HHS Public Access

Author manuscript

Nat Commun. Author manuscript; available in PMC 2015 September 20.

Published in final edited form as:

Nat Commun. ; 6: 6601. doi:10.1038/ncomms7601.

Genome-wide Identification of microRNA Expression Quantitative Trait Loci

Tianxiao Huan[1,2], **Jian Rong**[3], **Chunyu Liu**[1,2], **Xiaoling Zhang**[1,4], **Kahraman Tanriverdi**[5], **Roby Joehanes**[1,2,6,7,8], **Brian H. Chen**[1,2], **Joanne M. Murabito**[1,5], **Chen Yao**[1,2], **Paul Courchesne**[1,2], **Peter J. Munson**[6], **Christopher J. O'Donnell**[1,4], **Nancy Cox**[7], **Andrew D. Johnson**[1,4], **Martin G. Larson**[1,3], **Daniel Levy**[1,2,*], and **Jane E. Freedman**[5,*]

[1]The Framingham Heart Study, Framingham, MA, USA

[2]The Population Sciences Branch, Division of Intramural Research, National Heart, Lung, and Blood Institute, Bethesda, MD, USA

[3]Department of Mathematics and Statistics, Boston University, Boston, MA, USA

[4]Cardiovascular Epidemiology and Human Genomics Branch, Division of Intramural Research, National Heart, Lung and Blood Institute, Bethesda, MD, USA

[5]Department of Medicine, University of Massachusetts Medical School, Worcester, MA, USA; [5]Department of Medicine, Section of General Internal Medicine, Boston University School of Medicine, Boston, MA, USA

[6]Mathematical and Statistical Computing Laboratory, Center for Information Technology, National Institutes of Health, Bethesda, MD, USA

[7]Harvard Medical School, Boston, MA, USA

[8]Hebrew SeniorLife, Boston, MA, USA

[9]Department of Human Genetics, University of Chicago, Chicago, IL, USA

Abstract

[*]Denotes joint corresponding authors: **Correspondence should be addressed to:** Daniel Levy, MD, Framingham Heart Study, Population Sciences Branch, National Heart, Lung, and Blood Institute, 73 Mt. Wayte Avenue, Suite 2, Framingham, MA 01702, Levyd@nih.gov, Phone: 508-935-3458, Jane E. Freedman, MD, University of Massachusetts Medical School, AS7-1051, 368 Plantation St., Worcester, MA 01605-4319 USA, jane.freedman@umassmed.edu, Phone: (508) 856-6961.

Author Contributions

D. L. and J. F. designed, directed, and supervised the project. D. L. M. L, and J. F. were responsible for funding of the project. T. H, and D. L. drafted the manuscript. J. F. and K. T. directed and supervised the miRNA experiment. P. C. organized the experiment material and data exchange. All authors participated in revising and editing the manuscripts. All authors have read and approved the final version of the manuscript.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability

The mRNA and microRNA expression data of patient samples have been deposited in dbGaP (http://www.ncbi.nlm.nih.gov/gap) under the accession number phs000007.

Identification of microRNA expression quantitative trait loci (miR-eQTL) can yield insights into regulatory mechanisms of microRNA transcription, and can help elucidate the role of microRNA as mediators of complex traits. Here we present a miR-eQTL mapping study of whole blood from 5239 individuals, and identify 5269 *cis*-miR-eQTLs for 76 mature microRNAs. Forty-nine percent of *cis*-miR-eQTLs are located 300–500kb upstream of their associated intergenic microRNAs, suggesting that distal regulatory elements may affect the interindividual variability in microRNA expression levels. We find that *cis*-miR-eQTLs are highly enriched for *cis*-mRNA-eQTLs and regulatory SNPs. Among 243 *cis*-miR-eQTLs that were reported to be associated with complex traits in prior genome-wide association studies, many *cis*-miR-eQTLs miRNAs display differential expression in relation to the corresponding trait (e.g., rs7115089, miR-125b-5p, and HDL cholesterol). Our study provides a roadmap for understanding the genetic basis of miRNA expression, and sheds light on miRNA involvement in a variety of complex traits.

Introduction

MicroRNAs (miRNAs), a class of small noncoding RNAs, serve as key post-transcriptional regulators of gene expression and mRNA translation [1, 2]. miRNAs are increasingly recognized as mediators in a variety of biological processes including cardiovascular development and disorders [3, 4]. Highly specific miRNA expression patterns have been reported in association with heart failure [5, 6], myocardial infarction [7], and cancer [8]. However, the influence of genetic variation on miRNA expression and function still remains unclear.

Recently, many genome-wide expression quantitative trait locus (eQTL) mapping studies have revealed common genetic loci associated with mRNA expression levels of many genes [9, 10, 11, 12]. These eQTL analyses have demonstrated that transcript levels of many mRNAs behave as heritable quantitative traits. In contrast to more extensive investigations of mRNA eQTLs in multiple tissues [13] such as blood [9], brain [10], fat [11], and liver [12], there are relatively few studies of miRNA eQTLs (miR-eQTLs) and those that have been published to date are based on modest sample sizes (n<200) [14, 15, 16, 17, 18]. These studies have identified relatively few *cis*-miR-eQTLs; uncertainty persists regarding the number of miR-eQTLs in humans and their relations to regulatory elements in the human genome.

We conduct a genome-wide miR-eQTL study by utilizing genome-wide genotypes and miRNA expression profiling of whole blood derived RNA from 5239 Framingham Heart Study (FHS) participants. We analyze the associations of approximately 10 million 1000 Genomes Project [19] imputed SNPs (at minor allele frequency [MAF] >0.01 and imputation quality ratio >0.1) with whole blood-derived miRNA expression levels of 280 mature miRNAs expressed in >200 individuals, representing 11% of all discovered human miRNAs to date (2576 mature miRNAs have been reported in miRbase v20: www.mirbase.org). We calculate both *cis*- and *trans*- miR-eQTLs genome wide, and identify *cis*-miR-eQTLs with concordant effects in two pedigree independent study groups. By cross-linking *cis*-miR-eQTLs SNPs with regulatory SNPs annotated by the ENCODE project [20] and with complex trait-associated SNPs identified in prior genome-wide association studies (GWAS) [21, 22], and by linking *cis*-miR-eQTL miRNAs with differentially expressed miRNAs for complex

traits, we sought to dissect the genetic regulation of miRNA expression and explore the extent to which *cis*-miR-eQTLs may affect interindividual phenotype variability.

Results

Heritability of global miRNA expression in peripheral blood

The demographic and clinical characteristics of the 5239 FHS participants included in our analysis are shown in **Supplementary Data 1**. The pedigree structure formed by these participants is shown in **Supplementary Data 2**. We detected 280 mature miRNAs that were expressed in >200 participants (these miRNAs were used for identification of miR-eQTLs and unless specifically stated, miRNAs mentioned in the results and discussion refers to mature miRNAs), of these, 247 miRNAs were expressed in >1000 participants (Supplementary Fig 1). The distribution of narrow-sense heritability of miRNA expression for the 247 miRNAs expressed in >1000 participants is shown in Supplementary Fig 2, with an average heritability estimate (h^2_{miR}) of 0.11; 133 miRNAs (54%) had h^2_{miR}>0.1, and 9 miRNAs (miR-100-5p, miR-668, miR-133a, miR-127-3p, miR-409-3p, miR-20a-3p, miR-941, miR-191-3p, miR-1303) had h^2_{miR}>0.3 (details in **Supplementary Data 3**).

Cell type effects and reproducibility of miR-eQTLs

To evaluate whether blood cell type proportions significantly influence miR-eQTLs, we compared miR-eQTLs identified in 2138 FHS third generation cohort participants (in whom differential cell counts and proportion data were available) with and without adjustment for measured blood cell counts and cell type proportions (see **Methods**). Cell types did not appreciably influence miR-eQTLs (Supplementary Fig 3), however, we cannot exclude the possibility of small cell type effects. In the subsequent sections, we focus on miR-eQTLs from analyses that were unadjusted for cell counts (**Supplementary Data 4**). The miR-eQTLs from the model that adjusted for imputed cell counts in the larger set of 5024 participants are provided in **Supplementary Data 5**.

To evaluate the reproducibility of detected miR-eQTLs, we split our overall sample set 1:1 into two sets by pedigrees creating separate discovery and replication sets, and identified *cis*- and *trans*-miR-eQTLs in each set. At discovery FDRs of <0.1, <0.05, and <0.01, the replication rates for *cis*-miR-eQTLs were 53%, 56%, and 68% respectively, at a replication FDR<0.1, and 100% showed allele-specific directional effect concordance between the discovery and replication sets (Fig 1A–B). In contrast, no *trans*-miR-eQTLs replicated (at FDR<0.1), although 91% of *trans*-miR-eQTLs showed allele-specific directional effect concordance in the discovery and replication sets (Supplementary Fig 4). Therefore, in the subsequent sections, we mainly report *cis*-miR-eQTLs identified in the overall FHS set (unadjusted for cell counts).

