

1 **An atlas of genetic regulation and disease associations of** 2 **microRNAs**

3

4 Rima Mustafa, MD MSc^{1,2}, Michelle M.J. Mens, PhD^{3,4}, Arno van Hilten, MSc⁵, Jian Huang,
5 PhD^{1,6}, Gennady Roshchupkin, PhD^{3,5}, Tianxiao Huan, PhD⁷, Linda Broer, PhD⁸, Paul Elliott,
6 MD PhD^{1,2,9,10,11}, Daniel Levy, MD^{12,13}, M. Arfan Ikram, MD PhD³, Marina Evangelou, PhD¹⁴,
7 Abbas Dehghan, MD PhD^{1,2,9*}, Mohsen Ghanbari, MD PhD^{2*}

8

9 Affiliations

10 ¹Department of Epidemiology and Biostatistics, Imperial College London, London, UK.

11 ²Dementia Research Institute, Imperial College London, London, UK.

12 ³Department of Epidemiology, Erasmus MC, Rotterdam, The Netherlands.

13 ⁴Department of Epidemiology, Harvard T.H Chan School of Public Health, Boston, USA.

14 ⁵Department of Radiology and Nuclear Medicine, Erasmus MC, Rotterdam, The Netherlands.

15 ⁶Singapore Institute for Clinical Sciences (SICS), the Agency for Science, Technology and
16 Research (A*STAR), Singapore.

17 ⁷Department of Ophthalmology and Visual Sciences, University of Massachusetts Medical
18 School, Worcester, MA, USA.

19 ⁸Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands.

20 ⁹MRC Centre for Environment and Health, Imperial College London, London, UK.

21 ¹⁰Health Data Research (HDR) UK, Imperial College London, London, UK.

22 ¹¹BHF Centre for Research Excellence, Imperial College London, London, UK.

23 ¹²Framingham Heart Study, Framingham, MA, USA.

24 ¹³Population Sciences Branch, National Heart, Lung, and Blood Institute, National Institutes of
25 Health, Bethesda, MD, USA.

26 ¹⁴Department of Mathematics, Imperial College London, London, UK.

27

28 *Abbas Dehghan and Mohsen Ghanbari are joint senior authors

29

30

31 **Corresponding authors**

32 Mohsen Ghanbari, MD PhD

33 Department of Epidemiology

34 Erasmus MC University Medical Center

35 Wytemaweg 80, 3015 CN

36 Rotterdam, The Netherlands

37 Email: m.ghanbari@erasmusmc.nl

38 **Competing interests**

39 None.

40

41

42

43

44 **Abstract**

45 MicroRNAs (miRNAs) are small non-coding RNAs that post-transcriptionally regulate gene
46 expression. Identification of genetic variants influencing the transcription of miRNAs can
47 provide an understanding of their genetic regulation and implication in human disease. Here
48 we present genome-wide association studies of 2,083 plasma circulating miRNAs measured
49 by next-generation sequencing in 2,178 participants of the Rotterdam Study to identify miRNA-
50 expression quantitative trait loci (miR-eQTLs). We report 4,310 cis- and trans-miR-eQTLs for
51 64 miRNAs that have been replicated across independent studies. Many of these miR-eQTLs
52 overlap with gene expression, protein, and metabolite-QTLs and with disease-associated
53 variants. The consequences of perturbation in miRNA transcription on a wide range of clinical
54 conditions are systematically investigated in phenome-wide association studies, with their
55 causality tested using Mendelian randomization. Integration of genomics and miRNAs enables
56 interrogation of the genetic architecture of miRNAs, revealing their clinical importance, and
57 providing valuable resources for future studies of miRNAs in human disease.

58

59 Introduction

60 MicroRNAs (miRNAs) are small non-coding RNAs of approximately 22 nucleotides that
61 regulate gene expression at the post-transcriptional level and play critical roles in determining
62 whether genes are (in)active and how much a particular protein is translated (1,2). Over 2,500
63 miRNAs have been identified in humans (3), which altogether regulate more than half of
64 protein-coding genes through cleavage or translation repression of messenger(m)-RNAs
65 (4,5). miRNAs have shown their potential as disease biomarkers (6) and, to a lesser extent,
66 therapeutic targets (7). Identification of the role of miRNAs in regulating the expression of
67 specific genes and their effects in clinical conditions has been a subject of extensive work in
68 recent years. However, the genetic regulation of miRNAs remains less well understood.

69 Circulatory miRNAs are released from cells into circulation via extracellular vesicles such as
70 exosomes (8). Genetic variants are known to regulate the level of miRNAs in the circulation
71 (9-11) or tissues and cells (12-14), referred to as miRNA expression quantitative trait loci (miR-
72 eQTLs). Previous studies showed that each miR-eQTL contributed to a relatively small
73 proportion of variation in miRNA levels (9,10), with a tiny proportion of miR-eQTLs replicated
74 across studies thus far (9). The identified miR-eQTLs have been used to study the effect of
75 perturbation of miRNA levels on disease risk (9,10,14). However, such an effect on a wide
76 range of clinical conditions remains to be elucidated. Unravelling the genetic regulation of
77 nearly all high confidence miRNAs can provide insights into their roles in affecting disease risk
78 and discover candidates for therapeutic targets.

79 This study measured plasma levels of 2,083 circulatory miRNAs in the population-based
80 Rotterdam Study cohort using a next-generation sequencing platform (HTG EdgeSeq miRNA
81 Whole Transcriptome Assay). This assay allows simultaneous, quantitative detection of
82 miRNAs with a sensitivity and specificity of 97% (15,16). Subsequently, genome-wide
83 association studies (GWAS) were conducted for 2,083 miRNAs to identify miR-eQTLs
84 (N=2,178), followed by replication in two independent cohorts (9,10). We conducted
85 downstream analyses to elucidate functional characteristics of the findings through cis and
86 trans mapping of miR-eQTLs, cross-phenotype, and multi-omics QTLs look-up. A systematic
87 investigation of the effects of genetically determined miRNA levels on a wide range of clinical
88 conditions was conducted using phenome-wide association studies (PheWAS) in the UK
89 Biobank (N=423,419) (17,18). We performed Mendelian randomisation (MR) to assess
90 causality between miRNAs and clinical conditions (19). Potential downstream target genes
91 that might be involved in the disease processes were highlighted, generating testable
92 hypotheses for further functional studies to dissect the underlying molecular pathways.

93 Results

94 Genome-wide identification and replication of miR-eQTLs

95 An overview of the study workflow is presented in [Fig. 1](#). The list of 2,083 miRNAs
96 characterised in this study is shown in [Supplementary Table 1](#). The miRNA expression
97 profiling was performed for 2,754 participants from three sub-cohorts (RS-I-4, RS-II-2, and
98 RS-IV-1) in the Rotterdam Study (20). Genotype data were available for 2,435 of 2,754
99 participants. Participants of non-European ancestries and relatives based on kinship
100 coefficient > 0.088 were excluded, resulting in 2,178 participants in the analysis ([Extended](#)
101 [Data Fig. 1](#)). The clinical characteristics of participants are summarised in [Supplementary](#)
102 [Table 2](#).

103 We identified 3,292 associations between 1,289 SNPs and 63 miRNAs at $P < 2.4 \times 10^{-11}$
104 (genome-wide threshold of $P < 5 \times 10^{-8}$ and Bonferroni-corrected for 2,083 miRNAs)
105 ([Supplementary Table 3](#), [Extended Data Fig. 2](#)). After pruning with a series of linkage
106 disequilibrium (LD) thresholds, the number of associations reduced to 75 ($r^2 < 0.1$), 142 ($r^2 < 0.3$),
107 and 297 ($r^2 < 0.6$). The 3,292 identified associations included 1,733 cis associations for 32
108 miRNAs and 1,559 trans associations for 33 miRNAs. The overall proportion of variance
109 explained by each miR-eQTL ranged from 2% to 11% (median=2.9%). Eighteen miR-eQTLs
110 ($r^2 < 0.6$) were explaining more than 5% of the variation of corresponding miRNA levels
111 ([Supplementary Table 4](#)). The highest proportion of variance explained was observed for miR-
112 625-5p (11%) by rs2127868 ($P = 1.63 \times 10^{-60}$), identified as cis-miR-eQTL.

113 We sought to replicate our findings in a previous study using the same HTG EdgeSeq platform
114 by Nikpay et al. (9). Bonferroni correction was applied to address multiple testing ([Online](#)
115 [Methods](#)). There were 2,254 associations for 57 miRNAs with identical SNPs available in the
116 replication cohort by Nikpay et al. (9). Additionally, 27 associations for six miRNAs were tested
117 for replication using proxy SNPs ([Online Methods](#)). There were 1,462 associations for 27
118 miRNAs replicated at Bonferroni-threshold ($P < 0.05/58$) ([Extended Data Fig. 2](#), [Supplementary](#)
119 [Tables 5-6](#)). The effect estimates of replicated associations were strongly correlated ($r = 0.82$,
120 $P < 2.2 \times 10^{-16}$) ([Extended Data Fig. 3a](#)). Additionally, 1,719 associations were nominally
121 significant in Nikpay et al. (9), of which 1,685 (98%) were in a concordant direction
122 ([Supplementary Table 6](#)).

123

124 Replication of previously identified miR-eQTLs

125 In an alternative approach, we also attempted to replicate miR-eQTLs identified in previous
126 studies (9,10). Associations reaching $P < 2.4 \times 10^{-11}$ in Nikpay et al. (9) were tested for replication
127 in our study ([Extended Data Fig. 2](#)), where 2,957 associations corresponding to 1,973 SNPs

128 and 34 miRNAs were replicated ($P < 0.05/52$) with a significant correlation between effect
129 estimates ($r = 0.90$, $P < 2.2 \times 10^{-16}$) (Extended Data Fig. 3b). Additionally, we tested 8,820 cis and
130 trans associations for 87 miRNAs that were previously reported by the Framingham Heart
131 Study (10) in our data. Of those, 1,320 associations for 29 miRNAs were replicated
132 ($P < 0.05/195$) (Extended Data Fig. 3c). Collectively, 69% of associations in Nikpay et al. (20)
133 and 15% of associations in the Framingham Heart Study (19) were replicated in our study.
134 Altogether, our approaches successfully reported 4,310 associations pertaining to 64 miRNAs
135 which have been replicated across studies (Extended Data Fig. 2, Supplementary Table 7).

136

137 Functional annotation of miR-eQTLs

138 Using the Functional Mapping and Annotation (FUMA) (21), the identified and replicated miR-
139 eQTLs were mapped into 22 loci (Fig. 2a). Over 70% of miR-eQTLs were located in the intronic
140 and intergenic regions (Fig. 2b, Supplementary Table 8). Out of 22 loci, 11 loci were
141 pleiotropic, i.e., linked to the level of multiple miRNAs. Fourteen loci regulated two miRNAs
142 and eight loci regulated more than two miRNAs (Supplementary Table 9). Highly pleiotropic
143 loci were identified in locus chr14:100655022-101244293, regulating 23 miRNAs (also known
144 as 14q32 miRNA cluster), the majority of which were cis-miR-eQTLs (98%). This locus was
145 mapped to *RP11-566J3.2*, *RP11-638I2.4*, *YY1*, *YY1:RP11-638I2.2*, *SLC25A29*, *WDR25*,
146 *BEGAIN*, *DLK1*, *CTD-2644I21.1*, *LINC00523*, *RP11-566J3.4*, and *MEG3* (Extended Data Fig.
147 4). The locus in chr9:136128546-136296530 was mapped to *ABO*, *ABO:RP11-430N14.4*,
148 *Y_RNA*, and *LCN1P2* and regulated 18 miRNAs. This locus contained shared trans-miR-
149 eQTLs for several well-known miRNAs, including miR-10, let-7, and miR-30 families
150 (Supplementary Table 9).

151 Associations between 296 SNPs residing in seed, mature, or precursor genes of miRNAs and
152 corresponding miRNAs were extracted from our GWAS results. Twelve associations were
153 significant at Bonferroni-threshold considering the number of unique SNPs ($P < 0.05/296$),
154 consisting of three SNPs in the seed region of miR-4707-3p, miR-4482-5p, and miR-6891-3p,
155 three SNPs in mature region of miR-3130-3p, miR-6891-3p, and miR-6839-5p, and six SNPs
156 in precursor genes of six miRNAs. Additionally, 27 SNPs in 26 miRNAs were nominally
157 significant (Supplementary Table 10).

158

159 Heritability analysis

160 SNP-based heritability estimates for the plasma levels of 2,083 circulatory miRNAs were
161 obtained using massively expedited genome-wide heritability analysis (MEGHA) (22). The
162 distribution of heritability estimates is shown in Fig. 2c. Two miRNAs had a narrow-sense
163 heritability estimate greater than 0.7, namely miR-30e-5p (0.72) and miR-6511a-5p (0.70).

164 Twenty miRNAs had a narrow-sense heritability greater than 0.5, and 166 miRNAs had greater
165 than 0.3. Heritability estimates for all miRNA are shown in [Supplementary Table 11](#). Repeating
166 the heritability analysis, including the first five principal components as covariates, resulted in
167 SNP-based heritability estimates with a Pearson correlation of 0.99 with estimates without the
168 principal components ([Extended Data Fig. 5](#)).

169 [Cross-phenotype and multi-omics QTLs look-up](#)

170 As miRNAs dictate their role in biological processes by regulating the expression of their target
171 genes, it is interesting to know whether miR-eQTLs are linked to the expression of other
172 genes, including their host and target genes. To explore this, we sought overlaps between
173 replicated miR-eQTLs and gene expression (eQTLs), protein (pQTLs), and metabolite-QTLs
174 (met-QTLs). We further checked if any of the genes or proteins that shared QTLs were target
175 genes of miRNAs.

176 Summary statistics for genetic variants that influence expression levels of mRNA transcripts
177 (cis and trans-eQTLs) in whole blood (N=31,684) were used to identify miR-eQTLs that affect
178 the expression of other genes (23). In this dataset, trans-eQTL analysis was conducted only
179 for SNPs previously identified in GWAS (23). Cis-miR-eQTLs for 39 miRNAs overlapped with
180 cis-eQTLs for 146 genes, 123 of which were protein-coding genes ([Supplementary Table 12](#)).
181 Twelve intragenic miRNAs shared cis-miR-eQTLs with their host genes ([Supplementary Table](#)
182 [13](#)). In addition, miR-136-5p resides within the intronic region of *RTL1* but had overlapping
183 cis-eQTLs with nearby genes, such as *DLK1*, *WARS*, *BEGAIN*, *MEG3*, and *SLC25A29*
184 ([Supplementary Table 12](#)).

