years, medical research will increasingly focus on personalized medicine. The ultimate goal is to provide an optimal presentation of a person's health status that can illustrate and predict the impact of modifiable factors on aging and disease. To this end, the diagnosis and treatment will be tailored for each individual and adverse side effects will be minimized based on their genetic makeup. Multiple effort in genomics has already contributed to risk profiles that can be applied to support clinical decision making. More recently, the importance of factors that can change over the course of a person's life has been recognized in optimizing individual health care. Among these, gene-environment interactions that can be measured in epigenomics data are extremely important. This thesis provides evidence on the importance of miRNAs as genetic fine-tuners and regulators that should be considered as a key element in such personalized profile. The breakthrough of personalized medicine will with all certainty benefit humankind, but will also bring challenges. One of these challenges regards the large and complex datasets that require innovative tools such as artificial intelligence. Moreover, we face ethical and moral issues such as discrimination by healthcare companies, but also the burden of truth. Looking beyond the current limitations, which are only a matter of time to resolve, there is a bright future ahead for miRNAs.

References

1. Childs BG, Durik M, Baker DJ, van Deursen JM. Cellular senescence in aging and age-related disease: from mechanisms to therapy. *Nat Med* 2015; **21**(12): 1424-35.
2. Kennedy BK, Berger SL, Brunet A, et al. Geroscience: Linking Aging to Chronic Disease. *Cell* 2014; **159**(4): 709-13.
3. Sen P, Shah PP, Nativio R, Berger SL. Epigenetic Mechanisms of Longevity and Aging. *Cell* 2016; **166**(4): 822-39.
4. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. The Hallmarks of Aging. *Cell* 2013; **153**(6): 1194-217.
5. Yang J, Dungrawala H, Hua H, et al. Cell size and growth rate are major determinants of replicative lifespan. *Cell Cycle* 2011; **10**(1): 144-55.
6. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol* 2013; **14**(10): R115.
7. Huan T, Chen G, Liu C, et al. Age-associated microRNA expression in human peripheral blood is associated with all-cause mortality and age-related traits. *Aging Cell* 2018; **17**(1): e12687.
8. Ko D, Rahman F, Schnabel RB, Yin X, Benjamin EJ, Christophersen IE. Atrial fibrillation in women: epidemiology, pathophysiology, presentation, and prognosis. *Nat Rev Cardiol* 2016; **13**(6): 321-32.
9. Huang H-Y, Lin Y-C-D, Li J, et al. miRTarBase 2020: updates to the experimentally validated microRNA–target interaction database. *Nucleic Acids Research* 2019; **48**(D1): D148-D54.
10. Gurzov EN, Eizirik DL. Bcl-2 proteins in diabetes: mitochondrial pathways of β-cell death and dysfunction. *Trends in Cell Biology* 2011; **21**(7): 424-31.
11. Uhlén M, Fagerberg L, Hallström BM, et al. Tissue-based map of the human proteome. *Science* 2015; **347**(6220): 1260419.
12. Malik R, Chauhan G, Traylor M, et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nature Genetics* 2018; **50**(4): 524-37.
13. Mick E, Shah R, Tanriverdi K, et al. Stroke and Circulating Extracellular RNAs. *Stroke* 2017; **48**(4): 828-34.
14. Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. *Genome Biology* 2017; **18**(1): 83.
15. Lotta LA, Pietzner M, Stewart ID, et al. A cross-platform approach identifies genetic regulators of human metabolism and health. *Nature Genetics* 2021; **53**(1): 54-64.
16. Mitchell PS, Parkin RK, Kroh EM, et al. Circulating microRNAs as stable blood-based markers for cancer detection. *Proceedings of the National Academy of Sciences* 2008; **105**(30): 10513.
17. Chen X, Ba Y, Ma L, et al. Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. *Cell Res* 2008; **18**(10): 997-1006.
18. Biomarkers Definitions Working G. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther* 2001; **69**(3): 89-95.
19. Mayeux R. Biomarkers: potential uses and limitations. *NeuroRx* 2004; **1**(2): 182-8.
20. Ge Q, Zhou Y, Lu J, Bai Y, Xie X, Lu Z. miRNA in Plasma Exosome is Stable under Different Storage Conditions. *Molecules* 2014; **19**(2).

