

Integration of genetics and miRNA–target gene network identified disease biology implicated in tissue specificity

Saori Sakaue^{1,2,3}, Jun Hirata^{1,4}, Yuichi Maeda^{5,6,7}, Eiryō Kawakami^{8,9}, Takuro Nii^{5,6,7}, Toshihiro Kishikawa^{1,10}, Kazuyoshi Ishigaki², Chikashi Terao^{11,12,13}, Ken Suzuki^{1,2}, Masato Akiyama^{2,14}, Naomasa Suita^{1,15}, Tatsuo Masuda^{1,16}, Kotaro Ogawa^{1,17}, Kenichi Yamamoto^{1,18}, Yukihiko Saeki¹⁹, Masato Matsushita^{20,21}, Maiko Yoshimura²¹, Hidetoshi Matsuoka²¹, Katsunori Ikari^{22,23}, Atsuo Taniguchi²², Hisashi Yamanaka²², Hideya Kawaji^{24,25,26,27}, Timo Lassmann^{24,25,28}, Masayoshi Itoh^{24,25,26}, Hiroyuki Yoshitomi^{29,30}, Hiromu Ito³⁰, Koichiro Ohmura¹³, Alistair R R Forrest^{24,25,31}, Yoshihide Hayashizaki^{25,26}, Piero Carninci^{24,25,32}, Atsushi Kumanogoh⁵, Yoichiro Kamatani^{2,33}, Michiel de Hoon^{24,25,34}, Kazuhiko Yamamoto³⁵ and Yukinori Okada^{1,2,36,*}

¹Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita 565-0871, Japan, ²Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama 230-0045, Japan, ³Department of Allergy and Rheumatology, Graduate School of Medicine, the University of Tokyo, Tokyo 113-8655, Japan, ⁴Pharmaceutical Discovery Research Laboratories, TEIJIN PHARMA LIMITED, Hino 191-8512, Japan, ⁵Department of Respiratory Medicine and Clinical Immunology, Osaka University Graduate School of Medicine, WPI Immunology Frontier Research Center, Osaka University, Suita 565-0871, Japan, ⁶Laboratory of Immune Regulation, Department of Microbiology and Immunology, Graduate School of Medicine, WPI Immunology Frontier Research Center, Osaka University, Suita 565-0871, Japan, ⁷Japan Agency for Medical Research and Development-Core Research for Evolutional Science and Technology, Tokyo 100-0004, Japan, ⁸Healthcare and Medical Data Driven AI based Predictive Reasoning Development Unit, Medical Sciences Innovation Hub Program (MIH), RIKEN, Yokohama 230-0045, Japan, ⁹Laboratory for Developmental Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama 230-0045, Japan, ¹⁰Department of Otorhinolaryngology - Head and Neck Surgery, Osaka University Graduate School of Medicine, Suita 565-0871, Japan, ¹¹Clinical Research Center, Shizuoka General Hospital, Shizuoka 420-0881, Japan, ¹²The School of Pharmaceutical Sciences, University of Shizuoka, Shizuoka 422–8526, Japan, ¹³Department of Rheumatology and Clinical Immunology, Graduate School of Medicine, Kyoto University, Kyoto 606-8507, Japan, ¹⁴Department of Ophthalmology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Fukuoka 812-8582, Japan, ¹⁵Discovery Technology Research Laboratories, Tsukuba Research Institute, Ono Pharmaceutical Co., Ltd., Tsukuba 300-4247, Japan, ¹⁶Department of Obstetrics and Gynecology, Osaka University Graduate School of Medicine, Suita 565-0871, Japan, ¹⁷Department of Neurology, Osaka University Graduate School of Medicine, Suita 565-0871, Japan, ¹⁸Department of Pediatrics, Osaka University Graduate School of Medicine, Suita 565-0871, Japan, ¹⁹Department of Clinical Research, National Hospital Organization (NHO) Osaka Minami Medical Center, Japan, ²⁰Department of Rheumatology and Allergology, Saiseikai Senri Hospital, Suita 565-0862, Japan, ²¹Department of Rheumatology and Allergology, National Hospital Organization (NHO) Osaka Minami Medical Center, Japan, ²²Institute of Rheumatology, Tokyo Women's Medical University, Tokyo 162-0054, Japan, ²³Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Agency (JST), Tokyo 162-0054, Japan, ²⁴Division of Genomic Technologies, RIKEN Center for Life Science Technologies, Yokohama 230-0045, Japan, ²⁵RIKEN Omics Science Center (OSC), Yokohama 230-0045, Japan, ²⁶RIKEN Preventive Medicine and Diagnosis Innovation Program (PMI), Saitama 351-0198, Japan, ²⁷Preventive Medicine and Applied Genomics Unit, RIKEN Center for Integrative Medical Sciences, Yokohama

*To whom correspondence should be addressed. Tel: +81 6 6879 3971; Fax: +81 6 6879 3975; Email: yokada@sg.med.osaka-u.ac.jp

230-0045, Japan, ²⁸Telethon Kids Institute, the University of Western Australia, Western Australia 6008, Australia, ²⁹Laboratory of Tissue Regeneration, Department of Regeneration Science and Engineering, Institute for Frontier Life and Medical Sciences, Kyoto University, Kyoto 606-8507, Japan, ³⁰Department of Orthopaedic Surgery, Graduate School of Medicine, Kyoto University, Kyoto 606-8507, Japan, ³¹Harry Perkins Institute of Medical Research, QEII Medical Centre and Centre for Medical Research, the University of Western Australia, Western Australia 6009, Australia, ³²Laboratory for Transcriptome Technology, RIKEN Center for Integrative Medical Sciences, Yokohama 230-0045 Japan, ³³Kyoto-McGill International Collaborative School in Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto 606-8507, Japan, ³⁴Laboratory for Applied Computational Genomics, RIKEN Center for Integrative Medical Sciences, Yokohama 230-0045 Japan, ³⁵Laboratory for Autoimmune Diseases, RIKEN Center for Integrative Medical Sciences, Yokohama 230-0045, Japan and ³⁶Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Suita 565-0871, Japan

Received July 03, 2018; Revised October 12, 2018; Editorial Decision October 13, 2018; Accepted October 22, 2018

ABSTRACT

MicroRNAs (miRNAs) modulate the post-transcriptional regulation of target genes and are related to biology of complex human traits, but genetic landscape of miRNAs remains largely unknown. Given the strikingly tissue-specific miRNA expression profiles, we here expand a previous method to quantitatively evaluate enrichment of genome-wide association study (GWAS) signals on miRNA–target gene networks (MIGWAS) to further estimate tissue-specific enrichment. Our approach integrates tissue-specific expression profiles of miRNAs (~1800 miRNAs in 179 cells) with GWAS to test whether polygenic signals enrich in miRNA–target gene networks and whether they fall within specific tissues. We applied MIGWAS to 49 GWASs ($n_{\text{Total}} = 3\,520\,246$), and successfully identified biologically relevant tissues. Further, MIGWAS could point miRNAs as candidate biomarkers of the trait. As an illustrative example, we performed differentially expressed miRNA analysis between rheumatoid arthritis (RA) patients and healthy controls ($n = 63$). We identified novel biomarker miRNAs (e.g. hsa-miR-762) by integrating differentially expressed miRNAs with MIGWAS results for RA, as well as novel associated loci with significant genetic risk (rs56656810 at *MIR762* at 16q11; $n = 91\,482$, $P = 3.6 \times 10^{-8}$). Our result highlighted that miRNA–target gene network contributes to human disease genetics in a cell type-specific manner, which could yield an efficient screening of miRNAs as promising biomarkers.

INTRODUCTION

Complex human traits are products of orchestration of high dimensional multi-omics layers, including genome, epigenome, transcriptome, proteome and metabolome. Genome-wide association study (GWAS) has discovered thousands of genomic loci associated with these traits (1,2).

Integration of such large-scale genetic data with multi-layered omics information has successfully identified cell-type specific and context-dependent regulatory mechanism of diseases (3). While previous trans-omics approach mostly focused on integration with transcriptome (e.g. RNA-seq) (4) and epigenome data (e.g. ChIP-seq and Hi-C) (3,5), innovative construction of the analytic pipeline to integrate additional omics layers has been warranted to further elucidate complex biology of the traits.

