






RESEARCH PAPER



Epigenome-wide association study of DNA methylation and microRNA expression highlights novel pathways for human complex traits

Tianxiao Huan ^{a,b}, Michael Mendelson ^{a,b}, Roby Joehanes ^{a,b}, Chen Yao ^{a,b}, Chunyu Liu^{a,b,c}, Ci Song^{a,b}, Anindya Bhattacharya^d, Jian Rong^e, Kahraman Tanriverdi^f, Joshua Keefe^{a,b}, Joanne M. Murabito^{a,g}, Paul Courchesne ^{a,b}, Martin G. Larson^{a,e}, Jane E. Freedman^f, and Daniel Levy^{a,b}

^aThe National Heart, Lung, and Blood Institute, Boston University's Framingham Heart Study, Framingham, MA, USA; ^bThe Population Sciences Branch, Division of Intramural Research, National Heart, Lung, and Blood Institute, Bethesda, MD, USA; ^cDepartment of Biostatistics, Boston University School of Public Health, Boston University, Boston, MA, USA; ^dDepartment of Computer Science and Engineering, University of California, San Diego, CA, USA; ^eDepartment of Mathematics and Statistics, Boston University, Boston, MA, USA; ^fDepartment of Medicine, University of Massachusetts Medical School, Worcester, MA, USA; ^gDepartment of Medicine, Section of General Internal Medicine, Boston University School of Medicine, Boston, MA, USA

ABSTRACT

DNA methylation (DNAm) and microRNAs (miRNAs) have been implicated in a wide-range of human diseases. While often studied in isolation, DNAm and miRNAs are not independent. We analyzed associations of expression of 283 miRNAs with DNAm at >400K CpG sites in whole blood obtained from 3565 individuals and identified 227 CpGs at which differential methylation was associated with the expression of 40 nearby miRNAs (*cis*-miR-eQTMs) at FDR<0.01, including 91 independent CpG sites at $r^2 < 0.2$. *cis*-miR-eQTMs were enriched for CpGs in promoter and polycomb-repressed state regions, and 60% were inversely associated with miRNA expression. Bidirectional Mendelian randomization (MR) analysis further identified 58 *cis*-miR-eQTM CpG-miRNA pairs where DNAm changes appeared to drive miRNA expression changes and opposite directional effects were unlikely. Integration of genetic variants in joint analyses revealed an average partial between *cis*-miR-eQTM CpGs and miRNAs of 2% after conditioning on site-specific genetic variation, suggesting that DNAm is an important epigenetic regulator of miRNA expression. Finally, two-step MR analysis was performed to identify putatively causal CpGs driving miRNA expression in relation to human complex traits. We found that an imprinted region on 14q32 that was previously identified in relation to age at menarche is enriched with *cis*-miR-eQTMs. Nine CpGs and three miRNAs at this locus tested causal for age at menarche, reflecting novel epigenetic-driven molecular pathways underlying this complex trait. Our study sheds light on the joint genetic and epigenetic regulation of miRNA expression and provides insights into the relations of miRNAs to their targets and to complex phenotypes.

ARTICLE HISTORY

Received 18 March 2019
Revised 24 June 2019
Accepted 28 June 2019

KEYWORDS

DNA methylation; microRNA; miR-eQTM; eQTM; complex traits; age at menarche


Introduction

DNA methylation (DNAm) and microRNAs (miRNAs) influence protein-coding gene expression independent of and as a consequence of changes in DNA sequence. Recent population-based studies have demonstrated that variation in DNAm and miRNA expression are associated with interindividual phenotypic variation. For example, locus-specific differences in DNAm are associated with a wide range of cardiometabolic phenotypes such as body mass index^[1], lipid traits [2,3], glycaemic traits [4], and inflammatory biomarkers [5]. Aging and environmental exposures, such as

cigarette smoking and alcohol consumption, also have large effects on genome-wide DNAm [6–9]. In addition to DNAm, specific miRNA expression patterns have been associated with various cardiovascular disorders and cancer [10–16]. Both DNAm and miRNA expression are often studied in isolation, but they are not independent [17,18] and few studies have integrated DNAm and miRNA in relation to human diseases.

DNAm, the covalent binding of a methyl group to the 5' carbon of cytosine, occurs primarily at cytosine-phosphate-guanine (CpG) dinucleotide sequences in the genome and influences the pre-transcriptional regulation of protein-coding genes

CONTACT Daniel Levy  levydnhlbi.nih.gov; Tianxiao Huan  Tianxiao.huan@nih.gov  Framingham Heart Study, 73 Mount Wayte Avenue, Framingham, MA 01702, USA

 Supplemental data for this article can be accessed [here](#).

© 2019 Informa UK Limited, trading as Taylor & Francis Group

by affecting the binding of chromatin proteins and transcription factors (TFs) to DNA sequences [19,20]. For example, the methylation of CpG island promoters prevents binding of TFs, inhibits transcription initiation, and results in gene silencing. Methylation of the gene body region, however, can stimulate gene transcription elongation, and may also affect messenger RNA (mRNA) splicing [21]. miRNAs, which comprise a large family of single-stranded non-protein-coding RNAs of approximately 22 nucleotides, also influence protein-coding gene expression by causing degradation of complementary mRNAs or inhibiting translation [22,23]. There are over 2000 mature miRNAs [24] that control, in part, the expression of about two-thirds of human genes [25]. In addition, genetic variation strongly influences DNAm [26–32] and miRNA expression levels [33].

While the relationship between DNAm and protein-coding gene expression has been studied extensively [34,35], a comprehensive genome-wide interrogation of the relationship between DNAm and miRNAs is lacking. Integration of DNA sequence variation, DNAm, and miRNA expression may also provide vital insights into the biological processes. To this end, we performed an epigenome-wide association study of DNAm at over 400,000 CpGs with expression of 283 miRNAs in 3565 Framingham Heart Study (FHS) participants. A comprehensive map of CpGs linked to miRNAs (miRNA expression quantitative trait methylation loci; miR-eQTMs) as well as bi-directional Mendelian randomization (MR) analyses were used to unravel the influence of DNAm on miRNA expression. We additionally integrated genetic variants, miR-eQTMs, and protein-coding gene targets of miRNAs with human complex traits. Comprehensive two-step MR analyses were performed to explore the epigenetically-driven molecular cross-talk between DNAm, miRNAs, and human complex phenotypes (study design outlined in Figure 1).

Results

Clinical characteristics summary

A total of 3565 FHS participants had whole blood miRNA expression measurements (for 283 miRNAs

expressed in >20% of participants) and DNAm measurements at 415,318 CpGs were used to identify miR-eQTMs, including 2124 FHS Offspring samples (mean age 66, 55% female) and 1441 FHS 3rd generation samples (mean age 45, 52% female). The demographic and clinical characteristics of the study participants are shown in Supplementary Table 1.

mRNA expression was measured in 5626 FHS participants, of which 5357 participants who also had miRNA expression were used for subsequent miRNA-mRNA coexpression analysis. A total of 3345 participants had miRNA, mRNA, and DNAm measurements.

Evaluation of the reproducibility of detected *cis*- and *trans*-miR-eQTMs

To evaluate the reproducibility of detected *cis*- and *trans*-miR-eQTMs in our data, we used an internal validation strategy by splitting the overall sample set into two independent subgroups; a discovery set ($N = 1761$) and a replication set ($N = 1804$). The clinical characteristics of the discovery and replication sets did not show substantive differences (Supplementary Table 2). We identified *cis*- and *trans*-miR-eQTMs in each set in the same fashion as in the overall sample (see the following subsection). The association t-statistic values for *cis*- and *trans*-miR-eQTMs were consistent between the discovery and replication sets (Supplementary Figure 1). At successively stringent discovery FDR thresholds of <0.1, 0.05, and 0.01, the replication ratio of *cis*-miR-eQTMs at replication FDR<0.1 are 55% (88/160), 72% (83/116), and 83% (65/78), respectively. In contrast, the replication ratio of *trans*-miR-eQTMs at replication FDR values of <0.1 are much lower at 0.3% (51/15,363), 0.8% (7/911) and 2% (6/304), respectively. Due to the low replication rate of *trans*-miR-eQTMs, the remainder of this paper focuses on *cis*-miR-eQTMs identified in the overall study sample at FDR<0.01.