21. van Niel G, Porto-Carreiro I, Simoes S, Raposo G. Exosomes: A Common Pathway for a Specialized Function. *The Journal of Biochemistry* 2006; **140**(1): 13-21.
22. Doyle LM, Wang MZ. Overview of Extracellular Vesicles, Their Origin, Composition, Purpose, and Methods for Exosome Isolation and Analysis. *Cells* 2019; **8**(7): 727.
23. Valadi H, Ekström K, Bossios A, Sjöstrand M, Lee JJ, Lötvall JO. Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nature Cell Biology* 2007; **9**(6): 654-9.
24. Bartel DP. MicroRNAs: Target Recognition and Regulatory Functions. *Cell* 2009; **136**(2): 215-33.
25. Ratert N, Meyer HA, Jung M, et al. Reference miRNAs for miRNAome analysis of urothelial carcinomas. *PLoS One* 2012; **7**(6): e39309.
26. Schaefer A, Jung M, Mollenkopf HJ, et al. Diagnostic and prognostic implications of microRNA profiling in prostate carcinoma. *Int J Cancer* 2010; **126**(5): 1166-76.
27. Ludwig N, Leidinger P, Becker K, et al. Distribution of miRNA expression across human tissues. *Nucleic acids research* 2016; **44**(8): 3865-77.
28. Lutter D, Marr C, Krumsiek J, Lang EW, Theis FJ. Intronic microRNAs support their host genes by mediating synergistic and antagonistic regulatory effects. *BMC Genomics* 2010; **11**(1): 224.
29. Hinske LC, França GS, Torres HA, et al. miRIAD-integrating microRNA inter- and intragenic data. *Database (Oxford)* 2014; **2014**.
30. Kaur R, Kaur M, Singh J. Endothelial dysfunction and platelet hyperactivity in type 2 diabetes mellitus: molecular insights and therapeutic strategies. *Cardiovasc Diabetol* 2018; **17**(1): 121.
31. Ikram MA, Brusselle G, Ghanbari M, et al. Objectives, design and main findings until 2020 from the Rotterdam Study. *European Journal of Epidemiology* 2020; **35**(5): 483-517.
32. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003; **32**(1): 1-22.
33. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Human Molecular Genetics* 2014; **23**(R1): R89-R98.
34. Bowden J, Del Greco M F, Minelli C, Davey Smith G, Sheehan N, Thompson J. A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Stat Med* 2017; **36**(11): 1783-802.
35. Godoy PM, Barczak AJ, DeHoff P, et al. Comparison of Reproducibility, Accuracy, Sensitivity, and Specificity of miRNA Quantification Platforms. *Cell Reports* 2019; **29**(12): 4212-22.e5.
36. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. *Elife* 2015; **4**.
37. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity (Edinb)* 2005; **95**(3): 221-7.
38. Zou H, Hastie T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2005; **67**(2): 301-20.
39. Popkin BM, Du S, Green WD, et al. Individuals with obesity and COVID-19: A global perspective on the epidemiology and biological relationships. *Obesity Reviews* 2020; **21**(11): e13128.
40. Stefan N, Birkenfeld AL, Schulze MB. Global pandemics interconnected — obesity, impaired metabolic health and COVID-19. *Nature Reviews Endocrinology* 2021; **17**(3): 135-49.
41. Polack FP, Thomas SJ, Kitchin N, et al. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *New England Journal of Medicine* 2020; **383**(27): 2603-15.

42. Christopher AF, Kaur RP, Kaur G, Kaur A, Gupta V, Bansal P. MicroRNA therapeutics: Discovering novel targets and developing specific therapy. *Perspect Clin Res* 2016; **7**(2): 68-74.

43. Rupaimoole R, Slack FJ. MicroRNA therapeutics: towards a new era for the management of cancer and other diseases. *Nature Reviews Drug Discovery* 2017; **16**(3): 203-22.

44. van Rooij E, Kauppinen S. Development of microRNA therapeutics is coming of age. *EMBO Molecular Medicine* 2014; **6**(7): 851-64.

45. van Rooij E, Olson EN. MicroRNA therapeutics for cardiovascular disease: opportunities and obstacles. *Nature Reviews Drug Discovery* 2012; **11**(11): 860-72.

46. Janssen HLA, Reesink HW, Lawitz EJ, et al. Treatment of HCV Infection by Targeting MicroRNA. *New England Journal of Medicine* 2013; **368**(18): 1685-94.

47. Montgomery RL, Yu G, Latimer PA, et al. MicroRNA mimicry blocks pulmonary fibrosis. *EMBO Molecular Medicine* 2014; **6**(10): 1347-56.

48. Stenvang J, Petri A, Lindow M, Obad S, Kauppinen S. Inhibition of microRNA function by antimiR oligonucleotides. *Silence* 2012; **3**(1): 1.

49. Bonauer A, Carmona G, Iwasaki M, et al. MicroRNA-92a Controls Angiogenesis and Functional Recovery of Ischemic Tissues in Mice. *Science* 2009; **324**(5935): 1710.

50. Thum T, Gross C, Fiedler J, et al. MicroRNA-21 contributes to myocardial disease by stimulating MAP kinase signalling in fibroblasts. *Nature* 2008; **456**(7224): 980-4.

51. Stenvang J, Petri A, Lindow M, Obad S, Kauppinen S. Inhibition of microRNA function by antimiR oligonucleotides. *Silence* 2012; **3**(1): 1-.

52. van Rooij E, Purcell Angela L, Levin Arthur A. Developing MicroRNA Therapeutics. *Circulation Research* 2012; **110**(3): 496-507.

10

CHAPTER 10

Appendices

Summary of Research findings

The studies in this thesis are related to two main research objectives. The first aim was to elucidate the role of miRNAs on aging. I provided an overview of miRNAs involved in the cell cycle progression and investigated whether miRNAs are linked to biological aging in an advanced population. The results related to this objective are presented in **Part I** of this thesis. The second aim was to investigate the role of miRNAs in diseases. I shed light on the biomarker potential of circulatory miRNAs in disease and investigated the functional role these miRNAs in the pathophysiology. In this section I conducted observational research using data from the Rotterdam Study and examined the interplay of multiple omics layers in a systemic biological approach. The results related to this objective are presented in **Part II** of this thesis.