Non-coding regions in the human genome constitute one of the unrevealed layers in biology. MicroRNAs (miRNAs), short non-coding RNA molecules of 21–25 nucleotide long, are key players in post-transcriptional gene regulation (6,7). Numerous studies have shown their critical role in the pathogenesis of various human diseases (8,9) and application of miRNAs as a biomarker or a therapeutic target is ongoing and promising (10,11). Nevertheless, it has been difficult to detect comprehensive association signals of miRNAs as compared with those of protein-coding genes and mRNAs, because genomic region that encodes miRNAs is relatively small. Biological roles of a miRNA should also be interpreted in a tissue specific context along with its target gene. On the arrival of recent high throughput sequencing technologies, comprehensive catalog of miRNA expression profile was created and revealed that the expression levels of miRNAs varied greatly according to tissues and were highly skewed (12).

Harnessing this tremendous work, here we extended our method to quantitatively evaluate enrichment of GWAS polygenic signals on miRNA–target gene networks (MIGWAS; miRNA–target gene networks enrichment on GWAS) that we have previously reported (13) to further decipher tissue-specific contribution of miRNA function in each trait. The MIGWAS enables us to study the tissue-specific landscape of post-transcriptional regulation, and to identify candidate miRNAs that are essential in pathophysiology. Our method can also conduct *in silico* screening of the miRNAs that can be used as novel biomarkers or therapeutic targets on the traits, which was empirically validated by the subsequent case-control analysis of differentially expressed miRNAs obtained from clinical subjects and the large-scale genetic association analysis of the lead variants.

MATERIALS AND METHODS

Calculation of gene- and miRNA- P values from GWAS summary statistics

We converted GWAS SNP association signals into a gene- or miRNA- level P value (i.e. P_{Gene} and P_{miRNA} , respectively). Similarly to the method Segrè *et al.* used in MAGENTA software (14), the best P value of a set of SNP P values mapped onto each gene or miRNA was corrected for the confounding effects of physical and genetic properties of genes or miRNAs on the P value (Figure 1A). We excluded genes and miRNAs located in the major histocompatibility complex (MHC) region to avoid the influence from its long linkage disequilibrium and complex architecture (15).

Curation of miRNA expression data and miRNA–target gene network information

Tissue-specific miRNA expression data from the FANTOM5 consortium (12) excluding candidate novel ones was downloaded from a web site (see [URLs](#)). Expression count per million values of mature miRNAs derived from 1842 pre-miRNAs in 179 human cell types were quantile-normalized using preprocessCore v1.34.0 package in R software. These cell types were classified as 18 tissue type categories; bone ($n = 10$), brain ($n = 14$), cardiac ($n = 3$), eye ($n = 3$), fat ($n = 15$), fetal ($n = 14$), gastrointestinal (GI; $n = 7$), genitourinary (GU; $n = 6$), immune ($n = 22$), joint ($n = 1$), kidney ($n = 7$), liver ($n = 4$), lung ($n = 10$), muscle ($n = 8$), pancreas ($n = 1$), skin ($n = 6$), vascular ($n = 15$) and others ($n = 33$; Supplementary Table S1) based on the organs from which these cells were collected. The result of the principal component analysis describing the expression pattern of miRNAs in each cell is shown in Supplementary Figure S1. Next, we calculated a TSI as Ludwig *et al.* previously described (16). In brief, the TSI for j th miRNA was calculated as

$$tsi_j = \frac{\sum_{i=1}^N (1 - x_{j,i})}{N - 1},$$

where N corresponds to the total number of cells measured and $x_{j,i}$ is the expression amount of i th cell normalized by the maximal expression of any cell for j th miRNA.

In each cell type, we defined a set of highly and specifically expressed miRNAs that satisfy two conditions; (i) normalized expression value falls within the top 10 percentile of those obtained from all miRNAs and (ii) TSI > 0.7 (Figure 1B). TSI > 0.7 threshold made it possible to include miRNAs with a wide range of expression levels and a variety of miRNAs, while retaining the power (Supplementary Figure S2). This set of highly and specifically expressed miRNAs was used to partition miRNA's heritability signal into various human cell types.

We obtained four kinds of major target gene prediction algorithms; TargetScan Human, DIANA-TarBase, PITA, and miRDB (see [URLs](#); Supplementary Table S2). For each of the prediction algorithms, we made a prediction score matrix with a row corresponding to genes, and a column corresponding to miRNAs.

Quantifying the enrichment of miRNA–gene network to a trait in a specific cell type

In order to empirically evaluate cell type-specific enrichment of miRNA–target gene network to a trait, we extended MIGWAS, which we have previously described (Figure 1C).

Let $y_{A,S_i,j}$ be the set of miRNA and gene pairs that satisfy the following four conditions: (i) target prediction score of the miRNA and gene is above j th threshold in i th prediction algorithm, (ii) the above mentioned P_{miRNA} is below a nominal signal ($\alpha = 0.01$), (iii) the P_{gene} is also below α and (iv) the miRNA is included in a set of highly and specifically expressed miRNAs in cell A. $n(y_{A,S_i,j})$ denotes the number of pairs included in the set $y_{A,S_i,j}$. We estimated the enrichment signal in cell A by comparing this metric $n(y_{A,S_i,j})$ with the empirical null distribution using a permutation procedure (20 000 times). Under this null distribution, we assumed that there is no association among tissue-specific miRNA expression, miRNA–target gene network and GWAS signal. In each iteration step, we randomly shuffled P values of both all genes and all miRNAs while retaining the miRNA–target gene relationships, and calculated the metric of the k th permutation step as $n(y'_{(k)A,S_i,j})$ (i.e. the number of miRNA–gene pairs that met all the four condition (i)–(iv) with the shuffled P values). The significance $P_{\text{enrichment}, A, S_i, j}$ of the metric was defined as $P(n(y') \geq n(y))$. We then sequentially slid the threshold j of the i th prediction algorithm from top one percentile to 0.1 percentile with eight partitions. When integrating the statistics, we adopted the results obtained only if the mean of $n(y'_{(k)A,S_i,j})$ is more than five because the estimation of the null distribution can be biased when $n(y')$ s are sparse. In order to compensate for the uncertainty existing within a single target prediction algorithm, we also integrated the results of four prediction algorithms to obtain one robust enrichment P value in cell A by meta-analyzing the enrichment significance as follows:

$$P_{\text{enrichment}, A} = \Phi \left(\frac{1}{\sqrt{n_{\text{algorithm}}}} \sum_{i=1}^{n_{\text{algorithm}}=4} \left(\frac{1}{\sum_j k_{S_i,j}} \sum_{j=1}^9 k_{S_i,j} \Phi^{-1} (P_{\text{enrichment}, A, S_i, j}) \right) \right),$$

where $k_{S_i,j} = \begin{cases} 1, & \text{when mean } y' \geq 5 \\ 0, & \text{when mean } y' < 5 \end{cases}$

Φ represents the cumulative distribution function of the normal distribution. Permutation P value < 0.05 was interpreted as statistically significant in all the following analyses.

In addition to quantifying tissue specific enrichment signal, we recorded the pairs of a miRNA and a gene included in the set $y_{A,S_i,j}$ to be considered as tissue-specific trait associated pairs of a miRNA and a gene.

We have released a python source code together with formatted miRNA expression materials to allow users to perform MIGWAS using any GWAS summary statistics on their own computers (see [URLs](#)). A multithreading option using multiprocessing module in python is implemented to shorten the computing time (~6.50 h when applied to summary statistics of one GWAS with ~7 million variants for 179 cell types using four CPU cores with 2.3 GHz). Selection of the algorithms to predict miRNA–target genes, TSI