185 Next, predicted target genes for miRNAs from TargetScan v7.2 (5) and experimentally
186 validated target genes with robust validation methods (such as reporter assay, western blot or
187 qRT-PCR) from miRTarBase (4) were retrieved to check whether miR-eQTLs overlap with
188 eQTLs of their putative target genes. Twelve miRNAs shared either cis/trans-miR-eQTLs or
189 cis/trans-eQTLs of their putative target genes ([Supplementary Table 14](#)). We identified shared
190 trans-regulation by rs612169, located intronic to *ABO*, for miR-126-3p and its validated target
191 *TCF4*.

192 Protein-QTLs (pQTLs) summary results were used to identify miR-eQTLs affecting protein
193 levels in the blood (24,25). Cis-miR-eQTLs of 18 miRNAs overlapped with pQTLs for nine
194 proteins. Cis-miR-eQTLs for 14q32 miRNA cluster were shared with pQTLs of *DLK1* located
195 in the nearby genomic region and *SEMG2* in a distant region of chromosome 20 ([Fig. 3a](#),
196 [Supplementary Table 15](#)). These included shared miR-eQTLs of six intragenic miRNAs, of
197 which four miRNAs (miR-127-3p, miR-136-5p, miR-431-5p, and miR-433-5p) reside in *RTL1*.

198 Cis-miR-eQTLs of miR-625-5p overlapped with pQTLs for Alpha-(1,6)-fucosyltransferase, an
199 enzyme encoded by *FUT8* gene where miR-625-5p resides. Cis-miR-eQTLs for miR-335-5p
200 that resides in *MEST* overlapped with Carboxypeptidase A4, encoded by *CBPA4* in a nearby
201 genomic region in chromosome 7. Trans-miR-eQTLs for 11 miRNAs overlapped with pQTLs
202 for 103 proteins (Fig. 3b, Supplementary Table 16). Of these, *GNS* was identified as common
203 predicted targets of let-7b-5p and let-7c-5p. Additionally, trans-miR-eQTLs of miR-145-5p and
204 miR-195-5p overlapped with eQTLs of their predicted target genes. For cardiovascular
205 proteins (25), cis-miR-eQTLs for miR-130a-3p overlapped with pQTLs of Pappalysin-1
206 (*PAPPA*). Overlap for trans-miR-QTLs was found with pQTLs of validated target genes of miR-
207 126-3p (*TEK*) and miR-145-5p (*MMP1* and *VEGFA*) (Supplementary Table 17). The overlap
208 of trans-miR-eQTLs between miR-143-3p and miR-145-5p with Dickkopf-related protein 1
209 (*DKK1*) were identified in both pQTLs datasets (24,25).

210 We used summary statistics from two common metabolomics platforms, Metabolon and
211 Nightingale, to identify miR-eQTLs affecting metabolic pathways by investigating plasma
212 levels of metabolites. Metabolon covers 529 metabolites (N=7,824) (26), while Nightingale
213 covers 123 metabolites in Kettunen et al. (N>20,000) (27) and 249 metabolites in the UK
214 Biobank (N=115,078) (28). Cis-miR-eQTLs for miR-1908-5p, miR-148a-3, miR-339-5p, and
215 miR-130a-3p overlapped with met-QTLs for 218 metabolites in both platforms. For example,
216 rs174561, located in the precursor gene of miR-1908-5p and intronic to *FADS1*, both known
217 to be associated with lipid and obesity traits, was associated with metabolites in Nightingale
218 that were also mainly lipid fractions. Trans-miR-eQTLs for nine miRNAs overlapped with four
219 unnamed metabolites in Metabolon (M32740, M33801, M36115, M36230), and 146
220 metabolites in the Nightingale platform (Supplementary Table 18).

221 Finally, GWAS Catalog was used to identify miR-eQTLs associated with complex traits (29).
222 At genome-wide significance, cis-miR-eQTLs were associated with GWAS traits, including
223 mental health, haematological indices, cancers, anthropometric measures, lipid levels, and
224 blood pressure (Supplementary Table 19). For example, cis-miR-eQTLs for miR-1908-5p
225 were associated with multiple traits, mainly lipid through *FADS1*, *FADS2*, or *MYRF*, supporting
226 the observed overlaps with eQTLs and pQTLs as described earlier. Trans-miR-eQTLs were
227 associated with different traits and diseases, including haematological indices,
228 cardiometabolic, cancer, and allergy. The pleiotropic regulatory region in chr9:136128546-
229 136296530 consisted of trans-miR-eQTLs which were associated with protein and metabolite
230 levels and complex traits. For example, rs687289, intronic to *ABO*, was identified as trans-
231 miR-eQTL for six miRNAs and overlapped with pQTLs, met-QTLs, and associated with GWAS
232 traits such as monocyte count, coagulation factor levels, and pancreatic cancer. Another

233 variant in *ABO*, rs644234, was associated with 40 protein biomarkers, including five
234 cardiovascular proteins ([Supplementary Table 20](#)). This observation might suggest the
235 relevance of miR-eQTLs in *ABO* to cardiovascular traits.

236 [Phenome-wide association studies \(PheWAS\) in the UK-Biobank](#)

237 We conducted phenome-wide association studies (PheWAS) in the UK Biobank (17) to
238 investigate associations between genetically determined miRNA levels and a wide range of
239 clinical conditions ([Fig. 4a](#)). The participants with genetic and hospital episode statistics data
240 were used in the analysis. After excluding participants of non-European ancestries and one
241 from each pair of related individuals, 423,419 individuals remained in the analysis ([Extended](#)
242 [Data Fig. 6](#)). We tested the associations between genetically determined miRNA levels and
243 905 phecodes with at least 200 cases across 16 disease groups ([Fig. 4b](#)).

244 To extract genetic instruments for miRNAs, a threshold for cis-instruments at $FDR < 0.1$ was
245 calculated across all SNPs residing ± 500 kb of each miRNA. This threshold was chosen to
246 enable covering a higher number of miRNAs tested in PheWAS. At $FDR < 0.1$, cis-instruments
247 were identified for 204 miRNAs. Weak instruments were filtered out using $F\text{-statistics} > 10$
248 ([Online Methods](#)). After LD clumping ($r^2 < 0.1$), a single cis-instrument was available for 85
249 miRNAs for which a single variant PheWAS were conducted ([Supplementary Table 21](#)).
250 Multiple cis-instruments were available to compute genetic risk scores (GRS) for 119 miRNAs
251 and test in GRS-PheWAS ([Fig. 4a](#)). Since each clinical diagnosis is not entirely independent
252 of the other, FDR correction was applied for each miRNA to account for multiple testing.

253 In the single variant PheWAS, 29 significant associations were identified between 9 SNPs and
254 23 clinical diagnoses at $FDR < 0.05$ ([Fig. 4a](#)). Among these, rs2270197 ($P = 6.84 \times 10^{-05}$),
255 rs55936521 ($P = 4.27 \times 10^{-06}$), rs5623708 ($P = 1.40 \times 10^{-05}$) and rs7130989 ($P = 7.16 \times 10^{-05}$) were
256 associated with hypertension. rs1254901, located 2KB upstream of *VAMP5* and cis-miR-eQTL
257 for miR-6701, was associated with ischemic heart disease-related conditions, including
258 ischemic heart disease ($P = 6.51 \times 10^{-06}$). Rs2270197, intronic to *ITIH1* and cis-miR-eQTL for
259 miR-135a-5p, was associated with a range of clinical conditions, including osteoarthritis
260 ($P = 4.22 \times 10^{-04}$), hypertension ($P = 6.84 \times 10^{-05}$), and bipolar disorders ($P = 4.02 \times 10^{-04}$) ([Fig. 4c](#),
261 [Supplementary Table 22](#)).

262 In the genetic risk score (GRS) PheWAS, 44 associations between 17 cis-GRS and 24
263 diagnoses were identified at $FDR < 0.05$ ([Fig. 4a](#)). The strongest association was identified
264 between miR-1908-5p and benign neoplasm of colon ($OR = 0.96$, $P = 1.99 \times 10^{-08}$) and
265 cholelithiasis ($OR = 1.04$, $P = 1.51 \times 10^{-05}$). Three miRNAs were associated with lower risk of
266 obesity, namely miR-323b-3p ($OR = 0.97$, $P = 2.28 \times 10^{-04}$), miR-329-3p ($OR = 0.97$, $P = 1.12 \times 10^{-04}$).

267 ⁰⁴), and miR-543 (OR=0.97, P=1.20x10⁻⁰⁴). Three miRNAs were associated higher risk of skin
268 cancer, including miR-323b-3p (OR=1.05, P=4.79x10⁻⁰⁶), miR-376b-3p (OR= 1.04,
269 P=5.81x10⁻⁰⁵), and miR-379-5p (OR=1.05, P=7.68x10⁻⁰⁶) (Fig. 4d, Supplementary Table 22).
270 Out of 44 associations, extended-GRS was computed for 17 associations of five miRNAs with
271 genome-wide significant trans-miR-eQTLs, where all remained statistically significant in a
272 concordant direction (FDR<0.05) (Supplementary Table 23). Our PheWAS identified 73
273 associations between 45 clinical conditions and 26 miRNAs. Among those, 11 miRNAs were
274 associated with circulatory disorders (Fig. 5a). Several miRNAs were associated with clinical
275 diagnoses across different disease groups, indicating their pleiotropic properties, such as miR-
276 323b-3p associated with endocrine/metabolic disease, infectious disease, and neoplasms
277 (Fig. 5b).

278

279 Mendelian randomisation

280 Mendelian randomisation (MR) was conducted for miRNAs with at least three independent
281 instruments (Online Methods). Thirty-seven associations for 14 miRNAs with at least three
282 instruments were tested in MR-PheWAS (Fig. 4a). The same cis or extended instruments for
283 miRNAs were used in MR-PheWAS with effect estimates from the Rotterdam Study
284 (N=2,178), whereas genetic associations with the outcomes were taken from the UK-Biobank
285 (N=423,419). All 37 tested associations were significant in MR Inverse variance weighted
286 (IVW) (FDR<0.05), had no indication of pleiotropy or heterogeneity and were in concordant
287 direction with MR-Egger or Weighted Median (WM) (P<0.05) (Supplementary Table 24).
288 Among those, the strongest association was observed between miR-1908-5p and the risk of
289 benign neoplasm of the colon (MR-IVW estimate=-0.40, P=3.9x10⁻¹⁰). Six miRNAs were
290 associated with a higher risk of melanoma, including miR-329-3p (MR-IVW estimate=0.37,
291 P=5.4x10⁻⁹), miR-376b-3p (MR-IVW estimate=0.39, P=3.0x10⁻⁸), miR-323b-3p (MR-IVW
292 estimate=0.44, P=1.4x10⁻⁷), and miR-379-5p (MR-IVW estimate=0.47, P=1.5x10⁻⁵). All 37
293 associations were significant when the correlation matrix between instruments was added,
294 with no outliers detected by MRPRESSO. Of 37, 13 associations had genome-wide significant
295 trans-miR-eQTLs to conduct extended-MR as a replication. For the remaining 24 associations,
296 concordant direction across different MR methods was observed (Fig. 5c).

297 Extended-MR was conducted for 13 associations by adding genome-wide significant trans-
298 miR-eQTLs. Twelve associations were significant (MR-IVW FDR<0.05) with no indication of
299 pleiotropy and supported by MR-Egger or WM (P<0.05) (Supplementary Table 25). Nine
300 associations from MR-PheWAS had genetic association data available in large GWAS for
301 coronary artery disease, body mass index (BMI) and waist to hip ratio (WHR) (30-32) (Fig.
302 5a). Two associations were replicated between miR-543 and WHR (MR-IVW estimate=-0.02,

303 $P=1.72 \times 10^{-02}$) and between miR-329-3p and BMI (MR-IVW estimate=-0.03, $P=1.89 \times 10^{-02}$),
304 both in the same protective effects as in MR-PheWAS (Table 1, Extended Data Fig. 7,
305 Supplementary Table 26). Reverse MR analyses were conducted for BMI and WHR as risk
306 factors on miR-543 and miR-329-3p as the outcomes, where no significant effects in the
307 opposite directions were observed (Supplementary Table 27).

308

309 Target genes and enrichment analysis

310 An in-silico search of target genes using TargetScan v7.2 and miRTarBase (5,33) identified
311 eighty-two predicted and eighteen validated target genes of miR-543 associated with BMI or
312 WHR. Forty-three predicted and fifty-eight validated target genes for miR-329-3p were also
313 associated with BMI or WHR. Eighteen target genes were in common between miR-543 and
314 miR-329-3p, where eleven genes were validated targets for at least one of them, including
315 *BRCH1* and *TNRC6B* as validated targets of both miRNAs (Supplementary Table 28).
316 Enrichment analysis was conducted as described in our previous work (34), resulting in a
317 significant enrichment for BMI or WHR-related genes among validated targets of miR-543
318 ($P=9.00 \times 10^{-03}$) and predicted targets of miR-329-3p ($P=3.18 \times 10^{-02}$).

319

320 Discussion

321 We present a genome-wide identification of miR-eQTLs using the next-generation sequencing
322 method in 2,178 individuals in the population-based Rotterdam Study cohort. This study is
323 currently the most extensive single-site analysis of 2,083 circulatory miRNA levels in a
324 population of European ancestries. We discovered 3,292 genetic associations for 63 miRNAs.
325 The highest proportion of variance explained was observed for miR-625-5p (11%) by
326 rs2127868 ($P=1.63 \times 10^{-60}$), in perfect LD with rs2127870 associated with the level of miR-625-
327 5p in plasma ($P=2.9 \times 10^{-260}$) and whole blood ($P=2.98 \times 10^{-09}$) (9,10). Altogether, 4,310 genetic
328 associations for 64 miRNAs were replicated across different studies, including trans-miR-
329 eQTLs, whose replication was previously minimal. Genetically proxied miRNAs were tested
330 against a wide range of clinical conditions in the UK-Biobank, suggesting the pleiotropic
331 properties of several miRNAs by being associated with clinical outcomes. Such observation is
332 expected, given that miRNAs could potentially regulate many genes that are involved in
333 different molecular pathways (4,5). Our MR analysis identified the potentially causal role of
334 miRNAs in various complex traits and disorders.

335 We showed that miR-eQTLs overlap with gene expression QTLs and protein QTLs of their
336 target genes, supporting their role in translational repression. Since target genes tend to be
337 clustered to miRNAs according to their function (34,35), these shared miR-eQTLs might have

338 biological relevance. Cis-miR-eQTLs that overlap with trans-mRNA-eQTLs might point to the
339 downstream regulatory effect from miRNAs to their (direct or indirect) target genes. When cis-
340 mRNA-eQTLs overlap with trans-miR-eQTLs, the effect might be going from the genes to
341 miRNAs, pointing to bidirectional interaction between miRNAs and target genes as a feedback
342 mechanism (36,37). However, when trans-miR-eQTLs overlap with trans-mRNA-eQTLs, a
343 third factor may have contributed to simultaneous changes in miRNA and gene expression.
344 As an example, a genetic variant could affect the regulatory region shared between miRNA
345 and a gene that are co-expressed. We also hypothesise that cis- and trans-miR-eQTLs might
346 have different clinical relevance. The magnitude of associations between miRNAs and
347 complex traits appeared closer to the null when trans-miR-eQTLs were added as instruments.
348 Trans-miR-eQTLs might affect the stability of mature miRNAs, whereas cis-miR-eQTLs
349 influence the hairpin structure and regulate the expression of primary miRNAs (9).

