Genome-wide identification of microRNA expression quantitative trait loci

Tianxiao Huan1,2, Jian Rong3, Chunyu Liu1,2, Xiaoling Zhang1,4, Kahraman Tanriverdi5, Roby Joehanes1,2,6,7,8, Brian H. Chen1,2, Joanne M. Murabito1,9, Chen Yao1,2, Paul Courchesne1,2, Peter J. Munson6, Christopher J. O’Donnell1,4, Nancy Cox10, Andrew D. Johnson1,4, Martin G. Larson1,3, Daniel Levy1,2 & Jane E. Freedman5

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 5,239 individuals, and identify 5,269 cis-miR-eQTLs for 76 mature microRNAs. Forty-nine per cent of cis-miR-eQTLs are located 300–500 kb 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 single nucleotide polymorphisms. 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 (for example, rs7115089, miR-125b-5p and high-density lipoprotein 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.

1 The Framingham Heart Study, Framingham, Massachusetts 01702, USA. 2 The Population Sciences Branch, Division of Intramural Research, National Heart, Lung, and Blood Institute, Bethesda, Maryland 20824, USA. 3 Department of Mathematics and Statistics, Boston University, Boston, Massachusetts 02118, USA. 4 Cardiovascular Epidemiology and Human Genomics Branch, Division of Intramural Research, National Heart, Lung and Blood Institute, Bethesda, Maryland 20824, USA. 5 Department of Medicine, University of Massachusetts Medical School, Worcester, Massachusetts 01655, USA. 6 Mathematical and Statistical Computing Laboratory, Center for Information Technology, National Institutes of Health, Bethesda, Maryland 20824, USA. 7 Harvard Medical School, Harvard University, Boston, Massachusetts 02115, USA. 8 Hebrew SeniorLife, Boston, Massachusetts 02131, USA. 9 Department of Medicine, Section of General Internal Medicine, Boston University School of Medicine, Boston, Massachusetts 02118, USA. 10 Department of Human Genetics, University of Chicago, Chicago, Illinois 60637, USA. Correspondence and requests for materials should be addressed to D.L. (email: Levyd@nih.gov) or to J.E.F. (email: jane.freedman@umassmed.edu).
MicroRNAs (miRNAs), a class of small noncoding RNAs, serve as key post-transcriptional regulators of gene expression and miRNA translation. miRNAs are increasingly recognized as mediators in a variety of biological processes including cardiovascular development and disorders. Highly specific miRNA expression patterns have been reported in association with heart failure, myocardial infarction and cancer. 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 miRNA expression levels of many genes. These eQTL analyses have demonstrated that transcript levels of many miRNAs behave as heritable quantitative traits. In contrast to more extensive investigations of mRNA eQTLs in multiple tissues, such as blood, brain, fat, and liver, 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). 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 5,239 Framingham Heart Study (FHS) participants. We analyze the associations of ~1 million 10 Genomes Project imputed single nucleotide polymorphisms (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. 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 with regulatory SNPs annotated by the ENCODE project and with complex trait-associated SNPs identified in prior genome-wide association studies (GWAS), 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 5,239 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. The demographic and clinical characteristics of the 5,239 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. The demographic and clinical characteristics of the 5,239 FHS participants included in our analysis are shown in Supplementary Data 1. The demographic and clinical characteristics of the 5,239 FHS participants included in our analysis are shown in Supplementary Data 1. The demographic and clinical characteristics of the 5,239 FHS participants included in our analysis are shown in Supplementary Data 1. The demographic and clinical characteristics of the 5,239 FHS participants included in our analysis are shown in Supplementary Data 1. The demographic and clinical characteristics of the 5,239 FHS participants included in our analysis are shown in Supplementary Data 1. The demographic and clinical characteristics of the 5,239 FHS participants included in our analysis are shown in Supplementary Data 1. The demographic and clinical characteristics of the 5,239 FHS participants included in our analysis are shown in Supplementary Data 1. The demographic and clinical characteristics of the 5,239 FHS participants included in our analysis are shown in Supplementary Data 1. The demographic and clinical characteristics of the 5,239 FHS participants included in our analysis are shown in Supplementary Data 1. The demographic and clinical characteristics of the 5,239 FHS participants included in our analysis are shown in Supplementary Data 1. The demographic and clinical characteristics of the 5,239 FHS participants included in our analysis are shown in Supplementary Data 1. The demographic and clinical characteristics of the 5,239 FHS participants included in our analysis are shown in Supplementary Data 1. The demographic and clinical characteristics of the 5,239 FHS participants included in our analysis are shown in Supplementary Data 1. The demographic and clinical characteristics of the 5,239 FHS participants included in our analysis are shown in Supplementary Data 1. The demographic and clinical characteristics of the 5,239 FHS participants included in our analysis are shown in Supplementary Data 1. The demographic and clinical characteristics of the 5,239 FHS participants included in our analysis are shown in Supplementary Data 1. The demographic and clinical characteristics of the 5,239 FHS participants included in our analysis are shown in Supplementary Data 1.