In **chapter 2** I reviewed the current knowledge regarding the role of miRNAs in cell cycle progression of stem cells. The cell cycle is a series of events that take place in a cell to grow and divide. The literature included in this chapter described key miRNAs involved in different phases of the cell cycle, including the clusters miR-290-295, miR-302, miR-17-92, miR-106b-25 and miR-106a-363. These phases are under control by many regulators, such as p21 and p53 that are responsible for cell arrest when there are defects in the cell cycle. I focused on embryonic stem cells, adult stem cells and cancer stem cells.

In **chapter 3** I created an miRNA-age signature based on plasma miRNA expression levels and tested that were associated with mortality and morbidities. I identified 291 miRNAs to be associated with chronological age among 1,930 individuals from the Rotterdam Study. For this, I used elastic net regularization and cross-validation and showed that miRNA-age was correlated with chronological age. Based on the difference between miRNA-age and chronological age, I defined miRNA-delta age. Accelerated aging equates to a positive miRNA-delta age, while slower aging equates to a negative miRNA-delta age. I found that miRNA-delta age is predictive of all-cause mortality, with each additional miRNA-delta age year increasing the risk of all-cause mortality with 9%. In contrast, I did not find an association between miRNA-delta age and morbidities.

In **chapter 4** I examined the association between miRNAs plasma levels and type 2 diabetes. I observed a positive relationship between four miRNAs (miR-99a-5p, miR-29a-3p, miR-122-5p and miR-125b-5p) and incident diabetes during follow-up. These findings can contribute to an early detection of individuals with high-risk profiles for type 2 diabetes. On the other hand, I found a miRNA (miR-4664-3p) associated with incident diabetes in a negative direction, suggesting a protective effect. In addition, I provided evidence that plasma levels of miR-139-5p and miR-193a-5p were significantly increased in baseline diabetic patients. I performed a GWAS on the seven identified miRNAs to obtain genetic variants. Based on Mendelian randomization using these GWAS summary statistics I was able to show a causal effect of miR-139-5p on type 2

diabetes. Downstream effects of miR-139-5p, including target genes have previously been reported to be involved in pathways underlying diabetes, supporting the possible etiological role of miR-139-5p in type 2 diabetes.

In **chapter 5** I described the association between miRNAs plasma levels and the prevalence of fatty liver disease. I tested the associations between miRNAs and fatty liver disease and liver enzymes (gamma-glutamyl transferase and alkaline phosphatase). I provided evidence that 61 miRNAs were associated with liver enzymes. Furthermore 17 miRNAs were associated with fatty liver disease, of which 14 were also associated with liver enzymes. Among them I found miR-122-5p, which is a well-established miRNA in liver tissue.

In **chapter 6** I described a study in which I integrated genetic data, DNA methylation data and miRNA expression data from the general population to investigate miRNAs associated with cardiometabolic traits. By GWAS using summary statistics I identified genetic variants annotated to 67 miRNAs associated with cardiometabolic diseases (e.g. coronary heart disease and type 2 diabetes) or risk factors (e.g. anthropometric traits, blood pressure traits, glycemic traits and lipid traits). Alterations in DNA methylation of CpG sites annotated to 38 of these miRNAs and plasma expression levels of 8 of them were also associated with the same trait. Integrating the results from different omics data showed that miR-10b-5p, miR-148a-3p, miR-125b-5p and miR-100-5p are strongly linked traits. This study provides evidence that miRNAs are related to cardiometabolic traits in more than two omics layers, which may enhance the potential to be involved in the etiology of disease.

In **chapter 7** I reported the results of the association between miRNA plasma levels and incident stroke. I was able to identify three circulatory miRNAs (miR-6124, miR-5196-5p and miR-4292) associated with the risk of any type of stroke. These findings suggest that higher plasma levels of miR-6124, miR-5196-5p and miR-4292 are linked to higher risk for developing stroke. Using GWAS summary statistics I observed that 10 of the potential miRNA target genes were associated with stroke. Among these 10 target genes, 8 were reported to be expressed in the brain, which is a primary tissue for stroke. All in all, this reinforces the hypothesis that miR-6124, miR-5196-5p and miR-4292 are not only a biomarker but are also involved in the pathology of stroke.

In **chapter 8** I reported the findings on the association between miRNAs and prevalent and incident atrial fibrillation in the Rotterdam Study. Since sex differences exist in the pathology of atrial fibrillation pathology, I stratified the analysis by sex. I identified lower expression levels of miR-4798-3p associated with prevalent atrial fibrillation in men, but not in women or the general population, indicating that miR-4798-3p may have a sex-specific role in atrial fibrillation. I described in this study that miR-4798-3p has a predicted number of more than 50 target genes it can regulate, several of which are related to electrical and structural remodeling of the heart known to be involved in the pathophysiology of atrial fibrillation.

Nederlandse samenvatting

De studies in dit proefschrift hebben betrekking op twee onderzoeksdoelstellingen. Het eerste doel was om de rol van miRNA's op veroudering op te helderen. Ik gaf een overzicht van miRNA's die betrokken zijn in de celcyclus en onderzocht of miRNA's verband houden met biologische veroudering in een gevorderde populatie. De resultaten met betrekking tot deze doelstelling worden gepresenteerd in het eerste deel van dit proefschrift. Het tweede doel was om de rol van miRNA's bij ziekten te onderzoeken. Ik wierp licht op het biomarkerpotentieel van circulatoire miRNA's bij ziekte en onderzocht de functionele rol van deze miRNA's in de pathofysiologie. In deze sectie heb ik observationeel onderzoek gedaan met behulp van gegevens uit de Rotterdam Studie en het samenspel van meerdere omics-lagen in een systemisch biologische benadering onderzocht. De resultaten met betrekking tot deze doelstelling worden gepresenteerd in het tweede deel van dit proefschrift.