Identification of *cis*-miR-eQTMs

We identified 227 CpGs (*cis*-miR-eQTMs) that were associated with expression levels of 40 miRNAs (299 *cis* miRNA-CpG pairs) at FDR<0.01 (corresponding $P < 1.43E-5$; Figure 2 and Supplementary Table 3) in

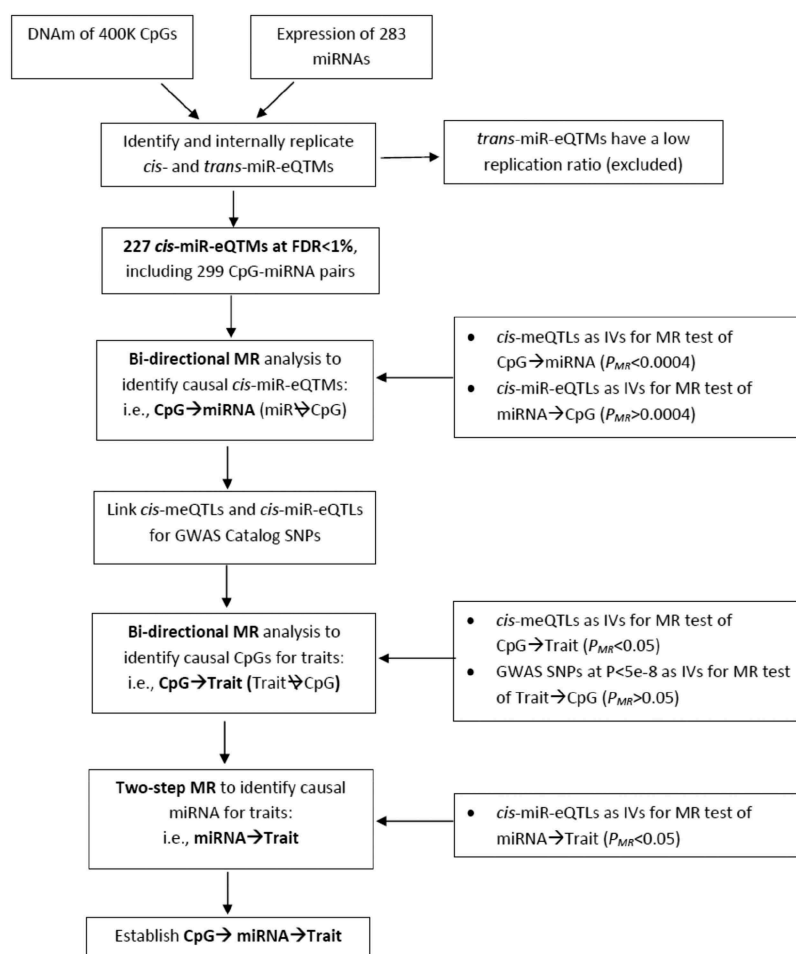


Figure 1. Analysis Flowchart.

Pair-wise associations between over 400K CpGs and 285 miRNAs were performed to identify *cis*- and *trans*-miR-eQTM. Internal replication suggested *cis*- but not *trans*-miR-eQTM are replicable. Therefore, the following analysis focused on *cis*-miR-eQTM. For each *cis*-miR-eQTM CpG-miRNA pair, bi-directional MR analysis was used to identify CpG→miRNA pairs (i.e. the changes of DNAm driving the changes of miRNAs but not vice versa). Then, we linked *cis*-meQTLs with human complex traits, bi-directional MR analysis was used to identify CpG→Trait. Two-step MR analysis was used to further identify miRNA→Trait to establish CpG→miRNA→Trait.

the overall study sample ($N = 3565$). A total of 117 (out of 227) of the CpG sites were retained after removing correlated CpGs at $r^2 > 0.5$ and 91 remained after removing those at $r^2 > 0.2$. The top 20 *cis*-miR-eQTM (selecting one peak CpG for each *cis* CpG-miR pair or clustered miRNAs at a single *cis* locus) are shown in Table 1. The most highly significant *cis*-miR-eQTM, cg06000878, was positively associated with both miR-100-5p (28Kb upstream) and miR-125b-5p (80Kb upstream) expression, with r^2 between the CpG and miRNAs 31% and 14% respectively.

Characterizing *cis*-miR-eQTM

To better understand *cis*-miR-eQTM, we further investigated positive vs. negative correlations

between CpGs and miRNAs, and if that direction of effects depends in part on their genomic locations in chromosomes. We found that 40% (121/299) of the *cis* CpG-miRNA pairs are positively correlated and 60% (178/299 pairs) are negatively correlated. The majority of *cis*-miR-eQTM (76%, 227/299 pairs) reside upstream (toward the 5' end) of the corresponding miRNAs. 14 *cis*-miR-eQTM CpGs are located within the promoter of the corresponding miRNAs (promoter regions were defined as 3Kb upstream and 500bp downstream of miRNA transcription start sites [TSS] [36,37]). Two *cis*-miR-eQTM CpGs are located within the precursor miRNA (pre-miRNAs), namely cg03295417 for miR-193b-3p, and cg16865908 for miR-125b-5p. Both are positively correlated. Enrichment analysis

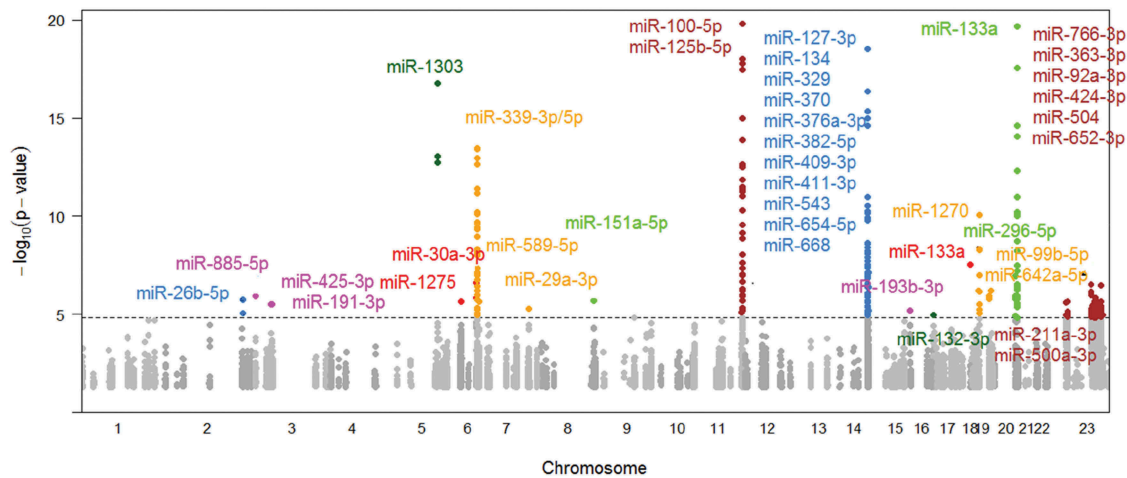


Figure 2. Manhattan plot of *cis*-miR-eQTMs.

The grey and dark grey colors denote *cis*-miR-eQTMs with $P < 0.05$, and the other colorful dots indicate *cis*-miR-eQTMs with $P < 1.43E-5$ (FDR<0.01). The dashed line indicates FDR = 0.01. 15 *cis*-miR-eQTMs for miR-125b-5p and miR-100-5p in Chromosome 11 and 3 *cis*-miR-eQTMs for miR-133a in Chromosome 20 were at $P < 1E-20$ and are not shown in this figure. miR-133a shows in both Chr18 and Chr20.

Table 1. Top 20 peak *cis*-miR-eQTMs identified in the 3656 FHS participants.

CpG	miRNA	Chromosome	Beta *	SD	T value	P value	FDR	Distance CpG to miRNA
cg06000878	miR-100-5p	11	-46.38	1.38	-33.72	1.91E-206	4.79E-201	28Kb Up
	miR-125b-5p		-19.63	0.82	-24.06	1.58E-118	1.98E-113	80Kb Up
cg11682508	miR-133a	20	-23.42	1.24	-18.90	6.94E-76	2.05E-71	12Kb Up
cg18089426	miR-370	14	4.81	0.53	9.03	2.90E-19	1.89E-15	201Kb Up
	miR-127-3p		3.82	0.45	8.45	4.39E-17	2.39E-13	173Kb Up
	miR-668		4.54	0.57	7.96	2.32E-15	1.15E-11	345Kb Up
	miR-543		3.26	0.48	6.83	1.03E-11	3.50E-08	322Kb Up
	miR-654-5p		3.65	0.55	6.58	5.57E-11	1.77E-07	330Kb Up
	miR-411-3p		2.86	0.43	6.58	5.80E-11	1.82E-07	313Kb Up
	miR-409-3p		2.71	0.42	6.54	7.15E-11	2.11E-07	355Kb Up
	miR-382-5p		3.47	0.58	5.97	2.64E-09	6.03E-06	344Kb Up
	miR-376a-3p		3.51	0.59	5.90	3.98E-09	8.95E-06	331Kb Up
cg17263206	miR-1303	5	5.96	0.70	8.57	1.78E-17	9.92E-14	6KB Down
cg22785556	miR-339-3p	7	1.66	0.22	7.62	3.28E-14	1.44E-10	18Kb Up
cg24642844	miR-339-5p	7	3.24	0.44	7.37	2.20E-13	8.89E-10	18Kb Up
cg13551269	miR-1270	19	4.35	0.67	6.51	8.58E-11	2.40E-07	601Kb Up
cg08509270	miR-134	14	11.57	2.15	5.39	7.74E-08	1.30E-04	7Kb Up
cg18125510	miR-329	14	-7.77	1.46	-5.33	1.07E-07	1.72E-04	651Kb Up
cg10546626	miR-296-5p	20	-8.51	1.60	-5.32	1.13E-07	1.77E-04	317Kb Up
cg11574469	miR-30a-3p	6	13.27	2.54	5.22	2.28E-07	3.47E-04	17Kb Up
cg23299641	miR-652-3p	X	2.27	0.44	5.14	2.97E-07	4.38E-04	430Kb Up
cg16807457	miR-363-3p	X	2.54	0.50	5.12	3.31E-07	4.84E-04	68Kb Up
	miR-92a-3p		0.91	0.20	4.43	9.50E-06	7.63E-03	68Kb Up
cg07060551	miR-99b-5p	19	-8.61	1.72	-4.99	6.26E-07	8.44E-04	997Kb Up
cg18735402	miR-642a-5p	19	-4.93	1.01	-4.88	1.11E-06	1.42E-03	3KB Down
cg11084334	miR-885-5p	3	7.06	1.45	4.87	1.16E-06	1.45E-03	842KB Down
cg07480829	miR-766-3p	X	1.42	0.30	4.82	1.48E-06	1.80E-03	91Kb Up
cg25925210	miR-26b-5p	2	-3.76	0.79	-4.79	1.75E-06	2.03E-03	309KB Down
cg25998745	miR-151a-5p	8	3.55	0.75	4.75	2.10E-06	2.40E-03	285Kb Up
cg09285672	miR-500a-3p	X	1.20	0.25	4.75	2.18E-06	2.45E-03	792Kb Up