Genome-wide identification of miR-eQTLs

At FDR<0.1 (corresponding *p* value threshold is 6.6×10^{-5}), we identified 5269 *cis*-miR-eQTLs for 76 miRNAs (27% of interrogated expressed miRNAs) (Fig 2). These *cis*-miR-eQTLs were further pruned by removing redundant *cis*-miR-eQTLs with high linkage

disequilibrium (LD). At a series of LD r^2 thresholds, i.e., r^2=0.2, 0.5, 0.7, 0.9 and 1, there were 283, 572, 982, 1602 and 2727 *cis*-miR-eQTLs retained. We further narrowed down the list to 52 peak *cis*-miR-eQTLs representing the single top *cis*-miR-eQTL for each miRNA or miRNA cluster shown in **Supplementary Data 6**. Table 1 shows 16 of the 52 peak *cis*-miR-eQTLs with GWAS SNPs in the NHGRI GWAS Catalog and the NHLBI GRASP data set [21, 22]. A miRNA cluster is defined as a group of miRNAs located within 10kb in the same chromosome (using the criteria described in www.mirbase.org). miRNAs with higher heritability estimates were more likely to have *cis*-miR-eQTLs. All of the 9 miRNAs with h^2_{miR}>0.3 were found to have *cis*-miR-eQTLs. The top *cis*-miR-eQTLs tended to explain a greater proportion of the variance in the respective miRNA transcript as a function of increasing h^2_{miR} (Fig 3). When the heritability of miRNA transcripts h^2_{miR} increased from (0 to 0.1) to (0.3 to 1), the proportion of variance of the miRNA transcript explained by single *cis*-miR-eQTLs ($h^2_{miR-eQTL}$) increased from 0.02 to 0.08 on average.

At FDR<0.10 (corresponding *p* value threshold is 1.0×10^{-8}), we identified 270 *trans*-miR-eQTLs for 15 miRNAs (5% of interrogated expressed miRNAs). Supplementary Fig 5 showed 2-D regional plot of *cis*- and *tran*-miR-eQTLs genome-widely (un-adjusted cell counts). **Supplementary Data 7–8** showed *trans*-miR-eQTLs at FDR<0.1 identified in the overall samples adjusted and un-adjusted cell counts respectively. We acknowledged those *trans*-miR-eQTLs need to be validated in independent cohorts.

cis-miR-eQTLs showing 5’ positional bias for miRNAs

Among the 76 mature miRNAs with *cis*-miR-eQTLs, 49 (64%) were intragenic, located within annotated protein-coding genes (located in exons, introns, or UTR regions of the host genes), and 27 (36%) were intergenic. We discovered a marked positional bias of *cis*-miR-eQTLs, with many *cis*-miR-eQTLs located in the 5’-upstream region of the corresponding miRNA rather than within miRNA coding regions or the 3’-downstream regions.

Among the 982 non-redundant (LD r^2<0.7) *cis*-miR-eQTLs (representing 1984 SNP-miRNA pairs), the relative distance of *cis*-miR-eQTLs to the corresponding mature miRNAs is shown in Fig 4 and the relative distance of *cis*-miR-eQTLs to the transcriptional start site (TSS) is shown in Supplementary Fig 6. Specifically, for intragenic miRNAs, 418 *cis*-miR-eQTLs (493 SNP-miRNA pairs, 58%) were located in the 5’-upstream region of the corresponding primary miRNAs and 432 *cis*-miR-eQTLs (536 SNP-miRNA pairs, 63%) were in the region defined by 200kb upstream to 100kb downstream of the TSS. In contrast, for intergenic miRNAs, 238 *cis*-miR-eQTLs (825 SNP-miRNA pairs, 83%) were located in the 5’-upstream region of the corresponding primary miRNAs, and 125 *cis*-miR-eQTLs (487 SNP-miRNA pairs, 49%) were in the region defined by 500kb to 300kb upstream of the TSS (**Supplementary Data 9**). There were 207 *cis*-miR-eQTLs (247 SNP-miRNA pairs, 29%) for intragenic miRNAs and 99 *cis*-miR-eQTLs (129 SNP-miRNA pairs, 13%) for intergenic miRNAs located within −/+ 50kb of the TSS of the corresponding miRNAs.

Genomic features of *cis*-miR-eQTLs

Most of the detected *cis*-miR-eQTLs are not located in protein-coding regions, i.e., 39% of eQTLs in intronic and 57% in intergenic regions (**Supplementary Data 10**). We found significant enrichment of *cis*-miR-eQTLs with expression regulatory elements (Table 2, **Supplementary Data 11** and Supplementary Fig 7), including CpG islands (2%), promoters (9%), enhancers (35%), and transcription factor (TF) binding regions (15%). We also found that *cis*-miR-eQTLs were enriched for miRNA mediated/targeted gene regulatory regions [23, 24].

There were 1066 (20%) *cis*-miR-eQTLs that overlapped with *cis*-mRNA-eQTLs identified in whole blood (enrichment p<1e-300 by hypergeometric test) [9, 25]. An example is shown in Supplementary Fig 8; 132 *cis*-miR-eQTLs (36%) for 12 intergenic mature miRNAs were also *cis*-mRNA-eQTLs for upstream protein-coding genes. We overlapped the 1 megabyte (Mb) region flanking the 132 *cis*-miR-eQTLs (chr14: 100.5Mb −102.5Mb) with the regulatory feature tracks download from UCSC Genome Browser (genome.ucsc.edu). Supplementary Fig 8 showed that the nearby regions of the 132 *cis*-miR-eQTLs for those 12 miRNAs overlap with Enhancer active region (chr14:101,100kb-101,200kb, H3K4Me1 and H2K27AC track, marked in lightyellow rectangle). The highly un-methylated status of GM12878, K562, HeLa-S3 and HepG2 cell lines are in chr14:101,400kb-101,600kb upstream of those *cis*-mRNA-eQTL miRNAs (CpG Methylation by Methy450K Bead Arrays from ENCODE/HAIB track, marked by pink color).

We also discovered eleven intragenic mature miRNAs share *cis*-eQTLs with their host mRNA genes (**Supplementary Data 12**). For *cis*-miR-eQTLs that overlapped with *cis*-mRNA-eQTLs, we performed conditional analysis to test if the associations between SNPs and miRNAs remained significant when conditioning on the corresponding mRNA expression levels using results from 5024 FHS participants with genotype, and miRNA, and mRNA expression data. As show in **Supplementary Data 13**, we found 923 *cis*-miR-eQTLs for 3384 miRNA-SNP association pairs (87%) that remained significant at FDR<0.1 (corresponding $p<6.6\times10^{-5}$) when conditioning on mRNA expression levels. These findings indicate that *cis* genetic variants may affect expression levels of neighboring miRNAs and mRNAs.

cis-miR-eQTLs and miRNA signatures for complex traits

We linked the *cis*-miR-eQTLs with GWAS SNPs in the NHGRI GWAS Catalog and the NHLBI GRASP data set [21, 22]. Among 5269 *cis*-miR-eQTLs, 243 *cis*-miR-eQTLs (for 31 miRNAs) overlapped with GWAS SNPs, including SNPs associated with multiple complex traits (Table 1, Fig 2, and regional association plots for several traits including height, menarche, platelet count, and lipid levels are shown in Supplementary Fig 9).

For miRNAs with *cis*-miR-eQTLs showing association with complex traits in GWAS, we further tested if expression of these miRNAs in FHS participants was associated with the corresponding traits. We discovered a number of miRNAs that showed differential expression in relation to the complex traits that correspond to the traits associated with their eQTLs in GWAS (Table 1). For example (Fig 5A–B), among *cis*-miR-eQTLs of

miR-100-5p and miR-125b-5p, we found 28 *cis*-miR-eQTLs (i.e., GWAS SNPs) that were associated in GWAS with lipid traits (HDL cholesterol, LDL cholesterol, total cholesterol [TC], and triglycerides), 1 (rs7941030) with multiple sclerosis, and 1 (rs1216554) with rheumatoid arthritis. These eQTLs are located ∼519kb upstream of their two associated miRNAs. We also found that miR-125b-5p showed differential expression in relation to plasma total cholesterol (p=0.005, by linear regression tests, see **Methods**) and HDL cholesterol (p=1.68e-5), and miR-100-5p showed differential expression in relation to HDL cholesterol (p=0.039). Another example (Fig 5C–D) is for miR-339-3p and miR-339-5p, which are located in an intron of *c7orf50*. Among the 282 *cis*-miR-eQTLs SNPs of miR-339-3p and 279 *cis*-miR-eQTLs of miR-339-5p, 8 were associated with TC and 3 with LDL cholesterol. We also found that expression of miR-339-3p was associated with TC (p=2.5e-7). These results establish links between SNPs affecting both miRNA expression levels and complex traits. Mendelian randomization tests provided evidence that four *cis*-miR-eQTLs SNPs (rs6951245, rs11763020, rs1997243 and rs2363286) alter the expression levels of miR-339-3p and miR-339-5p, and in turn affect interindividual variability of TC levels (causal p<0.05).

Discussion

On the basis of extensive integrated analyses of miRNA expression and genetic variants genome wide in 5239 individuals, we established a clear pattern of heritability of blood miRNA expression, and identified a substantial number of miRNAs that are controlled by *cis* genetic regulatory elements. Our results for *cis*-miR-eQTLs were highly replicable; in contrast, *trans*-miR-eQTLs were not replicable. Previously reported miR-eQTLs were identified in studies with small sample sizes (n<200) and revealed a few miR-eQTLs. For example, Borel et al. using umbilical cord blood from 180 newborns, identified only 12 *cis*-miR-eQTLs at FDR<0.5 [14]. In another study, no *cis*-miR-eQTLs were found in 176 lymphoblastoid cell lines from European and African ancestry samples [15]. Proxy SNPs of two *cis*-miR-eQTLs that we identified (rs2187519 for miR-100 and rs7797405 for miR-550) were reported by Borel et al.[14] (rs10750218 as a proxy for rs2187519 and rs12670233 for rs7797405 are in modest LD at r^2=0.29 and r^2=0.48 respectively).