In **hoofdstuk 2** heb ik de huidige kennis over de rol van miRNAs in celcyclus van stamcellen besproken. De celcyclus is een reeks gebeurtenissen die plaatsvinden in een cel om te groeien en te delen. De literatuur in dit hoofdstuk beschrijft de belangrijkste miRNA's die betrokken zijn bij verschillende fasen van de celcyclus, waaronder de clusters miR-290-295, miR-302, miR-17-92, miR-106b-25 en miR-106a-363. Deze fasen worden gecontroleerd door veel regulatoren, zoals p21 en p53, die verantwoordelijk zijn voor bij defecten in de celcyclus. Ik concentreerde me op embryonale stamcellen, volwassen stamcellen en kankerstamcellen.

In **hoofdstuk 3** heb ik een miRNA-leeftijdshandtekening gemaakt op basis van plasma-miRNA-expressieniveaus en de associatie met mortaliteit en morbiditeit getest. Ik identificeerde 291 miRNA's die geassocieerd zijn met chronologische leeftijd onder 1,930 individuen uit de Rotterdam Studie. Hiervoor gebruikte ik elastic net regularization en cross- validatie en liet zien dat miRNA-leeftijd gecorreleerd was met chronologische leeftijd. Op basis van het verschil tussen miRNA-leeftijd en chronologische leeftijd, heb ik miRNA-delta-leeftijd gedefinieerd. Versnelde veroudering staat gelijk aan een positieve miRNA-delta-leeftijd, terwijl langzamere veroudering gelijk staat aan een negatieve miRNA-delta-leeftijd. Ik ontdekte dat miRNA-delta-leeftijd voorspellend is voor sterfte, waarbij elk extra miRNA-delta-leeftijdsjaar het risico op sterfte door alle oorzaken met 9% verhoogt. Daarentegen vond ik geen verband tussen miRNA-delta-leeftijd en morbiditeit.

In **hoofdstuk 4** heb ik de associatie tussen miRNAs plasma-expressieniveaus en type 2 diabetes onderzocht. Ik heb een positieve relatie waargenomen tussen vier miRNA's (miR-99a-5p, miR-29a-3p, miR-122-5p en miR-125b-5p) en incidentele diabetes tijdens de follow-up. Deze bevindingen kunnen bijdragen aan een vroege

detectie van personen met een hoog risicoprofiel voor diabetes type 2. Aan de andere kant vond ik een miRNA (miR-4664-3p) geassocieerd met incidentele diabetes in een negatieve richting, wat een beschermend effect suggereert. Bovendien leverde ik bewijs dat de plasma-expressieniveaus van miR-139-5p en miR-193a-5p significant verhoogd waren bij prevalente diabetespatiënten. Ik heb een GWAS uitgevoerd op de zeven geïdentificeerde miRNA's om genetische varianten te verkrijgen. Op basis van Mendeliaanse randomisatie met behulp van deze GWAS-samenvattende statistieken kon ik een causaal effect van miR-139-5p op diabetes type 2 aantonen. Eerder is gerapporteerd dat target genen van miR-139-5p betrokken zijn bij diabetes, wat de mogelijke etiologische rol van miR-139-5p bij type 2-diabetes ondersteunt.

In **hoofdstuk 5** beschreef ik de associatie tussen miRNAs plasma-expressieniveaus en de prevalentie van leververvetting. Ik heb de associaties tussen miRNA's en leververvetting en leverenzymen (gamma-glutamyltransferase en alkalische fosfatase) getest. Ik leverde bewijs dat 61 miRNA's geassocieerd waren met leverenzymen. Bovendien waren 17 miRNA's geassocieerd met leververvetting, waarvan 14 ook geassocieerd waren met leverenzymen. Onder hen vond ik miR-122-5p, een veel bestudeerd miRNA in leverweefsel.

In **hoofdstuk 6** beschreef ik een studie waarin ik genetische data, DNA-methyleringsdata en miRNA-expressiedata van de algemene bevolking integreerde om miRNA's te onderzoeken die geassocieerd zijn met cardiometabole eigenschappen. Met behulp van GWAS summary statistics identificeerde ik genetische varianten die waren geannoteerd op 67 miRNA's en waren geassocieerd met cardiometabole ziekten (bijv. Coronaire hartziekte en type 2 diabetes) of risicofactoren (bijv. Antropometrische kenmerken, bloeddrukkenmerken, glykemische kenmerken en lipidenkenmerken). Ik vond veranderingen in DNA-methylatie op CpG-sites, geannoteerd op 38 van de eerder geïdentificeerde miRNA's die geassocieerd zijn met dezelfde cardiometabole eigenschap. Ook vond ik expressieniveaus van 8 van de eerder geïdentificeerde miRNA's geassocieerd met dezelfde cardiometabole eigenschap. Door de resultaten van de drie verschillende omics-gegevens te integreren, bleek dat miR-10b-5p, miR-148a-3p, miR-125b-5p en miR-100-5p waren geassocieerd met lipiden. Deze studie levert het bewijs dat miRNA's gerelateerd zijn aan cardiometabole eigenschappen in meer dan twee omics-lagen, wat het potentieel om betrokken te zijn bij de etiologie van ziekte kan vergroten.