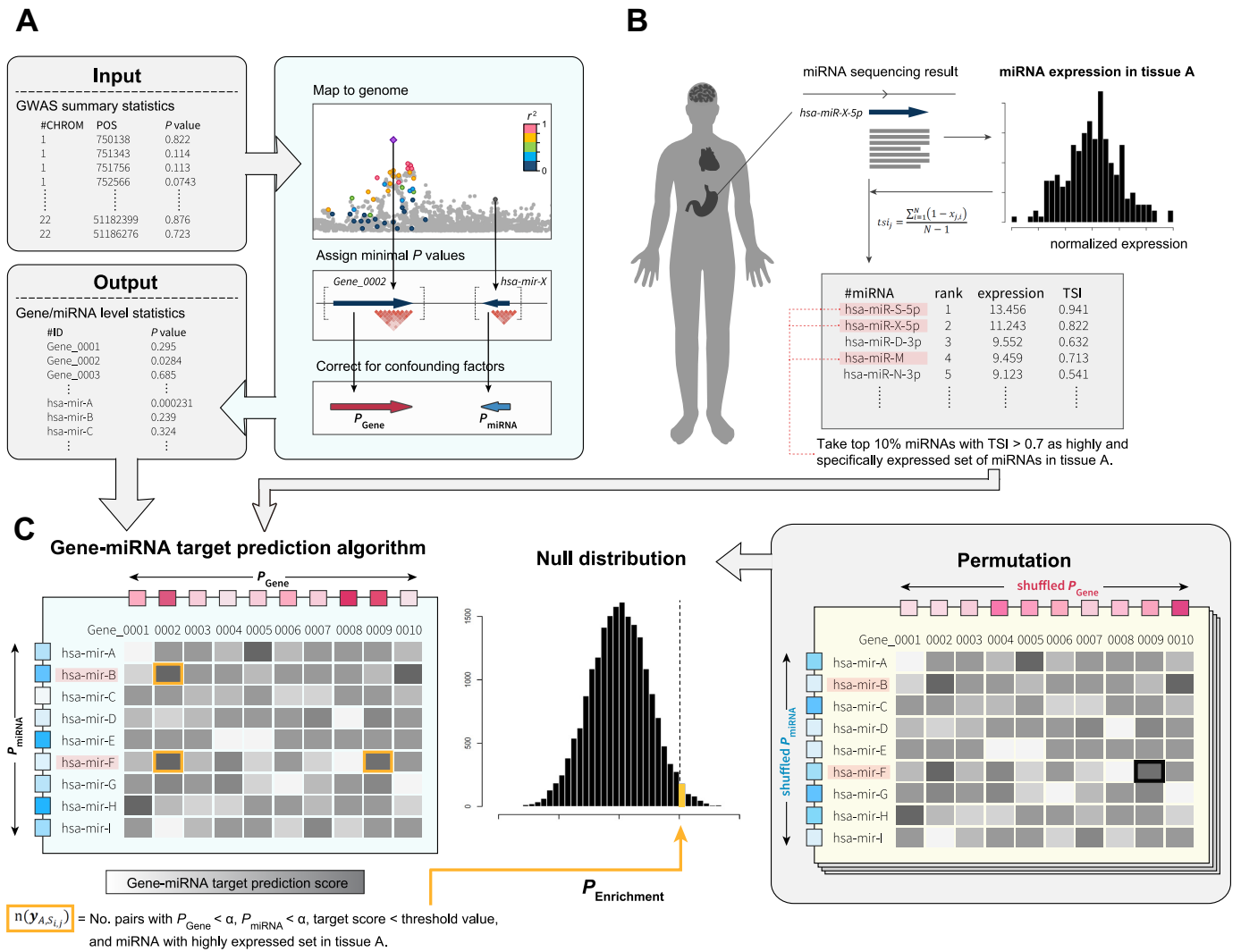


Figure 1. Overview of MIGWAS approach. (A) GWAS summary statistics are converted to gene- and miRNA- level P values (P_{Gene} and P_{miRNA}). (B) MiRNA expression data from various human tissues was quantile-normalized, and a set of highly and tissue-specifically expressed miRNAs was defined in each cell type. (C) Cell-type specific enrichment of GWAS signal in miRNA–target gene network was assessed by a permutation procedure (see Materials and Methods).

thresholds, and permutation numbers can be optimized by the users.

GWAS summary statistic data sets

We collected 49 GWAS summary statistics (Supplementary Table S3). Twenty-one of them were provided from public websites or collaborators. They include diseases consisting of four major categories (immune/allergy-related [$n = 4$], neuropsychiatric [$n = 2$], cardiovascular [$n = 1$] and genitourinary [$n = 1$]) and quantitative traits consisting of five major categories (anthropometric [$n = 3$], metabolic [$n = 3$], musculoskeletal [$n = 1$], cardiovascular [$n = 2$], kidney-related [$n = 2$] and hematological [$n = 2$]). The remaining 28 summary statistics are obtained from ongoing BioBank Japan project (17), >170 000 genome and health-related phenotype biobank of Japanese. These include diseases consisting of nine major categories (immune/allergy-related [$n = 4$], metabolic [$n = 1$], musculoskeletal [$n = 2$], neuropsychiatric [$n = 1$], eye-related [$n = 2$], cardiovascular [$n = 5$], lung-related [$n = 1$], liver-related [$n = 2$], genitourinary [$n = 2$], and malignancy [$n = 7$]), and quantitative traits consisting of 2 anthropometric traits (adult height, and body mass index [BMI]). Definition of the diseases and the process of patient registration are described elsewhere (17,18).

From these summary statistics, SNP positions based on the UCSC hg19 reference and their association P values from linear regression (quantitative trait) or logistic regression (binary trait) are collected and re-formatted.

From these summary statistics, SNP positions based on the UCSC hg19 reference and their association P values from linear regression (quantitative trait) or logistic regression (binary trait) are collected and re-formatted.

Study populations for high throughput miRNA expression sequencing

In this case-control study, Japanese patients with early rheumatoid arthritis (RA) were enrolled together with healthy Japanese volunteers. The diagnosis of RA was based on 2010 ACR/EULAR criteria (19). Written informed consent was obtained prior to the enrollment from all the in-

dividuals. In total, 30 RA patients and 33 healthy control (HC) participants were enrolled. Detailed information on the participants is in Supplementary Table S4. This study was approved by the ethical committee of Osaka University.

MiRNA expression sequencing and differential expression analysis in RA patients

Whole blood from the subject was collected into an ethylenediaminetetraacetic acid (EDTA) tube. PBMCs were isolated using Ficoll–Paque density gradient medium. Total RNA from PBMCs was isolated using miRNeasy Micro Kit (Qiagen). Libraries for miRNA-seq were prepared using SMARTer smRNA-Seq Kit (Takara) following manufacturer instructions. RNA-seq was performed using a HiSeq 2500 (Illumina, read length of 100 bp, single-end). For QC of the sequencing data, adaptor sequences were trimmed using Cutadapt, and reads with low quality score (Phred quality score < 20 in > 80% of total bases) and reads with length > 50 bp were removed. All the samples have total read counts > 1.0×10^6 , and thus proceeded to further analysis. Read count information is summarized in Supplementary Table S4. Reads were aligned against the known miRNA sequences in miRBase database, using Bowtie with recommended options described in the literature (20). Differential miRNA expression analysis between RA and HC was carried out with the R package DESeq2.

Overlap enrichment analysis of differentially expressed miRNAs with the MIGWAS result

To test whether differentially expressed miRNAs between RA patients and HC significantly overlap with candidate RA-associated miRNAs obtained from the MIGWAS, we performed a permutation procedure. We randomly shuffled sample IDs of the miRNA expression table, and defined differentially expressed miRNAs (FDR- $q < 0.05$). The number of overlapped miRNAs included both in differentially expressed miRNAs and in the MIGWAS candidates was recorded in each permutation step. After 5000 permutation steps, empirical P value was calculated as the number of permutation steps where the number of overlapped miRNAs was equal to or exceeded that from real dataset divided by the total number of permutation steps (= 5000 iterations).

The *in silico* replication study of the biomarker miRNA loci with RA risk

For each genetic locus that harbors a biomarker miRNA identified by the MIGWAS of RA and the differentially expressed miRNA analysis of the RA patients, we selected the lead SNP with the most significant association within the locus from the original trans-ethnic RA meta-analysis (19 234 RA cases and 61 565 controls) (21). By looking-up the additional two RA GWAS of Japanese (3308 RA cases and 8357 controls; Supplementary Table S5) (22), we conducted the *in silico* replication study of the lead SNPs with RA risk. Meta-analysis of the GWAS and replication study was conducted by an inverse-variance method assuming a fixed-effects model.

Integrative analysis with eQTL summary data and summary-level Mendelian Randomization (SMR) analysis

We hypothesized that miRNAs identified by the MIGWAS pipeline exert their effects on the trait by regulating the transcripts of their target genes, and that the target genes' expression level should also have a dose dependent association with the trait. To test this hypothesis, we undertook two ways of analysis using publicly available eQTL summary data. First, we tested whether target genes of the identified miRNA are enriched in *eGenes* within whole blood when compared with those across all tissues using summary eQTL data from GTEx consortium (23). *eGenes* in each tissue were defined as genes that harbor *cis*-eQTL variants which associate with their expression level with FDR- $q < 0.05$. The enrichment was assessed by the binomial test, with the prior probability being the number of *eGenes* within whole blood divided by the number of *eGenes* across all tissues that were included within the scope of eQTL analysis, GWAS, and miRNA's target prediction. Second, we performed SMR to show that the expression level of target genes that were identified through the above-mentioned analysis associates with the trait by causality or pleiotropy, and not by linkage. By integrating trans-ethnic GWAS summary statistics of RA (21) with CAGE eQTL data of peripheral blood (24), we did SMR analysis for 41 target genes of hsa-miR-762. We considered $P_{\text{HEIDI}} \geq 0.05$ as a threshold that we could not reject the null hypothesis that there was a single causal variant affecting both gene expression and trait variation (25).