As our data are from a well powered multi-generation study, we were able to assess narrow sense heritability (h^2_{miR}) of each miRNA expression trait. By comparing the overall heritability of the miRNAs and single *cis*-miR-eQTLs, we discovered that miRNAs with higher heritability were more likely to have *cis*-miR-eQTLs. When the heritability of miRNA transcripts h^2_{miR} increased, the proportion of variance of the miRNA transcript explained by single *cis*-miR-eQTLs ($h^2_{miR-eQTL}$) increased as well. Our heritability study of mRNA expression traits revealed single *cis*-mRNA-eQTLs explained 33–53% of variances in corresponding mRNA expression levels [26]. In contrast, single *cis*-miR-eQTLs explained much less proportion of variances in corresponding miRNA expression levels (∼1.3% on average).

In contrast to the functional annotation of *cis*-mRNA eQTLs, most of which are within ∼250kb of mRNA transcription start sites and without 5’ or 3’ positional bias [13], we

discovered that most *cis*-miR-eQTLs (58% for intragenic miRNAs and 83% for intergenic miRNAs) are located upstream of mature/primary miRNAs. For intergenic miRNAs, a significant fraction of *cis*-miR-eQTLs are quite far upstream (∼300–500kb). Distal regulatory elements can interact with the proximal elements that regulate miRNA expression [27]. In our results, we found that a significant fraction of *cis*-miR-eQTLs are distal, suggesting that variants in far upstream regions may play important roles in miRNA transcription. In addition, our results revealed that distal *cis*-miR-eQTLs explained a modest proportion (∼1.3% on average) of variance in miRNA expression levels. We speculate that the mild effects of *cis*-miR-eQTLs on miRNA expression result from evolutionary selection to stabilize the biological functions mediated by miRNAs.

Genetic variants that modify chromatin accessibility and transcription factor binding are a major mechanism through which genetic variation leads to expression differences for protein-coding genes in humans [28]. The investigation of regulatory mechanisms of miRNA transcription is still evolving. Genomic feature analyses of *cis*-miR-eQTLs reveal that a large proportion of *cis*-miR-eQTLs are located in regulatory elements such as CpG islands (2%), promoters (9%), enhancers (35%), and transcription factor (TF) binding regions (15%). We also discovered that *cis*-miR-eQTLs show a significant enrichment for mRNA-eQTLs and 87% of *cis*-miR-eQTLs that also are mRNA-eQTLs remained significant when conditioning on the corresponding mRNA expression levels. For example, as shown in Supplementary Fig 8, 132 *cis*-miR-eQTLs (36%) for 12 intergenic miRNAs were also *cis*-mRNA-eQTLs for upstream protein-coding genes. This finding suggests that genetic variants may influence the expression of both miRNAs and nearby protein-coding genes. These eQTL regulatory effects may act via modified chromatin accessibility, transcription factor binding affinity, or DNA methylation.

The mechanisms of transcriptional regulation of intragenic miRNAs are more complex than intergenic miRNAs, as intragenic miRNAs may mirror the regulatory mechanisms of their host genes, or be transcribed independently as a consequence of their unique promoter regions [29]. We identified 11 mature miRNAs from intragenic miRNAs that share *cis* eQTLs with their host protein coding genes (**Supplementary Data 12**). Among the *cis*-mRNA-eQTL miRNAs, 15 miRNAs having alternative intronic promoters (alternative intronic promoters were from [29]). We overlapped the *cis*-miR-eQTLs and expression regulatory elements annotations from ENCODE nearby regions of each miRNA (+/−50kb). We found, in some examples (Supplementary Fig 10), *cis*-miR-eQTLs near alternative intronic promoter regions demonstrated promoter and enhancer activities and were highly un-methylated in some cell lines. Our findings provide a guide for further functional studies of transcriptional elements of miRNAs.

We identified numerous *cis*-miR-eQTLs that are associated with complex diseases/traits in GWAS (Table 1). Equally noteworthy, we found several examples in which the miRNAs associated in *cis* with these GWAS SNPs were associated with the corresponding trait (e.g., three-way association of HDL cholesterol with its GWAS SNP, rs7115089, and with the corresponding miR-125b-5p). A single miRNA may target hundreds of protein-coding genes. Therefore, the effect of genetic variants on miRNAs can play an important regulatory role in mediating the targeted protein-coding genes, as well as complex phenotypes. We

speculate that some of the protein-coding genes targeted by miRNAs may also be involved in the cellular pathways related to the trait. For example, miR-125b-5p expression was associated with HDL cholesterol (p=1.7×10^{-5}, by a linear regression test). In a parallel project focusing on differentially expressed mRNAs in association with lipid levels, we found 17 genes targeted by miR-125b-5p (9% of miR-125b-5p targeted genes in miRTarBase [24]) that showed differential expression in association with HDL cholesterol (at p<0.05 corrected for ∼18,000 genes, by a linear regression test)[30]. Some of these genes are involved in metabolic processes, e.g., *PRDX2*, which was down-regulated in association with HDL cholesterol (p=1.1×10^{-15}, by a linear regression test). Further studies and biological experiments are needed to investigate whether these *cis*-miR-eSNPs affect the corresponding miRNA targeting genes.

In summary, our genome-wide miR-eQTL mapping study provides new insights into the genetic regulation of miRNA transcription and the roles of miRNAs in complex diseases. Our findings may help to identify new opportunities for drug treatment or diagnosis of human diseases.

Methods

Study populations

The FHS is a community-based study that began enrolling participants in 1948. In 1971, the offspring and offspring spouses (the Offspring cohort) of original FHS cohort participants were recruited and they have been examined every four to eight years [31]. From 2002 to 2005, the adult children of the offspring cohort participants (the third generation cohort) were recruited and examined [32]. In this study, we investigated 2272 offspring cohort attendees at examination cycle 8 (2005–2008) and 3057 third generation cohort attendees at examination cycle 2 (2008–2010). This study was approved under Boston University Medical Center protocol H-27984. Written informed consent was obtained from each participant.

miRNA expression profiling

miRNAs were measured from venous blood samples obtained from participants after overnight fasting. Whole blood samples (2.5ml) were collected in PAXgene Blood RNA™ tubes (Qiagen, Valencia, CA) and frozen at –800C. Total RNA was isolated from the frozen PAXgene Blood RNA tubes (Asuragen, Inc. Austin, TX) and a 2100 Bioanalyzer Instrument (Agilent, Santa Clara, CA) was used for RNA quality assessment. Isolated RNA samples were converted to complementary DNA (cDNA) using TaqMan miRNA Reverse Transcription Kit and MegaPlex Human RT Primer Pool Av2.1 and Pool Bv3.0. (Life Technologies, Foster City, CA) in a 384 well Thermal Cycler. The cDNA samples were PreAmplified using TaqMan PreAmp Master Mix and PreAmp Primers, Human Pool A v2.1 and Pool B v3.0 (Life Technologies, Foster City, CA).

qRT-PCR reactions were performed with the BioMark System using (Fluidigm, South San Francisco, CA) TaqMan miRNA Assays(Life Technologies, Foster City, CA). As described in the published literature, measurement of RNA by qRT-PCR is reliable and has high

specificity and sensitivity [33, 34, 35, 36]. The initial miRNA list encompassed all commercially available TaqMan miRNA assays obtainable at the start of the project (754 mature miRNAs). These miRNAs were initially assayed for measurement feasibility in RNA samples from 450 FHS participants. All qRT-PCR reactions were performed in the BioMark Real-Time PCR system using the following protocol: 10 min at 95°C, 15 sec at 95°C and 1 min at 60°C for 30 cycles. Single copy can be detected with BioMark system at 26–27 Cycle Thresholds. For replicates >95% of the data points had coefficients of variation <10% (mean ~4%).

miRNA normalization

We normalized miRNA expression using a model that adjusts raw miRNA cycle threshold Ct values for 4 technical variables: isolation batch (50 batches), RNA concentration, RNA quality (defined as RNA integrity number [RIN]), and RNA 260/280 ratio (ratio of absorbance at 260 and 280nm; measured using a spectrophotometer). Histograms (Supplementary Fig 11) show that this model explains 20% to 60% of variability of raw miRNA measurements for 80% of miRNAs

Genotyping

DNA was isolated from buffy coat or from immortalized lymphoblast cell lines. Genotyping was conducted with the Affymetrix 500K mapping array and the Affymetrix 50K gene-focused MIP array, using previously described quality control procedures [37]. Genotypes were imputed to the 1000 Genomes Project panel 19 of approximately 36.3 million variants using MACH [38]. We filtered out SNPs with MAF<0.01 and imputation quality ratio<0.1 (the imputation quality ratio is denoted by the ratio of the variances of the observed and the estimated allele counts), resulting in 9.8×10^6 SNPs (approximately 10 million SNPs) that were eligible for further miRNA eQTL testing.

miR-eQTL mapping

Because of the computational burden of running linear mixed effects (LME) models for approximately 10 million (SNPs) × 280 miRNAs (miRNAs expressed in >200 samples), we adapted a two-step analysis strategy. *Step 1*: linear regression was used to model the association between miRNA Ct values (miR) and the imputed SNP genotypes – adjusted for age, sex, cohort, and technical covariates – yielding results for roughly 280 miRNAs × 10 million SNPs, as shown in Equation 1. Associated SNP-miRNA pairs residing within 1Mb of the mature miRNA (*cis*) and those residing more than 1 Mb away (*trans*) were identified separately. We chose liberal p value thresholds to pre-filter the miR-eQTLs, at $p<1\times10^{-3}$ for cis and $p<1\times10^{-5}$ for *trans*. These *p* value thresholds were chosen to ensure that miR-eQTLs at a false discovery rate (FDR) <0.1 were not omitted as a result of this pre-filtering step. *Step 2*: we used a linear mixed model [39] to re-calculate the associations of SNPs and miRNA expression levels for the pre-selected eQTLs from *step 1*, adjusted for age, sex, and technical covariates as fixed effects and a familial correlation matrix (FAM) as the random effect using the *lmekin()* function of Kinship Package (http://cran.r-project.org/web/packages/kinship/) [39], as shown in Equation 2. In Equation 1 and 2, ε is the error term for each independent observation.