In **hoofdstuk 7** rapporteerde ik de resultaten van de associatie tussen miRNA plasma-expressieniveaus en een beroerte. Ik was in staat om drie circulatoire miRNA's (miR-6124, miR-5196-5p en miR-4292) te identificeren die verband houden met het risico op beroerte. Deze bevindingen suggereren dat hogere plasma-expressieniveaus van miR-6124, miR-5196-5p en miR-4292 verband houden met een hoger risico op het ontwikkelen van een beroerte. Met behulp van GWAS-summary statistics observeerde

ik dat 10 van de potentiële miRNA-target genen geassocieerd waren met een beroerte. Van deze 10 target genen werd eerder gemeld dat er 8 tot expressie werden gebracht in de hersenen, het primaire weefsel voor een beroerte. Al met al versterkt dit de hypothese dat miR-6124, miR-5196-5p en miR-4292 niet alleen een biomarker zijn, maar ook betrokken zijn bij de pathologie van een beroerte.

In **hoofdstuk 8** rapporteerde ik de bevindingen over de associatie tussen miRNA's en prevalente en incidentele atriumfibrilleren in de Rotterdam Studie. Omdat er sekseverschillen bestaan in de pathologie van atriale fibrillatiepathologie, heb ik de analyse gestratificeerd naar geslacht. Ik identificeerde dat lagere expressieniveaus van miR-4798-3p geassocieerd waren met prevalente atriale fibrillatie bij mannen, maar niet bij vrouwen of de algemene bevolking, wat aangeeft dat miR-4798-3p mogelijk een geslacht-specifieke rol heeft bij atriumfibrilleren. Ik beschreef in deze studie dat miR-4798-3p meer dan 50 target genen heeft die het kan reguleren, waarvan er verschillende gerelateerd zijn aan elektrische en structurele hermodellering van het hart waarvan bekend is dat ze betrokken zijn bij de pathofysiologie van atriumfibrilleren.

List of publiations

Mens MMJ, Ghanbari M. Cell Cycle Regulation of Stem Cells by MicroRNAs. Stem Cell Rev Rep. 2018 Jun;14(3):309-322. doi: 10.1007/s12015-018-9808-y. PMID: 29541978; PMCID: PMC5960494.

Wu JW, Mens MMJ, Goudsmit J, Ma Y, Liang L, Hofman A, Ikram MA, Ghanbari M. Distinct Plasma MicroRNA Aging Signature Predicts Lifespan but not Health-span. *Submitted*

Mens MMJ, Mustafa R, Ahmadizar F, Ikram MA, Evangelou M, Kavousi M, Dehghan A, Ghanbari M. MiR-139-5p is a causal biomarker for type 2 diabetes; Results from genome-wide microRNA profiling and Mendelian randomization analysis in a population-based study. *Submitted*

Zhang X, Mens MMJ, Abozaid YJ, Bos D, Darwish Murad S, de Knegt RJ, Ikram MA, Pan Q, Ghanbari M. Circulatory microRNAs as potential biomarkers for fatty liver disease: the Rotterdam study. Aliment Pharmacol Ther. 2021 Feb;53(3):432-442. doi: 10.1111/apt.16177. Epub 2020 Nov 27. PMID: 33244812; PMCID: PMC7839694.

Mens MMJ, Maas SCE, Klap J, Weverling GJ, Klatser P, Brakenhoff JPJ, van Meurs JBJ, Uitterlinden AG, Ikram MA, Kavousi M, Ghanbari M. Multi-Omics Analysis Reveals MicroRNAs Associated With Cardiometabolic Traits. Front Genet. 2020 Feb 27;11:110. doi: 10.3389/fgene.2020.00110. PMID: 32174972; PMCID: PMC7056871.

Mens MMJ, Heshmatollah A, Fani L, Ikram MA, Ikram MK, Ghanbari M. Circulatory MicroRNAs as Potential Biomarkers for Stroke Risk: The Rotterdam Study. Stroke. 2021 Mar;52(3):945-953. doi: 10.1161/STROKEAHA.120.031543. Epub 2021 Feb 10. PMID: 33563011.

Geurts S, Mens MMJ, Bos MM, Ikram MA, Ghanbari M, Kavousi M. Association between Circulatory microRNAs in Plasma and Atrial Fibrillation among Men and Women from the General Population: the Rotterdam Study. *Submitted*

Maas SCE, Mens MMJ, Kühnel B, van Meurs JBJ, Uitterlinden AG, Peters A, Prokisch H, Herder C, Grallert H, Kunze S, Waldenberger M, Kavousi M, Kayser M, Ghanbari M. Smoking-related changes in DNA methylation and gene expression are associated with cardio-metabolic traits. Clin Epigenetics. 2020 Oct 22;12(1):157. doi: 10.1186/s13148-020-00951-0. PMID: 33092652; PMCID: PMC7579899.