* Higher Ct values indicate lower miRNA expression levels. Therefore, positive beta values indicate negative associations between miRNA expression and DNAm.

(Table 2) revealed that *cis*-miR-eQTM CpGs are enriched at promoter and polycomb-repressed state regions (marked by chromatin markers H3K27me3 at Bonferroni-corrected $P < 0.05/11$), suggesting

important regulatory roles of these CpGs on miRNA expression levels.

For the top result (see Table 1), cg06000878 resides in the gene body of non-coding RNA *LOC399959*,

Table 2. Genomic features of *cis*-miR-eQTMs.

Genomic Features	<i>cis</i> -miR-eQTMs	Total CpGs	Fold Change	<i>P</i> value
CpG Island	82	150,254	1.06	0.49
Low-CpG region	6	4954	2.34	0.03
High-CpG region	15	27,104	1.08	0.76
Enhancer	45	102,559	0.85	0.23
Enhancer (Encode gm12878)	12	23,988	1.99	0.01
Promoter	71	98,090	1.41	9.7E-4
Promoter (Encode gm12878)	44	71,180	2.46	1.49E-11
Heterochrom/lo (Encode gm12878)	30	79,312	1.51	0.01
Insulator (Encode gm12878)	4	4008	NA	NA
Repetitive/CNV (Encode gm12878)	1	674	NA	NA
Repressed (Encode gm12878)	22	24,355	3.60	3.89E-11

* The significant terms were defined as Fold change > 1 and Bonferroni corrected *P* < 0.05/11.

* CpG annotation were from Illumina annotation platform and Encode ChromHMM Track of gm12878 cell line.

and associated with both miR-100-5p and miR-125b-5p. There were 35 additional *cis*-miR-eQTM CpGs (27 CpG sites at $r^2 < 0.2$) for either miR-100-5p, miR-125b-5p, or both. Expression levels of miR-100-5p and miR-125b-5p were moderately correlated ($r^2 = 0.32$). Scatter plots and correlation heatmaps of *cis*-miR-eQTMs for both miRNAs are presented in Figure 3(a-d). Another example is on chromosome 14 where multiple CpGs were associated with expression of multiple miRNAs (Figure 3(e-h)). The top CpG (cg18089426) was associated with 9 miRNAs (miR-127-3p, miR-370, miR-668, miR-543, miR-654-5p, miR-411-3p, miR-409-3p, miR-382-5p, and miR-376a-3p) in this region including 49 *cis*-miR-eQTM CpG-miRNA pairs. This region previously identified as a genomic imprinted region, includes the imprinted genes *DLK1* and *RTL1* [38,39]. The observed associations between CpGs and miRNAs at this locus did not change after adjusting for the expression levels of *DLK1* and *RTL1*.

***cis*-miR-eQTMs for intragenic miRNAs conditioning on their host mRNAs**

Intragenic miRNAs, where miRNAs are located within a protein-coding gene, can be transcribed either along with their host mRNAs or independent of them. We further evaluated whether the association between DNAm and intragenic miRNA

expression was independent of the association between DNAm and expression of the host mRNA expression. The 40 miRNAs with *cis*-miR-eQTMs were located both within (intragenic) and outside (intergenic) coding gene bodies; 23 were intergenic miRNAs (97 *cis*-miR-eQTMs, 149 CpG-miRNA pairs) and 17 were intragenic miRNAs (136 *cis*-miR-eQTMs, 150 CpG-miRNA pairs). For the 17 intragenic miRNAs, we conditioned the association analyses between CpG methylation and miRNA expression on host mRNA expression levels. We found that each of the 150 intragenic *cis* miRNA-CpG pairs for intragenic miRNAs remained significantly associated after conditioning on host mRNA expression at Bonferroni corrected *P* < 3.3E-4 (0.05/150). Only two *cis*-miR-eQTMs were also associated with expression of the host mRNA at Bonferroni corrected *P* < 3.3E-4, including cg02355304 for miR-589-5p and *FBXL18*, cg01349480 for miR-504 and *FGF13*. The associations between cg02355304 and miR-589-5p, and between cg01349480 and miR-504 remained significant when conditioning on their host mRNAs (Supplementary Table 4). These results suggest that the relationship between DNAm and miRNA expression levels can be independent of their host mRNA expression levels.

Dissecting genetic and epigenetic effects on miRNAs

To evaluate the joint genetic and epigenetic effects on miRNAs with *cis*-miR-eQTMs and *cis*-miR-eQTLs, we tested the proportion of variation in expression of each miRNA that was attributed to: 1) a single *cis*-miR-eQTM; 2) a single *cis*-miR-eQTL; 3) multiple *cis*-miR-eQTMs for the miRNA; 4) multiple *cis*-miR-eQTLs for the miRNA; and 5) a combination of *cis*-miR-eQTLs and *cis*-miR-eQTMs for the miRNA (see **Methods**). Our results show that the partial R^2 between individual *cis*-miR-eQTM CpG and a miRNA is 0.02 ± 0.03 (mean \pm SD), and between individual *cis*-miR-eQTL SNP and a miRNA is 0.01 ± 0.02 . For the 27 miRNAs having both *cis*-miR-eQTMs and *cis*-miR-eQTLs, the partial R^2 between multiple *cis*-miR-eQTM CpGs and a miRNA is 0.05 ± 0.08 , whereas between multiple *cis*-miR-eQTL SNPs and a miRNA is 0.13 ± 0.09 . The partial R^2 of *cis*-