350 The current analysis identified co-expression of miRNAs and host genes, which could occur
351 through modification of promoter activity, chromatin accessibility, transcription factor binding,
352 or DNA methylation (10). However, many miRNAs also have their own promoters (38), and
353 the association could be independent of the host genes (10,39). This finding deepens our
354 understanding that the relationship between miRNAs and gene expressions is more
355 commonly driven by genetics (40). The genetic effect might be less strong for miRNAs than
356 mRNAs (12,13), as shown by the small variation explained by miR-eQTLs, which could act as
357 a mechanism to maintain biological function during evolution.

358 Given that each miRNA potentially regulates multiple target genes and pathways (1,2,5), even
359 small changes in miRNA expression could result in considerable consequences. This concept
360 aligns with the strong evolutionary constraint on miRNAs and their binding sites in gene 3-
361 UTRs in humans and other species (5). Moreover, the seed, mature, and precursor regions of
362 miRNA genes are known to have a lower density of genetic variation than the whole genome
363 (41). Our study shows that the genetic variants in those regions could have functional
364 importance, such as affecting miRNA expression. This functional consequence occurs by
365 interfering with the processing of precursor to mature miRNA or the interaction between
366 mature miRNA and target genes, resulting in gain and loss of function, which could deregulate
367 biological pathways (42,43).

368 Human miRNAs can be categorised into families with similar functions due to their conserved
369 structures in the mature or seed sequences (44) and clusters when they are encoded from the
370 same region in our genome (3). Here, we showed that the 14q32 miRNA cluster shares cis-
371 regulatory variants. We also showed that multiple miRNAs are regulated by shared miR-
372 eQTLs (45), such as the pleiotropic trans-miR-eQTLs in the *ABO* gene. Several families

373 sharing trans-regulatory variants in *ABO*, such as miR-10 family, miR-30 family, let-7 family,
374 and miR-139-5p, were well-known in cardiometabolic traits and cancers (46-48). This finding
375 agrees with the concept that several miRNAs can work in networks to control gene expression
376 and pathways underlying diseases (49).

377 Several associations with complex traits highlighted in this study were reported in the
378 literature. For example, miR-543 was released in plasma following a high-fat diet (50), which
379 could be a physiological response to reduce the risk of obesity. Target genes of miR-329-3p
380 were involved in lipid and glucose metabolism in rats (51). Low miR-329 expression was
381 observed in melanoma cells, while miR-329 mimics could suppress the progression of
382 melanoma (52). The effect in tumour tissue for miR-329 and miR-1908-5p (14,52) was
383 opposite compared to our MR analysis which better captures the lifetime effect of miRNAs.
384 This suggests the changes in the level of miRNAs in tumour tissue might be the consequence
385 of disease processes and supports the hypothesis that the dysregulation of miRNA in diseased
386 tissue might arise from negative feedback by downstream genes (36,37). It is also possible
387 that the genetic effects have been buffered by canalisation (19), where people with a
388 genetically higher level of miRNAs since the intra-uterine period might be resistant to the effect
389 of higher miRNAs throughout life.

390 Here we would like to underline several aspects to be considered when attempting to replicate
391 miR-eQTLs across studies. First, we found fewer trans were replicated than cis-miR-eQTLs,
392 as observed in the large eQTL analysis (53). Trans-eQTLs are known to have weaker effects,
393 are less replicable, and are more tissue-specific (54-56) than cis-eQTLs. Second, the
394 concordant direction with those reported by Nikpay et al. (9) suggested that the type of
395 biological sample and profiling method could have an effect. The lower replication rate in the
396 Framingham Heart Study is likely due to differences in type of sample (whole blood vs
397 plasma), as previously reported (57). Third, one should consider any systematic difference in
398 participants' characteristics across studies. This study came from a population-based cohort
399 which makes the findings more generalisable. Other studies were in obese individuals (9) or
400 enriched for a specific disease (11), making it particularly useful for investigating the relevant
401 disease but not for a wide range of complex traits and disorders. Finally, since the overall
402 proportion of variation explained by each miR-eQTL is relatively small, larger GWAS for miR-
403 eQTLs identification will be a valuable resource to enrich the genetic studies on miRNAs. In
404 particular, incorporating diverse ancestries could generate more transferrable findings for a
405 wider population.

406 Collectively, the integration of genomics, molecular, and clinical data in this study has provided
407 a better understanding of the genetic regulation of miRNAs and allowed us to perform a

408 systematic investigation on the effect of perturbations of plasma miRNA levels on a wide range
409 of clinical conditions. As an example, we highlight miR-543 and miR-329-3p associated with
410 obesity-related traits with potential downstream targets. Although it is unlikely a single miRNA
411 or its target genes will be entirely responsible for the disease mechanisms, it is plausible that
412 the effect of identified miRNAs to be mediated at least in part through those target genes. Our
413 approach allows generating testable hypotheses for further functional and clinical studies to
414 dissect the underlying molecular and cellular pathways of various traits and diseases.

415 The summary statistics for miR-eQTLs identified in our study and their link to other omics
416 layers and their associations with various clinical outcomes will be available through a web
417 tool called miRNomics Atlas (www.mirnomicsatlas.com). This web tool allows the use of
418 genetic association data of miR-eQTLs, serving as valuable resources for future research to
419 decipher the association and causal role of miRNAs in human diseases and their regulatory
420 pathways.

421

422

423 **Online Methods**

424 **Cohort description**

425 The Rotterdam Study (RS) is a large prospective population-based cohort study among
426 middle-aged and elderly in the suburb Ommoord in Rotterdam, the Netherlands. In 1990,
427 7,983 inhabitants aged 55 years old and older were recruited to participate in the first cohort
428 (RS-I). In 2000, the study was extended with a second cohort of 3,011 participants (RS-II) who
429 became 55 years old or moved into the study district since the beginning of the study. In 2006,
430 a further extension of the cohort (RS-III) was initiated, including 3,932 participants aged 45–
431 54 years. In 2016, the recruitment of another extension started (RS-IV), targeting participants
432 aged 40 years and over, adding 3,005 new participants. A detailed description of the
433 Rotterdam Study can be found elsewhere (20).

434 **Circulatory miRNA levels**

435 Plasma miRNA levels were determined using the HTG EdgeSeq miRNA Whole Transcriptome
436 Assay (WTA) to quantitatively detect the expression of 2,083 human miRNAs transcripts (HTG
437 Molecular Diagnostics, Tuscon, AZ, USA) and using the Illumina NextSeq 500 sequencer
438 (Illumina, San Diego, CA, USA). This method characterises miRNA expression patterns and
439 measures the expression of 13 housekeeping genes to allow flexibility during data
440 normalisation and analysis. Quantification of miRNA expression was based on counts per
441 million (CPM). Log₂ transformation of CPM was used as standardisation and adjustment for
442 total reads within each sample. MiRNAs with log₂ CPM <1.0 were indicated as not expressed
443 in the samples.

444 **Genotype data**

445 At baseline, blood was drawn for genotyping from 6,291 participants in RS-I, 2,157 in RS-II,
446 and 2,654 in RS-IV. Genotyping in RS-I-II was performed using the HumanHap550 Duo
447 BeadChip (Illumina, San Diego, California) for RS-I-II and the Global Screening Array
448 (GSAMD-v3) Illumina array for RS-IV. Samples with a call rate below 97.5%, gender
449 mismatches, excess autosomal heterozygosity, duplicates or family relations, and ethnic
450 outliers were excluded. Variants with call rates below 95.0%, failing missingness test, Hardy-
451 Weinberg equilibrium $P < 10^{-6}$, and allele frequency below 1% were removed. Genotypes were
452 imputed using the MaCH/minimac software to the 1000 Genomes phase I version 3 reference
453 panel or phase 3 version 5 reference panels (for RS-IV). Genetic variants with minor allele
454 frequency < 0.05 and imputation quality < 0.7 were filtered out after genotype imputation.

455 [Genome-wide association studies](#)

456 Genome-wide association studies (GWAS) were conducted for 2,083 miRNAs in 2,178
457 participants randomly selected from three sub-cohorts of the Rotterdam Study to identify
458 miRNA-expression quantitative trait loci (miR-eQTLs). Given the high number of miRNAs,
459 GWAS was performed within the high-dimensional analysis framework (HASE) to reduce the
460 computational burden and enable efficient implementation of GWAS on thousands of
461 phenotypes (58). Multiple linear regression was used to test for association between each
462 genetic variant and miRNA level, with miRNA level as the outcome and expected genotype
463 count from imputation as predictors, with adjustment for age, sex, sub-cohort, and the first five
464 principal components to account for population stratification.

465 We used the genome-wide threshold of $P < 5 \times 10^{-08}$ and Bonferroni-corrected for 2,083 miRNAs
466 ($P < 2.4 \times 10^{-11}$) to identify significant associations. Associations reaching significance in the
467 Rotterdam Study were taken forward for replication in a published miR-eQTLs study by Nikpay
468 et al. (9) using the SNPs or their proxy SNPs ($r^2 > 0.7$ within 500kb on either side of lead SNP
469 position) obtained using LDlinkR (59). Linkage disequilibrium (LD) pruning was used to identify
470 the number of independent SNPs for each miRNA ($r^2 < 0.01$). Similarly, associations identified
471 in previous GWAS by Nikpay et al. (at $P < 2.4 \times 10^{-11}$) (9) and Huan et al. in the Framingham
472 Heart Study (at $FDR < 0.1$) (10) were also tested for replication. The Bonferroni threshold was
473 used for replication ($\alpha < 0.05/n$, where n is the total number of SNP-miRNA pairs after pruning).
474 Replication was defined when the associations between SNP and miRNA were Bonferroni-
475 significant in an independent cohort with a concordant direction of effect.

476 [Functional annotation of miR-eQTLs](#)

477 Genomic coordinates of miRNAs were extracted from miRBase v20
478 (<ftp://mirbase.org/pub/mirbase/20/genomes/has.gff3>) (3). Both SNPs and mature miRNA
479 positions were based on Genome Reference Consortium Human Build 37 (GRCh37). The
480 position of each miR-eQTL was mapped as cis or trans with respect to the miRNA position.
481 SNPs located ± 500 kb upstream and downstream of the start position of mature miRNAs were
482 identified as cis, and those located more than ± 500 kb away were identified as trans. To identify
483 SNPs in seed, mature, or precursor genes of miRNA, the database was downloaded from
484 <http://bioinfo.life.hust.edu.cn/miRNASNP/#!/download> (60).

485 The web-based tool Functional Mapping and Annotation (FUMA) was used to annotate
486 identified miR-eQTLs. A detailed description of the FUMA workflow is described elsewhere
487 (21). Independent significant miR-eQTLs were defined as those with $P < 5 \times 10^{-08}$ in the
488 discovery GWAS or those replicated in independent cohorts and moderate LD with each other
489 at $r^2 < 0.6$. LD calculation was referenced based on the 1000 Genomes phase 3 panel. These

490 SNPs were further clumped to lead SNPs ($r^2 < 0.1$). Genomic risk loci were then defined based
491 on the lead SNPs when they overlap with a maximum distance of 250kb between LD blocks.
492 The major histocompatibility complex (MHC) region was excluded using the default region
493 between *MOG* and *COL11A2* genes (21,61).

494 [Heritability analysis](#)

495 The SNP-based heritability estimates for 2,083 circulatory miRNAs were obtained using
496 massively expedited genome-wide heritability analysis (MEGHA) (22). A genetic relationship
497 matrix was constructed from 1000 Genome imputed genotypes filtered on imputation quality
498 (< 0.5) and allele frequency (< 0.1) using GCTA (62). After applying a stringent cut-off of 0.025
499 for genetic relatedness, 1,506 individuals were used for heritability estimation. Using MEGHA,
500 the genetic relationship matrix, and age and sex as covariates, we computed the heritability
501 and uncertainty (p-values based on 1000 permutations).

503 [Cross-phenotype and quantitative trait loci look-up](#)

504 Cross-phenotype and quantitative trait loci look-up leveraged replicated miR-eQTLs with
505 summary statistics for gene expression (eQTLs), protein-QTLs (pQTLs), metabolite-QTLs
506 (met-QTLs), and complex traits. Summary for cis and trans-eQTLs analysis in whole blood
507 were downloaded from <https://www.eqtngen.org/index.html> (23). Summary statistics for pQTLs
508 were from <https://www.phpc.cam.ac.uk/ceu/proteins/> (24) and the SCALLOP consortium
509 available through <https://zenodo.org/record/2615265/> (25). Summary statistics for met-QTLs
510 were from Metabolon (<http://metabolomics.helmholtz-muenchen.de/gwas/>) (26) and
511 Nightingale from a published study (http://www.computationalmedicine.fi/data/NMR_GWAS/)
512 (27) and in the UK Biobank available through OpenGWAS project (<https://gwas.mrcieu.ac.uk>)
513 (28). The database for intragenic miRNAs was from [https://bmi.ana.med.uni-](https://bmi.ana.med.uni-muenchen.de/miriad/)
514 [muenchen.de/miriad/](https://bmi.ana.med.uni-muenchen.de/miriad/) (63). GWAS Catalog was downloaded from
515 <https://www.ebi.ac.uk/gwas/docs/file-downloads> (29).

516 [Phenome-wide association studies](#)

517 To investigate associations between genetically determined circulatory miRNA and a wide
518 range of clinical diagnoses, a phenome-wide association study (PheWAS) was performed
519 using hospital episode statistics data in the UK Biobank, a large prospective cohort study with
520 over 500,000 individuals aged 40-69 years old recruited between 2006-2010 (17). In brief,
521 participants with genotype and phenotype data were considered in the analysis. Quality control
522 steps taken in the UK Biobank has been described elsewhere. Our analysis was restricted to
523 participants who identified themselves as “White”. One from each pair of relatives and
524 withdrawn individuals as of August 2021 were excluded ([Extended Data Fig. 3](#)). ICD (ninth

525 and tenth editions) codes from the hospital episode statistics data were aligned into phecodes
526 to identify clinically related phenotypes. The analysis was limited to phecodes with at least 200
527 cases to allow sufficient power for MR analysis (64). PheWAS was conducted using the
528 PheWAS package in R (65).