RESULTS

Overview of our MIGWAS statistical methods

The overview of our MIGWAS method is shown in Figure 1. The principal hypothesis of MIGWAS is that if the miRNA is associated with a genetic risk of a trait, the target genes of the miRNA are also associated with the trait as a miRNA functionally works by interacting with its target genes. To test this hypothesis, we first converted SNP-level association statistics in the GWAS into gene- and miRNA-level association P values (i.e. P_{Gene} and P_{miRNA} , respectively; Figure 1A). Then we quantitatively evaluated enrichment of pairwise association signals of miRNAs and their target genes. We note that we set the significance threshold of P_{Gene} and P_{miRNA} ($\alpha = 0.01$) less stringent than the typical genome-wide significant level (26), to enhance sensitivity and to comprehensively incorporate the polygenic nature of complex traits (27,28).

In order to quantify tissue-specific contribution of miRNAs in each trait, we next partitioned enrichment of miRNA–target gene signals into each cell type separately. After normalization of miRNA expression data released by the FANTOM5 consortium (12), we defined a set of highly and tissue-specifically expressed miRNAs in a total of 179 human cells (Figure 1B). We performed cell type-based partitioning analysis within the defined set of miRNAs of each of the 179 cell types (Figure 1C), and thus obtained cell-type specific enrichment P value through permutation-based parallel computing. In addition our MIGWAS pipeline automatically identifies a set of trait-associated miRNA–

target gene pairs for both specific tissue categories and in all tissue categories combined together.

Heterogeneous enrichment of miRNA–target gene network to human complex traits

We applied the extended MIGWAS pipeline to the 49 GWAS summary statistics to evaluate the overall contribution of miRNA–target gene networks to various complex human traits, firstly without considering tissue-specificity. The GWASs covered a wide range of traits including anthropometric ($n = 4$), immune/allergy ($n = 8$), metabolic ($n = 5$), musculoskeletal ($n = 2$), neuropsychiatric ($n = 3$), eye-related ($n = 2$), cardiovascular ($n = 8$), lung-related ($n = 1$), kidney-related ($n = 2$), liver-related ($n = 2$), genitourinary ($n = 3$), hematological ($n = 2$) and malignancy traits ($n = 7$). The detailed information of the GWASs is shown in Supplementary Table S3 and the Materials and Methods section. We observed nominally significant ($P < 0.05$) miRNA–target gene enrichment signal in three traits for five GWASs; height, RA and type 2 diabetes (Figure 2A), which was consistent and robust with our previous results despite of the update on miRNA–target gene prediction algorithms (13). We note that we observed trans-ethnically consistent results for height and type 2 diabetes ($P < 0.05$ in both populations), two traits for which independent summary statistics were obtained from European and Japanese populations separately (29,30), which empirically validates the robustness of the MIGWAS results.

Tissue specific MIGWAS results successfully identified the disease-relevant tissue and associated miRNAs

Motivated by the robustness of our MIGWAS pipeline supported by trans-ethnic consistency, next we performed the partitioned enrichment analysis using the tissue specific miRNA expression data. Each trait was analyzed considering 179 different human cell types in parallel, and we annotated each cell into 18 tissue categories from which it was collected from (Supplementary Table S1). We here highlighted enrichment signal to immune-related cells in each trait as an example (Figure 2B). Tissue specificity of the traits in immune-related cells showed the different spectrum of enrichment with the most significant enrichment in Graves' disease ($P = 2.6 \times 10^{-6}$). Relatively high enrichment signals were also found in type 2 diabetes as well ($P < 5.9 \times 10^{-5}$ in both populations), where obesity-associated chronic tissue inflammation is reported to play a key contributing role (31).

Figure 2C shows the strongest associated cell in each trait. These results successfully illustrated the diverse tissue-specific nature of miRNA–target gene contribution to a trait, which was not obvious through the tissue-naïve approach. In many cases the analysis identified biologically relevant cell types in each trait (e.g. immune cells in Graves' disease, chronic hepatitis C, systemic lupus erythematosus, Crohn's disease and asthma, adipocytes in LDL cholesterol and coronary artery disease, brain tissue in age at menarche (32)). Detailed cell names of top association in all the traits are described in Supplementary Table S6.

Considering the abundance of accumulated literature on miRNA's associations in immune- and allergy- related traits

(33), we highlighted detailed enrichment signals in immune- and allergy- related traits (Figure 3). Here again, we observed disease relevant cell types in each trait. In RA, miRNA–target gene contribution was enriched in lung, bone, and immune cells ($P = 3.6 \times 10^{-3}$, 9.6×10^{-3} and 1.5×10^{-2} , respectively), which is consistent with biological understanding of the disease that pathogenic autoimmunity is first triggered in the lungs faced with environmental stimuli such as cigarette smoking and then causes bone erosion (34). We note that these disease-relevant tissues have not been identified by any previous methods to integrate with mRNA transcriptome and epigenome data despite their pivotal roles in RA pathophysiology (3,4,5), thereby demonstrating validity of our method to incorporate tissue-specific miRNA profiles. All the cell names with significant enrichment in immune-related traits are described in Supplementary Table S6.

Differential expression of miRNAs among rheumatoid arthritis patients enriches in MIGWAS result and pinpoints novel causal mechanisms

Through the MIGWAS pipeline, we also obtained a list of candidate trait-associated miRNAs and target genes as candidate biomarkers, which was defined systematically to satisfy the criteria used in the enrichment analysis (i.e. both miRNA and its predicted target gene are associated with the trait with a nominal significance [P value < 0.01]) with all tissue categories combined together. Given the previous studies showing the association of miRNAs in the biology of RA (13,35,36), we decided to focus on RA as a target of *in vivo* validation. To validate candidate novel miRNAs of which association with RA was identified by MIGWAS, we profiled miRNA expression in peripheral blood mononuclear cells (PBMCs) from 30 patients with early RA and 33 control subjects using high throughput miRNA sequencing. PBMCs were selected for the miRNA expression profiling because they have been shown to harbor much of heritability in the pathology of RA (3,32), and also because they can be collected in the clinical setting. We found 94 differentially expressed miRNAs with a false discovery rate (FDR) of 0.05 (Figure 4A). Of these, four miRNAs (hsa-miR-93-5p, hsa-miR-106b-5p, hsa-miR-301b-3p and hsa-miR-762) overlapped those identified by the MIGWAS pipeline applied for the RA GWAS. Permutation procedures in which sample IDs of the miRNA expression data were randomly shuffled 5000 times revealed that the observed overlap of as many as four miRNAs was much larger than would be expected by chance (98.0-fold enrichment [four miRNAs divided by the mean simulated overlap, 0.0408], permutation P value = 0.0010 [five simulation steps with equal or more overlap than the observed overlap divided by total simulation steps], Materials and Methods for permutation procedure; Table 1 and Supplementary Figure S4). One of the overlapped miRNAs, hsa-miR-762, was significantly highly expressed in RA patients (\log_2 fold change = 1.15 and FDR-q = 0.043) and has a prominent expression specific to immune-related cells (tissue specificity index [TSI] = 0.982, Figure 4B and C). This would implicate the utility of hsa-miR-762 as a future target of clinical validation for a biomarker of RA. On the other hand, the other three miR-