$$miR = SNP + age + sex + cohort + \sum_{i=1}^{n} techical\ covariates + \varepsilon \quad (1)$$

$$miR = SNP + age + sex + \sum_{i=1}^{n} techical\ covariates + FAM + \varepsilon \quad (2)$$

Genome coordinate annotation for miRNAs used miRbase v20 (mirbase.org), and for SNPs we used the February 2009 assembly of the human genome (hg19, GRCh37 Genome Reference Consortium Human Reference 37). Based on the coordinates of 280 mature miRNAs and 9.8×10^6 SNPs, we estimated there were 13,935,272 (1.4×10^7) potential SNP-miRNA pairs where the SNP was located within 1Mb on either side of the corresponding mature miRNA. We estimated there were 1.4×10^7 potential *cis* SNP-miRNA pairs, and 2.7×10^9 (i.e., $280 \times 9.8 \times 10^6 - 1.4 \times 10^7$) potential *trans* SNP-miRNA pairs. We used the Benjamini-Hochberg method [40] to calculate FDR for *cis*- and *trans*-miR-eQTLs by correcting for 1.4×10^7 potential *cis* SNP-miRNA pairs and 2.7×10^9 potential trans SNP-miRNA pairs, respectively. We selected an FDR threshold of 0.1, corresponding to $p<6.6\times10^{-5}$ for *cis*- and $p<1.0\times10^{-8}$ for *trans*-miR-eQTLs.

For identified *cis*-miR-eQTLs at FDR<0.1, we used FESTA (Fragmented Exhaustive Search for TAgSNPs) [41] to select non-redundant miR-eQTLs based on a series of LD r^2 thresholds, 0.2, 0.5, 0.7, 0.9 and 1. FESTA used a mixture of search techniques to partition the whole SNP set into disjointed precincts and selected a tag SNP for each SNP block, which represented a set of SNPs at a LD r^2 >threshold [41].

To estimate the replicability of miR-eQTLs, we split the overall sample set 1:1 into discovery and replication sets. The discovery and replication sets represent independent pedigrees to ensure that individuals in the two sets were unrelated. We used the methods described above to identify miR-eQTLs in the discovery and replication sets, separately. We evaluated the concordance of effect sizes of *cis*- and *trans*-miR-eQTLs in the discovery and replication sets. We identified eQTLs at FDR<0.1 in the discovery set, and attempted to replicate them in the replication set.

mRNA expression data

Whole blood samples (2.5ml) were collected in PAXgene™ tubes by Asuragen, Inc. (PreAnalytiX, Hombrechtikon, Switzerland). Total RNA was isolated according to the company's standard operating procedures for automated isolation of RNA from 96 samples in a single batch on a KingFisher® 96 robot. Then 50ng RNA samples were amplified using the WT-Ovation Pico RNA Amplification System (NuGEN, San Carlos, CA) as recommended by the manufacturer in an automated manner using the genechip array station (GCAS). RNA expression was conducted using the Affymetrix Human Exon Array ST 1.0 (Affymetrix, Inc., Santa Clara, CA). The core probe sets were annotated using the Affymetrix annotation files from Netaffx (www.netaffx.com, HuEx-1_0-st-v2.na29.hg18.probeset.csv).

The raw gene expression data were at first preprocessed by quartile normalization. Then the RMA (robust multi-array average) values of every gene (17,318 measured genes) were adjusted for a set of technical covariates, e.g. chip batch, by fitting linear mixed regression (LME) models. Imputed blood cell counts (i.e. white blood cell [WBC], red blood cell [RBC], platelet, lymphocyte, monocyte, eosinophil, and basophil) (Joehanes R, in preparation) were also evaluated as covariates and adjusted if deemed significant, as detailed below. The residuals were retained for further analysis.

Matching *cis*-miR-eQTLs with *cis*-mRNA-eQTLs

We overlapped the *cis*-miR-eQTLs at FDR<0.1 reported in this study with *cis*-mRNA-eQTLs at FDR<0.1identifed by [9, 25], hyergeometirc test was used to evaluate if *cis*-miR-eQTLs were significantly enriched for *cis*-mRNA-eQTLs. For those overlap eQTLs, i.e., *cis*-miR-eQTLs that were also *cis*-mRNA-eQTLs, we used the same linear mixed regression model as described in "*miR-eQTL mapping*" section to re-analyze the associations between genotypes and miRNA expression levels but conditional regression on corresponding mRNA expression levels.

Estimating effects of cell counts in the miRNA eQTLs

Since the miR-eQTLs in whole blood may be driven by cellular composition, we compared the miR-eQTLs in 2138 individuals with measured cell counts before and after correction for cell count effects (Supplementary Fig 3). Differential cell counts and proportions in whole blood were measured in 2138 individuals in the FHS third generation cohort, including seven cell types, white blood cell [WBC], red blood cell [RBC], platelet, neutrophil, lymphocyte, monocyte, eosinophil and basophil. The cell counts and proportions for 5024 FHS participants were estimated using mRNA expression values by partial least squares (PLS) regression prediction. The estimated cell count proportion values are highly consistent with the measured cell counts proportion values (Joehanes R, PhD, unpublished data, 2014).

We did not find any evidence that cell counts affected the miR-eQTLs; however, we cannot exclude small effects from cell counts. Therefore, we report miR-eQTLs unadjusted for cell counts in our main results, and secondarily report miR-eQTLs adjusted for imputed cell counts (i.e. WBC, RBC, platelets, lymphocytes, monocytes, eosinophils, and basophils) in **Supplementary Data 5**. Please note that there were 215 samples without mRNA expression data; therefore, the maximum sample size of unadjusted for cell counts is 5239 and the maximum sample size of analyses adjusted for cell counts is 5024.

Estimating the heritability of miRNA expression levels

To estimate the narrow-sense heritability of the expression for a specific miRNA (denoted as h^2_{miR}), we used the model as shown in Equation 3.

$$miR = age + sex + \sum_{i=1}^{n} techical\ covariates + FAM + \varepsilon \quad (3)$$

Here, age, sex, and technical covariates were included as fixed effects, *FAM* was the familial correlation matrix included as the random effect. *FAM* represented additive polygenic genetic effects [39]. ε is the error term for each independent observation. h^2_{miR} was the proportion of the additive polygenic genetic variance (σ^2_{FAM}) among the total phenotypic variance (σ^2_{miR}) of miRNA expression: $h^2_{miR}=\sigma^2_{FAM}/\sigma^2_{miR}$. We estimated h^2_{miR} for every miRNA expression trait (247 miRNAs expressed in more than 1000 samples) using the *lmekin()* function of Kinship package (http://cran.r-project.org/web/packages/kinship/) [39].

Estimating proportion of variance in miRNAs attributable to miR-eQTLs

To estimate the proportion of variance in a single miRNA trait that is attributable to a single miR-eQTL (denoted as $h^2_{miR-eQTL}$), we used the following two models: Full model:

$$miR=SNP+age+sex+cohort+\sum_{i=1}^{n} techical\ covariates+FAM+\varepsilon \quad (4)$$

Null model:

$$miR=age+sex+cohort+\sum_{i=1}^{n} techical\ covariates+FAM+\varepsilon \quad (5)$$

Here, *age*, *sex*, *cohort* (offspring cohort and the third generation cohort in the FHS) and *technical covariates* were included as fixed effects, *FAM* was the familial correlation matrix included as the random effect. ε is the error term for each independent observation. The proportion of variance in a single miRNA trait that is attributable to a single miR-eQTL was denoted as $h^2_{miR-eQTL}$ and was calculated as follows:

$$h^2_{miR-eQTL}=\max\left(0, \frac{\sigma^2_{QTL.null}+\sigma^2_{err.null}-\sigma^2_{QTL.full}-\sigma^2_{err.full}}{\sigma^2_{miR}}\right) \quad (6)$$

where σ^2_{miR} was the total phenotypic variance of a miRNA expression trait; $\sigma^2_{QTL.full}$ and $\sigma^2_{err.full}$ were the polygenic and error variances, respectively, when modeling with the tested miR-eQTL; $\sigma^2_{QTL.null}$ and $\sigma^2_{err.null}$ were the polygenic and error variances, when modeling without the tested miR-eQTL. The *lmekin()* function in the Kinship package [39] was used to estimate $h^2_{miR-eQTL}$.

Identification of differentially expressed miRNAs for complex traits

We used the NHGRI GWAS Catalog (http://www.genome.gov/gwastudies/) [21] and NHLBI GRASP database (http://apps.nhlbi.nih.gov/grasp/) [22] to annotate complex trait associated miR-eQTLs. The *cis*-miR-eQTLs identified in this study were compared with SNPs in the NHGRI GWAS Catalog and NHLBI GRASP GWAS results for SNPs at $p<1\times10^{-5}$.