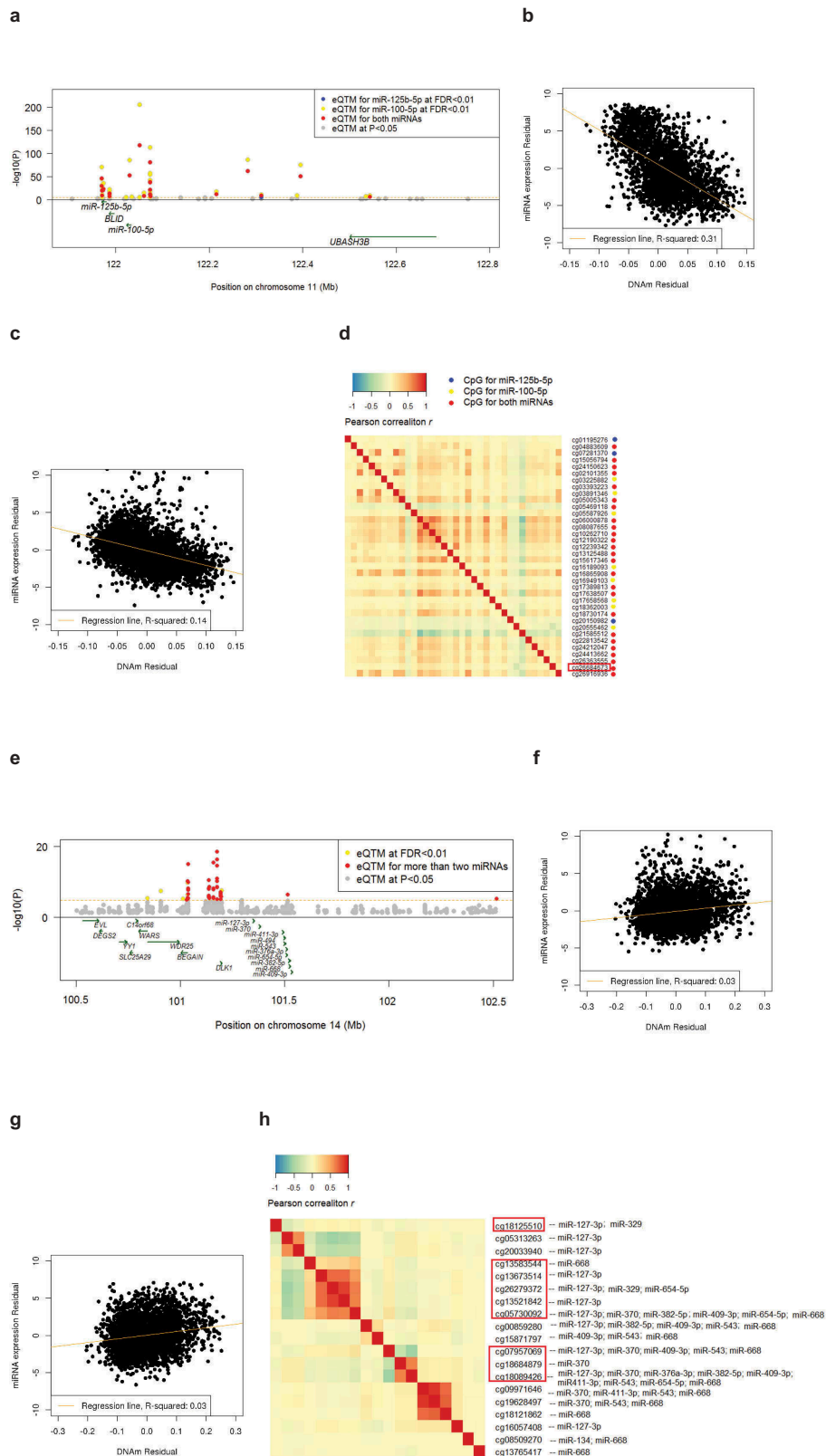


Figure 3. Examples of top *cis*-miR-eQTM for miRNAs.

(a) Regional association plot of *cis*-miR-eQTM and miRNAs (miR-100-5p, miR-125b-5p); (b) Scatter plot of DNAm values of the top CpG, cg06000878 with miR-100-5p; (c) Scatter plot of cg06000878 with miR-125b-5p; (d) Heatmap of Spearman correlation matrix of *cis*-miR-eQTM for miR-100-5p and/or miR-125b-5p; (e) Regional association plot of *cis*-miR-eQTM and miRNAs (miR-127-3p, miR-370, miR-668, miR-543, miR-654-5p, miR-411-3p, miR-409-3p, miR-382-5p, and miR-376a-3p); (f) Scatter plot of DNAm values of the top CpG, cg18089426 with miR-127-3p; (g) Scatter plot of DNAm values of the top CpG, cg18089426 with miR-127-3p; (h) Heatmap of Spearman correlation matrix of *cis*-miR-eQTM for miR-127-3p, miR-370, miR-668, miR-543, miR-654-5p, miR-411-3p, miR-409-3p, miR-382-5p, and miR-376a-3p. Highlighted red rectangle in (d) shows putatively causal CpGs for HDL cholesterol. Highlighted red rectangle in (e) shows putatively causal CpGs for age at menarche. The x-axis and y-axis in (b), (c), (g) and (h) are DNAm and miRNA expression residuals after adjusting for covariates (see Methods).

miR-eQTM CpGs for each miRNA was 0.02 ± 0.03 when conditioning on the *cis*-miR-eQTL SNPs (Supplementary Table 5). The top example is for miR-100-5p; the partial R^2 of the 33 *cis*-miR-eQTM CpGs and miR-100-5p is 0.14 when conditioning on the 339 *cis*-miR-eQTL SNPs.

***cis*-miR-eQTM effects on miRNA targets**

To investigate the effects of *cis*-miR-eQTMs on the expression of miRNA protein-coding gene targets (i.e., mRNA expression), we performed miRNA-mRNA co-expression analyses and used *in silico* prediction models to identify miRNA targets (see **Methods**). For the 40 miRNAs having *cis*-miR-eQTMs, we identified 1739 mRNAs that were co-expressed and were predicted to be miRNA targets, representing 2621 miRNA-mRNA pairs (Supplementary Table 6). We found that *cis*-miR-eQTM CpGs tended to be associated with the expression of miRNAs targeting mRNAs (by binomial test, $P < 2.2E-16$), suggesting that *cis*-miR-eQTMs may affect miRNA targets through miRNAs. Due to the complex interactions between miRNAs and their targets (e.g., pleiotropic effects of miRNAs on multiple targets, multiple miRNAs targeting the same mRNAs, and other genes/proteins driving expression changes of mRNAs), we were unable to estimate the exact proportion of CpG effect on miRNA targets mediated by miRNAs.

Bidirectional mendelian randomization testing to identify putatively causal CpGs for miRNAs

Due to the nature of the association analyses between CpGs and miRNAs, we were unable to determine whether DNAm drives miRNA expression changes or vice versa. Therefore, we performed bidirectional MR testing to further distinguish whether CpGs are causal for miRNA or vice versa (Figure 1). In our previous studies, we identified genetic variants associated with miRNA expression (miR-eQTLs [33]) and with methylation of CpGs (meQTLs [40]). Among the 299 CpG-miRNA pairs, there were 116 pairs for which the CpGs and miRNAs have both *cis*-meQTLs ($n \geq 3$ variants at LD $r^2 < 0.1$) and *cis*-

miR-eQTLs ($n \geq 3$ variants at LD $r^2 < 0.1$) and were eligible for robust MR testing. The other pairs were not suitable for MR analysis. Bidirectional MR results are presented in Supplementary Table 7. We identified 89 pairs where the CpG was putatively causal for miRNA expression (CpG→miRNA, at $P_{MR} < 0.05/116$), and 34 pairs where the miRNA expression was putatively causal for CpG (miRNA→CpG, at $P_{MR} < 0.05/116$). Thirty-one pairs overlapped, i.e., both CpG→miRNA and miRNA→CpG MR tests were significant. We excluded these 31 bidirectionally causal pairs and focused on the 58 pairs for which the CpG tested causal for the miRNA but the miRNA did not test causal for the CpG; this represented 36 CpGs and 17 miRNAs. There were only three pairs for which the miRNA tested causal for the CpG but not vice versa, i.e., miR-1270→cg03570035/cg07227744/cg03317517. Bidirectional MR results showed that the majority of CpG-miRNA pairs are CpG causal for miRNA rather than miRNA causal for CpG. This result, coupled with the results reported above that *cis*-miR-eQTM CpGs are enriched for promoter-associated and polycomb-repressed state regions, suggests that changes in DNAm of *cis*-miR-eQTM CpGs may affect miRNA expression rather than vice versa.

***cis*-miR-eQTMs helping to identify novel pathways underlying human complex traits**

To further demonstrate the utility of our *cis*-miR-eQTMs resource, we linked *cis*-meQTLs and *cis*-miR-eQTLs with GWAS Catalog SNPs for human complex diseases and traits. We further performed MR to identify putatively causal CpGs for human complex traits where the causal pathways may or may not act through miRNAs, i.e., CpG→trait, or CpG→miRNA→trait. The analysis flowchart is shown in Figure 1. Among the 58 CpG→miRNA pairs, 52 pairs having both *cis*-meQTLs and *cis*-miR-eQTLs perfectly matched GWAS SNPs for eight complex traits in the GWAS Catalog, including acute graft versus host disease in bone marrow transplantation, atrial fibrillation, age at menarche, C-reactive protein levels, facial morphology, HDL cholesterol, magnesium levels, and total cholesterol (TC). This overlap of SNP associations suggests that CpGs and miRNAs

may be involved in the gene regulatory pathways for the traits or that changes in CpGs and miRNAs reflect trait variability. Full GWAS summary results for five traits that are publicly available were suitable for robust MR testing. Bidirectional MR tests were performed on CpG-trait pairs to identify putatively causal CpGs for traits (i.e., CpG→trait, at $P_{MR} < 0.05$) for which reverse causality was unlikely (i.e., trait→CpG at $P_{MR} > 0.05$, Figure 1). In doing so, we identified cg25998745 as putatively causal for atrial fibrillation, cg26684673 for HDL cholesterol, 16 CpGs (3 CpG sites at $r^2 < 0.2$, e.g., cg24642844) for total cholesterol, and 9 CpGs (3 CpG sites at $r^2 < 0.2$, including cg18089426) for age at menarche (Supplementary Table 6). Two-step MR testing was performed to identify miRNAs in the causal pathways that were driven by CpG methylation, e.g., cg26684673→miR-125b-5p→HDL cholesterol (Supplementary Table 6). A notable example is the 9 CpGs and 3 miRNAs that tested causal for age at menarche (also see in Figure 3 (d-f)). By utilizing two GWAS resources for age at menarche from independent sample sets including 182,416 individuals from 57 studies [41] and 272,937 individuals from UK Biobank (www.nealelab.is/uk-biobank/) we cross-validated the MR results (Table 3). Many *cis*-miR-eQTM were found in this genomic imprinted region that were associated with age at menarche. This finding is consistent with the hypothesis that epigenetic regulation involving both DNAm and miRNAs is critical in genomic imprinting and results in wide variability of age at menarche. Further experiments and investigation are necessary to prove this hypothesis. The directions of associations in causal testing suggest that hypermethylation of

cg18089426 promotes decreased expression of miR-411-3p, miR-382-5p, and miR-127-3p and contributes to older age at menarche.