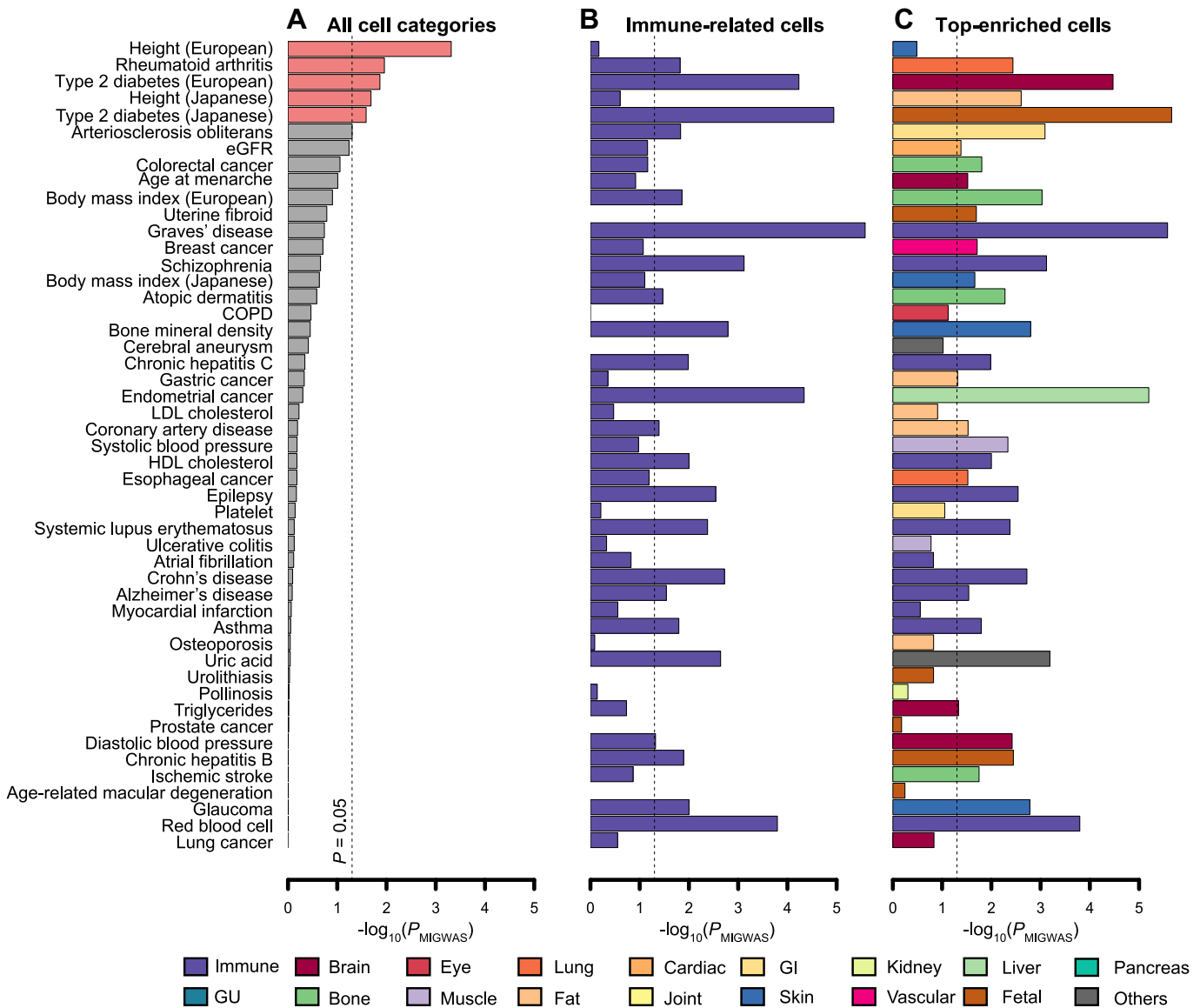


Figure 2. MIGWAS results of the GWASs on 49 complex human traits. An enrichment signal is shown by $-\log_{10}(P_{MIGWAS})$. (A) Overall polygenic contribution of miRNA–target gene network to the traits through the tissue-naïve approach. Traits that have bars colored with pink showed nominally significant enrichment ($P_{MIGWAS} < 0.05$). (B) An example of tissue-specific enrichment signals focusing on immune-related cells. A bar in each trait represents an enrichment signal of the most significant immune-related cell. (C) An enrichment signal in the strongest associated cell of each trait. The tissue category of the cell is shown by the color of the bars. Detailed descriptions on cell types and their categories are available at Supplementary Table S1. GI: gastrointestinal, GU: genitourinary.

NAs showed repressed expressions in the RA patients with ubiquitous expression profiles among the cell types ($TSI < 0.40$).

While genetic variants located nearby the identified biomarker miRNAs confer suggestive associations in the original RA case-control GWAS (19 234 RA cases and 61 565 controls, $P_{GWAS} \geq 3.2 \times 10^{-7}$) (21), overlap with the RA case-control miRNA-seq analysis strongly prioritizes true-positive associations of the lead variants in such loci. Thus, we conducted an *in silico* replication study using additional 3308 RA cases and 8357 controls (Supplementary Table S5) (22). Of these, two loci that harbor the three miRNAs satisfied the genome-wide significance threshold

when GWAS and the replication study was meta-analyzed (22 119 cases and 69 363 controls; $P_{META-GWAS} = 3.3 \times 10^{-8}$ for rs34130487 at *MIR95-MIR106B* at 7q22 and $P = 3.6 \times 10^{-8}$ for rs56656810 at *MIR762* at 16q11; Figure 4D, E, Table 2, and Supplementary Table S5), of which RA associations were novel findings.

We further listed target genes of these miRNAs (Table 1). As for hsa-miR-762, 41 genes were identified as potential target genes that work synergistically to cause RA (Figure 4D). To investigate whether those target genes have an eQTL effect in immune-related cells where hsa-miR-762 is supposed to be functioning, we referred to the summary eQTL statistics in human tissues released by GTEx Consor-

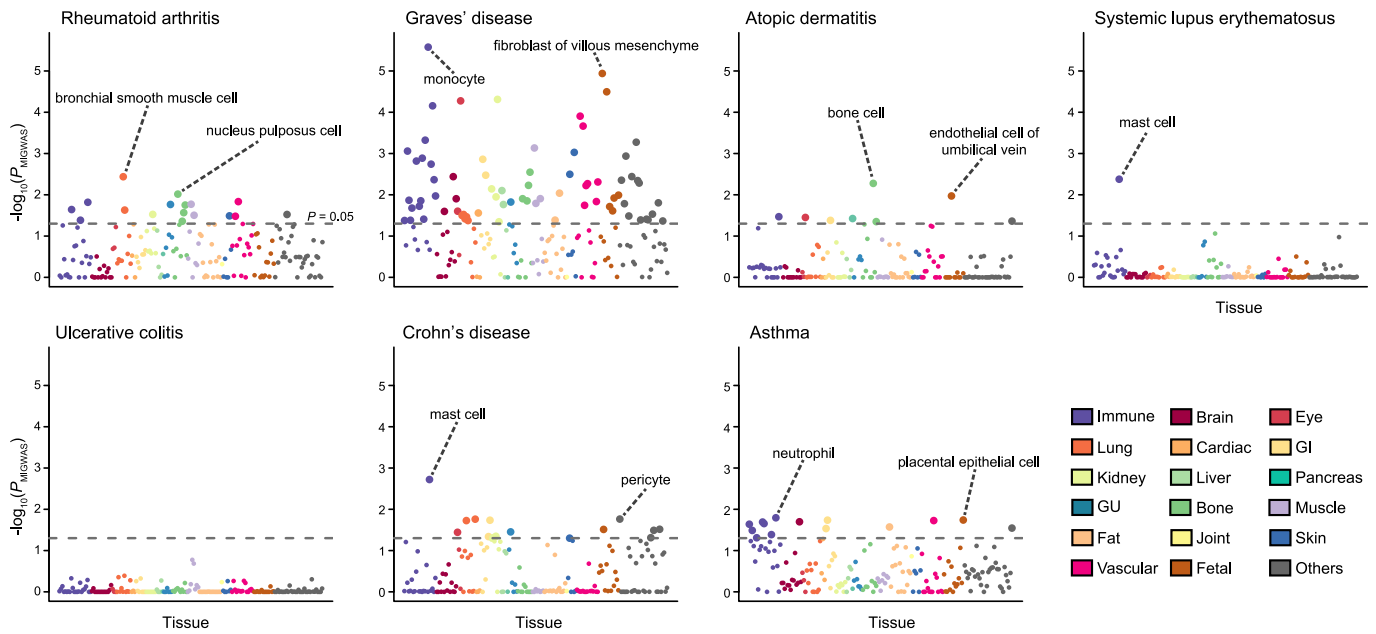


Figure 3. Tissue-specific enrichment signal of miRNA–target gene network in selected immune/allergy- related traits. Each circle represents the cell-specific enrichment signal in $-\log_{10}(P_{\text{MIGWAS}})$, colored with tissue categories. Large circles pass the nominal threshold of enrichment P value < 0.05 , which was indicated as dotted lines. The names of top 2 enriched cells that passed the threshold are also shown. Detailed descriptions on cell types and their categories are available at Supplementary Table S1. GI; gastrointestinal, GU; genitourinary.