529 For each miRNA, associations for genetic variants residing in 500kb on either side of the
530 miRNA position (cis-SNPs) were extracted. The false discovery rate (FDR) was calculated
531 across all cis-SNPs for each miRNA, where those with $FDR < 0.1$ were selected as cis
532 instruments (18). Trans-SNPs associated with circulatory miRNA at $P < 5 \times 10^{-08}$ were added as
533 trans instruments in the sensitivity analysis. Instruments were filtered for $F\text{-statistics} > 10$ to
534 avoid weak instrument bias (66). Linkage disequilibrium (LD) clumping for the instruments was
535 conducted using a threshold of $r^2 < 0.1$ and a window of 10,000 kb.

536 For miRNA with single cis miR-eQTL satisfying the instrument criteria, cis-SNP was used as
537 the proxy for corresponding miRNA in single-variant PheWAS. For miRNAs with multiple
538 independent miR-eQTLs, weighted genetic risk scores (GRS) were computed for individuals
539 in the UK Biobank as the sum score of miRNA-increasing alleles of miR-eQTLs identified in
540 the Rotterdam Study using effect sizes as their weights as implemented using PLINK (36).
541 The weighted GRS was rescaled by subtracting GRS from its mean and dividing by its
542 standard deviation to express the association per-SD of the miRNA-increasing allele.

543 In the main analysis, GRS for each miRNA (miRNA-GRS) was computed from cis-miR-eQTLs
544 (cis-GRS). Additionally, trans-miR-eQTLs at genome-wide significant ($P < 5 \times 10^{-08}$) were added,
545 in an extended analysis, to validate findings from cis-GRS. Multiple logistic regression was
546 performed in the UK Biobank for each miRNA-GRS with adjustment for age, sex, genotyping
547 array, and the first five principal components to account for population stratification. Given
548 each phecode is not independent of the other, the false discovery rate (FDR) was calculated
549 for each miRNA-GRS to account for multiple testing (67).

550 [Mendelian randomisation](#)

551 Following PheWAS, two-sample Mendelian randomisation (MR) analysis was conducted to
552 assess the causal relationship between candidate miRNAs and outcomes of interest identified
553 from PheWAS. MR-PheWAS considered miRNAs with three or more independent instruments
554 to enable performing robust MR methods as sensitivity analysis. The same set of genetic
555 instruments used in PheWAS contributed to the exposure data in MR-PheWAS. The level of
556 each candidate miRNA was rescaled by subtracting the value from its mean and dividing by
557 its standard deviation (SD) to express the association per-SD increase of the miRNA level.
558 The genetic association between the instruments and the outcome was taken from the UK
559 Biobank. Associations that were significant at $FDR < 0.05$ from MR-PheWAS were taken

560 forward for validation using genetic association estimates (outcome data) from large GWAS
561 consortia or by adding genome-wide significant trans-miR-eQTLs in an extended-MR.

562 The multiplicative random effect inverse variance weighted method (IVW) was used in the
563 main analysis to combine the effect estimates of the genetic instruments assuming all
564 instruments are valid (68). FDR adjustment was calculated for each miRNA based on p-values
565 of MR-IVW since it was considered the most powerful method when all instruments are valid
566 (68). Robust MR methods which allow the inclusion of pleiotropic variants were used as a
567 sensitivity analysis, including weighted median (WM) or MR-Egger (69-71). WM estimate is
568 valid if less than half of the weight of the genetic instrument is free from horizontal pleiotropy.
569 MR-Egger does not force the regression line through an intercept of zero, making it statistically
570 inefficient but provides a causal estimate corrected for directional horizontal pleiotropy. A non-
571 null intercept in MR-Egger indicates evidence of pleiotropy (70). The agreement among
572 different MR methods was examined to support a robust estimation of causal effects.

573 Since a liberal LD threshold ($r^2 < 0.1$) was used for clumping, a further sensitivity analysis was
574 conducted by incorporating the correlation matrix between genetic instruments in the fixed
575 effect IVW method. MRPRESSO was used to detect outliers (72) and MR analysis was
576 repeated after excluding outliers. Results for MR analysis using different MR methods were
577 presented as forest plots. For replicated associations, reverse MR was conducted to assess
578 the directionality of associations. Independent genetic instruments for complex traits
579 ($r^2 < 0.001$) were identified from large GWAS consortia for the outcomes of interest.
580 Associations between the genetic instruments for complex traits with candidate miRNA levels
581 were extracted from the Rotterdam Study.

582 [Target genes and enrichment analysis](#)

583 To identify putative target genes for miRNAs, predicted and validated target genes of miRNAs
584 were retrieved from TargetScan v7.2 and miRTarBase (4,5). For enrichment analysis, genes
585 with predicted miRNA-target interaction (MTI) in TargetScan or validated MTI, including weak
586 and strong validation methods in miRTarBase, were considered. SNPs reaching genome-
587 wide significance in large consortia GWAS for significant traits were mapped into protein-
588 coding genes where they reside. Of those, predicted and validated target genes of miRNAs
589 were identified. The enrichment analysis was conducted to test if the target genes of candidate
590 miRNAs are enriched for the associated traits, as described in our previous work (34).
591 Enrichment analysis was performed separately for predicted and validated target genes.

592

593

594 **Acknowledgements**

595 We would like to thank all participants of the Rotterdam Study and the UK Biobank. The
596 Rotterdam Study is supported by the Erasmus MC University Medical Center and Erasmus
597 University Rotterdam; The Netherlands Organisation for Scientific Research (NWO); The
598 Netherlands Organisation for Health Research and Development (ZonMw); the Research
599 Institute for Diseases in the Elderly (RIDE); The Netherlands Genomics Initiative (NGI); the
600 Ministry of Education, Culture and Science; the Ministry of Health, Welfare and Sports; the
601 European Commission (DG XII); and the Municipality of Rotterdam. The UK Biobank has
602 approval from the North-West Multi-centre Research Ethics Committee (MREC) as a
603 Research Tissue Bank (RTB) approval. Explicit informed consent was obtained from all
604 participants when they enrolled in the UK Biobank. Access to the UK Biobank was provided
605 through application 52569. This work was enabled by the computing resources and support
606 from the Imperial College Research Computing Service and Erasmus MC. We thank Loukas
607 Zagkos for helping with the visualisation of the results.

608 RM is supported by the President's PhD Scholarship from Imperial College London. AD is
609 funded by a Wellcome Trust seed award (206046/Z/17/Z). This project is partly supported by
610 the Erasmus MC Fellowship (EMCF20213) and Alzheimer Nederland (WE.03-2021-10) grants
611 of MG. PE acknowledges support from the Medical Research Council (MR/S019669/1) for the
612 MRC Centre for Environment and Health, the British Heart Foundation (RE/18/4/34215) for
613 the Imperial BHF Centre for Research Excellence, the UK Dementia Research Institute
614 (MC_PC_17114) and the National Institute for Health Research Imperial College Biomedical
615 Research Centre for infrastructure support.

616

617 **Code availability**

618 For statistical analyses, we used the following software High-Dimensional Analysis
619 Framework (HASE) (<https://github.com/roshchupkin/hase>), R version 3.6.3 (<https://www.r-project.org>),
620 Plink v1.9 (<https://www.cog-genomics.org/plink/>), FUMA v1.4.1
621 (<https://fuma.ctglab.nl/>).

622

623

624

625 **References**

- 626 (1) Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*. 2004; 116
627 (2): 281-297.
- 628 (2) Bartel DP. MicroRNAs: target recognition and regulatory functions. *Cell*. 2009; 136 (2):
629 215-233.
- 630 (3) Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences to
631 function. *Nucleic acids research*. 2019; 47 (D1): D155-D162.
- 632 (4) Huang H, Lin Y, Li J, Huang K, Shrestha S, Hong H, et al. miRTarBase 2020: updates to
633 the experimentally validated microRNA–target interaction database. *Nucleic acids research*.
634 2020; 48 (D1): D148-D154.
- 635 (5) Agarwal V, Bell GW, Nam J, Bartel DP. Predicting effective microRNA target sites in
636 mammalian mRNAs. *elife*. 2015; 4 e05005.
- 637 (6) Grasedieck S, Sorrentino A, Langer C, Buske C, Döhner H, Mertens D, et al. Circulating
638 microRNAs in hematological diseases: principles, challenges, and perspectives. *Blood, The
639 Journal of the American Society of Hematology*. 2013; 121 (25): 4977-4984.
- 640 (7) Diener C, Keller A, Meese E. Emerging concepts of miRNA therapeutics: From cells to
641 clinic. *Trends in Genetics*. 2022; .
- 642 (8) Garcia-Martin R, Wang G, Brandão BB, Zanotto TM, Shah S, Kumar Patel S, et al.
643 MicroRNA sequence codes for small extracellular vesicle release and cellular retention.
644 *Nature*. 2022; 601 (7893): 446-451.
- 645 (9) Nikpay M, Beehler K, Valsesia A, Hager J, Harper M, Dent R, et al. Genome-wide
646 identification of circulating-miRNA expression quantitative trait loci reveals the role of several
647 miRNAs in the regulation of cardiometabolic phenotypes. *Cardiovascular research*. 2019;
648 115 (11): 1629-1645.
- 649 (10) Huan T, Rong J, Liu C, Zhang X, Tanriverdi K, Joehanes R, et al. Genome-wide
650 identification of microRNA expression quantitative trait loci. *Nature communications*. 2015; 6
651 (1): 1-9.
- 652 (11) Akiyama S, Higaki S, Ochiya T, Ozaki K, Niida S, Shigemizu D. JAMIR-eQTL: Japanese
653 genome-wide identification of microRNA expression quantitative trait loci across dementia
654 types. *Database*. 2021; 2021 (2021): baab072.
- 655 (12) Civelek M, Hagopian R, Pan C, Che N, Yang W, Kayne PS, et al. Genetic regulation of
656 human adipose microRNA expression and its consequences for metabolic traits. *Human
657 molecular genetics*. 2013; 22 (15): 3023-3037.
- 658 (13) Lappalainen T, Sammeth M, Friedländer MR, Ac't Hoen P, Monlong J, Rivas MA, et al.
659 Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*.
660 2013; 501 (7468): 506-511.
- 661 (14) Sonehara K, Sakaue S, Maeda Y, Hirata J, Kishikawa T, Yamamoto K, et al. Genetic
662 architecture of microRNA expression and its link to complex diseases in the Japanese
663 population. *Human molecular genetics*. 2021; .

- 664 (15) Brown RA, Epis MR, Horsham JL, Kabir TD, Richardson KL, Leedman PJ. Total RNA
665 extraction from tissues for microRNA and target gene expression analysis: not all kits are
666 created equal. *BMC biotechnology*. 2018; 18 (1): 1-11.
- 667 (16) Godoy PM, Barczak AJ, DeHoff P, Srinivasan S, Etheridge A, Galas D, et al.
668 Comparison of reproducibility, accuracy, sensitivity, and specificity of miRNA quantification
669 platforms. *Cell reports*. 2019; 29 (12): 4212-4222. e5.
- 670 (17) Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an
671 open access resource for identifying the causes of a wide range of complex diseases of
672 middle and old age. *Plos med*. 2015; 12 (3): e1001779.
- 673 (18) Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, Brown-Gentry K, et al.
674 PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene–disease
675 associations. *Bioinformatics*. 2010; 26 (9): 1205-1210.
- 676 (19) Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology
677 contribute to understanding environmental determinants of disease? *International journal of*
678 *epidemiology*. 2003; 32 (1): 1-22.
- 679 (20) Ikram MA, Brusselle G, Ghanbari M, Goedegebure A, Ikram MK, Kavousi M, et al.
680 Objectives, design and main findings until 2020 from the Rotterdam Study. *European journal*
681 *of epidemiology*. 2020; 1-35.
- 682 (21) Watanabe K, Taskesen E, Van Bochoven A, Posthuma D. Functional mapping and
683 annotation of genetic associations with FUMA. *Nature communications*. 2017; 8 (1): 1-11.
- 684 (22) Ge T, Nichols TE, Lee PH, Holmes AJ, Roffman JL, Buckner RL, et al. Massively
685 expedited genome-wide heritability analysis (MEGHA). *Proceedings of the National*
686 *Academy of Sciences*. 2015; 112 (8): 2479-2484.
- 687 (23) Vösa U, Claringbould A, Westra H, Bonder MJ, Deelen P, Zeng B, et al. Unraveling the
688 polygenic architecture of complex traits using blood eQTL metaanalysis. *BioRxiv*. 2018;
689 447367.
- 690 (24) Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, et al. Genomic
691 atlas of the human plasma proteome. *Nature*. 2018; 558 (7708): 73-79.
- 692 (25) Folkersen L, Gustafsson S, Wang Q, Hansen DH, Hedman ÅK, Schork A, et al.
693 Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals.
694 *Nature metabolism*. 2020; 2 (10): 1135-1148.
- 695 (26) Shin S, Fauman EB, Petersen A, Krumsiek J, Santos R, Huang J, et al. An atlas of
696 genetic influences on human blood metabolites. *Nature genetics*. 2014; 46 (6): 543-550.
- 697 (27) Kettunen J, Demirkan A, Würtz P, Draisma HH, Haller T, Rawal R, et al. Genome-wide
698 study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA.
699 *Nature communications*. 2016; 7 (1): 1-9.
- 700 (28) Elsworth B, Lyon M, Alexander T, Liu Y, Matthews P, Hallett J, et al. The MRC IEU
701 OpenGWAS data infrastructure. *BioRxiv*. 2020; .