Discussion

In this investigation, we report results of the first systematic epigenome-wide association study of miRNA expression and identified 299 correlated *cis* CpG-miRNA pairs. Bidirectional MR analysis identified 58 pairs for which the DNAm appears to drive changes in miRNA expression. Joint analyses of genetic variants, DNAm, and miRNA expression demonstrated a partial R^2 between DNAm and miRNA expression of 2% after conditioning on genetic variants. Our study constitutes an important step toward a more comprehensive understanding of the regulation of miRNA expression by genetic and epigenetic mechanisms. In addition, we identified examples of DNAm and miRNA expression that are putatively causally in relation to human complex traits. For instance, we found that hypermethylation of cg18089426 and expression of miR-411-3p, miR-382-5p, and miR-127-3p contribute to variation in age at menarche. Moreover, this CpG and these miRNAs are located at a region of the genome that has previously been reported to contribute to imprinting [38,39].

DNAm influences pre-transcriptional regulation of protein-coding genes. Several studies have revealed differential methylation of thousands of CpGs in association with protein-coding mRNAs, i.e., *cis*-mRNA-eQTMs [34,35]. To our knowledge, our study is the first to provide genome-wide

Table 3. Mendelian Randomization examining causal CpGs and miRNAs for age at menarche.

CpG	MR: CpG→Trait (Perry JR, et al. Nature 2014)			MR: CpG→Trait (Nealelab, UKBiobank)			Reverse MR: Trait→CpG (Perry JR, et al. Nature 2014)			miRNA	MR: miRNA→Trait (Perry JR, et al. Nature 2014)		
	meQTLs	Beta	Pval	meQTLs	Beta	Pval	GWAS SNPs	Beta	Pval		miR-eQTLs	Beta	Pval
cg05730092	10	-0.94	5.34E-06	15	-0.23	6.82E-04	86	-0.03	0.09	miR-382-5p	13	-0.02	1.15E-03
										miR-127-3p	15	-0.02	0.05
cg13521842	7	-0.86	5.22E-04	15	-0.19	6.11E-03	86	-0.02	0.15	miR-127-3p	15	-0.02	0.05
cg13583544	7	-2.77	1.58E-04	11	-0.76	7.25E-04	86	-0.01	0.08				
cg13673514	6	-0.87	9.22E-04	10	-0.19	0.03	86	-0.02	0.17	miR-127-3p	15	-0.02	0.05
cg26279372	8	-0.94	1.16E-04	16	-0.21	2.94E-03	86	-0.03	0.10	miR-127-3p	15	-0.02	0.05
cg18089426	8	0.25	0.02	18	0.07	0.02	86	0.01	0.61	miR-411-3p	7	-0.05	6.09E-07
										miR-382-5p	13	-0.02	0.00
										miR-127-3p	15	-0.02	0.05
cg07957069	9	-0.58	6.41E-04	16	-0.13	0.02	86	-0.01	0.27	miR-127-3p	15	-0.02	0.05
cg18684879	8	1.06	6.70E-03	14	0.20	0.06	86	2.32E-03	0.44				
cg18125510	6	-1.80	1.25E-04	10	-0.23	0.07	86	-0.01	0.16				

* cg05730092, cg13521842, cg13583544, cg13673514, and cg26279372 were at $r^2 > 0.2$; cg18089426, cg07957069 and cg18684879 were at $r^2 > 0.2$.

mapping of DNAm and miRNA expression. We found that *cis*-miR-eQTM are enriched in promoter and polycomb-repressed regions. Higher methylation at 60% of *cis*-miR-eQTMs is associated with lower miRNA expression. These results suggest that methylation of CpGs residing adjacent to miRNAs may silence the transcription of those miRNAs, similar to observed relations of differential methylation of CpGs to expression of nearby protein-coding genes [42]. We identified 2 *cis*-miR-eQTMs located within pre-miRNAs or mature miRNAs that are positively associated with miRNA expression. This suggests that methylation in miRNA body regions may foster the transcription of miRNAs, similar to what has been reported for protein-coding genes [43].

We found genetic variants that are associated with both *cis*-DNAm (*cis*-meQTLs) and *cis*-miRNA expression (*cis*-miR-eQTLs). We hypothesize that given the close physical proximity of interrelated SNPs, CpGs, and miRNAs, the observed associations may be driven by changes in TF binding or by direct influences on the three-dimensional conformational structure of chromatin, perhaps by impairing CTCF binding sites. We further demonstrated that inter-individual variation in miRNA expression is influenced by *cis* genetic and epigenetic contributions. After conditioning on shared SNPs (miR-eQTLs and meQTLs), our results show that the associations of CpGs and miRNAs remain significant and often display substantial effect sizes. The strongest association of *cis*-miR-eQTMs (CpG with nearby miRNA) were for miR-100-5p and miR-125b-5p. These two miRNAs are located close to each other on chromosome 11 and are moderately correlated with each other ($r^2 = 0.3$). The partial R^2 of the 33 *cis*-miR-eQTM CpGs for miR-100-5p expression was 14%, and the partial R^2 of the 28 *cis*-miR-eQTMs for miR-125b-5p expression is 6%, both after conditioning on *cis*-miR-eQTL SNPs. We identified 25 shared *cis*-miR-eQTMs for both miRNAs, with 14 of these CpGs lying within known enhancers (Supplementary Table 3). Additional epigenetic regulatory effects may exist from elements that we did not assay. The full list of *cis*-miR-eQTMs includes miRNAs that warrant further functional exploration.

Among the 40 miRNAs harboring *cis*-miR-eQTM, there are 17 intragenic miRNAs. Intragenic miRNAs lying within protein-coding

genes can be transcribed together with their host mRNAs or independently. The conditional analyses revealed that 98% of DNAm-miRNA expression pairs remain significantly associated after adjusting for expression of the host mRNA, thus suggesting that the identified epigenetic relations of DNA methylation to miRNA expression is independent of the host mRNA expression. We identified methylation at two CpGs to be associated with both intragenic miRNA and host mRNA expression; specifically, cg02355304 for miR-589-5p and *FBXL18*, and cg01349480 for miR-504 and *FGF13*. These results support the hypothesis that miRNAs can be transcribed together with their host mRNAs or independently.

To further demonstrate the utility of our *cis*-miR-eQTM resource, we crosslinked *cis*-miR-eQTM with GWAS Catalog traits by linking *cis*-meQTLs and *cis*-miR-eQTLs with GWAS Catalog SNPs [44]. We were able to link 90% (52/58) of causal CpG→miRNA pairs to complex traits via perfect overlap with *cis*-meQTLs and *cis*-miR-eQTLs. These findings suggest that DNAm and miRNAs may be important contributors to human diseases and traits. For the five traits for which full GWAS results were publicly available, we performed MR analyses to further assess causal relations between CpGs and complex traits. A similar study by Richardson et al. used *cis*-meQTLs as instrumental variables to identify causal CpGs for cardiovascular disease traits by MR [45]. However, they focused on a different set of CpGs than the *cis*-miR-eQTMs in this study. By linking *cis*-miR-eQTM miRNAs with putatively causal CpGs, we provide a more complete picture of CpG-related causal pathways involving miRNAs. Our most biologically compelling example is that of 49 *cis* CpG-miRNA pairs for 19 CpGs and 9 miRNAs residing in an imprinted region on 14q32. Prior studies identified paternally inherited SNPs in association with human complex traits [46], e.g., rs941576 in this region in relation to type I diabetes [47]. Imprinted genes were defined as genes that exhibit activity differences in the offspring depending on which parent passes on that gene. Imprinted genes were also identified in relation with human complex traits. For example, *DLK1* was reported to be associated with age at menarche [41] and with many cancers [48–50].

About ten imprinted genes in the orthologous mouse region (12qF1), including *DLK1*, were reported [38]. miRNAs aligning to this region in mice were reported to exhibit a pattern of maternal inheritance [51]. Genomic imprinting was thought to be involved in epigenetic regulation, however, the precise molecular mechanisms underlying genomic imprinting are largely unknown [51]. We identified many *cis*-miR-eQTMs in the 14q32 imprinted region. Our findings lead to a new hypothesis that the CpG-miRNAs pairs in this region may be involved in the observed imprinting. Many of these interlinked CpGs and miRNAs at this locus are associated with age at menarche and MR testing suggests causal relations. Age at menarche varies considerably between individuals and this variation reflects both genetic and environmental influences. Perry et al. reported on the association of SNPs and expression levels of *DLK1* in 14q32 with age at menarche [41]. Their research provides an example of imprinted genes that may influence later life health outcomes. Subsequent research further established links between age at menarche with other diseases such as cancer [52]. Our findings in concert with previous studies reveal a wider and more complex network involving both DNAm and miRNA, and suggest that DNAm and miRNAs are key epigenetic determinants of age at menarche and may contribute to related disease risk. We acknowledge that a limitation of this study is that array-based data is incapable of measuring DNAm levels in an allele-specific manner, therefore, we were unable to confirm if specific CpGs show allele-specific DNAm inherited from the father and/or mother. Further experiments are warranted to test this hypothesis.