Table 1. Candidate biomarker miRNAs identified both in the differentially expressed miRNA study and in the MIGWAS analysis (all cell types combined)

Candidate biomarker miRNA	RA cases versus controls			Target genes
	\log_2 fold change in expression	FDR-q	TSI	
hsa-miR-93-5p	-0.522	0.00022	0.254	<i>ANKH</i> , <i>ANKRD52</i> , <i>ARCNI</i> , <i>BCL2L15</i> , <i>C7orf43</i> , <i>CASP8</i> , <i>CDKN1A</i> , <i>DENND1B</i> , <i>FAM126B</i> , <i>FAM133B</i> , <i>FYCO1</i> , <i>IKZF4</i> , <i>KIAA1109</i> , <i>KLF2</i> , <i>LDLR</i> , <i>MAGI3</i> , <i>PEX13</i> , <i>PHTF1</i> , <i>PRDX5</i> , <i>RAB5B</i> , <i>RAG1</i> , <i>RSBN1</i> , <i>SAR1B</i> , <i>SEC24A</i> , <i>SLC12A5</i> , <i>SNN</i> , <i>STK38</i> , <i>TAGAP</i> , <i>TRAK2</i> , <i>ZBTB10</i>
hsa-miR-106b-5p	-0.811	0.00022	0.269	<i>ANKH</i> , <i>ANKRD52</i> , <i>BCL2L15</i> , <i>C7orf43</i> , <i>CASP8</i> , <i>CDK6</i> , <i>CDKN1A</i> , <i>CEP76</i> , <i>FAM126B</i> , <i>FAM167A</i> , <i>FAM65B</i> , <i>FYCO1</i> , <i>ICOS</i> , <i>KLF2</i> , <i>LDLR</i> , <i>MAGI3</i> , <i>PA2G4</i> , <i>PAPOLG</i> , <i>PDGFB</i> , <i>RAB5B</i> , <i>RSBN1</i> , <i>RTKN2</i> , <i>SAR1B</i> , <i>STK38</i> , <i>TMEM151B</i> , <i>TNFAIP3</i> , <i>VPS37C</i> , <i>ZFP36L1</i>
hsa-miR-301b-3p	-0.954	0.0090	0.397	<i>C5orf30</i> , <i>CDK6</i> , <i>DDX6</i> , <i>ITSN1</i> , <i>LDLR</i> , <i>PAN3</i> , <i>SAR1B</i> , <i>SERBP1</i> , <i>SNRPE</i> , <i>SRSF3</i> , <i>TMEM50B</i> , <i>ZBTB10</i>
hsa-miR-762	1.154	0.043	0.982	<i>ARHGAP20</i> , <i>C11orf20</i> , <i>C11orf9</i> , <i>C1orf93</i> , <i>C7orf59</i> , <i>CPNE5</i> , <i>DAP</i> , <i>FADS2</i> , <i>GATS</i> , <i>GNAI2</i> , <i>HDAC5</i> , <i>INPP5B</i> , <i>IQGAP1</i> , <i>IRF5</i> , <i>ITSN1</i> , <i>PADI2</i> , <i>PFKL</i> , <i>PGAP3</i> , <i>PHF15</i> , <i>PHKG2</i> , <i>PHLDB1</i> , <i>PKNOX2</i> , <i>PPIL2</i> , <i>RAF1</i> , <i>RAVER1</i> , <i>SLC25A23</i> , <i>SLC44A2</i> , <i>STAC2</i> , <i>SYNGRI</i> , <i>TABI</i> , <i>TMEM151B</i> , <i>TMPRSS3</i> , <i>TNFRSF14</i> , <i>TRAF1</i> , <i>TSPAN33</i> , <i>UBASH3A</i> , <i>UPK2</i> , <i>UTP11L</i> , <i>VPS37C</i> , <i>YDJC</i> , <i>ZNF594</i>

Each miRNA's target genes were defined to have top one percentile of target prediction score in at least two prediction algorithms. Genes in bold face are those overlapped with *eGenes* of GTEx eQTL data of whole blood. TSI; tissue specificity index, RA; rheumatoid arthritis.

tium (23). We observed that target genes of hsa-miR-762 harbor *cis*-eQTL variants within whole blood as a proxy of immune-related cells more frequently than expected when compared to those in all available human tissues (1.46-fold enrichment with binomial P value = 0.014). This consistently suggests that miRNAs should modulate the association with the trait through interfering the transcriptome, and that the targeted gene's association with the trait should

be driven by regulation on expression within a specific trait-associated tissue.

In order to clarify the connection between eQTL signals of the variants which hsa-miR-762's target genes harbor and their neighboring association signals in RA GWAS, we performed the summary data-based Mendelian randomization (SMR) (25). This analysis can test whether the expression level of the target genes associates with RA by causality or pleiotropy, and not by linkage. In many of the

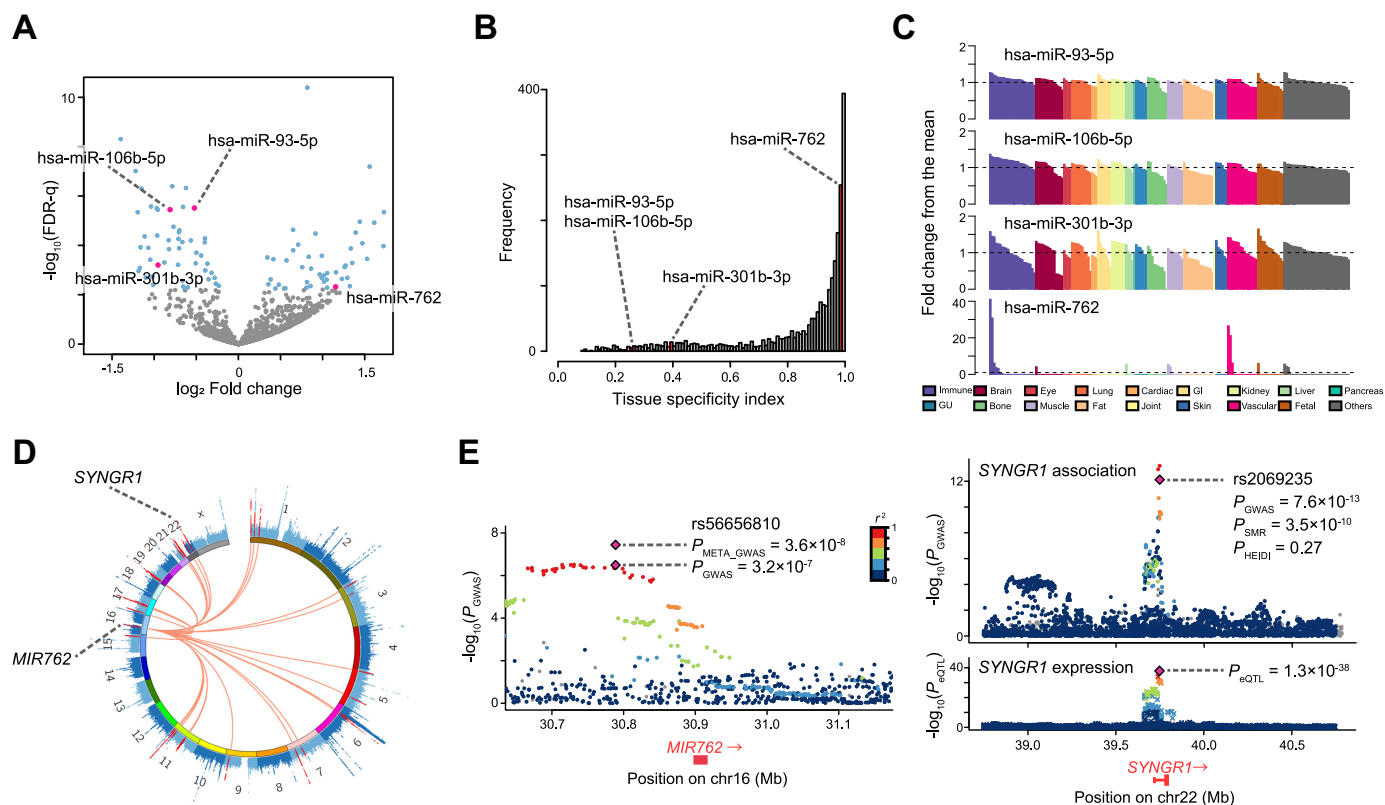


Figure 4. Results of differentially expressed miRNA analysis between RA patients and healthy controls. (A) Volcano plot displaying differentially expressed miRNAs between RA and healthy controls. The x-axis corresponds to the \log_2 fold change value and the y-axis corresponds to $-\log_{10}(\text{FDR}-q)$. The blue dots represent the differentially expressed miRNAs ($\text{FDR}-q < 0.05$). The pink dots represent differentially expressed miRNAs that overlapped with those identified in MIGWAS analysis. (B) The distribution of tissue specificity index (TSI) among all the miRNAs in expression data by FANTOM5. Bars colored with pink show the bins which the overlapped candidate miRNAs belong to. (C) Quantile-normalized expression values of the overlapped miRNAs in 179 cell types (x-axis), scaled as fold changes of the expression levels from the mean levels among all the cell types (y-axis). Bar colors represent the tissue into which each cell is categorized. GI; gastrointestinal, GU; genitourinary. (D) CIRCOS plot that shows a Manhattan plot of the RA GWAS marked by *MIR762* and its target genes in red. (E) Regional plots of the RA GWAS illustrating variants neighboring *MIR762* (left) and *SYNGR1* (right top). Diamonds represent the leading variants, and also shown are their identifiers, linkage disequilibrium structure and association statistics. Two diamonds in the left panel represent the P value of rs2069235 from the original RA GWAS summary statistics (P_{GWAS}) (21) and one from the new meta-analysis described in this paper ($P_{\text{META.GWAS}}$). Right bottom, eQTL P values from CAGE dataset for a probe tagging *SYNGR1*.