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 2,138 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. In the subsequent sections, we focus on miR-eQTLs from analyses that were unadjusted for cell counts of 5,239 individuals (Supplementary Data 4). The miR-eQTLs from the model that adjusted for imputed cell counts in the larger set of 5,024 participants are provided in Supplementary Data 5.

To evaluate the reproducibility of detected miR-eQTLs, we split our overall sample set:1 into two sets by pedigrees creating separate discovery and replication sets, and identified cis- and trans-miR-eQTLs in each set. At discovery false discovery rates (FDRs) of < 0.1, 133 miR-eQTLs were 54% 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...
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 untranslated 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 1,984 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 200 kb upstream to 100 kb 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 123 cis-miR-eQTLs (487 SNP-miRNA pairs, 49%) were in the region defined by 500 to 300 kb 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 ±50 kb 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, that is, 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%), promotors (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.

There were 1,066 (20%) cis-miR-eQTLs that overlapped with cis-mRNA-eQTLs identified in whole blood (enrichment $P < 1e - 300$ by hypergeometric test). 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 Mb region flanking the 132 cis-miR-eQTLs (chr14: 100.5–102.5 Mb) with the regulatory feature tracks downloaded from UCSC Genome Browser (genome.ucsc.edu). Supplementary Figure 8 showed that the nearby regions of the 132 cis-miR-eQTLs for those 12 miRNAs overlap with Enhancer active region (chr14:101,100 kb–101,200 kb, H3K4Me1 and H2K27AC track, marked in lightyellow rectangle). The highly unmethylated status of GM12878, K562, HeLa-S3 and HepG2 cell lines are in chr14:101,400 kb–101,600 kb upstream of those cis-mRNA-eQTL miRNAs (CpG Methylation by Methy450K Bead Arrays from ENCODE/HAIB track, marked by pink colour).

We also discovered 11 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 shown in Supplementary Data 13, we found 923 cis-miR-eQTLs for 3,384 mRNA-SNP association pairs (87%) that remained significant at FDR < 0.1 (corresponding \( P < 6.6 \times 10^{-5} \)) when conditioning on mRNA expression levels. These findings indicate that cis genetic variants may affect expression levels of neighbouring miRNAs and mRNAs.

\[ \text{cis-miR-eQTLs and miRNA signatures for complex traits.} \]

We linked the cis-miR-eQTLs with GWAS SNPs in the NHGRI GWAS Catalogue and the NHLBI GRASP data set\(^{21,22}\). Among