We acknowledge several other limitations of our study. First, to our knowledge, no other study has published extensive association analyses of miRNA expression and DNAm, and therefore we were unable to perform independent external replication of our results. The internal validation strategy suggested that *cis*-miR-eQTMs are replicable but *trans*-miR-eQTMs are not replicable. *trans*-miR-eQTMs may require much larger sample sizes as well as robust independent external replication. Second, we assessed DNAm and miRNA expression in whole blood, which consists of multiple cell types. We were unable to

identify cell type-specific miR-eQTMs using our data. We were also unable to describe the tissue-specific relations between DNA methylation and miRNA expression that may vastly differ from what we observed in whole blood. Third, even though the Illumina DNAm platform we used covers about 450,000 CpG sites, it does not assay every CpG site in the genome. In addition, we did not measure every possible miRNA, and therefore may have missed important regulatory miR-eQTMs. The diversity and complexity of the multifaceted relationship between miRNA and methylation warrants additional studies.

Experimental procedures

Study participants

The FHS Offspring cohort was recruited in 1971 and consisted of 5124 children (and their spouses) of the FHS Generation 1 cohort [53]. Between 2002 to 2005, 4095 children of the Offspring cohort participants were recruited, comprising the Third Generation cohort [54]. Eligible participants from the Offspring cohort who attended their eighth examination cycle (Exam 8, 2005–2008) and Third generation who attended their second examination cycle (Exam 2, 2008–2011) were included in this study. For internal validation, the entire FHS study sample was split 1:1 by pedigrees into independent discovery and replication sets.

miRNA expression assays (N = 5221) were conducted on whole blood samples from 2295 FHS Offspring (Exam 8) and 2926 Third Generation (Exam 2) participants. DNAm assays (N = 3943) were conducted on whole blood samples from 2427 Offspring cohort participants (Exam 8) and 1516 Third Generation (Exam 2) participants. miRNA expression and DNAm data were both available for 3565 participants. This study was approved under the Boston University Medical Center protocol H-27,984. All participants provided written consent for genetic research.

DNAm profiling and data normalization

Buffy coats were prepared from whole blood samples. Genomic DNA was extracted using the Genra Puregene DNA extraction kit (Qiagen, Venlo, Netherlands), and bisulfite conversion was performed using the EZ DNAm kit (Zymo Research, Irvine, CA). Samples underwent DNA amplification,

fragmentation, array hybridization, and single-base pair extension. DNA methylation levels were measured using the Illumina Infinium Human Methylation450 BeadChip (450K). DNAm arrays of FHS Offspring cohort participants were run in two laboratory batches at the Johns Hopkins Center for Inherited Disease Research (lab batch #1) and University of Minnesota Biomedical Genomics Center (lab batch #2). DNAm arrays were run on 1516 FHS Third Generation cohort participants (lab batch #3) by Illumina, Inc (City, State).

DNAm results were normalized by laboratory batch using the DASEN methodology implemented in the watermelon package [55] in R (version 3.0.2). For sample quality control, samples with a missing rate >1% at $P < 0.01$, poor single nucleotide polymorphism (SNP) matching with the 65 SNP control probe locations, and outliers from multi-dimensional scaling techniques were excluded. For quality control at the probe level, we excluded probes that had a missing rate >20% at $P < 0.01$, were previously identified to map to multiple locations [56], or had an underlying SNP (minor allele frequency >5% in European ancestry (EUR) 1000 genomes project data) at the CpG site or within 10 bp of the single base extension. A total of 415,318 CpG probes were retained for further analyses.

We used surrogate variable analyses (SVA) to eliminate unwanted variation in the DNAm data [57]. SVs were generated in each lab batch separately, totaling 71 (Lab 1), 170 (Lab 2), and 173 (Lab 3) SVs from the DNAm data in the three laboratory batches, respectively. DNAm beta values were regressed on batch-specific SVs, and the DNAm residual was taken forward. All three lab batches were merged together for analysis. In our previous studies using the same data, the SVA method was demonstrated to adequately account for latent batch effects, variations in blood cell proportions, and technical covariates [1,6]. Therefore, to avoid over adjustment, we did not further adjust for technical covariates (i.e. plates, rows and columns) or cell types in DNAm data.

miRNA expression profiling and data normalization

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 -80°C . 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 to assess RNA quality (i.e., RNA integrity number [RIN], ranging from 1 to 10). A RIN threshold of 4 was considered to reflect adequate quality. This threshold was determined based on comparison between the samples with $\text{RIN} \geq 9$ and $\text{RIN} \geq 4$ by principal component tests as described by previously Joehanes et al [58]. 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 highly specific and sensitive [59–61]. 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. Among the 70 replicate samples, more than 95% of the data points had coefficients of variation that were less than 10% (mean ~4%).

The initial miRNA list encompassed all TaqMan miRNA assays available at the start of the study, including 754 miRNAs that were profiled in 455 FHS individuals. 346 miRNAs expressed in > 20% samples were further profiled in 2445 FHS Offspring and 3245 Third Generation cohort participants. Quantification of miRNA expression was based on cycle threshold (Ct), where lower Ct values signified higher miRNA expression levels. Ct values ≥ 27 indicated that the respective miRNAs were not expressed in the sample. Outlier miRNAs with Ct values ≥ 5 standard deviations from the mean Ct value were categorized as missing. We excluded miRNAs expressed in <100 samples. A total of 283 miRNAs remained for analysis.

As previously described, raw miRNA Ct values were adjusted for four 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 260nm and 280nm from spectrophotometry) [33]. This normalization model explained 20% to 60% of the variability in raw miRNA measurements for 80% of miRNAs.

mRNA expression profiling and data normalization

Whole blood samples (2.5ml) were collected in PAXgene™ tubes by Asuragen, Inc. (PreAnalytiX, Hombrechtikon, Switzerland). Total RNA was isolated in a single batch on a KingFisher® 96 robot according to the company's standard operating procedures for automated isolation of RNA from 96 samples. Then, 50ng RNA samples were amplified using the WT-Ovation Pico RNA Amplification System (NuGEN, San Carlos, CA) in an automated manner using the genechip array station (GCAS) as recommended by the manufacturer. mRNA expression profiling was obtained from the Affymetrix Human Exon 1.0 ST GeneChip platform, and the raw gene expression data was first preprocessed by quartile normalization. The robust multi-array average (RMA) values of every gene (17,318 measured genes) were adjusted for a set of technical covariates, i.e. chip batch, by fitting linear mixed regression (LME) models as described previously [58].

mRNA expression data (N = 5626) were available for FHS Offspring (Exam 8, N = 2446) and Third Generation (Exam 2, N = 3180) participants. mRNA and miRNA expression were both available for 5357 participants.

Identification of differential DNAm in association with miRNA expression (miR-eQTLs)

DNAm and miRNA were separately regressed on technical and batch covariates or SVs, as described above, and residuals taken forward for analyses (see **miRNA expression and data normalization**, and **DNAm profiling and data normalization**). Linear mixed models (R package 'lme4', <https://cran.r-project.org/web/packages/kinship/>, [62]) were used to conduct pairwise DNAm – miRNA expression association analyses for residuals of all profiled CpGs (N = 415,318) and 283 miRNAs expressed in at least 100 samples. All analyses were adjusted for age and sex

as fixed effects, and family structure as a random effect. In previous studies, we found that adjusting for blood cell proportions did not materially change the results for miRNAs in relation to genotypes [33] or across multiple phenotypes [14]. We investigated cell types effects using 932 individuals with quantification of cell type proportions. The Pearson correlation coefficients of the beta values of *cis*- miRNA-CpG associations with versus without adjustment for blood cell types was $r > 0.99$. We also compared beta values of *cis*- miRNA-CpGs associations with versus without adjustment for cell type compositions imputed from DNAm data via the Houseman methods in all 3565 samples. The Pearson correlation coefficients of beta values was $r > 0.99$. Therefore, we did not adjust for cell types in the miRNA data normalization model. CpG-miRNA pairs in which the CpG resided within 1Mb of the mature miRNA were categorized as *cis*-miR-eQTLs, and those residing more than 1Mb away were considered *trans*-miR-eQTLs.