Table 2. RA case-control associations of the SNPs located at the biomarker miRNA loci

rsID	Chr	Position	Allele 1/2	Gene	No. subjects		Allele 1 frequency		OR (95%CI)	P
					RA cases	Controls	RA cases	Controls		
rs34130487	7	99 759 205	C/T	<i>MIR95-</i> <i>MIR106B</i>	16 633	54 807	0.638	0.623	1.08 (1.05–1.11)	3.3×10^{-8}
rs56656810	16	30 788 759	A/C	<i>MIR762</i>	22 119	69 363	0.210	0.196	1.09 (1.05–1.12)	3.6×10^{-8}
rs11089637	22	21 979 096	C/T	<i>MIR301B</i>	22 119	69 363	0.280	0.273	1.07 (1.05–1.10)	3.7×10^{-7}

RA, rheumatoid arthritis; OR, odds ratio; 95%CI, 95% confidence interval. Detailed results are available at Supplementary Table S5.

target genes of hsa-miR-762, we could finemap the potentially causal or pleiotropic eQTL variants that not only existed nearby GWAS genes but also both mediated gene-expression levels and were associated with RA risk (Supplementary Table S7 for the summary statistics). As an illustrative example, genetic variants at *SYNGR1* locus, which was one of the target genes of hsa-miR-762 and showed the strongest association statistics in the SMR analysis, are

associated with RA risk (upper right panel at Figure 4E). In expression data from Consortium for the Architecture of Gene Expression (CAGE) in PBMCs (24), *SYNGR1* locus harbors *cis*-eQTL variants (lower right panel at Figure 4E). The SMR analysis reveals that GWAS signal of the SNP (rs2069235, $P_{\text{GWAS}} = 7.6 \times 10^{-13}$, $P_{\text{eQTL}} = 1.3 \times 10^{-38}$) in *SYNGR1* is strongly and significantly mediated by eQTL effect ($P_{\text{SMR}} = 3.5 \times 10^{-10}$), which was not

driven by linkage ($P_{\text{HEIDI}} = 0.27$). The MIGWAS pipeline and miRNA expression profiles from clinical subjects successfully pinpointed the causal mechanism of *MIR762*'s association with RA, where increased expression of hsa-miR-762 in immune-related cells interferes with transcripts such as *SYNGRI*, whose expression levels are associated with causing RA.

DISCUSSION

By integrating the large scale GWAS data and miRNA–target gene prediction algorithms together with comprehensive tissue-specific miRNA expression profiles, here we successfully extended a method MIGWAS to evaluate enrichment of GWAS signals on miRNA–target gene networks and to partition them into a tissue specific context. The application of the MIGWAS pipeline to a wide range of genetics of complex human traits depicted cell type-specificity of the trait biology, as well as identification of biomarker miRNAs and novel genetic risk loci such as hsa-miR-762 encoded in the *MIR762* locus at 16q11 for RA. The implementation of genetic information on miRNAs will certainly serve to help investigators elucidate pathophysiology and forward clinical application, as well as further development of novel therapeutic targets.

Our MIGWAS pipeline highlights four innovative features: (i) quantitatively assessing the miRNA–target gene contribution to genetics of a trait, (ii) deciphering tissue specificity to identify causal tissues on a trait, (iii) efficiently prioritizing genetic loci for a follow-up study that harbor true-positive associations with disease risk and (iv) providing candidate biomarker miRNA–target gene lists that might contribute to the understanding of disease biology. Our work firmly supports the idea that the miRNA's contribution to genetics of complex human traits occurs in a cell type-specific manner in the same way as gene expression (4) and epigenetic regulation (3,32). A list of candidate miRNA–target gene pairs provides an ideal resource for experimental validation aiming at the discovery of biomarkers or potential therapeutics. The MIGWAS enables us to pinpoint a causal miRNA–target gene network to a trait, which was not achieved by GWAS alone or previous approaches to integrate GWAS and epigenome or transcriptome data.

We acknowledge that our study has several limitations. First, tissue specific MIGWAS analysis was performed based on an expression data from healthy individuals, while miRNA expression profile might be different depending on the disease status of individuals. We consider that a promising approach to use summary expression data from healthy individuals as ours has been shown to successfully capture enrichment of the trait heritability to biologically-relevant tissue (37). However, to have further insights into the disease biology, future application of expression profiles from pathologically altered tissues or inclusion of specifically depleted miRNAs should be warranted. Second, the *in silico* prediction of a miRNA and a target gene confers ambiguity dependent on the algorithms. In order to obtain unbiased prediction scores and to assure the robustness of the analysis, we integrated multiple different algorithms. When compared with a recently updated database miRTarBase (38), which provides the largest amount of miRNA–

target gene interactions with experimental validations to date, our approach of multiple thresholding and multiple algorithms indeed contributed to the better functional prediction *in vitro* (Supplementary Figure S3). Further integration of such experiment-based target gene information into the MIGWAS pipeline would be warranted. Third, since our method is based on the GWAS results, enrichment signal could be inflated due to the inflated GWAS summary statistics. To address this point, we took a permutation approach to estimate the null distribution, which made the MIGWAS results robust against the inflation of the GWAS itself (13). Fourth, we did not incorporate tissue-specific expression profile of genes in our method. The interaction of both miRNAs' and their target genes' expression levels is beyond this study because it suffers from exponential combinations and computational intensiveness. Nevertheless, future integration would contribute to deciphering fine nature of tissue specificity.

In conclusion, MIGWAS demonstrated that miRNA–target gene networks contribute to human disease genetics in the context of cell type-specific expressions, and successfully identified miRNAs as promising biomarkers.

URLs

The URLs for data presented herein are as follows:

- MIGWAS source code, <https://github.com/saorisakaue/MIGWAS>
- FANTOM5 consortium, <http://fantom.gsc.riken.jp/5/>
- The BioBank Japan Project (BBJ), <https://biobankjp.org/english/index.html>
- miRBase, <http://www.mirbase.org/>
- miRDB, <http://www.mirdb.org/>
- TargetScan Human, http://www.targetscan.org/vert_72/
- DIANA-TarBase, <http://diana.imis.athena-innovation.gr/DianaTools/index.php>
- PITA, https://genie.weizmann.ac.il/pubs/mir07/mir07_data.html
- GTEx consortium, <https://www.gtexportal.org/home/>
- CAGE eQTL data, <http://cnsgenomics.com/shiny/CAGE/>
- R DESeq2 package, <https://bioconductor.org/packages/release/bioc/html/DESeq2.html>
- R preprocessCore package, <https://bioconductor.org/packages/release/bioc/html/preprocessCore.html>
- SMR software, cnsgenomics.com/software/smr/
- CIRCOS, <http://circos.ca/>
- Cutadapt, cutadapt.readthedocs.io/en/stable/
- Bowtie, <http://bowtie-bio.sourceforge.net/index.shtml>

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Prof. Kiyoshi Takeda and Dr Miho Kaneda for their kind supports on the study. We would like to thank all members of the FANTOM5 consortium for contributing to generation of samples and analysis of the data-set and thank GeNAS for data production.