For the complex traits that could be mapped with *cis*-miR-eQTLs (and also were measured in the FHS), including menarche, lipids (HDL cholesterol, triglycerides [TG], and total cholesterol [TC]), type II diabetes mellitus (T2D), and glucose level we used linear mixed

models to test their association with miR-eQTL miRNAs in FHS individuals. These phenotypes were ascertained at examinations 8 and 2 for the offspring and the third generation cohorts, respectively. We identified differentially expressed miRNAs associated with HDL cholesterol, TC, TG, T2D, and glucose after accounting for age, sex, cell counts, and technical covariates (see methods miRNA normalization) and family structure in LME models implemented in the *lmekin* function [39]. Differentially expressed miRNA associated with age at menarche were tested in LME models (*lmekin*) after accounting for birth year, cell counts, technical covariates and family structure.

miRNA transcription start site and promoter regions

The transcriptional regulatory mechanisms affecting miRNA expression are unclear. There are technical barriers to the precise identification of primary miRNAs, transcription start sites (TSSs), and promoter regions for most mature miRNAs [29]. Recently, Marsico et al. [29] and Chen et al.[42] predicted miRNAs TSSs. Their results were incorporated with the results from previous similar studies [43]. However, by comparing the TSS positions identified by these two studies, there was, on average, 55kb distance difference between TSSs positions to the corresponding mature miRNAs. Therefore, in our analysis, we annotated the miRNA TSSs collected and predicted by these two studies, respectively. The predicted promoter annotations for miRNAs were obtained from Marsico et al. which were screened within −/+ 50kb from the TSSs for each miRNA [29].

Functional annotation of *cis*-miR-eQTLs

We annotated the genomic features *cis*-miR-eQTLs (n=5269) using HaploReg [44], which integrates results from ENCODE [20]. The overlap of *cis*-miR-eQTLs with ENCODE annotated SNPs in promoter, enhancer and transcription factor (TF) binding sites were retrieved (**Supplementary Data 11**).

For enrichment tests of functional SNPs in *cis*-miR-eQTLs identified in this study, we downloaded regulatory tracks contained in the UCSC Genome Browser (genome.ucsc.edu), including ENCODE histone modification sites, and transcription factor and CTCF binding sites in lymphoblastoid cell lines (GM12878), ORegAnno (Open Regulatory Annotation) [45], UCSC CpG islands, and long intergenic non-coding RNA [46]. We also downloaded other regulatory tracks, including experimentally validated miRNA targets from TARbase [47], and experimentally supported miRNA-mediated gene regulatory sites from Patrocles [23]. Binominal tests were used to evaluate if the identified *cis*-miR-eQTLs set (5269 *cis*-miR-eQTLs) showed enrichment for regulatory SNPs for each track (methods described by [13]).

We further determined whether or not the detected *cis*-miR-eQTLs SNPs were enriched for promoter, enhancer, or protein binding regions on the genome. To do so, we annotated all *cis*-miR-eQTLs (n=5269) using HaploReg [44], which integrates results from ENCODE [20]. We examined enrichment in 9 different cell lines (i.e., GM12878, H1-hESC, HepG2, HMEC, HSMM, HUVEC, K562, NHEK and NHLF). The null distributions of eQTLs were generated using a permutation strategy by randomly selecting equal number of SNPs (n=5269) 100 times. The pools of candidate SNPs for the permutation were from 1000-genomes imputed SNPs with MAF>0.01 and imputation quality ratio>0.1. In order to match

the distribution of MAFs of the permutation SNPs (the permutation-SNPs set) with the *cis*-miR-eQTLs SNPs (the tested-SNPs set), we categorized MAF into four categories: MAF of (0.01, 0.05), (0.05, 0.1), (0.1, 0.2), and (0.2, 0.5). For each MAF category, we kept the proportion of SNPs in the permutation-SNPs set equal to the proportion of SNPs in the tested-SNPs set. In the four MAF categories, the proportions of SNPs are 3%, 7%, 19% and 71% respectively. The average of the overlap between permutation and regulatory region SNPs (i.e. SNPs in promoter, enhancer, and protein binding regions) was compared with the overlap between the tested-SNPs and regulatory region SNPs.

Mendelian randomization test

We used a two-stage least squares (2SLS) Mendelian randomization (MR) method [48] to estimate the causal relationships between miRNAs and complex traits measured in FHS participants; the traits analyzed included menarche, lipids (HDL, TG, and TC), T2D, and glucose, using *cis*-miR-eQTLs as instrumental variables (IV). MR was only performed in the pre-filtered SNP-miRNA-trait pairs, when a SNP was a *cis*-miR-eQTL and also present in NHGRI GWAS Catalog (http://www.genome.gov/gwastudies/) [21] or in the NHLBI GRASP database (http://apps.nhlbi.nih.gov/grasp/) [22], and the miRNA showed differential expression in relation to the corresponding trait at $p<0.05$ in FHS participants.

To determine the strength of the genetic instrument, an F-statistic in a linear regression model was derived from the proportion of variation in the miRNA expression levels (miRNA Ct values) that was explained by the corresponding *cis*-miR-eQTL, by modeling age, sex, family structure and 4 technical variables as covariates (see in the miRNA normalization section). *cis*-miR-eQTLs with an F-statistic less than 10, indicating a weak instrument, was excluded. The first stage of the 2SLS method involves using a linear regression of the modifiable exposure (miRNA) on the IV (SNP) and covariates, and saving the predicted miRNA values. In the second stage, the outcome (complex trait) is regressed on the predicted miRNA values. The regression coefficient obtained in the second stage can be interpreted as being the causal effect of the exposure (miRNA) on the outcome (complex trait). The Durbin–Wu-Hausman test [49] is used to estimate whether the estimates derived from the first and second stage of the 2SLS are consistent.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The Framingham Heart Study is funded by National Institutes of Health contract N01-HC-25195. The laboratory work for this investigation was funded by the Division of Intramural Research, National Heart, Lung, and Blood Institute, National Institutes of Health. The analytical component of this project was funded by the Division of Intramural Research, National Heart, Lung, and Blood Institute, and the Center for Information Technology, National Institutes of Health, Bethesda, MD.

This study utilized the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, MD. (http://biowulf.nih.gov).

References

1. Lee R, Feinbaum R, Ambros V. A short history of a short RNA. Cell. 2004; 116:S89–S92. 81 p following S96. [PubMed: 15055592]
2. Lee RC, Ambros V. An extensive class of small RNAs in Caenorhabditis elegans. Science. 2001; 294:862–864. [PubMed: 11679672]
3. Cordes KR, Srivastava D. MicroRNA regulation of cardiovascular development. Circ Res. 2009; 104:724–732. [PubMed: 19325160]
4. Small EM, Olson EN. Pervasive roles of microRNAs in cardiovascular biology. Nature. 2011; 469:336–342. [PubMed: 21248840]
5. Thum T, et al. MicroRNAs in the human heart: a clue to fetal gene reprogramming in heart failure. Circulation. 2007; 116:258–267. [PubMed: 17606841]
6. Tijsen AJ, et al. MiR423-5p as a circulating biomarker for heart failure. Circ Res. 2010; 106:1035–1039. [PubMed: 20185794]
7. Fiedler J, Thum T. MicroRNAs in myocardial infarction. Arterioscler Thromb Vasc Biol. 2013; 33:201–205. [PubMed: 23325477]
8. Lu J, et al. MicroRNA expression profiles classify human cancers. Nature. 2005; 435:834–838. [PubMed: 15944708]
9. Westra H-J, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nature genetics. 2013; 45:1238–1243. [PubMed: 24013639]
10. Zhang B, et al. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. Cell. 2013; 153:707–720. [PubMed: 23622250]
11. Emilsson V, et al. Genetics of gene expression and its effect on disease. Nature. 2008; 452:423–428. [PubMed: 18344981]
12. Schadt EE, et al. Mapping the genetic architecture of gene expression in human liver. PLoS Biol. 2008; 6:e107. [PubMed: 18462017]
13. Zhang X, et al. Synthesis of 53 tissue and cell line expression QTL datasets reveals master eQTLs. BMC genomics. 2014; 15:532. [PubMed: 24973796]
14. Borel C, et al. Identification of cis- and trans-regulatory variation modulating microRNA expression levels in human fibroblasts. Genome research. 2011; 21:68–73. [PubMed: 21147911]
15. Gamazon ER, et al. Genetic architecture of microRNA expression: implications for the transcriptome and complex traits. Am J Hum Genet. 2012; 90:1046–1063. [PubMed: 22658545]
16. Civelek M, et al. Genetic regulation of human adipose microRNA expression and its consequences for metabolic traits. Hum Mol Genet. 2013; 22:3023–3037. [PubMed: 23562819]
17. Somel M, et al. MicroRNA, mRNA, and protein expression link development and aging in human and macaque brain. Genome research. 2010; 20:1207–1218. [PubMed: 20647238]
18. Siddle KJ, et al. A genomic portrait of the genetic architecture and regulatory impact of microRNA expression in response to infection. Genome research. 2014; 24:850–859. [PubMed: 24482540]
19. Abecasis GR, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012; 491:56–65. [PubMed: 23128226]
20. Birney E, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature. 2007; 447:799–816. [PubMed: 17571346]
21. Hindorff LA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proceedings of the National Academy of Sciences of the United States of America. 2009; 106:9362–9367. [PubMed: 19474294]
22. Leslie R, O'Donnell CJ, Johnson AD. GRASP: analysis of genotype-phenotype results from 1390 genome-wide association studies and corresponding open access database. Bioinformatics. 2014; 30:i185–i194. [PubMed: 24931982]
23. Hiard S, Charlier C, Coppieters W, Georges M, Baurain D. Patrocles: a database of polymorphic miRNA-mediated gene regulation in vertebrates. Nucleic Acids Res. 2010; 38:D640–D651. [PubMed: 19906729]
24. Hsu SD, et al. miRTarBase: a database curates experimentally validated microRNA-target interactions. Nucleic Acids Res. 2011; 39:D163–D169. [PubMed: 21071411]