Genome coordinate annotation for miRNAs and DNAm were conducted using miRbase v20 (mirbase.org) and Illumina annotation file v1.2, respectively. miRbase v20 and Illumina annotation file v1.2 were based on the human February 2009 (GRCh37/hg19) assembly. Based on the coordinates of 283 mature miRNAs and ~400K CpGs, we estimated that there were 250,622 potential *cis* CpG-miRNA pairs, and 1.1×10^8 (i.e., 400,000 * 283--250,622) potential *trans* CpG-miRNA pairs. We used the Benjamini-Hochberg method [63] to calculate the false-discovery rate (FDR) for *cis*- and *trans*-miR-eQTLs, taking into account correlated *cis* and *trans* pairs. We selected an FDR threshold of 0.01, corresponding to $P < 1.43E-5$ for *cis*- and $P < 1.87E-6$ for *trans*-miR-eQTLs.

Cis-miR-eQTLs and cis-meQTLs

cis-miR-eQTLs (SNP located within 1Mb on either side of the corresponding mature miRNA) were generated using the same miRNA data in the FHS cohorts as described previously. There were 5269 *cis*-miR-eQTLs for 76 mature miRNAs [33] in total.

cis-meQTLs (SNP located within 1Mb on either side of the CpGs) were generated using the same DNAm data in the FHS cohorts [40]. Genotyping was conducted using the Affymetrix 500K mapping and Affymetrix 50K gene-focused MIP arrays

according to previously described quality control procedures [64]. Genotypes were imputed using the 1000 Genomes Project panel via MACH [65] with $MAF \geq 0.01$ and imputation quality ratio ≥ 0.1 . Linear mixed models were applied to calculate the associations of SNPs (predictors) and DNAm residuals (outcomes, see **DNAm profiling and data normalization**). Analyses were adjusted for age and sex as fixed effects, and a familial correlation matrix (FAM) as the random effect using the *lmekin()* function of Kinship Package (<https://cran.r-project.org/web/packages/kinship/>) [62]. Bonferroni-corrected $P < 2E-11$ was used to define significant *cis*-meQTLs.

miRNA – mRNA co-expression analysis

miRNA-mRNA co-expression analyses were performed on 5626 FHS participants with available miRNA and mRNA data. A linear mixed-effects model implemented in the *lmekin()* R function was used to conduct pair-wise association analyses between miRNAs and mRNAs. Analyses were adjusted for age, sex, technical covariates for miRNA and mRNA data, and SVs for mRNA data. The P value threshold for choosing miRNA-mRNA co-expressed pairs was determined using a Benjamini-Hochberg corrected $FDR < 0.05$ [63].

Predicted miRNA targets

For each significant miRNA-mRNA co-expressed pair, we used two miRNA target prediction methods TargetScan 7.0 and miRanda to predict if the co-expressed mRNAs were also miRNA targets. TargetScan predicts miRNA targets of miRNAs by searching for the presence of 8mer, 7mer, and 6mer sites that match the seed region of each miRNA [66]. miRanda three-step predicts miRNA targets with a threshold of score > 140 [67]. The sequences from 3'UTR, 5'UTR, and coding regions of each mRNA were downloaded from the University of California Santa Cruz (UCSC) Table Browser (<https://genome.ucsc.edu/>), and miRNA seed regions were downloaded from miRbase v21 (<http://www.mirbase.org/>).

Mendelian randomization

Bi-directional Mendelian randomization (MR) was used to test if changes in CpG methylation drive changes in miRNA expression (CpG→miRNA) or if changes in miRNA expression drive changes in CpG

methylation (miRNA→CpG). MR tests were performed in 116 CpG-miRNA pairs where both the CpG and miRNA have at least three independent instrument variables (IVs). *cis*-meQTLs (at LD $r^2 < 0.1$) were used as IVs to test if CpG→miRNA, and *cis*-miR-eQTLs (at LD $r^2 < 0.1$) were used as IVs to test if miRNA→CpG. Inverse variance-weighted (IVW) MR methods were used for all MR tests [68]. The significant threshold was a Bonferroni-corrected $P_{MR} < 0.05/116$.

Two-sample IVW-MR methods were used to identify if CpG methylation and miRNA expression were putatively causal for human complex traits, including atrial fibrillation, age at menarche, HDL cholesterol, magnesium levels, and total cholesterol. SNP-trait associations were taken from published meta-analysis GWAS results of age at menarche [41], total cholesterol and HDL cholesterol [69], and from UK Biobank GWAS results of age at menarche, magnesium levels and atrial fibrillation downloaded from (<http://www.nealelab.is/uk-biobank/>). To identify CpG→Trait, independent *cis*-meQTLs at LD $r^2 < 0.1$ were used as IVs. To identify miRNA→Trait, independent *cis*-miR-eQTLs at LD $r^2 < 0.1$ were used as IVs. We also perform MR to test if trait→CpG by utilizing trait-associated SNPs at $P < 5E-8$ in GWAS as IVs (independent SNPs were selected by LD $r^2 < 0.1$). A random effects model of IVW-MR method was used to minimize the heterogeneity effects of IVs [70]. MR analyses were conducted using the mrBase R package [70].

Functional annotations of cis-miR-eQTLs

We annotated *cis*-miR-eQTL CpGs using Illumina platform annotations and GM12878 Encode chromatin state segmentation by hidden Markov models (ChromHMM) [71]. The Illumina platform annotation provided the predicted functions for each CpG site (CpG Islands, low CpG dense, high CpG dense), enhancer and promoter regions. The GM12878 ChromHMM Track provided chromatin states identified by computationally integrating ChIP-seq data (H3K4me1, H3K4me2, H3K4me3, H3K27ac, H3K9ac, H3K27me3, H4K20me1, and H3K36me3) from GM12878, a lymphoblastoid cell line, including promotor, enhancer, polycomb-repressed, heterochromatin, and repetitive/copy number variation. Hypergeometric tests were used

to test if *cis*-miR-eQTM CpGs sites were enriched at functional genomic sites that were annotated as functional. The significance threshold was Bonferroni-corrected $P < 0.05/11$.

Data access

The microRNA expression, mRNA expression and DNAm data used in this investigation are available online in dbGaP (<http://www.ncbi.nlm.nih.gov/gap>; accession number phs000007).

Acknowledgments

The Framingham Heart Study is funded by National Institutes of Health contract N01-HC-25195 and HHSN268201500001I. 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 and R56AG029451 (M.G.L.).

The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or the U.S. Department of Health and Human Services.

Authors' contributions: D.L. designed, directed, and supervised the project. T.H., M.M., and D.L. drafted the manuscript. T.H., C.L., J.R., C.Y., C.S., J.R., and A.B. conducted the analyses. J.E.F. and K.T. conducted the miRNA expression experiments. All authors participated in revising and editing the manuscripts. All authors have read and approved the final version of the manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the National Heart, Lung, and Blood Institute [N01-HC-25195, HHSN268201500001I, and R56AG029451].

ORCID

Tianxiao Huan  <http://orcid.org/0000-0001-5840-9413>
 Michael Mendelson  <http://orcid.org/0000-0001-7590-3958>
 Roby Joehanes  <http://orcid.org/0000-0001-5549-9054>
 Chen Yao  <http://orcid.org/0000-0002-3944-7788>
 Paul Courchesne  <http://orcid.org/0000-0001-5691-6913>

References

- [1] Mendelson MM, Marioni RE, Joehanes R, et al. Association of body mass index with DNA methylation and gene expression in blood cells and relations to cardiometabolic disease: a mendelian randomization approach. *PLoS Med.* 2017;14:e1002215.
- [2] Hedman ÅK, Mendelson MM, Marioni RE, et al. Epigenetic patterns in blood associated with lipid traits predict incident coronary heart disease events and are enriched for results from genome-wide association studies clinical perspective. *Circ Cardiovasc Genet.* 2017;10:e001487.
- [3] Irvin MR, Zhi D, Joehanes R, et al. Epigenome-wide association study of fasting blood lipids in the genetics of lipid lowering drugs and diet network study. *Circulation.* 2014;114:009158.
- [4] Volkmar M, Dedeurwaerder S, Cunha DA, et al. DNA methylation profiling identifies epigenetic dysregulation in pancreatic islets from type 2 diabetic patients. *Embo J.* 2012;31:1405–1426.
- [5] Ligthart S, Marzi C, Aslibekyan S, et al. DNA methylation signatures of chronic low-grade inflammation are associated with complex diseases. *Genome Biol.* 2016;17:255.
- [6] Liu C, Marioni R, ÅK H, et al. A DNA methylation biomarker of alcohol consumption. *Mol Psychiatry.* 2016;23:422.
- [7] Hannum G, Guinney J, Zhao L, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell.* 2013;49:359–367.
- [8] Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol.* 2013;14:3156.
- [9] Joehanes R, Just AC, Marioni RE, et al. Epigenetic signatures of cigarette smoking. *Circ Cardiovasc Genet.* 2016;9:436–447.
- [10] Cordes KR, Srivastava D. MicroRNA regulation of cardiovascular development. *Circ Res.* 2009;104:724–732.
- [11] Small EM, Olson EN. Pervasive roles of microRNAs in cardiovascular biology. *Nature.* 2011;469:336–342.
- [12] Thum T, Galuppo P, Wolf C, et al. MicroRNAs in the human heart. *Circulation.* 2007;116:258–267.
- [13] Huan T, Rong J, Tanriverdi K, et al. Dissecting the roles of microRNAs in coronary heart disease via integrative genomic analyses. *Arterioscler Thromb Vasc Biol.* 2015;ATVBAHA:114.305176.
- [14] McManus DD, Rong J, Huan T, et al. Messenger RNA and MicroRNA transcriptomic signatures of cardiometabolic risk factors. *BMC Genomics.* 2017;18:139.
- [15] Jansson MD, Lund AH. MicroRNA and cancer. *Mol Oncol.* 2012;6:590–610.
- [16] Esteller M. Epigenetics in cancer. *N Engl J Med.* 2008;358:1148–1159.
- [17] Loginov V, Rykov S, Fridman M, et al. Methylation of miRNA genes and oncogenesis. *Biochemistry (Moscow).* 2015;80:145–162.
- [18] Chuang JC, Jones PA. Epigenetics and microRNAs. *Pediatr Res.* 2007;61:24R–9R.