- 702 (29) Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The
703 NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays
704 and summary statistics 2019. *Nucleic acids research*. 2019; 47 (D1): D1005-D1012.
- 705 (30) Nelson CP, Goel A, Butterworth AS, Kanoni S, Webb TR, Marouli E, et al. Association
706 analyses based on false discovery rate implicate new loci for coronary artery disease.
707 *Nature genetics*. 2017; 49 (9): 1385.
- 708 (31) Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, et al. Meta-
709 analysis of genome-wide association studies for height and body mass index in ~ 700000
710 individuals of European ancestry. *Human molecular genetics*. 2018; 27 (20): 3641-3649.
- 711 (32) Shungin D, Winkler TW, Croteau-Chonka DC, Ferreira T, Locke AE, Mägi R, et al. New
712 genetic loci link adipose and insulin biology to body fat distribution. *Nature*. 2015; 518
713 (7538): 187-196.
- 714 (33) Hsu S, Lin F, Wu W, Liang C, Huang W, Chan W, et al. miRTarBase: a database
715 curates experimentally validated microRNA–target interactions. *Nucleic acids research*.
716 2011; 39 (suppl_1): D163-D169.
- 717 (34) Mustafa R, Ghanbari M, Evangelou M, Dehghan A. An enrichment analysis for
718 cardiometabolic traits suggests non-random assignment of genes to microRNAs.
719 *International journal of molecular sciences*. 2018; 19 (11): 3666.
- 720 (35) Sakaue S, Hirata J, Maeda Y, Kawakami E, Nii T, Kishikawa T, et al. Integration of
721 genetics and miRNA–target gene network identified disease biology implicated in tissue
722 specificity. *Nucleic acids research*. 2018; 46 (22): 11898-11909.
- 723 (36) Melling GE, Flannery SE, Abidin SA, Clemmens H, Prajapati P, Hinsley EE, et al. A
724 miRNA-145/TGF- β 1 negative feedback loop regulates the cancer-associated fibroblast
725 phenotype. *Carcinogenesis*. 2018; 39 (6): 798-807.
- 726 (37) Aguda BD, Kim Y, Piper-Hunter MG, Friedman A, Marsh CB. MicroRNA regulation of a
727 cancer network: consequences of the feedback loops involving miR-17-92, E2F, and Myc.
728 *Proceedings of the National Academy of Sciences*. 2008; 105 (50): 19678-19683.
- 729 (38) Ozsolak F, Poling LL, Wang Z, Liu H, Liu XS, Roeder RG, et al. Chromatin structure
730 analyses identify miRNA promoters. *Genes & development*. 2008; 22 (22): 3172-3183.
- 731 (39) Borel C, Deutsch S, Letourneau A, Migliavacca E, Montgomery SB, Dimas AS, et al.
732 Identification of cis-and trans-regulatory variation modulating microRNA expression levels in
733 human fibroblasts. *Genome research*. 2011; 21 (1): 68-73.
- 734 (40) Gamazon ER, Ziliak D, Im HK, LaCroix B, Park DS, Cox NJ, et al. Genetic architecture
735 of microRNA expression: implications for the transcriptome and complex traits. *The*
736 *American Journal of Human Genetics*. 2012; 90 (6): 1046-1063.
- 737 (41) Cammaerts S, Strazisar M, De Rijk P, Del Favero J. Genetic variants in microRNA
738 genes: impact on microRNA expression, function, and disease. *Frontiers in Genetics*. 2015;
739 6 186.
- 740 (42) Ryan BM, Robles AI, Harris CC. Genetic variation in microRNA networks: the
741 implications for cancer research. *Nature Reviews Cancer*. 2010; 10 (6): 389-402.

- 742 (43) Goulart LF, Bettella F, Sønderby IE, Schork AJ, Thompson WK, Mattingsdal M, et al.
743 MicroRNAs enrichment in GWAS of complex human phenotypes. *BMC genomics*. 2015; 16
744 (1): 1-10.
- 745 (44) Kaczkowski B, Torarinsson E, Reiche K, Havgaard JH, Stadler PF, Gorodkin J.
746 Structural profiles of human miRNA families from pairwise clustering. *Bioinformatics*. 2009;
747 25 (3): 291-294.
- 748 (45) Somel M, Guo S, Fu N, Yan Z, Hu HY, Xu Y, et al. MicroRNA, mRNA, and protein
749 expression link development and aging in human and macaque brain. *Genome research*.
750 2010; 20 (9): 1207-1218.
- 751 (46) Duell EJ, Lujan-Barroso L, Sala N, Deitz McElyea S, Overvad K, Tjonneland A, et al.
752 Plasma microRNAs as biomarkers of pancreatic cancer risk in a prospective cohort study.
753 *International journal of cancer*. 2017; 141 (5): 905-915.
- 754 (47) Mens MM, Maas SC, Klap J, Weverling GJ, Klatser P, Brakenhoff JP, et al. Multi-omics
755 analysis reveals microRNAs associated with cardiometabolic traits. *Frontiers in genetics*.
756 2020; 11 110.
- 757 (48) Duisters RF, Tijssen AJ, Schroen B, Leenders JJ, Lentink V, van der Made I, et al. miR-
758 133 and miR-30 regulate connective tissue growth factor: implications for a role of
759 microRNAs in myocardial matrix remodeling. *Circulation research*. 2009; 104 (2): 170-178.
- 760 (49) Backes C, Kehl T, Stöckel D, Fehlmann T, Schneider L, Meese E, et al. miRPathDB: a
761 new dictionary on microRNAs and target pathways. *Nucleic acids research*. 2016; gkw926.
- 762 (50) Mantilla-Escalante DC, López de las Hazas, María-Carmen, Gil-Zamorano J, del Pozo-
763 Acebo L, Crespo MC, Martín-Hernández R, et al. Postprandial circulating miRNAs in
764 response to a dietary fat challenge. *Nutrients*. 2019; 11 (6): 1326.
- 765 (51) Li Y, Xiao L, Li J, Sun P, Shang L, Zhang J, et al. MicroRNA profiling of diabetic
766 atherosclerosis in a rat model. *European journal of medical research*. 2018; 23 (1): 1-10.
- 767 (52) Mo Y, Fang R, Wu J, Si Y, Jia S, Li Q, et al. MicroRNA-329 upregulation impairs the
768 HMGB2/ β -catenin pathway and regulates cell biological behaviors in melanoma. *Journal of*
769 *cellular physiology*. 2019; 234 (12): 23518-23527.
- 770 (53) Vösa U, Claringbould A, Westra H, Bonder MJ, Deelen P, Zeng B, et al. Large-scale
771 cis-and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that
772 regulate blood gene expression. *Nature genetics*. 2021; 1-11.
- 773 (54) Westra H, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al.
774 Systematic identification of trans eQTLs as putative drivers of known disease associations.
775 *Nature genetics*. 2013; 45 (10): 1238-1243.
- 776 (55) Yao C, Joehanes R, Johnson AD, Huan T, Liu C, Freedman JE, et al. Dynamic role of
777 trans regulation of gene expression in relation to complex traits. *The American Journal of*
778 *Human Genetics*. 2017; 100 (4): 571-580.
- 779 (56) GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature*.
780 2017; 550 (7675): 204-213.

- 781 (57) Shah R, Tanriverdi K, Levy D, Larson M, Gerstein M, Mick E, et al. Discordant
782 expression of circulating microRNA from cellular and extracellular sources. *PLoS one*. 2016;
783 11 (4): e0153691.
- 784 (58) Roshchupkin GV, Adams H, Vernooij MW, Hofman A, Van Duijn CM, Ikram MA, et al.
785 HASE: Framework for efficient high-dimensional association analyses. *Scientific reports*.
786 2016; 6 36076.
- 787 (59) Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-
788 specific haplotype structure and linking correlated alleles of possible functional variants.
789 *Bioinformatics*. 2015; 31 (21): 3555-3557.
- 790 (60) Gong J, Liu C, Liu W, Wu Y, Ma Z, Chen H, et al. An update of miRNASNP database
791 for better SNP selection by GWAS data, miRNA expression and online tools. *Database*.
792 2015; 2015 .
- 793 (61) MHC Sequencing Consortium. Complete sequence and gene map of a human major
794 histocompatibility complex. *Nature*. 1999; 401 (6756): 921-923.
- 795 (62) Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex
796 trait analysis. *The American Journal of Human Genetics*. 2011; 88 (1): 76-82.
- 797 (63) Hinske LC, Franca GS, Torres HA, Ohara DT, Lopes-Ramos CM, Heyn J, et al.
798 miRIAD—integrating microRNA inter-and intragenic data. *Database*. 2014; 2014 .
- 799 (64) Verma A, Bradford Y, Dudek S, Lucas AM, Verma SS, Pendergrass SA, et al. A
800 simulation study investigating power estimates in phenome-wide association studies. *BMC*
801 *bioinformatics*. 2018; 19 (1): 120.
- 802 (65) Carroll RJ, Bastarache L, Denny JC. R PheWAS: data analysis and plotting tools for
803 phenome-wide association studies in the R environment. *Bioinformatics*. 2014; 30 (16):
804 2375-2376.
- 805 (66) Burgess S, Thompson SG, CRP CHD Genetics Collaboration. Avoiding bias from weak
806 instruments in Mendelian randomization studies. *International journal of epidemiology*. 2011;
807 40 (3): 755-764.
- 808 (67) Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful
809 approach to multiple testing. *Journal of the Royal statistical society: series B*
810 *(Methodological)*. 1995; 57 (1): 289-300.
- 811 (68) Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with
812 multiple genetic variants using summarized data. *Genetic epidemiology*. 2013; 37 (7): 658-
813 665.
- 814 (69) Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using
815 the MR-Egger method. *European journal of epidemiology*. 2017; 32 (5): 377-389.
- 816 (70) Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid
817 instruments: effect estimation and bias detection through Egger regression. *International*
818 *journal of epidemiology*. 2015; 44 (2): 512-525.

819 (71) Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian
820 randomization with some invalid instruments using a weighted median estimator. *Genetic*
821 *epidemiology*. 2016; 40 (4): 304-314.

822 (72) Verbanck M, Chen C, Neale B, Do R. Detection of widespread horizontal pleiotropy in
823 causal relationships inferred from Mendelian randomization between complex traits and
824 diseases. *Nature genetics*. 2018; 50 (5): 693-698.

825

Tables

Table 1. The results of Mendelian randomisation (MR-IVW) for replicated associations.

exposure	MR-PheWAS						Replication							
	outcome	n	beta	SE	P	P-het	Egger int P	outcome	n	beta	SE	P	P-het	Egger int P
Replication in extended-MR														
miR-30d-5p	angina pectoris	5	-0.47	0.10	1.5×10 ⁻⁶	3.4×10 ⁻¹	2.9×10 ⁻¹	angina pectoris	6	-0.34	0.12	6.0×10 ⁻³	3.1×10 ⁻²	2.7×10 ⁻¹
	coronary atherosclerosis	5	-0.39	0.09	7.6×10 ⁻⁶	3.6×10 ⁻¹	2.5×10 ⁻¹	coronary atherosclerosis	6	-0.26	0.11	1.8×10 ⁻²	2.8×10 ⁻²	2.4×10 ⁻¹
	nephrotic syndrome	5	1.91	0.41	3.9×10 ⁻⁶	6.1×10 ⁻¹	5.7×10 ⁻¹	nephrotic syndrome	6	1.32	0.50	8.9×10 ⁻³	7.0×10 ⁻²	9.9×10 ⁻¹
	nonspecific chest pain	5	-3.07	0.56	4.9×10 ⁻⁸	7.9×10 ⁻¹	9.2×10 ⁻¹	nonspecific chest pain	6	-2.03	0.82	1.4×10 ⁻²	1.2×10 ⁻²	6.5×10 ⁻¹
miR-323b-3p	melanomas of skin	6	0.45	0.08	1.4×10 ⁻⁷	8.8×10 ⁻¹	5.7×10 ⁻¹	melanomas of skin	7	0.31	0.11	4.7×10 ⁻³	4.3×10 ⁻²	8.1×10 ⁻¹
	obesity	6	-0.16	0.03	6.6×10 ⁻⁶	7.7×10 ⁻¹	7.1×10 ⁻¹	obesity	7	-0.11	0.05	1.9×10 ⁻²	4.6×10 ⁻²	8.7×10 ⁻¹
	overweight or obesity	6	-0.15	0.03	8.5×10 ⁻⁶	7.1×10 ⁻¹	6.9×10 ⁻¹	overweight or obesity	7	-0.11	0.04	1.7×10 ⁻²	5.6×10 ⁻²	8.6×10 ⁻¹
	skin cancer	6	0.26	0.05	2.2×10 ⁻⁷	2.9×10 ⁻¹	9.6×10 ⁻¹	skin cancer	7	0.21	0.06	3.4×10 ⁻⁴	5.2×10 ⁻²	9.0×10 ⁻¹
	viral enteritis	6	0.72	0.15	1.7×10 ⁻⁶	8.7×10 ⁻¹	7.3×10 ⁻¹	viral enteritis	7	0.61	0.13	3.2×10 ⁻⁶	6.7×10 ⁻¹	6.6×10 ⁻¹
miR-409-3p	melanomas of skin	12	0.25	0.05	7.7×10 ⁻⁶	1.1×10 ⁻¹	5.0×10 ⁻¹	melanomas of skin	13	0.24	0.05	3.3×10 ⁻⁶	1.4×10 ⁻¹	4.7×10 ⁻¹
Replication using GWAS summary statistics														
miR-329-3p	overweight or obesity	10	-0.15	0.03	1.5×10 ⁻⁷	3.2×10 ⁻¹	2.4×10 ⁻¹	BMI	6	-0.03	0.01	1.9×10 ⁻²	4.9×10 ⁻⁷	9.9×10 ⁻²
	obesity	10	-0.15	0.03	3.9×10 ⁻⁸	3.5×10 ⁻¹	2.4×10 ⁻¹							
miR-543	overweight or obesity	8	-0.15	0.03	9.5×10 ⁻⁹	4.7×10 ⁻¹	8.4×10 ⁻¹	WHR	7	-0.02	0.01	1.7×10 ⁻²	7.7×10 ⁻¹	9.0×10 ⁻¹
	obesity	8	-0.15	0.03	1.1×10 ⁻⁸	4.9×10 ⁻¹	8.9×10 ⁻¹							

BMI: body mass index. WHR: waist-to-hip ratio. n is the number of genetic instruments used in the analysis. SE: standard error. P-het denotes P-value for heterogeneity of MR-IVW estimates. Egger int P: P-values for MR Egger intercept. The summary statistics presented are based on MR-IVW. Full results for other MR methods are presented in Supplementary Tables 25 and 26.

Figure Legends

Figure 1. Overview of the study workflow.

Figure 2. a. Fujiplot of identified and replicated miR-eQTLs mapped into 22 genomic loci. Each circular layer represents miRNA, and dots represent SNP-miRNA associations. The dots that form a radial pattern indicate loci associated with multiple miRNAs. The innermost histogram shows the number of miR-eQTLs identified in each locus. ** in chromosome 5 represents one locus mapped to *CTC-5298P.1*. The top-left inset is a barplot to show the number of independent cis and trans-miR-eQTLs in pleiotropic loci, each regulating the level of multiple miRNAs. Highly pleiotropic loci were identified in locus chr14:100655022-101244293, regulating 23 miRNAs, the majority of which were cis-miR-eQTLs. The locus in chr9:136128546-136296530 consisted of trans-miR-eQTLs and regulated 18 miRNAs. **b. Functional consequences of identified miR-eQTLs on nearby and far genes.** **c. SNP-based heritability estimates distribution for 2,083 miRNAs.**

Figure 3. a. Overlap between cis miR-eQTLs and proteins (pQTLs). The bottom half of the circle shows miRNAs in different colours, and the top half of the circle (grey coloured) shows the genes. **b. Overlap between trans-miR-eQTLs and proteins (pQTLs).** Trans-miR-eQTLs are shown to be more pleiotropic than cis-miR-eQTLs.