25. Joehanes, R., et al. Genome-wide Expression Quantitative Trait Loci: Results from the NHLBI's SABRe CVD Initiative; the American Society of Human Genetics (ASHG) conference; 2013.
26. Huan T, et al. A systematic heritability analysis of the human whole blood transcriptome. Human genetics. 2015:1–16. [PubMed: 25429801]
27. Cesana M, et al. A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. Cell. 2011; 147:358–369. [PubMed: 22000014]
28. Degner JF, et al. DNase I sensitivity QTLs are a major determinant of human expression variation. Nature. 2012; 482:390–394. [PubMed: 22307276]
29. Marsico A, et al. PROmiRNA: a new miRNA promoter recognition method uncovers the complex regulation of intronic miRNAs. Genome Biol. 2013; 14:R84. [PubMed: 23958307]
30. Chen BH, et al. Transcriptome-wide Association Study of Circulating Lipid Levels. Circulation. 2014; 129:A35–A35.
31. Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP. The Framingham Offspring Study Design and preliminary data. Prev Med. 1975; 4:518–525. [PubMed: 1208363]
32. Splansky GL, et al. The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. Am J Epidemiol. 2007; 165:1328–1335. [PubMed: 17372189]
33. Chen C, Tan R, Wong L, Fekete R, Halsey J. Quantitation of microRNAs by real-time RT-qPCR. Methods Mol Biol. 2011; 687:113–134. [PubMed: 20967604]
34. Chen C, et al. Real-time quantification of microRNAs by stem-loop RT-PCR. Nucleic Acids Res. 2005; 33:e179. [PubMed: 16314309]
35. Jensen SG, et al. Evaluation of two commercial global miRNA expression profiling platforms for detection of less abundant miRNAs. BMC genomics. 2011; 12:435. [PubMed: 21867561]
36. Jang J, et al. Quantitative miRNA expression analysis using fluidigm microfluidics dynamic arrays. BMC genomics. 2011; 12:144. [PubMed: 21388556]
37. Levy D, et al. Genome-wide association study of blood pressure and hypertension. Nat Genet. 2009; 41:677–687. [PubMed: 19430479]
38. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009; 25:1105–1111. [PubMed: 19289445]
39. Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. Am J Hum Genet. 1998; 62:1198–1211. [PubMed: 9545414]
40. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B (Methodological). 1995:289–300.
41. Qin ZS, Gopalakrishnan S, Abecasis GR. An efficient comprehensive search algorithm for tagSNP selection using linkage disequilibrium criteria. Bioinformatics. 2006; 22:220–225. [PubMed: 16269414]
42. Chen D, et al. Dissecting the chromatin interactome of microRNA genes. Nucleic acids research. 2014; 42:3028–3043. [PubMed: 24357409]
43. Chien CH, et al. Identifying transcriptional start sites of human microRNAs based on high-throughput sequencing data. Nucleic Acids Res. 2011; 39:9345–9356. [PubMed: 21821656]
44. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 2012; 40:D930–D934. [PubMed: 22064851]
45. Griffith OL, et al. ORegAnno: an open-access community-driven resource for regulatory annotation. Nucleic Acids Res. 2008; 36:D107–D113. [PubMed: 18006570]
46. Cabili MN, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. 2011; 25:1915–1927. [PubMed: 21890647]
47. Vergoulis T, et al. TarBase 6.0: capturing the exponential growth of miRNA targets with experimental support. Nucleic Acids Res. 2012; 40:D222–D229. [PubMed: 22135297]
48. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Stat Med. 2008; 27:1133–1163. [PubMed: 17886233]

49. Baum CF, Schaffer ME, Stillman S. IVENDOG: Stata module to calculate Durbin-Wu-Hausman endogeneity test after ivreg. Statistical Software Components. 2007

50. Teslovich TM, et al. Biological, clinical and population relevance of 95 loci for blood lipids. Nature. 2010; 466:707–713. [PubMed: 20686565]

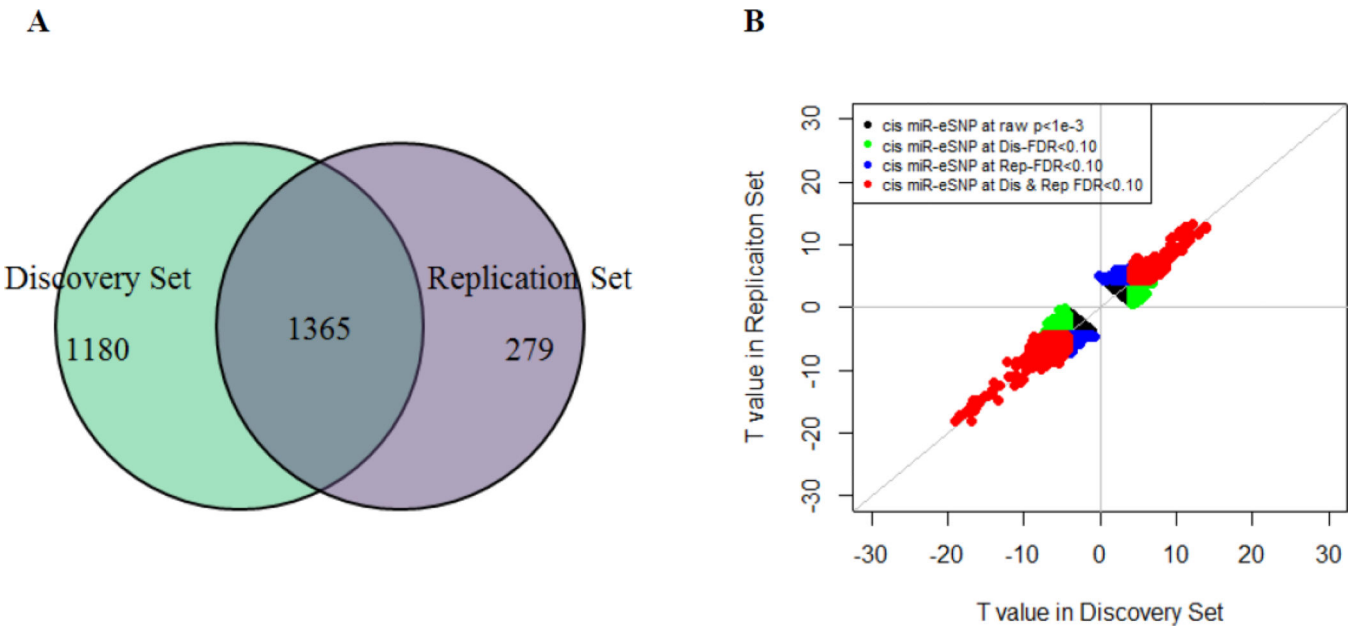

Figure 1. Genome-wide identification of *cis* miR-eQTLs

a) Venn diagram of *cis*-miR-eQTLs identified in pedigree independent discovery (n=2671) and replication sets (n=2658). The number indicated *cis*-miR-eQTLs identified in discovery, replication or both at FDR<0.1. b) T values of *cis*-miR-eQTLs between discovery and replication groups.

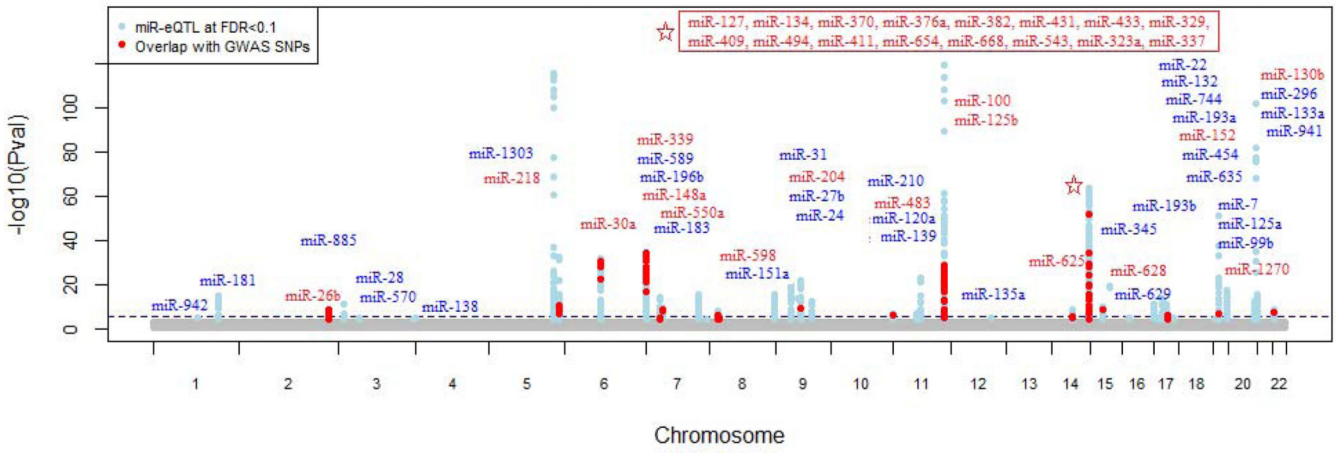

Figure 2. Manhattan plot of *cis* miR-eQTLs

Genome-wide -log10 (P) value plots are shown for every interrogated miRNA (280 expressed miRNAs). 76 miRNAs having *cis*-miR-eQTLs are labeled in this figure (for 52 unique peak loci). The horizontal dotted line indicates FDR<0.1 (corresponding to $p<6.6\times10^{-5}$). *cis*-miR-eQTL SNPs overlapping with GWAS SNPs reported in NHGRI GWAS Catalog (http://www.genome.gov/gwastudies/) [21] and NHLBI GRASP database (http://apps.nhlbi.nih.gov/grasp/) [22] are shown in red.