- [19] Ehrlich M, Gama-Sosa MA, Huang L-H, et al. Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells. *Nucleic Acids Res.* 1982;10:2709–2721.
- [20] Wu H, Zhang Y. Reversing DNA methylation: mechanisms, genomics, and biological functions. *Cell.* 2014;156:45–68.
- [21] Lister R, Pelizzola M, Dowen RH, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature.* 2009;462:315–322.
- [22] Lee R, Feinbaum R, Ambros V. A short history of a short RNA. *Cell.* 2004;116:S89–S92.
- [23] Lee RC, Ambros V. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science.* 2001;294:862–864.
- [24] Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* 2014;42:D68–D73.
- [25] Friedman RC, Farh KK-H, Burge CB, et al. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* 2009;19:92–105.
- [26] Lemire M, Zaidi SH, Ban M, et al. Long-range epigenetic regulation is conferred by genetic variation located at thousands of independent loci. *Nat Commun.* 2015;6:6326.
- [27] Hannon E, Spiers H, Viana J, et al. Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci. *Nat Neurosci.* 2016;19:48–54.
- [28] Bonder MJ, Luijk R, Zhernakova DV, et al. Disease variants alter transcription factor levels and methylation of their binding sites. *Nat Genet.* 2017;49:131.
- [29] Shi J, Marconett C, Duan J, et al. Characterizing the genetic basis of methylome diversity in histologically normal human lung tissue. *Nature communications.* 2014;6:3365.
- [30] Gaunt TR, Shihab HA, Hemani G, et al. Systematic identification of genetic influences on methylation across the human life course. *Genome Biol.* 2016;17:61.
- [31] McClay JL, Shabalina AA, Dozmorov MG, et al. High density methylation QTL analysis in human blood via next-generation sequencing of the methylated genomic DNA fraction. *Genome Biol.* 2015;16:291.
- [32] McRae A, Marioni RE, Shah S, et al. Identification of 55,000 replicated DNA methylation QTL. *Scientific Reports.* 2018;8:17605.
- [33] Huan T, Rong J, Liu C, et al. Genome-wide identification of microRNA expression quantitative trait loci. *Nat Commun.* 2015;6:6601.
- [34] Wagner JR, Busche S, Ge B, et al. The relationship between DNA methylation, genetic and expression inter-individual variation in untransformed human fibroblasts. *Genome Biol.* 2014;15:R37.
- [35] Liu Y, Ding J, Reynolds LM, et al. Methylomics of gene expression in human monocytes. *Hum Mol Genet.* 2013;22:5065–5074.
- [36] Marsico A, Huska MR, Lasserre J, et al. PROMiRNA: a new miRNA promoter recognition method uncovers the complex regulation of intronic miRNAs. *Genome Biol.* 2013;14:R84.
- [37] Chen D, Fu L-Y, Zhang Z, et al. Dissecting the chromatin interactome of microRNA genes. *Nucleic Acids Res.* 2014;42:3028–3043.
- [38] Hagan JP, O'Neill BL, Stewart CL, et al. At least ten genes define the imprinted *Dlk1-Dio3* cluster on mouse chromosome 12qF1. *PLoS One.* 2009;4:e4352.
- [39] Kagami M, Sekita Y, Nishimura G, et al. Deletions and epimutations affecting the human 14q32. 2 imprinted region in individuals with paternal and maternal upD (14)-like phenotypes. *Nat Genet.* 2008;40:237.
- [40] Huan T. Genome-wide identification of DNA methylation quantitative trait loci in human whole blood highlights novel pathways for cardiovascular diseases. American Society of Human Genetics 68th Annual Meeting; San Diego; 2018.
- [41] Perry JR, Day F, Elks CE, et al. Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature.* 2014;514:92.
- [42] Kass SU, Landsberger N, Wolffe AP. DNA methylation directs a time-dependent repression of transcription initiation. *Curr Biol.* 1997;7:157–165.
- [43] Jones PA. The DNA methylation paradox. *Trends Genet.* 1999;15:34–37.
- [44] MacArthur J, Bowler E, Cerezo M, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS catalog). *Nucleic acids res* 2016; 45: D896–D901.
- [45] Richardson TG, Zheng J, Smith GD, et al. Mendelian randomization analysis identifies CpG sites as putative mediators for genetic influences on cardiovascular disease risk. *Am J Hum Genet.* 2017;101:590–602.
- [46] Kong A, Steinthorsdottir V, Masson G, et al. Parental origin of sequence variants associated with complex diseases. *Nature.* 2009;462:868.
- [47] Wallace C, Smyth DJ, Maisuria-Armer M, et al. The imprinted *DLK1-MEG3* gene region on chromosome 14q32. 2 alters susceptibility to type 1 diabetes. *Nat Genet.* 2010;42:68.
- [48] Sakajiri S, O'Kelly J, Yin D, et al. *Dlk1* in normal and abnormal hematopoiesis. *Leukemia.* 2005;19:1404.
- [49] Astuti D, Latif F, Wagner K, et al. Epigenetic alteration at the *DLK1-GTL2* imprinted domain in human neoplasia: analysis of neuroblastoma, pheochromocytoma and Wilms' tumour. *Br J Cancer.* 2005;92:1574.
- [50] Khoury H, Suarez-Saiz F, Wu S, et al. An upstream insulator regulates *DLK1* imprinting in AML. *Blood* 2010. blood-2009-03-212746.
- [51] Gao Y-Q, Chen X, Wang P, et al. Regulation of *DLK1* by the maternally expressed miR-379/miR-

- 544 cluster may underlie callipyge polar overdominance inheritance. *Proc Nat Acad Sci*. 2015;112:13627–13632.
- [52] Day FR, Thompson DJ, Helgason H, et al. Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nat Genet*. 2017;49:834.
- [53] Feinleib M, Kannel WB, Garrison RJ, et al. The Framingham offspring study. Design and preliminary data. *Prev Med*. 1975;4:518–525.
- [54] Splansky GL, Corey D, Yang Q, 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.
- [55] Pidsley R, CC YW, Volta M, et al. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics*. 2013;14:293.
- [56] Chen YA, Lemire M, Choufani S, et al. Discovery of cross-reactive probes and polymorphic CpGs in the illumina Infinium HumanMethylation450 microarray. *Epigenetics*. 2013;8:203–209.
- [57] Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet*. 2007;3:e161.
- [58] Joehanes R, Ying S, Huan T, et al. Gene expression signatures of coronary heart disease significance. *Arterioscler Thromb Vasc Biol*. 2013;33:1418–1426.
- [59] Jensen SG, Lamy P, Rasmussen MH, et al. Evaluation of two commercial global miRNA expression profiling platforms for detection of less abundant miRNAs. *BMC Genomics*. 2011;12:435.
- [60] Jang JS, Simon VA, Feddersen RM, et al. Quantitative miRNA expression analysis using fluidigm microfluidics dynamic arrays. *BMC Genomics*. 2011;12:144.
- [61] Chen C, Ridzon DA, Broomer AJ, et al. Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Res*. 2005;33:e179–e.
- [62] Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet*. 1998;62:1198–1211.
- [63] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc Ser B (Methodological)*. 1995;57:289–300.
- [64] Levy D, Ehret GB, Rice K, et al. Genome-wide association study of blood pressure and hypertension. *Nat Genet*. 2009;41:677–687.
- [65] Li Y, Willer CJ, Ding J, et al. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol*. 2010;34:816–834.
- [66] Agarwal V, Bell GW, Nam J-W, et al. Predicting effective microRNA target sites in mammalian mRNAs. *elife*. 2015;4:e05005.
- [67] Enright AJ, John B, Gaul U, et al. MicroRNA targets in *Drosophila*. *Genome Biol*. 2003;5:R1.
- [68] Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol*. 2013;37:658–665.
- [69] Willer CJ, Schmidt EM, Sengupta S, et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet*. 2013;45:1274.
- [70] Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife*. 2018;7:e34408.
- [71] Ernst J, Kheradpour P, Mikkelsen TS, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*. 2011;473:43.