Licher S, Terzikhan N, Splinter M, Velek P, van Rooij F, Verkroost-van Heemst J, A.E.G. Haarman AEG, Thee EF, Geurts S, Mens MMJ, van der Schaft N, de Feijter M, Cortes L, Leening Kieboom B, Ikram MA. Design, implementation and initial findings of COVID-19 research in the Rotterdam Study: leveraging existing infrastructure for population-based investigations on an emerging disease. European Journal of Epidemiology. *Accepted for publication.*

Mens MMJ, Keijsers L, Dietvorst E, Koval S, Legerstee J, Hillegers MHJ. The Grow It! app reveals mood trajectories of Dutch adolescents during the COVID-19 pandemic. *In preparation.*

E.Dietvorst, Legerstee J, Vreeker A, Koval S, Mens MMJ, Keijsers L, Hillegers MHJ. The Grow It! Serious Gaming App, longitudional changes in adolescents well-being during COVID-19. *In preparation*

Mustafa R, Mens MMJ, Huang J, Roshchupkin G, Uitterlinden AG, Ikram MA, Evangelou M, Dehghan A, Ghanbari M. Understanding the genetic architecture of microRNAs and building a roadmap of their relationships with clinical traits using population data. *In preparation.*

Yaqub A, Mens MMJ, Ikram MK, Ghanbari M, Ikram MA. Plasma microRNAs associated with general cognitive function and dementia: a population-based study. *In preparation*

Word of thanks

August 2021

Finalizing this doctorate feels like a victory of many challenges I have faced. I am proud I have conquer them. I am grateful for everybody I have met during this period. You were an enormous inspiration for me. A few deserve special mention.

My foremost gratitude is to my promotor and co-promotor.

Beste Arfan Ikram, beste Arfan, ik herinner me nog goed onze eerste ontmoeting in maart 2018. Ik was toen laatstejaars master student Biomedische Wetenschappen en liep stage op de Epidemiologie afdeling. Je bood me aan om een PhD te gaan starten en daarnaast een masterprogramma te volgen in Health Sciences… in 2.5 jaar tijd. Met een achtergrond in fundamenteel onderzoek, beperkte kennis over termen zoals odds- en hazard ratios, en moeite met het uitspreken van het woord “epi-demi-ologie”, vond ik dit vrij uitdagend. Ik ben erg blij dat je me deze kans hebt gegeven. Ondanks dat deze 2.5 jaar zeker niet altijd even makkelijk waren, heb ik er geen moment spijt van gehad. Arfan, ik wil ik je heel erg bedanken voor je vertrouwen in mij en het feit dat je me hebt blijven uitdagen. Jouw vermogen om dingen in een breder perspectief te zien, heeft me enorm geholpen bij het stellen van prioriteiten en het verminderen van de werkdruk wanneer dat nodig was. Veel dank en ik hoop dat we in de toekomst kunnen blijven samenwerken.

Dear Mohsen Ghanbari, dear Mohsen, where do I begin… You are the one who made me fall in love with microRNAs. Thank you for this incredible opportunity you have given me. I admire your support over the years we have worked together. While I have to admit that your eye for detail was sometimes annoying for me as a not-so-patient person to keep working on those manuscripts until they were perfectly polished, I've come to appreciate this. You have certainly contributed to shaping me professionally and personally. Thank you for your patience, your faith in me and your guidance during my PhD.

I would like to express my gratitude to the reading committee, Prof.dr. Bas Heijmans, Prof.dr. Joyce van Meurs and Dr. Erik Wiemer, for their time and attention to read my dissertation. To the members of the large doctoral committee, thank you very much for participating in my PhD defense. It is a great honor to have you as members of my doctoral committee.

Ik wil graag de deelnemers van de Rotterdam Study bedanken. Zonder jullie was dit proefschrift er niet geweest. In het bijzonder wil ik de 2754 deelnemers bedanken waarbij de expressie van microRNAs is gemeten. Uiteraard gaat mijn dank ook uit naar de medewerkers van het ERGO centrum en het data-management team.

Since the work in this thesis is mainly based on collaborations, I would like to thank all the co-authors of the papers with whom I have collaborated in recent years.

Dear Julia Wu, dear Julia, we met in Rotterdam in 2019 to collaborate on a project in miRNA aging (chapter 3 in this thesis). Besides this is one of my favorite projects I have conducted, I really enjoyed our collaboration. We had always fun. I admire you in your mentorship, your personal vision and in your scientific work. I am more than grateful that you offered me a postdoctoral position at Harvard T.H. Chan to continue working with you on biological aging. I look forward to these exciting times to come.

Dear Rima, it was a pleasure to work with such a smart and diligent lady like you. You have a pure heart. Our common love for miR-139-5p and miR-193-5p was real. I'm curious what the future will bring you.

Querida Silvana, quiero agradecerles por toda su ayuda. Je hebt me enorm geholpen om R een beetje onder de knie te krijgen. Wat vond ik het jammer dat je naar Barcelona ging verhuizen, maar was zo blij voor je dat je je dromen achterna durfde te gaan! Jouw doorzettingsvermogen is enorm. Ik ben trots op je!

Dear Xiaofang, It was a pleasure to work with you! It was great watching you grow, both personally and professionally. There is a bright future ahead of you.

Beste Sven, bedankt voor de samenwerking aan het AF en microRNA project! Ik zal onze discussies en de middagwandelingen in het park gaan missen. Met jouw ijzeren discipline word jij vast en zeker een top thorax chirurg.