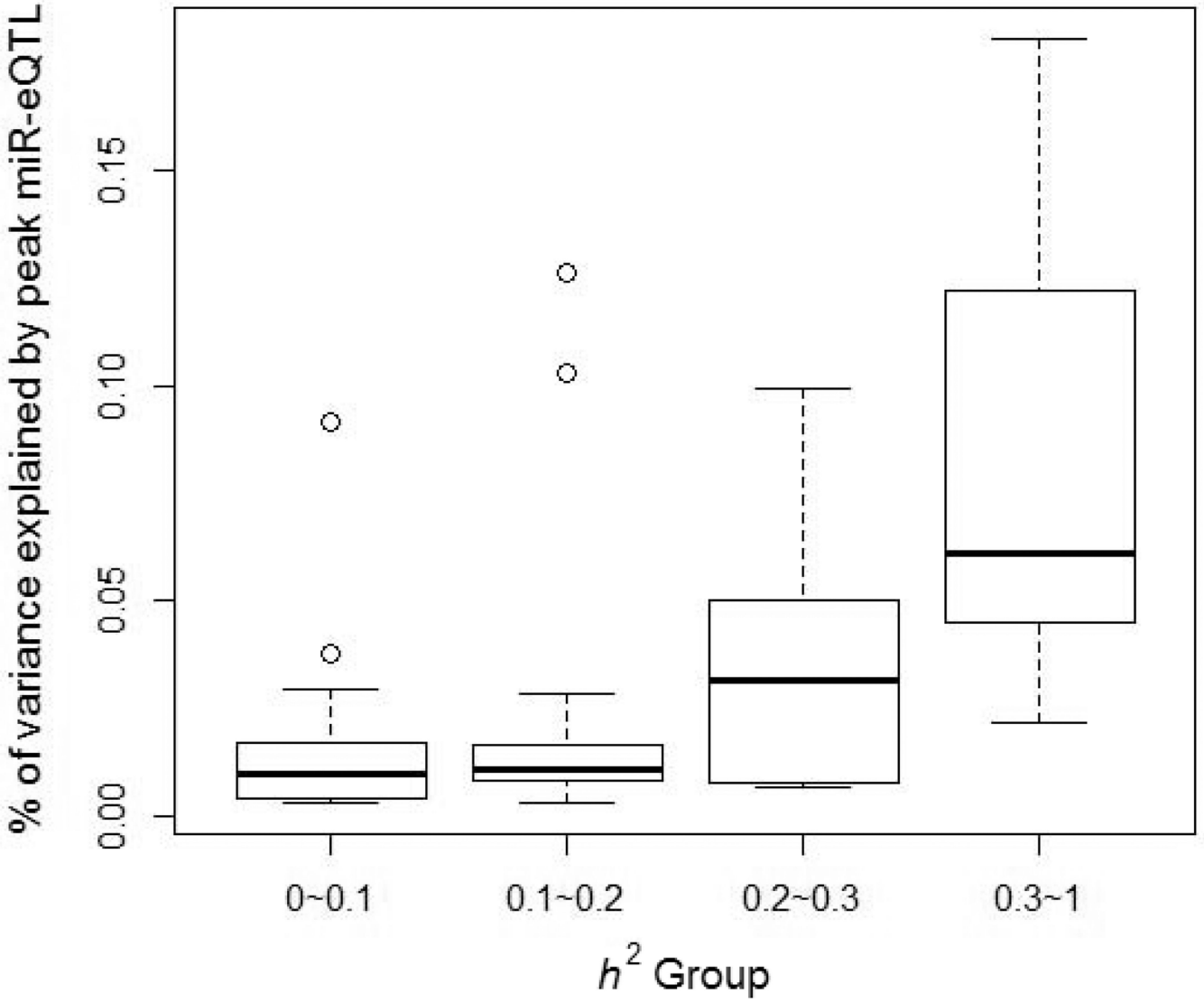

Figure 3. The variance proportion of miRNA expression explained by single *cis*- miR-eQTLs at different heritability levels

This figure was plotted by the boxplot function in the R library. The boxes indicate the interquartile range (IQR) of data between 75% (Q3) and 25% (Q1). The bars below and above each box indicate the data in Q1–1.5 × IQR and Q3+1.5 × IQR respectively.

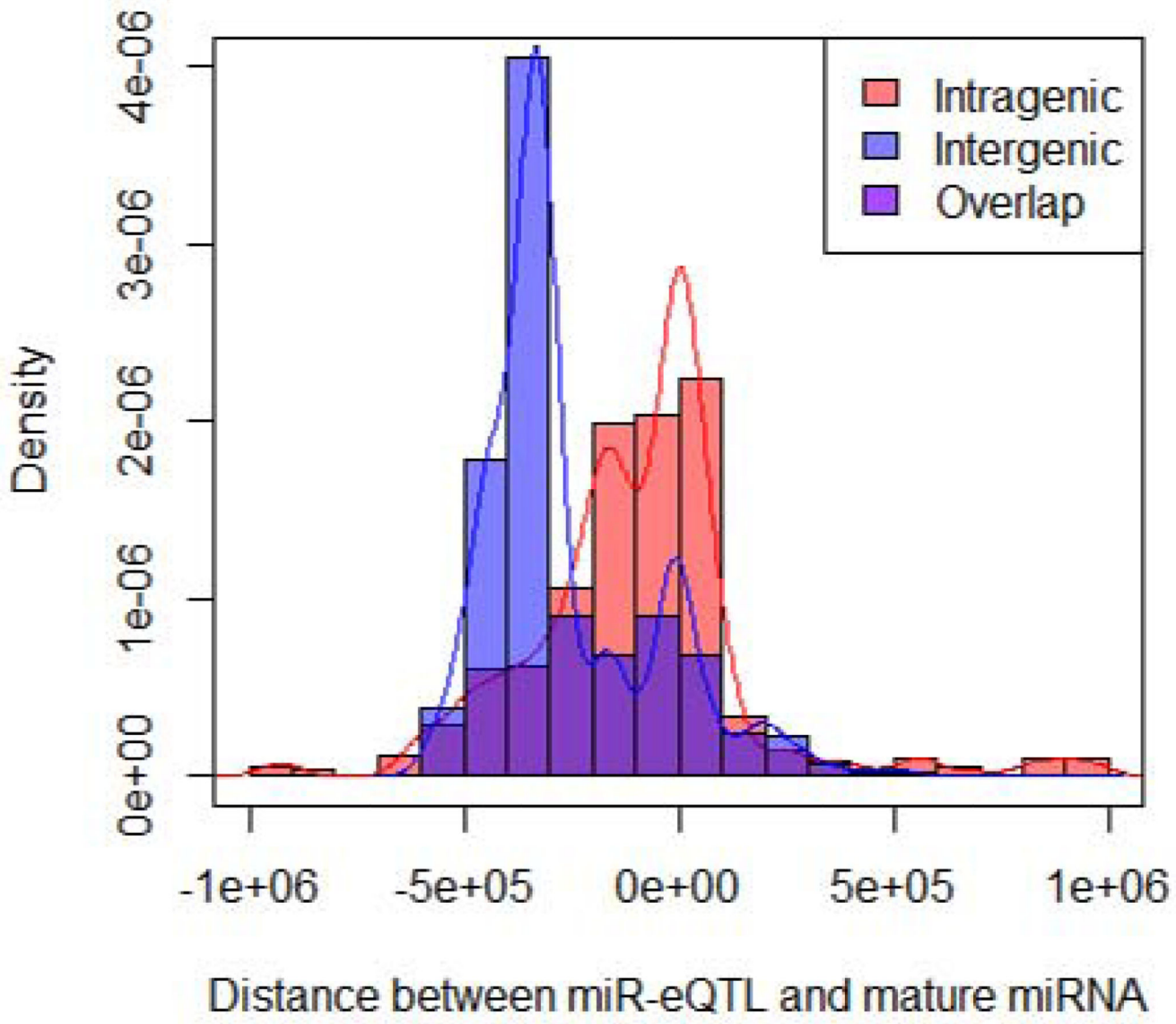

Figure 4. The distribution of distance between *cis*- miR-eQTLs and miRNA position

cis-miR-eQTLs for intergenic miRNAs are generally located further upstream than for intragenic miRNAs. The position of the first nuclear acid of the mature miRNA is marked as 0. The distribution statistics are based on 982 unique *cis*-miR-eQTLs with LD r^2<0.7.