### Table 1 | Summary association results for 16 peak cis-miR-eQTLs having supporting GWAS evidence.

| Peak cis- miR-eQTL | Chr. | miRNAs | Genomic 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 (CTDSPI) | 1.2e-5 | rs1541777 | Height | 8.6e-9 | — |
| rs13165104 chr5   | miR-218-5p, miR-218-2p | Intron (SFT3) | 7.6e-12 | rs4282339 | Severe statin-induced myopathy | 1.3e-6 | — |
| rs9342836 chr6    | miR-30a-3p | Intron (C6orf55) | 8.3e-28 | rs7349905 | Kawasaki disease | 6.4e-6 | — |
| rs11763835 chr7   | miR-339-3p, miR-339-5p | Intron (C7orf50) | 2.5e-30 | rs6951245 | Total cholesterol | 6.1e-8 | miR-339-3p (\( P = 2.5e^{-7} \)) |
| rs1839612 chr7    | miR-550a-3p | Intron (ZNF2) | 2.3e-6 | rs13242526 | LDL cholesterol | 7.7e-6 | — |
| rs7789194 chr7    | miR-148a-3p | Intergenic | 5.8e-3 | rs6951827 | Triglycerides | 7.9e-7 | — |
| rs1774733 chr8    | miR-598 | Intergenic (XKRG6) | 7.0e-5 | rs22446468 | Triglycerides | 3.0e-7 | miR-598 (\( P = 0.032 \)) |
| rs7836059         | | | | rs10090800 | Systemic lupus erythematosus (SLE) (females) | 4.0e-10 | — |
| rs28640110 chr9   | miR-204-5p | Intron (TRPM3) | 2.6e-18 | rs2993008 | Second to fourth digit length ratio | 5.3e-6 | — |
| rs2370747 chr11   | miR-100-5p, miR-125b-5p | Intergenic | 1.8e-130 | rs7115089 | Total cholesterol | 3.2e-10 | miR-125b-5p (\( P = 0.005 \)) |
| rs11042699 chr11  | miR-483-3p | Intron (IGF2) | 9.6e-4 | rs7941030 | HDL cholesterol | 8.4e-9 | — |
| 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 | LDL cholesterol | 7.6e-6 | — |
| rs2127868 chr14   | miR-625-5p, miR-625-3p | Intron (FUT8) | 3.6e-6 | rs7149242 | Platelet count (PLT) | 2.7e-8 | — |
| rs28483325 chr15  | miR-628-3p | Intron (CCPGD) | 8.1e-7 | rs7168869 | Desialylated Glycan Peak 1/Biglycan peak | 4.4e-18 | — |
| rs2373 chr17      | miR-152 | Intron (CP7Z2) | 3.1e-8 | rs1553754 | Desialylated Glycan Peak 2 | 4.7e-6 | miR-625-5p (\( P = 0.035 \)) |
| rs11079828         | | | | rs6504340 | Mean corpuscular volume (MCV) | 4.3e-6 | — |
| rs28576121 chr19  | miR-1270 | Intron (ZNF26P) | 2.7e-50 | rs7251204 | Body mass index (BMI) | 2.8e-6 | — |
| rs373001 chr22    | miR-130b-5p, miR-130b-3p | Exon (PPL2) | 1.1e-5 | rs2562664 | Fasting blood glucose | 4.0e-6 | miR-1270 (\( P = 0.002 \)) |

Chr., chromosome; eQTL, expression quantitative trait loci; FDR, false discovery rate; GWAS, genome-wide association studies; miR, microRNA.
miR-125b-5p showed differential expression in relation to plasma upstream of their two associated miRNAs. We also found that NATURE COMMUNICATIONS | DOI: 10.1038/ncomms7601 | www.nature.com/naturecommunications with rheumatoid arthritis. These eQTLs are located (TGs)), 1 (rs7941030) with multiple sclerosis and 1 (rs1216554) protein (LDL) cholesterol, total cholesterol (TC) and triglycerides (high-density lipoprotein (HDL) cholesterol, low-density lipoprotein (LDL) cholesterol, total cholesterol (TC) and triglycerides (TGs)), 1 (rs7941030) with multiple sclerosis and 1 (rs1216554) with rheumatoid arthritis. These eQTLs are located ~519 kb upstream of their two associated miRNAs. We also found that miR-125b-5p showed differential expression in relation to plasma TC (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 (MR) 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 5,239 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.14 using umbilical cord blood from 180 newborns, identified only 12 cis-miR-eQTLs at FDR < 0.5. In another study, no cis-miR-eQTLs were found in 176 lymphoblastoid cell lines from European and African ancestry samples15. 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 moderate 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 \( h^2_{\text{narrow}} \) 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 \( h^2_{\text{miR}} \) increased, the proportion of variance of the miRNA transcript