My colleagues from the MolSysEpi, Cardiometabolic and Nutrition & Lifestyle groups, many thanks to Amber, Shahzad, Ivana, Eliana, Yasir, Maryam, Elif, Fariba, Jaap, Martijn, Maxime, Mirjam, Carolina, Trudy, Mónica, Kim, Anh Nhi, Eralda, Natalie, Alis, Lana, Joyce, Janine, Maud, Silvan, Sadaf, Zhangling, Oscar, and to all whose name I might have missed (sorry). Merel en Cindy, bedankt voor de gezelligheid tijdens de ASHG conference in Houston. Maaike, heel erg bedankt voor je goede hulp! Amy, bedankt voor je goede humor tijdens de middagwandelingen. Annet, hebben we samen toch maar mooi die Moleculaire en Genetische Epidemiologie master in de pocket! Het was een genoegen om dit samen met jou te mogen doen. Bedankt voor de etentjes/koffietjes! Banafsheh joon, you were my R hero.. my solution to all those errors. Actually you should have been included in all acknowledgment parts of this thesis ;).

I had great fun working together with you. Nano, mijn andere R en IT-gerelateerde superheld! Heel erg bedankt voor je hulp. Je was altijd vrolijk en goedlachs. Marlou, jij was toch wel mijn epi-buddy. Deze PhD zou niet zo leuk geweest zijn zonder jou. Ik denk met veel plezier terug op onze koffie/thee breaks, het sparren over werk en niet-werk gerelateerde dingen, en onze uitjes (bowlen, bolderen, lunch-uitjes, friet-dates, koekela etc.). Ik ga je enorm missen en weet dat je altijd welkom bent in Boston.

Van februari 2021 tot augustus 2021 heb ik gewerkt als postdoc aan de afdeling Kinder en Jeugd Psychiatrie/Psychologie. Lief Grow It! Team, ik wil jullie heel erg bedanken voor de fijne samenwerkingen, mijn dank gaat uit naar Manon, Loes, Jeroen, Evelien, Leonie, Michelle en Soldado. Manon, bedankt voor je vertrouwen om mij aan te nemen als postdoc. Naast mijn enorme bewondering voor jou als onderzoeker, afdelingshoofd, psychiater, en saleswoman ;), wil ik je heel erg bedanken dat je zo'n goede mentor voor me geweest bent. Loes, wat ben jij een inspiratie voor mij geweest! Jouw passie voor de wetenschap, je enthousiasme en het feit dat ik je voor iedere vraag kon bellen, heb ik enorm gewaardeerd. Jeroen, bedankt voor de fijne gesprekken en het feit dat je altijd tijd voor me kon vrijmaken. Evelien, je bent een topper! Ik ken weinig mensen die zo van aanpakken houden als jij. En onthoud, keep shedding that light.

Mijn favoriete buurvouw, Lisa, you're a power woman. Bedankt voor de fijne afleiding tijdens mijn PhD. Onze dinner- en koffie dates, bubbels, wandelingen, bubbels, wijn, bubbels, cocktails, weekendjes weg, en bubbels (omdat we al onze overwinningen in stijl moeten vieren), hebben deze jaren een stuk leuker gemaakt ;)!

Ook wil ik Crossfit nultien bedanken voor alle WODs, de PR's, en de gezelligheid. Jullie hebben me zowel mentaal als fysiek fit gehouden. All I need to say: GET IT!!!

Lieve Ayda en Rocío, bedankt dat jullie mijn paranimfen willen zijn! Ayda, you have a beautiful soul with a pure heart. Ik zou letterlijk niet weten wat jij in Rotterdam moet zonder mij ;)! Ik ben zo blij dat ik jou als beastie bestie heb in mijn leven. Naast dat ik enorm met je kan lachen, blijven we elkaar uitdagen. Met name stel jij mijn richtingsgevoel met regelmaat op de proef wanneer we onderweg zijn met de Toyota, want voor je het weet zitten we in de Maastunnel. Ik ben blij en trots dat je naast me staat tijdens mijn verdediging. Rocío, hermosa y dulce hermana, ik ben enorm blij om jou als zus te hebben. I didn't hesitate to ask you to be my paranymph, aka bridesmaid, because we are a great team together. Although I have to admit that my sense of rhythm is already quite advanced, I wish I could someday be such a great dancer like you. You should know that I am more than proud of you coming to the Netherlands and I hope we can spend more time together soon. Te quiero <3.

Mijn lieve broertjes, Vic en Jip. Ik zal voor altijd jullie lievelingszus zijn, hoewel ik straks wel graag met doctor wil worden aangesproken (koplopertje mag ook). Ik ben ontzettend blij met jullie als broertjes. Lieve Vic, wat ben ik enorm trots op jou! De echte zakenman en carrièretijger van de familie. Jij rockt hem bij de Rabo en wie weet waar in de toekomst allemaal nog meer. Ik ben benieuwd waar ik jou en Rocío (en Po) later kan komen opzoeken, misschien in Portugal, Spanje of Peru. Lieve Jip, jij bent een echte strijder. Je bent een talent op de tennisbaan, superslim, grappig en ook nog eens lief. Ik hoop stiekem dat je volgend jaar gaat tennissen en studeren in Boston zodat ik je veel kan zien.