Figure 4. a. Summary of PheWAS and MR. Cis-miR-eQTLs were used as proxies for miRNAs. When multiple cis-miR-eQTLs were available, cis-GRS was computed for PheWAS. Otherwise, single variant PheWAS was conducted. MR were conducted for miRNA with at least three instruments. When available, large GWAS data were used to replicate the findings. Otherwise, genome-wide trans-miR-eQTLs were added in the extended-MR. **b.** Number of cases available within each disease group. The figure corresponds to clinical diagnoses with at least 200 cases. **c. Enhanced volcano plots for single variant PheWAS and d. GRS PheWAS.** The X-axis denotes effect estimates for corresponding SNP or GRS. Y-axis indicates $-\log_{10}$ of the association p-values between each SNP or GRS and clinical condition. Different colours of the dots represent different SNPs. Different shapes show different disease groups. Thresholds of significance are indicated by dashed blue (nominal), red (FDR), and purple (Bonferroni) lines. Plots were only created for SNP and GRS with at least one FDR-significant finding.

Figure 5. a. Number of miRNAs associated with diagnoses in each disease group as identified in PheWAS. **b. Schematic network showing miRNAs and disease groups associations.** Each line corresponds to association between miRNA and clinical diagnosis that belong to a particular disease group. The colour of the circles indicates miRNA (orange) or disease groups (green). **c. Forest plots for 24 associations in MR-PheWAS with no genome-wide significant trans-miR-eQTLs.** Different colours correspond to different MR methods, as labelled.

Figure 1

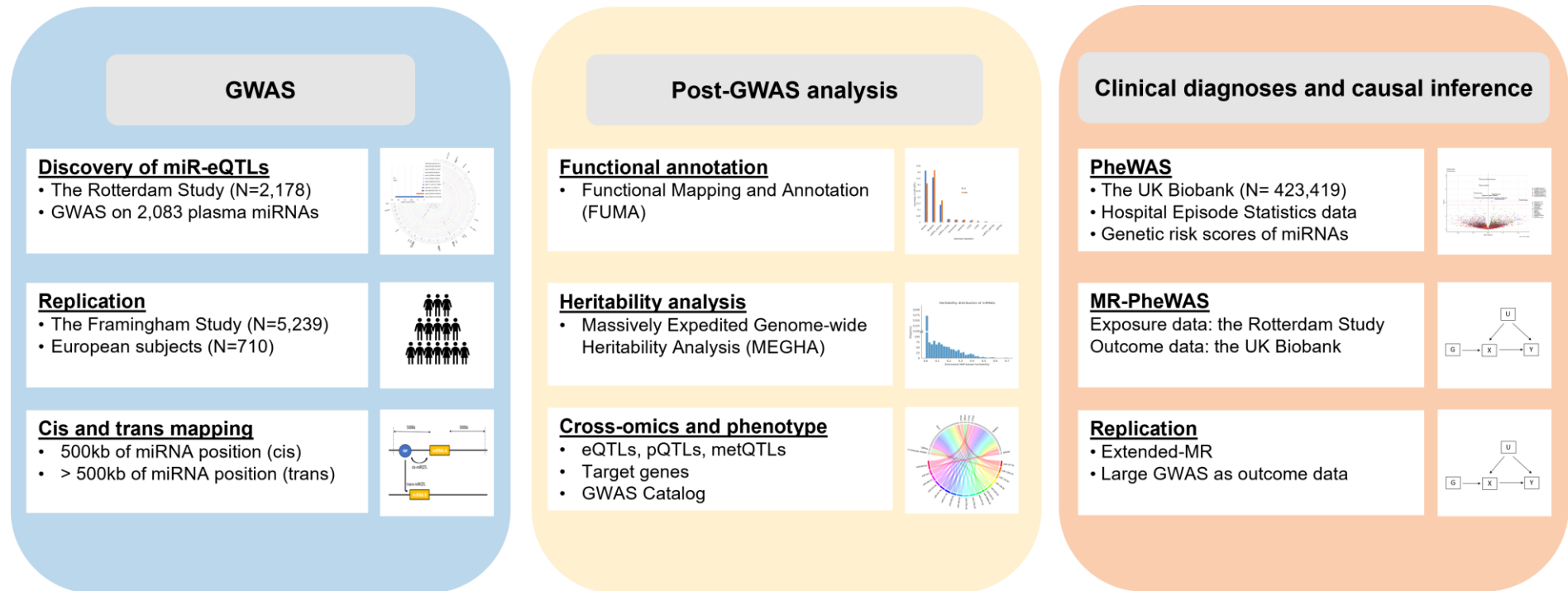
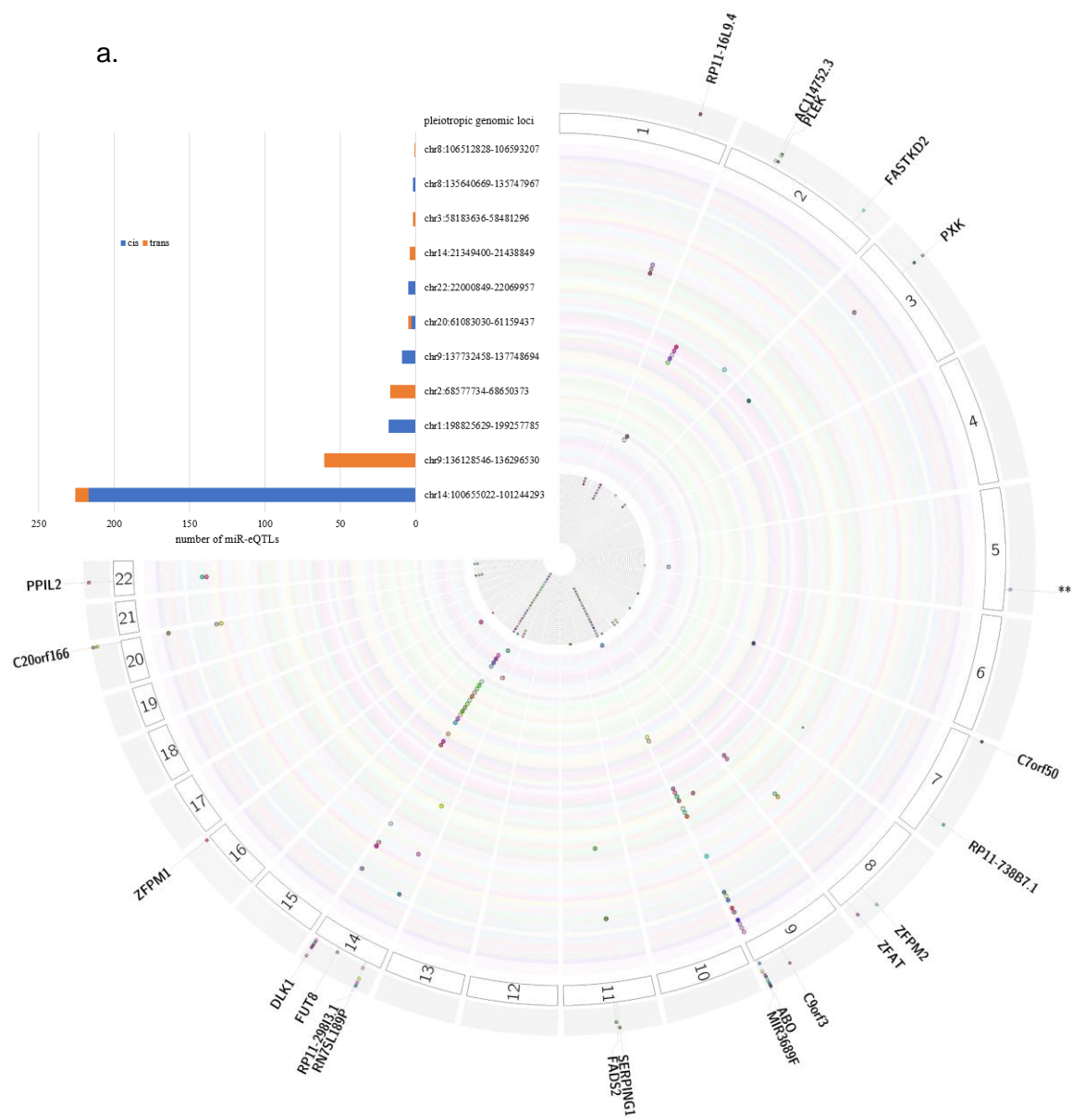
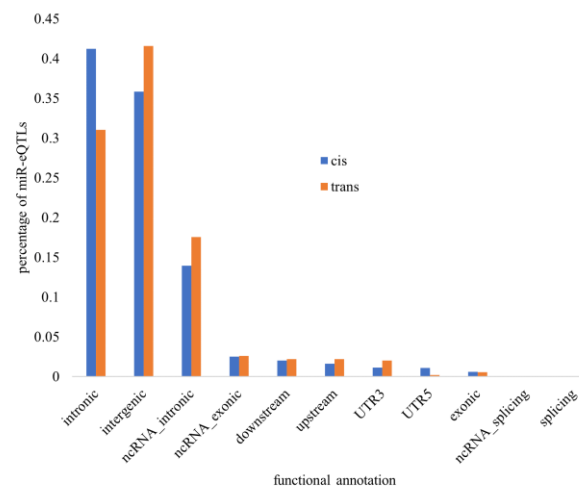


Figure 2

a.



b.



c.

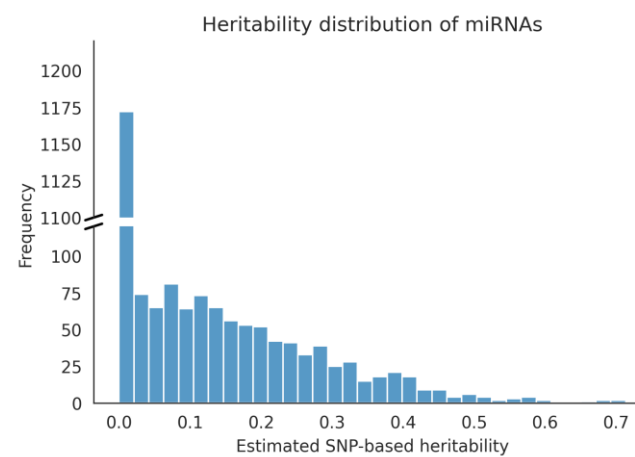


Figure 3

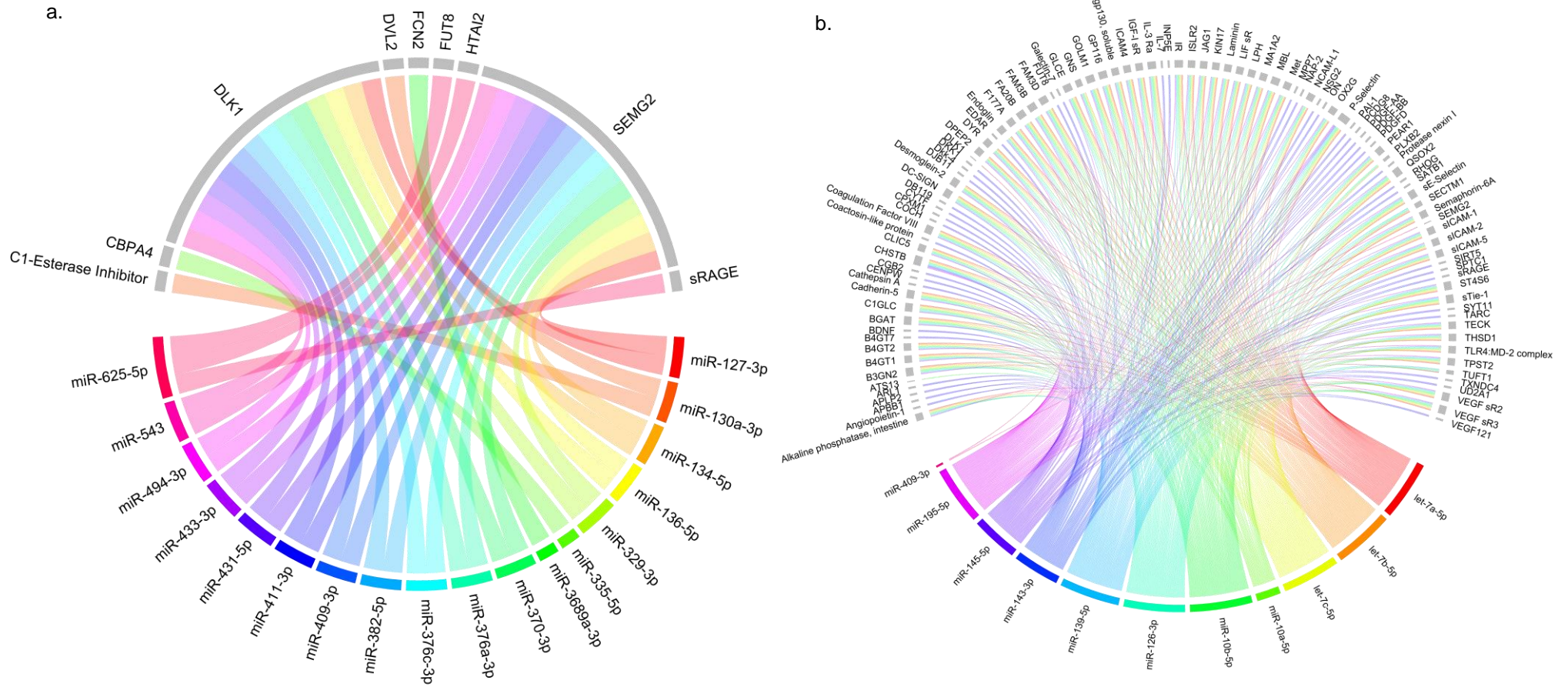
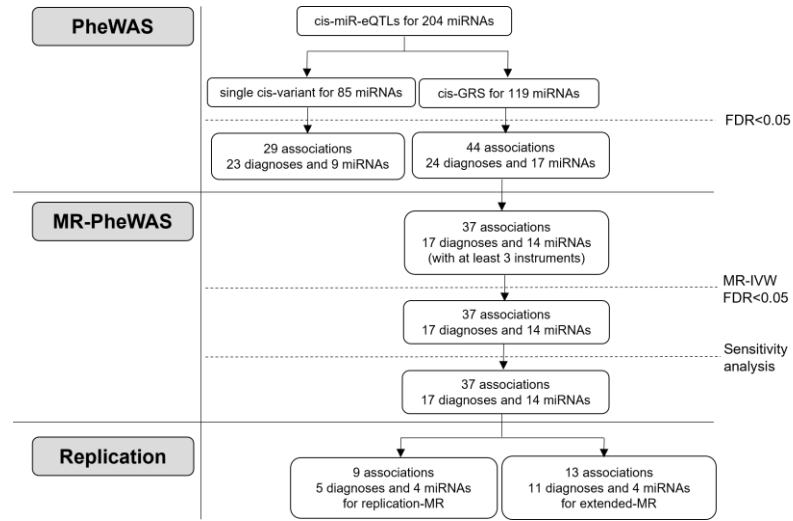
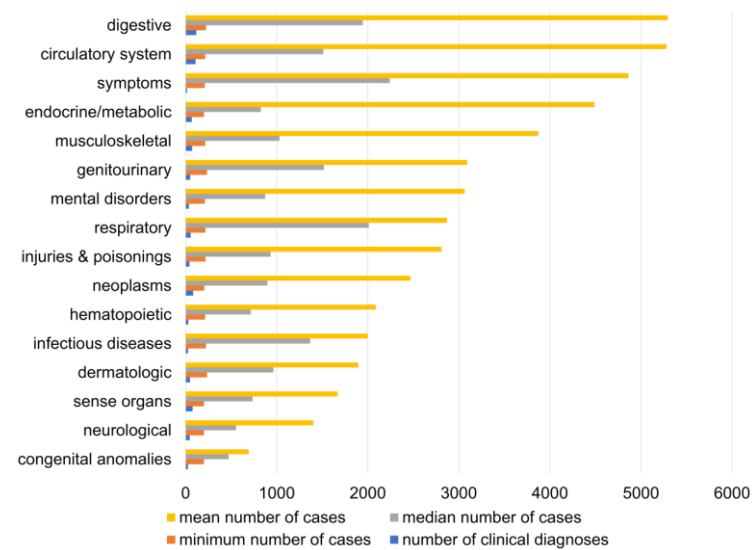