| 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                | 12719148             | 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            | 1.4         | 1.17e − 92    |
| GM12878 H3k27me3        | 1136357520           | 1.6         | 2.80e − 106   |
| GM12878 H3k36me3        | 631024019            | 1.9         | 6.59e − 63    |
| GM12878 H3k4me3         | 242340600            | 2.0         | 2.90e − 37    |

*P values are for binomial tests for enrichment of observed over expected; GM12878 is a lymphoblastoid cell line; CTCF marks CTCF Binding Sites by ChIP-seq from ENCODE; H3k27ac and H3k4me1 mark active/ poised enhancers; H3K4me3 marks active/ poised promoters; and H3K36me3 marks actively transcribed regions.
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. We identified 11 mature miRNAs from intragenic miRNAs that share cis-eQTLs with their host protein-coding genes (Supplementary Data 12). Among the cis-miRNA-eQTL miRNAs, 15 miRNAs have alternative intronic promoters (alternative intronic promoters were from ref. 29). We overlapped the cis-miRNA-eQTLs and expression regulatory elements annotations from ENCODE nearby regions of each miRNA (+50 kb). We found, in some examples (Supplementary Fig. 10), cis-miRNA-eQTLs near alternative intronic promoter regions demonstrated promoter and enhancer activities and were highly unmethylated 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/trait 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 (for example, 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 (\(P = 1.7 \times 10^{-5}\), by a linear regression test). In a parallel project focusing on differentially expressed mRNAs in a cohort of original FHS cohort participants were recruited and they have been examined every 4 to 8 years. From 2002 to 2005, the adult children of the offspring cohort participants (the third generation cohort) were recruited and examined in the examination cycle 2 (2008–2010). This study was approved under the 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 (2.5 ml) were collected in PAXgene Blood RNA tubes (Qiagen, Valencia, CA, USA) and frozen at \(-80^\circ\mathrm{C}\). Total RNA was isolated from the frozen PAXgene Blood RNA tubes (Asuragen, Inc. Austin, TX, USA) and a 2100 Bioanalyzer Instrument (Agilent, Santa Clara, CA, USA) was used for RNA quality assessment. Isolated RNA samples were converted to complementary DNA using TaqMan miRNA Reverse Transcription Kit and MegaFlex Human RT Primer Pool Av1.1 and Pool Bv3.1. (Life Technologies, Foster City, CA, USA) in a 384-well Thermal Cycler. The complementary DNA samples were preAmplified using TaqMan PreAmp Master Mix and PreAmp Primers, Human Pool A v2.1 and Pool B v3.0. (Life Technologies).