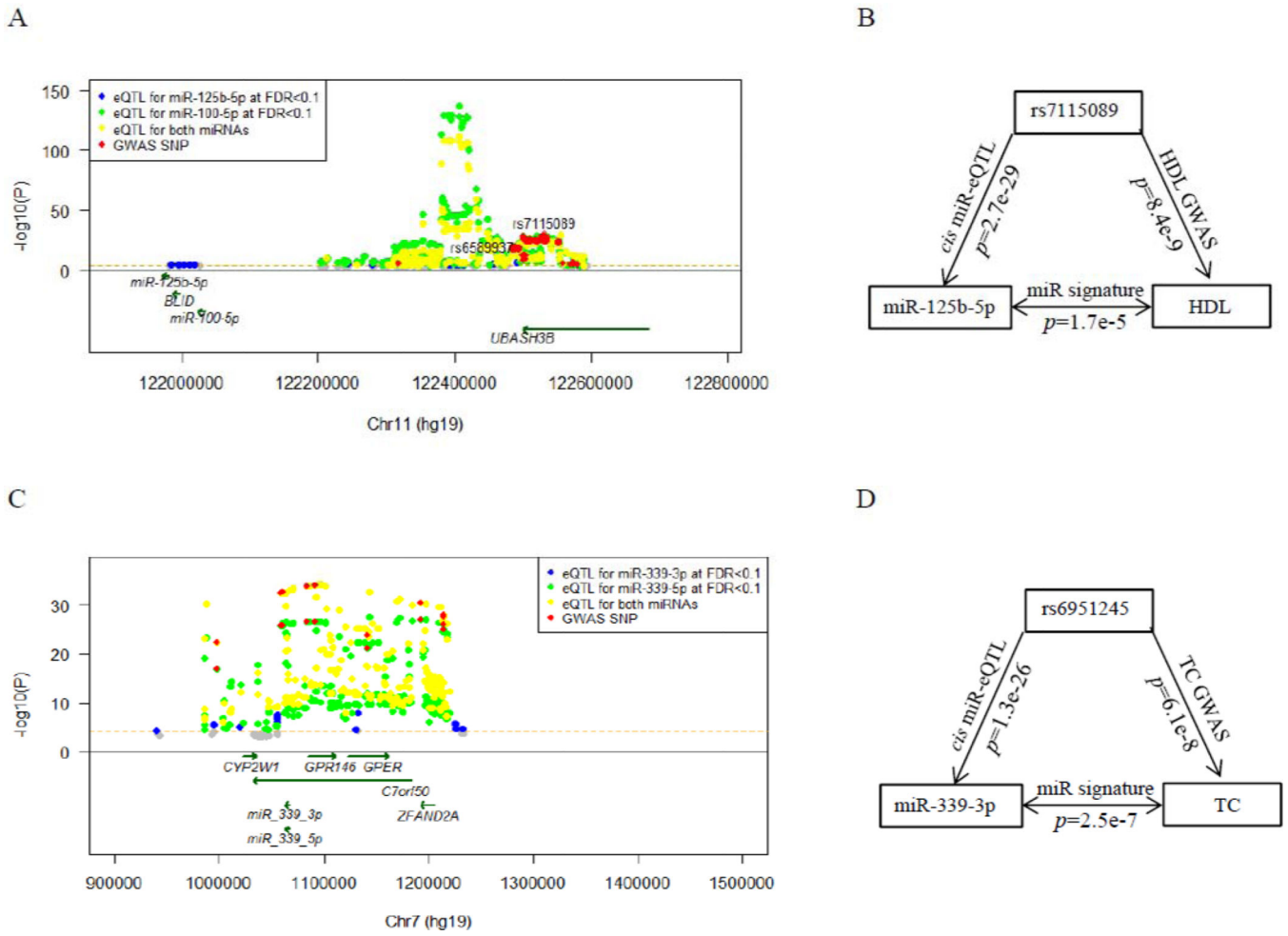

Figure 5. Regional association plot of *cis*- miR-eQTLs that were associated with GWAS SNPs

a) miR-eQTLs for intergenic miRNAs miR-100-5p and miR-125b-5p, with GWAS SNPs for lipid traits, multiple sclerosis, and rheumatoid arthritis. The highlighted SNP, rs7115089, is associated with both HDL and total cholesterol at GWAS $p<5\times10^{-8}$ by linear regression tests [50] b) the triangular relationships between SNP (i.e., rs7115089), miRNA (i.e., miR-125b-5p) and HDL cholesterol; c) miR-eQTLs for intragenic miRNAs miR-339-3p and miR-339-5p, with GWAS SNPs for TC and LDL; d) the triangular relationships between SNP (i.e., rs6951245), miRNA (i.e., miR-339-3p) and TC. –log10 (p) indicates the –log10 transformed miRNA-SNP association *p* values.

Nat Commun. Author manuscript; available in PMC 2015 September 20.

Table 1

Summary association results for 16 peak *cis*-miR-eQTLs having supporting GWAS evidence

Peak cis-miR-eQTL	Chr.	miRNAs	Genome context (miRNA)	miR-eQTL FDR	Proxy miR-eQTLs overlap with GWAS SNPs	GWAS traits	GWAS p value	Trait signature miRNAs
rs7607369	chr2	miR-26b-5p	Intron (*CTDSP1*)	1.2E-05	rs1541777	Height	8.6E-9	
					rs2241527	Severe statin-induced myopathy	1.3E-6	
					rs17572485	Kawasaki disease (with coronary artery lesions)	6.4E-6	
rs13165104	chr5	miR-218-5p, miR-218-2-3p	Intron (*SLIT3*)	7.6E-121	rs4282339	Height	6.6E-16	
rs9342836	chr6	miR-30a-3p	Intron (*C6orf155*)	8.3E-28	rs7349905	Kawasaki disease	5.4E-6	
rs11763835	chr7	miR-339-3p, miR-339-5p	Intron (*C7orf50*)	2.5E-30	rs6951245	Total cholesterol	6.1E-8	miR-339-3p (p=2.5E-7)
					rs13242526	LDL cholesterol	7.7E-6	
rs1839612	chr7	miR-550a-3p	Intron (*ZNRF2*)	2.3E-06	rs2060708	White blood cell count (WBC)	3.9E-6	
rs7789194	chr7	miR-148a-3p	Intergenic	5.8E-03	rs6951827	Triglycerides	7.9E-7	
rs17747335	chr8	miR-598	Intron (*XKR6*)	7.0E-05	rs2244648	Triglycerides	3.0E-7	miR-598 (p=0.032)
					rs7836059	Systemic lupus erythematosus (SLE) (females)	4.0E-10	
					rs10090800	Second to fourth digit length ratio	5.3E-6	
rs28640110	chr9	miR-204-5p	Intron (*TRPM3*)	2.6E-18	rs2993008	Common variable immunodeficiency (Splenectomy)	2.5E-6	
rs2370747	chr11	miR-100-5p, miR-125b-5p	Intergenic	1.8E-130	rs7115089	Total cholesterol	3.2E-10	miR-125b-5p (p=0.005)

Peak cis-miR-eQTL	Chr.	miRNAs	Genome context (miRNA)	miR-eQTL FDR	Proxy miR-eQTLs overlap with GWAS SNPs	GWAS traits	GWAS p value	Trait signature miRNAs
						HDL cholesterol	8.4E-9	miR-100-5p (p=0.039); miR-125b-5p (p=1.68E-5)
						LDL cholesterol	7.6E-6	
					rs7941030	Multiple sclerosis	8.2E-6	
rs11042699	chr11	miR-483-3p	Intron (*IGF2*)	9.6E-04	rs6578985	Coronary artery disease (CAD)	1.6E-6	
rs4905998	chr14	miR-127-3p, miR-134, miR-370, miR-376a-3p, miR-382-5p, miR-431-5p, miR-433, miR-329, miR-409-3p, miR-494, miR-411-3p, miR-654-5p, miR-668, miR-543, miR-323a-3p, miR-337-3p	Intergenic	2.7E-59	rs6575793	Age at menarche	1.7E-10	miR-376a-3p (p=0.007); miR-382-5p (p=0.046)
					rs7149242	Platelet count (PLT)	2.7E-8	
rs2127868	chr14	miR-625-5p, miR-625-3p	Intron (*FUT8*)	3.6E-06	rs1269068	Desialylated Glycan Peak 1 /Biantennary nongalactosylated glycans /Glycan Peak 1	4.4E-18	
					rs1256526	Type 2 diabetes	4.7E-6	miR-625-5p (p=0.035)
rs28483325	chr15	miR-628-3p	Intron (*CCPG1*)	8.1E-07	rs7168869	Mean corpuscular volume (MCV)	4.3E-6	
rs2737	chr17	miR-152	Intron (*COPZ2*)	3.1E-08	rs1553754	Body mass index (BMI)	2.8E-6	
					rs11079828	Attention-deficit/hyperactivity disorder (ADHD)	6.5E-6	
					rs6504340	Primary Tooth Development during Infancy (Number of teeth by one year of age)	6.1E-7	

Peak cis-miR-eQTL	Chr.	miRNAs	Genome context (miRNA)	miR-eQTL FDR	Proxy miR-eQTLs overlap with GWAS SNPs	GWAS traits	GWAS p value	Trait signature miRNAs
rs28576121	chr19	miR-1270	Intron (*ZNF826P*)	2.1E-50	rs7251204	Fasting blood glucose	4.0E-6	miR-1270 (p=0.002)
					rs2562664	Fasting insulin	8.7E-6	
rs373001	chr22	miR-130b-5p, miR-130b-3p	Exon (*PPIL2*)	1.1E-05	rs861844	Myocardial infarction (MI), sudden cardiac arrest in patients with coronary artery disease (CAD)	5.3E-6	

Table 2

Summary of human genome regulatory features of *cis*-miR-eQTLs

Genome Regulatory Track	Nucleotides per track	Fold Change	P-value[+]
UCSC CpG Islands	21575631	2.6	2.97E-16
lincRNAs	127119148	0.8	1
Known regulatory elements (Oreganno)	11265267	3.2	7.48E-15
miRNA targets (TARbase)	49662027	6.9	5.15E-289
miRNA-mediated gene regulatory sites (Patrocles)	3375454	10.0	1.64E-37
GM12878 CTCF	44516245	2.0	7.42E-15
GM12878 H3k27ac	125879335	1.9	4.11E-35
GM12878 H3k27me3	1136357520	1.4	1.17E-92
GM12878 H3k36me3	631024019	1.6	2.80E-106
GM12878 H3k4me1	242340600	1.9	6.59E-63
GM12878 H3k4me3	120458965	2.0	2.90E-37

[+]P-values are for binomial tests for enrichment of observed over expected;

GM12878 is a lymphoblastoid cell line; CTCF mark CTCF Binding Sites by ChIP-seq from ENCODE; H3k27ac and H3K4me1 mark active/poised enhancers; H3K4me3 mark, active/poised promoters; and H3K36me3 mark actively transcribed regions. me, methylation.