Figure 4

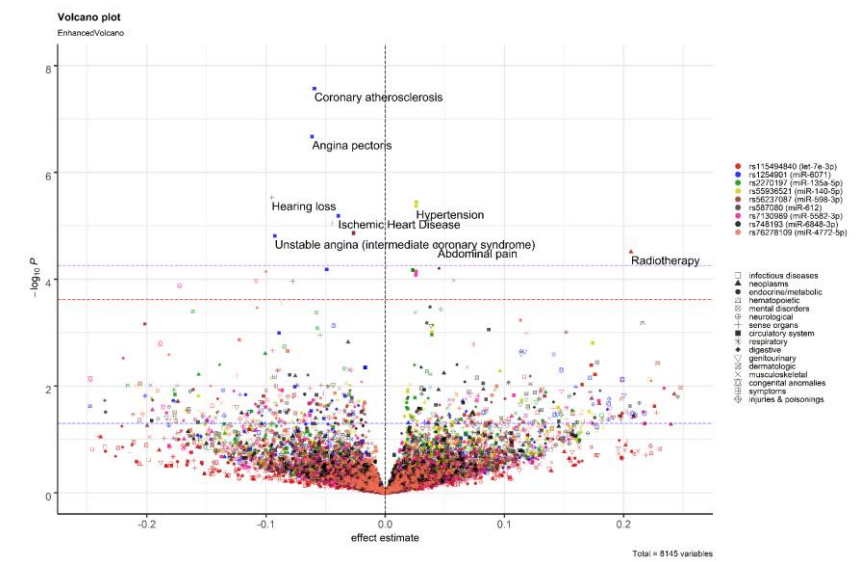
a.



b.



c.



d.

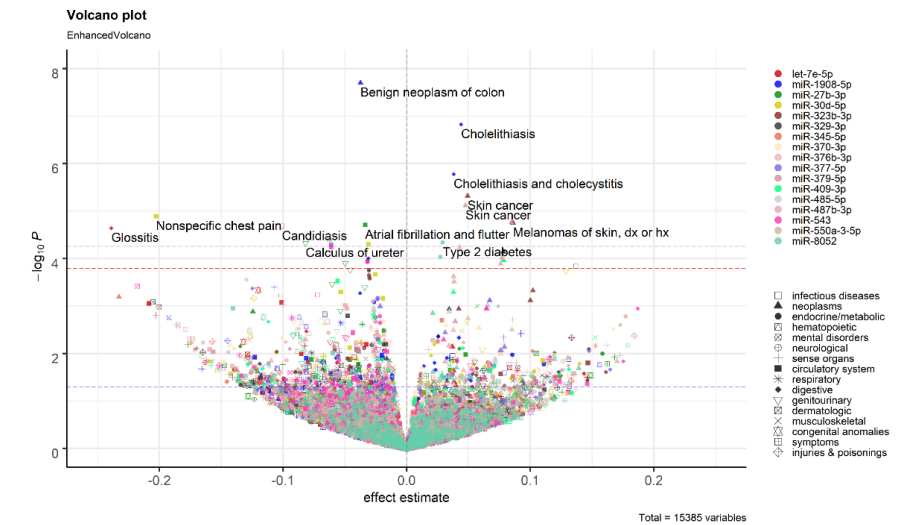
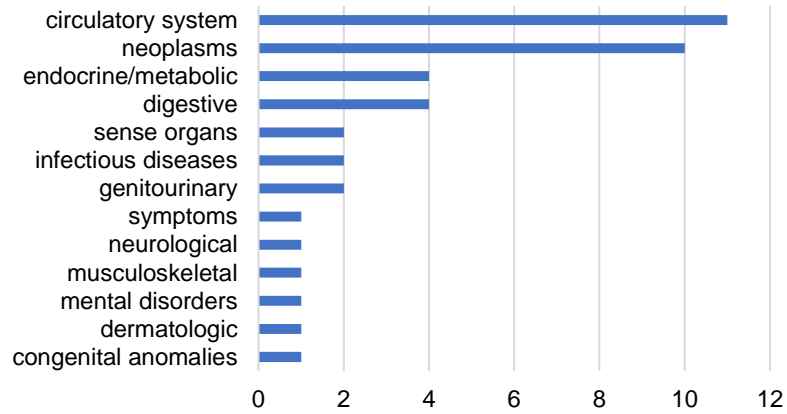


Figure 5

a.

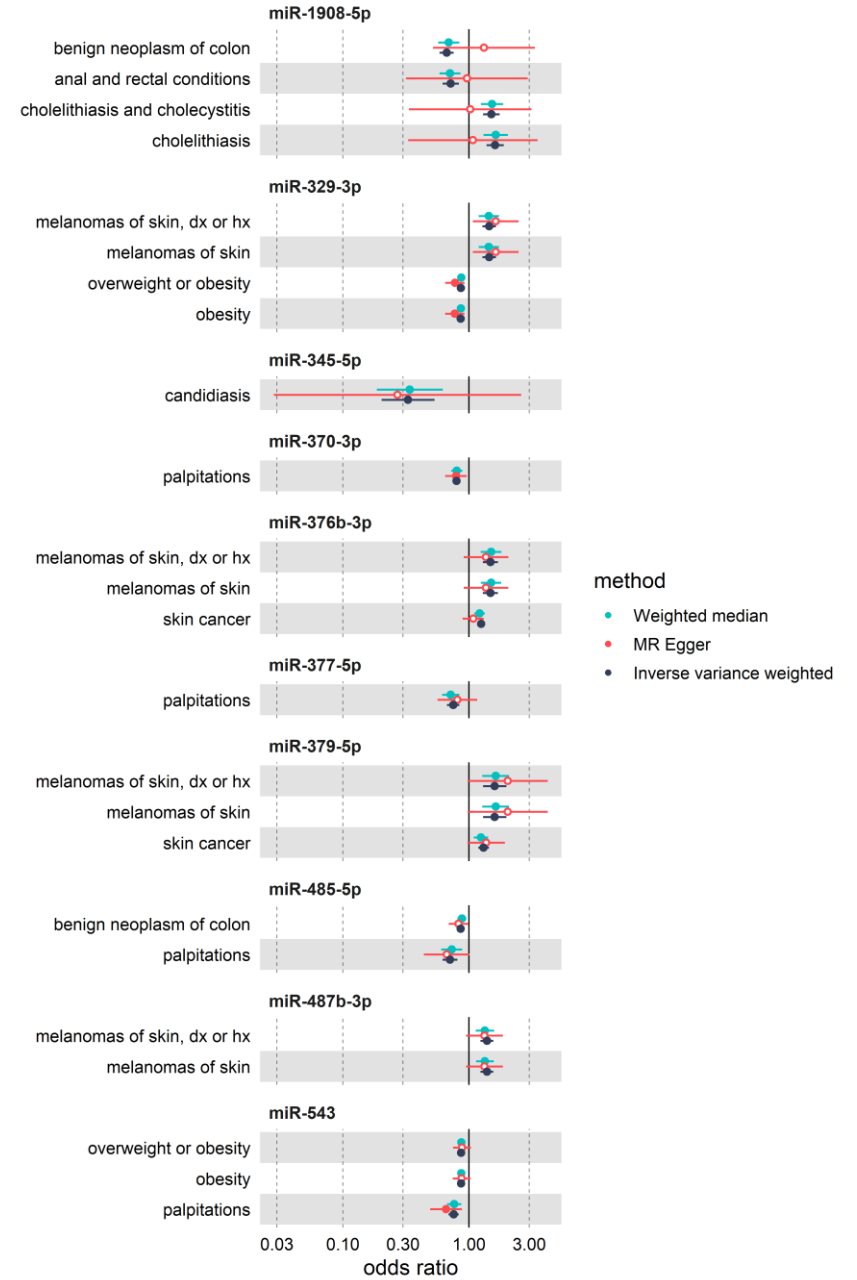
Number of miRNAs associated with diagnoses in each disease group



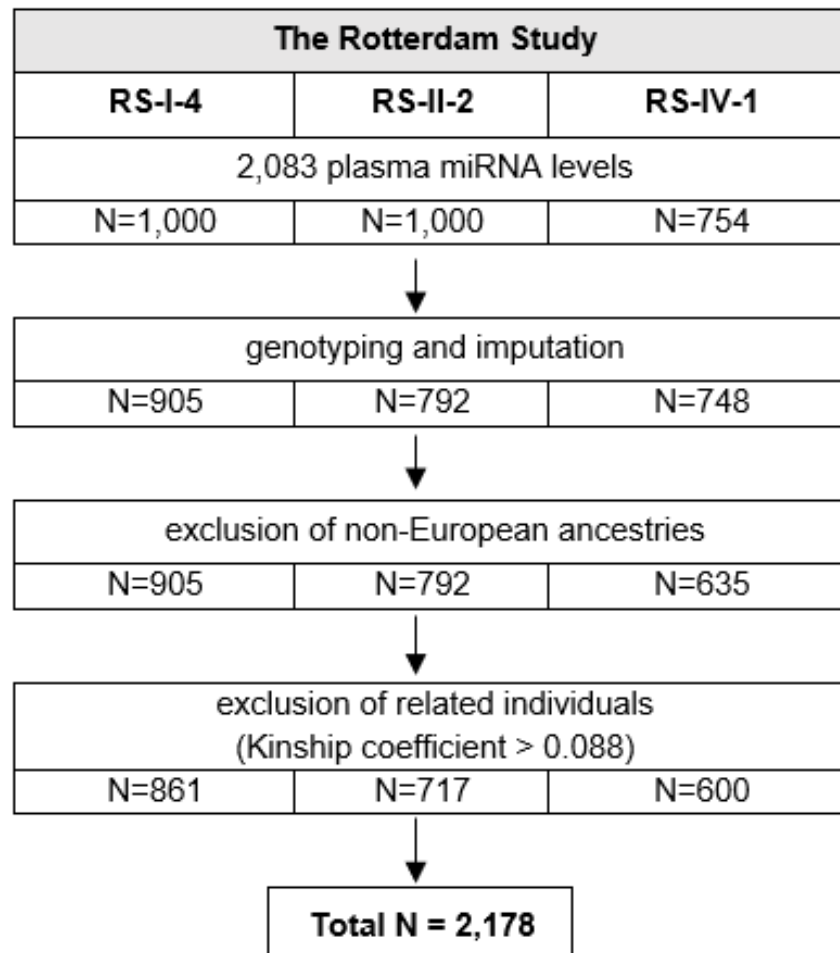
b.



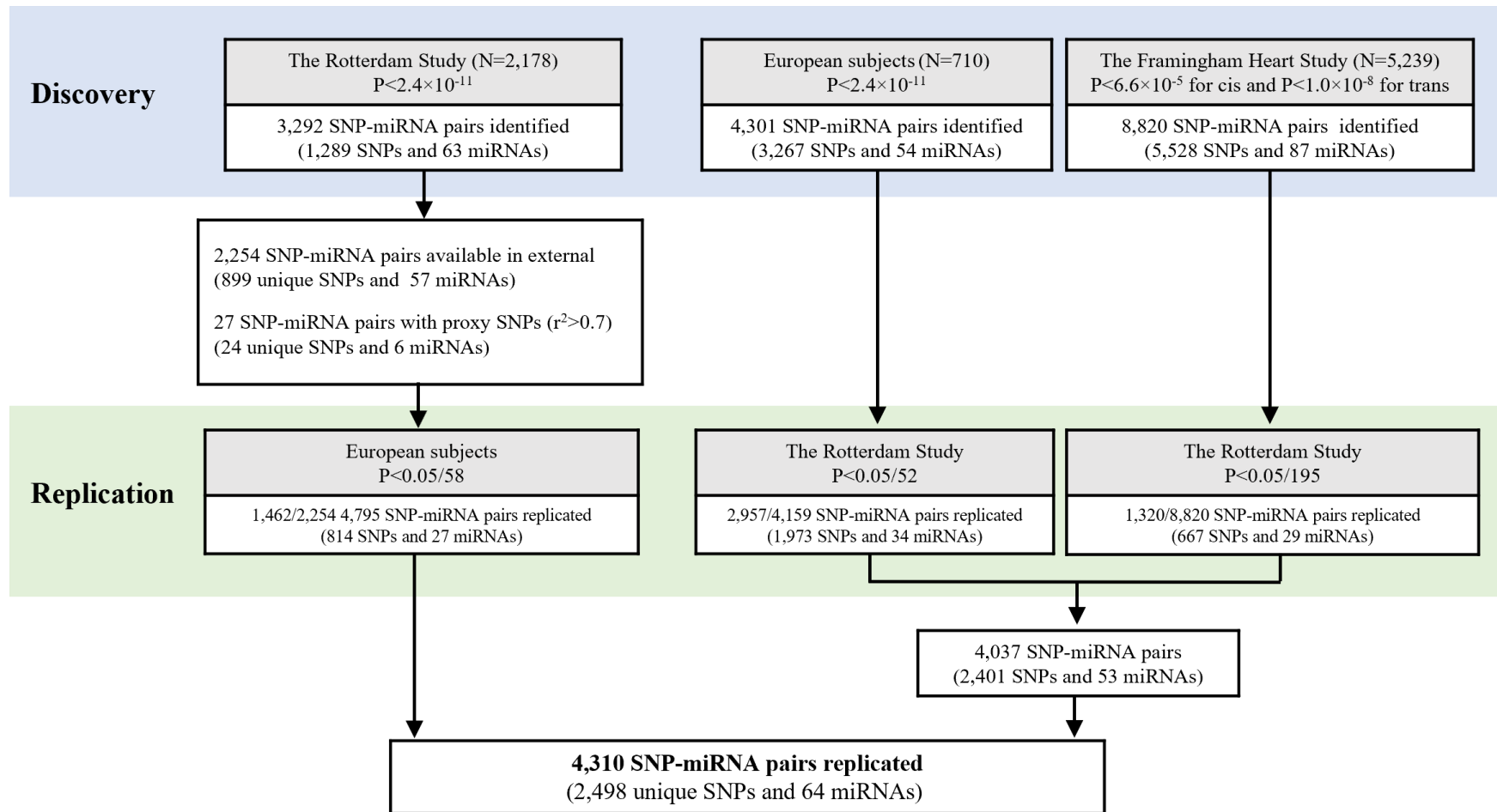
c.