Quantitative reverse transcription PCR (qRT–PCR) reactions were performed with the BioMark System using (Fluidigm, South San Francisco, CA, USA) TaqMan miRNA Assays (Life Technologies). As described in the published literature, measurement of RNA by qRT–PCR is reliable and has high specificity and sensitivity. 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 s 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 (Ci) values for four technical variables: isolation batch (50 batches), RNA concentration, RNA quality (defined as RNA integrity number) and RNA 260/280 ratio (ratio of absorbance at 260 and 280 nm; measured using a spectrophotometer). Histograms (Supplementary Fig. 11) show that this model explains 20–60% of variability of raw miRNA measurements for 80% of miRNAs.

miR-eQTL mapping. Because of the computational burden of running linear mixed effects (LMEs) models for ~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 1 Mb 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 \(P < 1 \times 10^{-5}\) for cis and \(P < 1 \times 10^{-5}\) for trans. These P value thresholds were chosen to ensure that miR-eQTLs at a FDR <0.1 were not omitted as a result of this pre-filtering step. Step 2: we used a linear mixed model 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 lme4() function of Kinship Package (http://cran.r-project.org/web/packages/kinship) as shown in Equation (2). In Equations (1) and (2), \(r\) is the error term for each independent observation.

\[
m_{\text{miR}} = \text{SNP} + \text{age} + \text{sex} + \text{cohort} + \sum_{i=1}^{n_{\text{tech}}} \text{technical covariates} + \varepsilon
\]

\[
m_{\text{miR}} = \text{SNP} + \text{age} + \text{sex} + \sum_{i=1}^{n_{\text{tech}}} \text{technical covariates} + \text{FAM} + \varepsilon
\]

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). Based on the coordinates of 280 mature miRNAs and 9.8 × 106 SNPs, we estimated there were 13,955,272 (1.4 × 106) potential SNP–miRNA pairs, where the SNP was located within 1 Mb on either side of the corresponding mature miRNA. We estimated there were 1.4 × 106 potential cis SNP–miRNA pairs, and 2.7 × 106 (that is, 280 × 9.8 × 106 − 1.4 × 106) potential trans SNP–miRNA pairs. We used the Benjamini–Hochberg method to calculate FDR for cis- and trans-miR-eQTLs by correcting for 1.4 × 106 potential cis SNP–miRNA pairs and 2.7 × 106 potential trans SNP–miRNA pairs, respectively. We selected an FDR threshold of 0.1, corresponding to \(P < 6.6 \times 10^{-5}\) for cis- and \(P < 1.0 \times 10^{-8}\) for trans-miR-eQTLs.

For identified cis-miR-eQTLs at FDR<0.1, we used Fragmented Exhaustive Search for TagSNPs (FESTA) to select non-redundant miR-eQTLs based on a set of LD thresholds \(r^2 = 0.5, 0.7, 0.9\) and 1. FESTA used a mixture of search techniques to partition the whole SNP set into disjointed precints and selected a single SNP for each SNP block, which represented a set of SNPs at a LDL threshold.

To estimate the replicability of miR-eQTLs, we split the overall sample set at a 1:1 ratio 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.

miRNA expression data. Whole blood samples (2.5 ml) were collected in PAXgene tubes by Asuragen, Inc. (PreAnalytiX, Hombrechtickon, 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 30 mg RNA samples were amplified using the WT-Ovation Pico RNA Amplification System (NuGEN, San Carlos, CA, USA) as recommended by the manufacturer in an automated manner using the genchip array station. RNA expression was conducted using the Affymetrix Human Exon Array ST 1.0 (Affymetrix, Inc., Santa Clara, CA, USA). The core probe sets were annotated using the Affymetrix annotation files from NetAffx (www.netaffx.com, HuEx-1.0-st- v2.0z29.18.probeset.csv).

The raw gene expression data were at first preprocessed by quartile normalization. Then the robust multi-array average values of each gene (17,318
measured genes) were adjusted for a set of technical covariates, for example, chip batch by fitting LME models. Imputed blood cell counts (that is 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 identified in this study with cis-mRNA-eQTLs at FDR<0.1 identified by refs 9, 25. Hypergeometric test was used to evaluate if cis-miR-eQTLs were significantly enriched for cis-miRNA-eQTLs. For those overlap eQTLs, that is, 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 and mRNA expression levels but conditional regression on corresponding mRNA expression levels.