Lieve pap & mam, hoewel ik jullie sowieso tekort doe als ik geen heel boek wijd om jullie te bedanken is deze laatste alinea voor jullie. De onvoorwaardelijke support en liefde die jullie mij hebben gegeven, zijn de redenen geweest dat ik dit proefschrift succesvol heb afgerond. Ik heb me geen betere ouders kunnen wensen. Lieve mam, bedankt dat je er altijd voor me bent. Je bent een rolmodel voor me. Er is niemand op de wereld die zo goed kan zorgen als jij. Hoewel het vrijwel onmogelijk is, hoop dat ik ooit aan jou kan tippen. Lieve pap, ook wel bekend als mijn klusjesman, jurist, medeauteur, reismaatje, maar bovenal Rafiki, je hebt me altijd in alles gesteund en aangemoedigd. Ik ben blij dat ik een beetje van jouw onuitputtelijke doorzettingsvermogen heb meegekregen. Bedankt dat jullie me de vrijheid hebben gegeven, maar me ook (heel vaak) hulp aanbieden wanneer dat nodig is. Bedankt voor alles. ❤ ❤ ❤

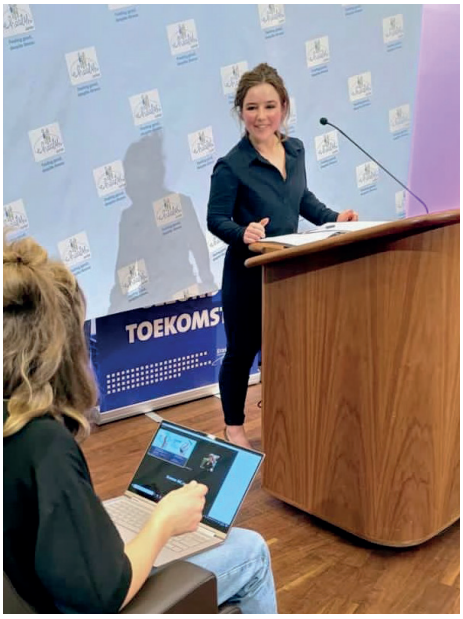
PhD Portfolio

PhD candidate:	Michelle M.J. Mens
Erasmus MC department:	Epidemiology
Research school:	Netherlands Institute of Health Sciences (NIHES)
PhD period:	August 2018 – January 2021
Promotor:	Prof.dr. M. Arfan Ikram
Copromotor:	Dr. Mohsen Ghanbari

	Year	ECTs
General research skills		
Master of Science in Health Sciences	2018 - 2019	70
Specialization: Genetic and Molecular Epidemiology		
Principles of Research in Medicine and Epidemiology		0.7
Markers and Prediction Research		0.7
Genome-wide Association Studies		0.7
Principles of Genetic Epidemiology		0.7
Genetic-epidemiological Research Methods		5.1
Genomics in Molecular Medicine		1.4
Advances in Genomics Research		0.4
SNPs and Human Diseases		2.0
Human Epigenomics		0.7
An introduction to the analysis of the NGS data		1.4
Study Design		4.3
Biostatical Methods I: Basic Principles		5.7
Biostatical Methods II: Classical Regression Models		4.3
Linux for Scientists		0.6
Scientific writing		2.0
Scientific integrity		0.3
Advanced courses		
Mendelian Randomization		0.9
Family-based Genetic Analysis		1.4
Intermediate course in R		1.4
Cardiovascular Epidemiology		0.9
Joint Models for Longitudinal and Survival Data		0.7

	Year	ECTs
Health Economics		0.7
Genomics & Transcriptomics, LUMC		0.3
Seminars and meetings		
Molecular Systems Epidemiology, Erasmus MC	2019 - 2021	2.0
Cardiometabolic Epidemiology, Erasmus MC	2018 - 2021	2.0
Lifestyle & Nutritional Epidemiology, Erasmus MC	2018 - 2021	2.0
2020 meetings, Erasmus MC	2018 - 2021	2.0
Seminars at department of Epidemiology, Erasmus MC	2018 - 2021	2.0
Molecular Epidemiology, Erasmus MC	2018 - 2021	2.0
Conferences		
American Society of Human Genetics, Houston	2019	1.5
Poster presentation		0.5
Scholarships		
Erasmus Trustfonds - Travel grant	2019	
Others		
Seminar committee: organization of weekly seminars at the department of Epidemiology, Erasmus MC	2019-2021	1.0
Reviewer of international peer-reviewed journals: Pediatric Obesity, BMC Cardiovascular Disorders	2021	1.0
Teaching acitivities: Advances in Clinical Epidemiology	2021	1.0

About the author



Michelle Mens was born on December 10th, 1992 in Bergen op Zoom. In 2000, she moved with her family to Schagen where she attended to the Regius College, from which she graduated in 2012. After her graduation she moved to Sydney, Australia to study English. Upon her return in 2013, she started her studies in Biomedical Sciences at the University of Amsterdam and obtained her master degree in 2018.

In August 2018 she went on to study Health Sciences from which she graduated in 2019. Subsequently, in August 2018 Michelle started her PhD in the group of Mohsen Ghanbari and Arfan Ikram at the Department of Epidemiology at the Erasmus Medical Center, where she conducted the research described in this thesis.

From Februari 2021 till August 2021, she worked as a postdoctoral researcher at the Department of Child and Adolescent Psychiatry/Psychology, Erasmus Medical Center in the group of Manon Hillegers. In August 2021, Michelle moved to Boston, USA to work as a postdoctoral researcher at the Department of Epidemiology, Harvard T.H. Chan School of Public Health to continue her research on biological aging together with Julia Wu.

A