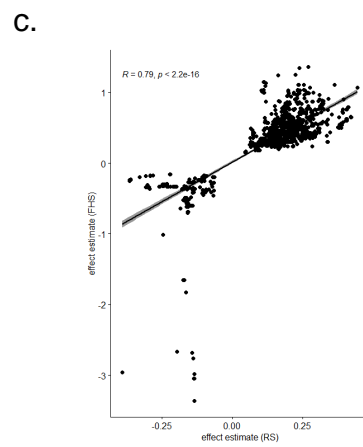
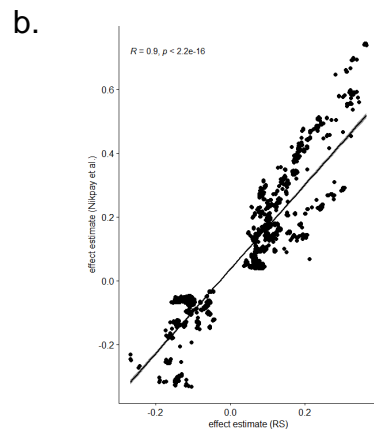
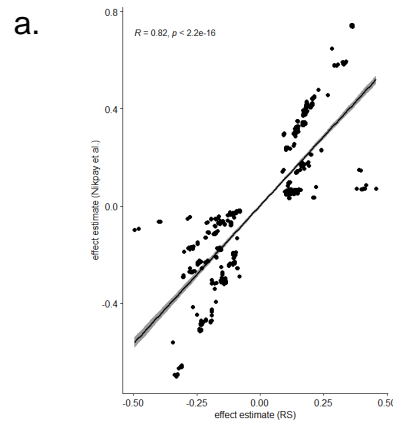
Extended Data Figures



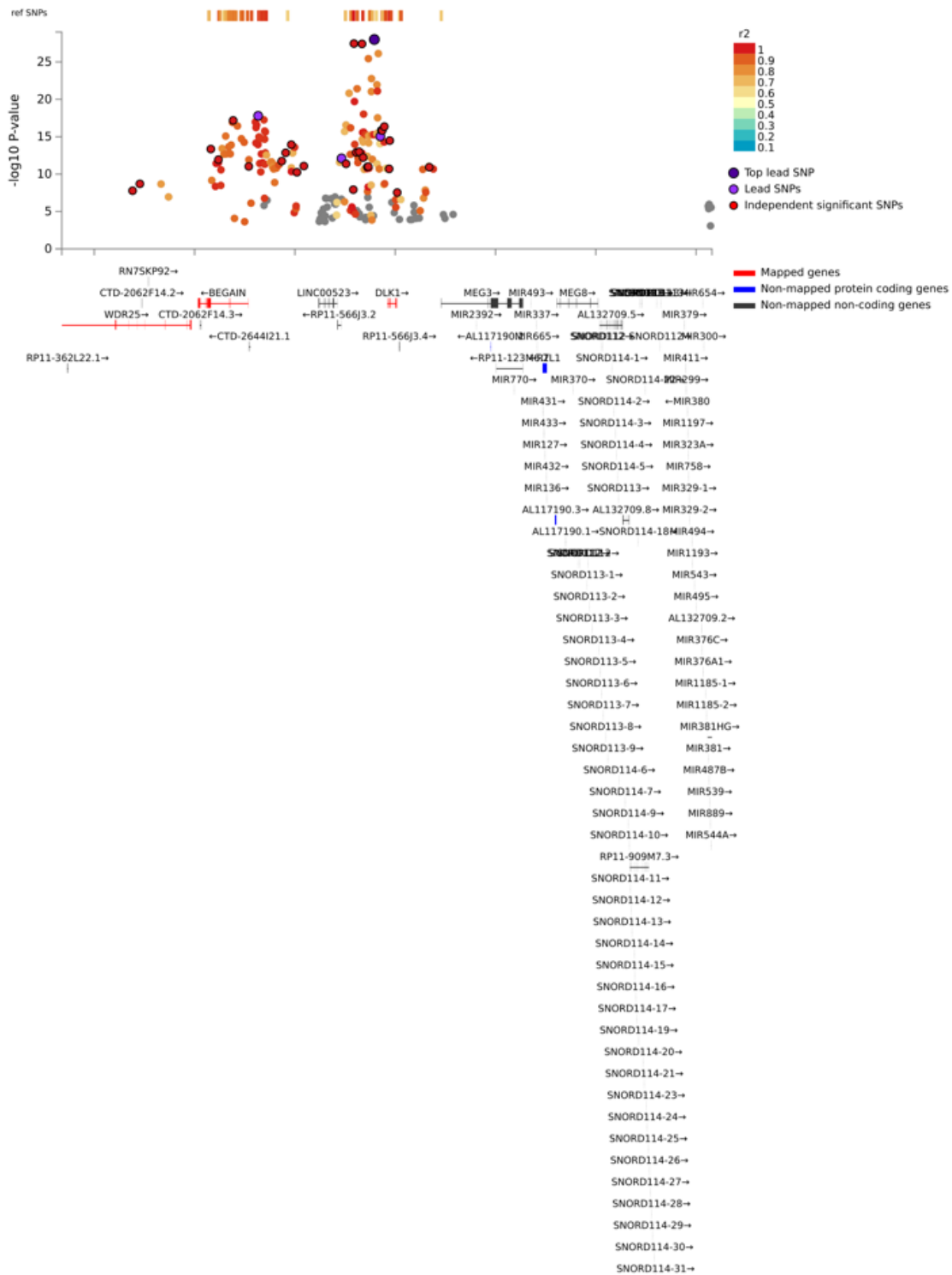
Extended Data Fig. 1. Selection of study participants in the Rotterdam Study.



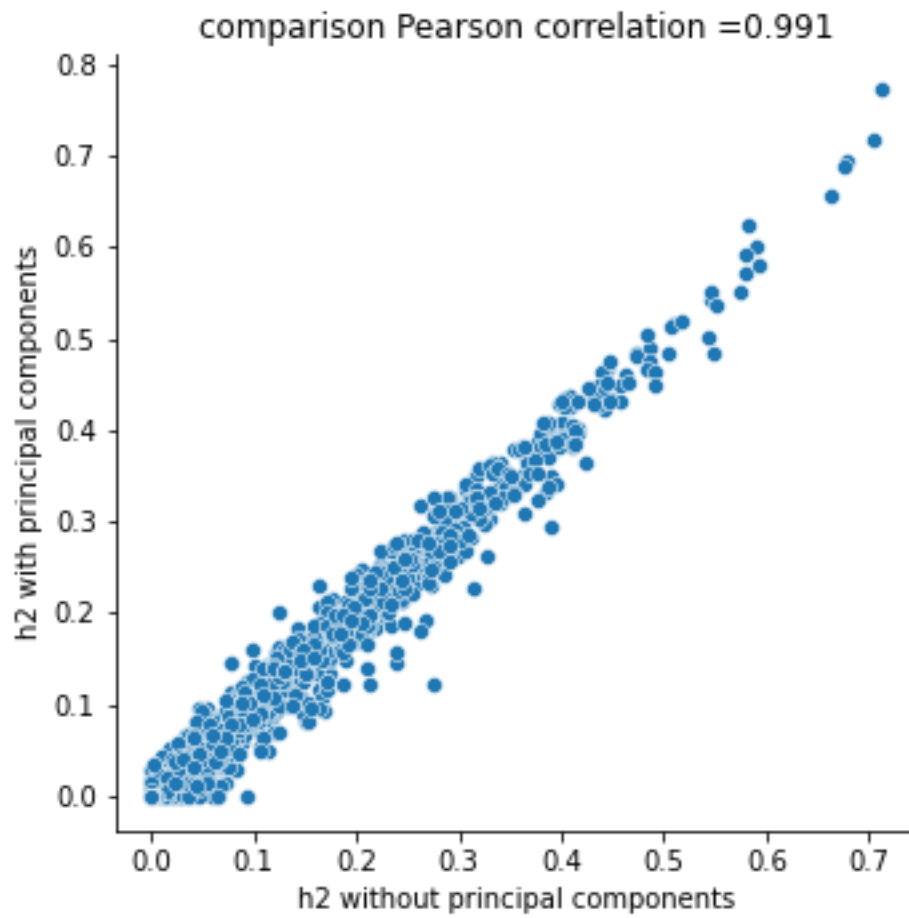
Extended Data Fig. 2. Identification of miR-eQTLs and replication in independent cohorts. In total, 3,292 significant associations were discovered for 63 miRNAs. Of those 1,462 out of 2,254 associations available in Nikpay et al. (9) were replicated for 27 miRNAs (P < 0.05/58). On the other hand, 2,957 associations identified by Nikpay et al. (9) for 34 miRNAs were replicated (P < 0.05/52) and 1,320 associations for 29 miRNAs identified in the Framingham Heart Study (10) were also replicated (P < 0.05/195). Collectively, 4,310 associations for 64 miRNAs were successfully replicated across studies.



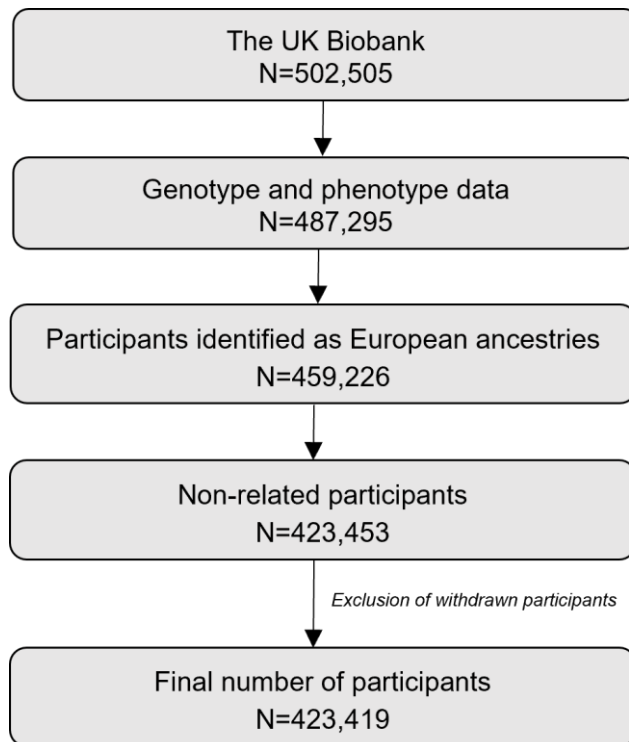
Extended Data Fig. 3. Correlation of effect estimates between discovery and replication of miR-eQTLs. a. Associations in the Rotterdam Study that were replicated in Nikpay et al.(9). b. Associations in Nikpay et al. (9) that were replicated in the Rotterdam Study. c. Associations in the Framingham Study that were replicated in the Rotterdam Study.



Extended Data Fig. 4. Regional plot for genomic risk loci in chr 14:100655022-101244293 harbouring cis-miR-eQTLs for 31 miRNAs that are clustered together. Plot was extracted from FUMA.

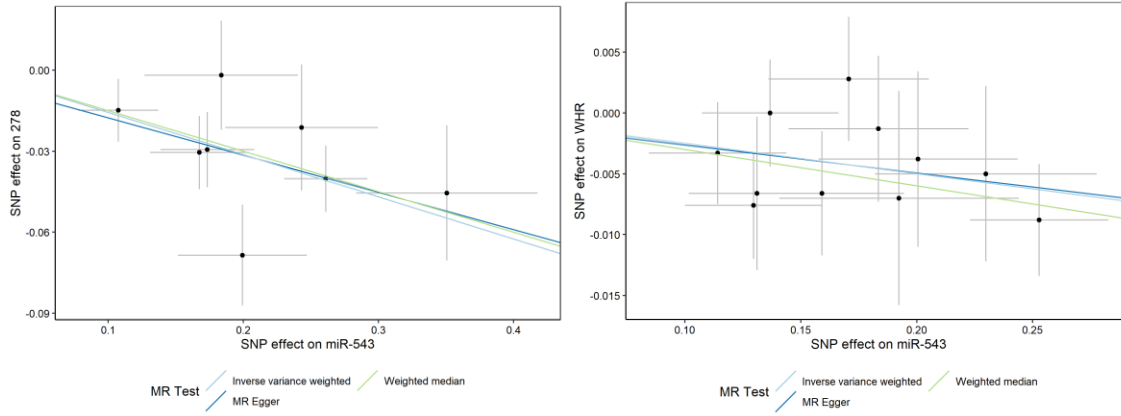


Extended Data Fig. 5. Pearson correlation of heritability estimates with and without principal components.

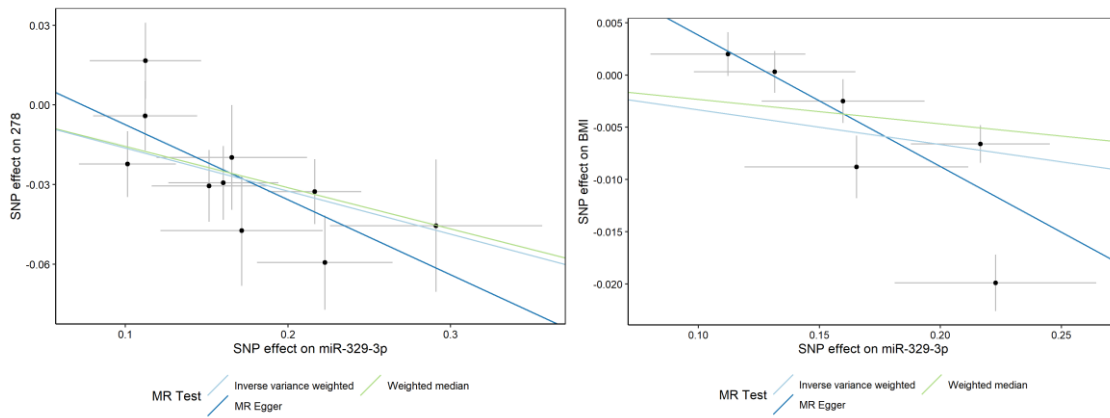


Extended Data Fig. 6. Selection of participants for PheWAS and MR-PheWAS in the UK Biobank.

a.



b.



Extended Data Fig. 7. Scatter plots for MR-PheWAS and replication MR (bottom). a. miR-543 and obesity (left) and waist to hip ratio (right). b. miR-329-3p and obesity (left) and body mass index (right).