Estimating effects of cell counts in the miR-eQTLs. Since the miR-eQTLs in whole blood may be driven by cellular composition, we compared the miR-eQTLs in 2,138 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 2,138 individuals in the FHS third generation cohort, including seven cell types, WBC, RBC, platelet, neutrophil, lymphocyte, monocyte, eosinophil and basophil. The cell counts and proportions for 5,024 FHS participants were estimated using mRNA expression values by partial least squares regression prediction. The estimated cell count proportion values are highly consistent with the measured cell counts proportion values (Joehanes R, 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 (that is, 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 analyses unadjusted for cell counts is 5,239 and the maximum sample size of analyses adjusted for cell counts is 5,024.

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

\[
\text{miR} = \text{age} + \text{sex} + \sum_{i=1}^{n} \text{technical covariates} + \text{FAM} + \epsilon
\]

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 \( \epsilon \). \( \epsilon \) is the error term for each independent observation. \( h^2_{\text{miR}} \) was the proportion of the additive polygenic genetic variance (\( \sigma^2_{\text{FAM}} \)) among the total phenotypic variance (\( \sigma^2_{\text{miR}} \)) of miRNA expression:

\[
h^2_{\text{miR}} = \frac{\sigma^2_{\text{FAM}}}{\sigma^2_{\text{miR}}}
\]

We estimated \( h^2_{\text{miR}} \) for every mRNA expression trait (247 miRNAs expressed in >1,000 samples) using the lmekin() function of Kinship package (http://cran.r-project.org/web/packages/kinship/).26

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 \( h^2_{\text{miR\_eQTL}} \)), we used the following two models: Full model:

\[
\text{miR} = \text{SNP} + \text{age} + \text{sex} + \text{cohort} + \sum_{i=1}^{n} \text{technical covariates} + \text{FAM} + \epsilon
\]

Null model:

\[
\text{miR} = \text{age} + \text{sex} + \text{cohort} + \sum_{i=1}^{n} \text{technical covariates} + \text{FAM} + \epsilon
\]

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. \( \epsilon \) 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 \( h^2_{\text{miR\_eQTL}} \), and was calculated as follows:

\[
h^2_{\text{miR\_eQTL}} = \max \left( 0, \frac{\sigma^2_{\text{miR\_eQTL,full}} - \sigma^2_{\text{miR\_eQTL,full} - \text{SNP}}}{\sigma^2_{\text{miR}}} \right)
\]

where \( \sigma^2_{\text{miR\_eQTL,full}} \) and \( \sigma^2_{\text{miR\_eQTL,full} - \text{SNP}} \) were the polygenic and error variances, respectively, when modelling with the tested miR-eQTL \( \text{miR\_eQTL} \), \( \sigma^2_{\text{miR}} \) and \( \sigma^2_{\text{error}} \) were the polygenic and error variances, respectively, when modelling without the tested miR-eQTL. The lmekin() function in the Kinship package39 was used to estimate \( h^2_{\text{miR\_eQTL}} \).

Identification of differentially expressed miRNAs for complex traits. We used the NHGRI GWAS Catalogue (http://www.genome.gov/gwastudies/)34 and NHLBI GRASP database (http://apps.nhlib.nih.gov/grasp/)17 to annotate cis-eQTLs that are present in cis-eQTLs identified in this study were compared with SNPs in the NHGRI GWAS Catalogue and NHLBI GRASP GWAS results for SNPs at \( P \leq 1 \times 10^{-5} \). For the complex traits that could be mapped with cis-miR-eQTLs (and also were measured in the FHS), including menarche, lipids (HDLC, cholesterol, TG and TC), type II diabetes mellitus (T2D) and glucose level, we used linear mixed models to test their association with miR-eQTL mRNA 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 HDLC, cholesterol, TG, TC, T2D and glucose after accounting for age, sex, cell counts and technical covariates (see ‘miRNA normalization’ in methods) and family structure in LME models implemented in the lmekin function39. Differentially expressed miRNA associated with age at menarche were tested in LME models (lmekin) after accounting for birth year, sex, cell counts, technical covariates and family structure.

miRNA TSS and promoter regions. The transcriptional regulatory mechanisms affecting miRNA expression are unclear. There are technical barriers to the precise identification of primary miRNAs, TSSs and promoter regions for most mature miRNAs29. Recently, Marsico et al.31 and Chen et al.32 predicted miRNAs TSSs. Their results were incorporated with the results from previous similar studies33,34. However, by comparing the TSS positions identified by these two studies, there was a 53 kb discordance. 53 kb average between TSSs was predicted. Therefore, we annotated the corresponding mature miRNAs. In our analysis, we annotated the miRNA TSSs collected and predicted by these two studies. The predicted promoter annotations for miRNAs were obtained from Marsico et al.31 which were screened within ±50 kb from the TSSs for each miRNA.

Functional annotation of cis-miR-eQTLs. We annotated the genomic features cis-miR-eQTLs (n = 5,269) using HaploReg45, which integrates results from ENCODE20. The overlap of cis-miR-eQTLs with ENCODE annotated SNPs in promoter, enhancer and 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, including ENCODE histone modification sites, and TF and CTCF binding sites in lymphoblastoid cell lines (GM12878), Open Regulatory Annotation (ORegAnno)46, UCSC CpG islands and long intergenic non-coding RNA46. We also downloaded other regulatory tracks, including experimentally validated miRNA targets from TARbase27, and experimentally supported miRNA-mediated gene regulatory sites from Patrocles28. Binomial tests were used to evaluate if the identified cis-miR-eQTLs or SNPs with MAF > 0.01 and imputation quality ratio > 0.1. To match the distribution of MAFs of the permutation SNPs (the permutation-SNPs set) with the cis-miR-eQTL 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, 0.3). 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 (that is SNPs in promoter, enhancer and protein binding regions) was compared with the overlap between the tested-SNPs and regulatory region SNPs.

MR test. We used a two-stage least squares (2SLS) MR method46 to estimate the causal relationships between miRNAs and complex traits measured in FHS participants; the traits analyzed included menarche, lipids (HDLC, TG and TC), T2D and glucose, using cis-miR-eQTLs as instrumental variables. 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 Catalogue (http://www.genome.gov/gwastudies/)31 or in the NHLBI GRASP database (http://apps.nhlbi.nih.gov/grasp/32), and the miRNA showed differential expression in 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. For each variable included in the model, we conducted a separate analysis. We ran the model on the whole sample or sex-stratified samples, including age, sex, family structure and four technical variables as covariates (see in the miRNA normalization section). cis-miR-eQTLs with an F-statistic <10, indicating a weak instrument, were excluded. The first stage of the
2SLS method involves using a linear regression of the modifiable exposure (miRNA) on the instrumental variable (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 is used to estimate whether the estimates derived from the first and second stage of the 2SLS are consistent.

References

1. Lee, R., Feinbaum, R. & Ambros, V. A short history of a short RNA. Cell 116, 589–592 (2004).
2. Lee, R. C. & Ambros, V. An extensive class of small RNAs in Caenorhabditis elegans. Science 294, 862–864 (2001).
3. Cordes, K. R. & Srivastava, D. MicroRNA regulation of cardiovascular development. Circ. Res. 104, 724–732 (2009).
4. Small, E. M. & Olson, E. N. Pervasive roles of microRNAs in cardiovascular biology. Nature 469, 336–342 (2011).
5. Thum, T. et al. MicroRNAs in the human heart: a clue to fetal gene reprogramming in heart failure. Circulation 116, 258–267 (2007).

The Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

How to cite this article: Huan, T. et al. Genome-wide identification of microRNA expression quantitative trait loci. Nat. Commun. 6:6601 doi: 10.1038/ncomms7601 (2015).