FUNDING

Japan Society for the Promotion of Science (JSPS) KAKENHI [15H05670, 15H05911, 15K14429, 16KT0196]; Japan Agency for Medical Research and Development [17ek0410047h0001, 18gm6010001h0003, 18ek0410041h0002]; Takeda Science Foundation; Nakajima Foundation; SECOM Science and Technology Foundation, Integrated Frontier Research for Medical Science Division, Institute for Open and Transdisciplinary Research Initiatives, Osaka University, Bioinformatics Initiative of Osaka University Graduate School of Medicine, and Inoue Foundation for Science. FANTOM5 was made possible by a Research Grant for RIKEN Omics Science Center from the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) (to Y.H.); MEXT for the RIKEN Preventive Medicine and Diagnosis Innovation Program (to Y.H.); Grant of the Innovative Cell Biology by Innovative Technology (Cell Innovation Program) from the MEXT, Japan (to Y.H.); Grant from MEXT to the RIKEN Center for Life Science Technologies and Grant from MEXT to RIKEN Center for Integrative Medical Sciences; J.H. is an employee of TEIJIN PHARMA LIMITED; N.S. is an employee of Ono Pharmaceutical Co., Ltd. Funding for open access charge: Japan Agency for Medical Research and Development [18ek0410041h0002].

Conflict of interest statement. None declared.

REFERENCES

- Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorf, L. *et al.* (2014) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.*, **42**, D1001–D1006.
- Visscher, P.M., Brown, M.A., McCarthy, M.I. and Yang, J. (2012) Five years of GWAS discovery. *Am. J. Hum. Genet.*, **90**, 7–24.
- Trynka, G., Sandor, C., Han, B., Xu, H., Stranger, B.E., Liu, X.S. and Raychaudhuri, S. (2013) Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat. Genet.*, **45**, 124–130.
- Ongen, H., Brown, A.A., Delaneau, O., Panousis, N.I., Nica, A.C. and Dermitzakis, E.T. (2017) Estimating the causal tissues for complex traits and diseases. *Nat. Genet.*, **49**, 1676–1683.
- Martin, P., McGovern, A., Orozco, G., Duffus, K., Yarwood, A., Schoenfelder, S., Cooper, N.J., Barton, A., Wallace, C., Fraser, P. *et al.* (2015) Capture Hi-C reveals novel candidate genes and complex long-range interactions with related autoimmune risk loci. *Nat. Commun.*, **6**, 10069.
- Lee, R., Feinbaum, R. and Ambros, V. (2004) A short history of a short RNA. *Cell*, **116**, S89–S92.
- Lee, R.C. and Ambros, V. (2001) An extensive class of small RNAs in *Caenorhabditis elegans*. *Science*, **294**, 862–864.
- Lu, J., Getz, G., Miska, E.A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B.L., Mak, R.H., Ferrando, A.A. *et al.* (2005) MicroRNA expression profiles classify human cancers. *Nature*, **435**, 834–838.
- Brest, P., Lapaquette, P., Souidi, M., Lebrigand, K., Cesaro, A., Vouret-Craviari, V., Mari, B., Barbry, P., Mosnier, J.-F., Hébuterne, X. *et al.* (2011) A synonymous variant in IRGM alters a binding site for miR-196 and causes deregulation of IRGM-dependent xenophagy in Crohn's disease. *Nat. Genet.*, **43**, 242–245.
- Krützfeldt, J., Rajewsky, N., Braich, R., Rajeev, K.G., Tuschl, T., Manoharan, M. and Stoffel, M. (2005) Silencing of microRNAs in vivo with 'antagomirs'. *Nature*, **438**, 685–689.
- Titze-de-Almeida, R., David, C. and Titze-de-Almeida, S.S. (2017) The race of 10 synthetic RNAi-based drugs to the pharmaceutical market. *Pharm. Res.*, **34**, 1339–1363.
- De Rie, D., Abugessaisa, I., Alam, T., Arner, E., Arner, P., Ashoor, H., Åström, G., Babina, M., Bertin, N., Burroughs, A.M. *et al.* (2017) An integrated expression atlas of miRNAs and their promoters in human and mouse. *Nat. Biotechnol.*, **35**, 872–878.
- Okada, Y., Muramatsu, T., Suita, N., Kanai, M., Kawakami, E., Iotchkova, V., Soranzo, N. and Inazawa, J. (2016) Significant impact of miRNA – target gene networks on genetics of human complex traits. *Sci. Rep.*, **6**, 22223.
- Ayellet, V.S., Groop, L., Mootha, V.K., Daly, M.J. and Altshuler, D. (2010) Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet.*, **6**, e1001058.
- Okada, Y., Momozawa, Y., Ashikawa, K., Kanai, M., Matsuda, K., Kamatani, Y., Takahashi, A. and Kubo, M. (2015) Construction of a population-specific HLA imputation reference panel and its application to Graves' disease risk in Japanese. *Nat. Genet.*, **47**, 798–802.
- Ludwig, N., Leidinger, P., Becker, K., Backes, C., Fehlmann, T., Pallasch, C., Rheinheimer, S., Meder, B., Stähler, C., Meese, E. *et al.* (2016) Distribution of miRNA expression across human tissues. *Nucleic Acids Res.*, **44**, 3865–3877.
- Nagai, A., Hirata, M., Kamatani, Y., Muto, K., Matsuda, K., Kiyohara, Y., Ninomiya, T., Tamakoshi, A., Yamagata, Z., Mushiroya, T. *et al.* (2017) Overview of the BioBank Japan Project: Study design and profile. *J. Epidemiol.*, **27**, S2–S8.
- Hirata, M., Kamatani, Y., Nagai, A., Kiyohara, Y., Ninomiya, T., Tamakoshi, A., Yamagata, Z., Kubo, M., Muto, K., Mushiroya, T. *et al.* (2017) Cross-sectional analysis of BioBank Japan clinical data: A large cohort of 200,000 patients with 47 common diseases. *J. Epidemiol.*, **27**, S9–S21.
- Aletaha, D., Neogi, T., Silman, A.J., Funovits, J., Felson, D.T., Bingham, C.O., Birnbaum, N.S., Burmester, G.R., Bykerk, V.P., Cohen, M.D. *et al.* (2010) 2010 Rheumatoid arthritis classification criteria: An American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis Rheum.*, **62**, 2569–2581.
- Ziemann, M., Kaspi, A. and El-Osta, A. (2016) Evaluation of microRNA alignment techniques. *RNA*, **22**, 1120–1138.
- Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S. *et al.* (2014) Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature*, **506**, 376–381.
- Okada, Y., Suzuki, A., Ikari, K., Terao, C., Kochi, Y., Ohmura, K., Higasa, K., Akiyama, M., Ashikawa, K., Kanai, M. *et al.* (2016) Contribution of a Non-classical HLA Gene, HLA-DOA, to the Risk of Rheumatoid Arthritis. *Am. J. Hum. Genet.*, **99**, 366–374.
- GTE Consortium. (2017) Genetic effects on gene expression across human tissues. *Nature*, **550**, 204–213.
- Lloyd-Jones, L.R., Holloway, A., McRae, A., Yang, J., Small, K., Zhao, J., Zeng, B., Bakshi, A., Metspalu, A., Dermitzakis, M. *et al.* (2017) The genetic architecture of gene expression in peripheral blood. *Am. J. Hum. Genet.*, **100**, 228–237.
- Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M.R., Powell, J.E., Montgomery, G.W., Goddard, M.E., Wray, N.R., Visscher, P.M. *et al.* (2016) Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.*, **48**, 481–487.
- Kanai, M., Tanaka, T. and Okada, Y. (2016) Empirical estimation of genome-wide significance thresholds based on the 1000 Genomes Project data set. *J. Hum. Genet.*, **61**, 861–866.
- Lee, S.H., Wray, N.R., Goddard, M.E. and Visscher, P.M. (2011) Estimating missing heritability for disease from genome-wide association studies. *Am. J. Hum. Genet.*, **88**, 294–305.
- Bulik-Sullivan, B., Loh, P.R., Finucane, H.K., Ripke, S., Yang, J., Patterson, N., Daly, M.J., Price, A.L., Neale, B.M., Corvin, A. *et al.* (2015) LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.*, **47**, 291–295.
- Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H., Gustafsson, S., Chu, A.Y., Estrada, K., Luan, J., Kutalik, Z. *et al.* (2014) Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.*, **46**, 1173–1186.
- DIAGRAM Consortium, Asian Genetic Epidemiology Network Type 2 Diabetes (AGEN-T2D) Consortium, South Asian Type 2 Diabetes (SAT2D)

- Consortium, Mexican American Type 2 Diabetes (MAT2D) Consortium, Type 2 Diabetes Genetic Exploration by Next-generation sequencing in multi-Ethnic Samples (T2D-GENES) Consortium, Mahajan, A., Go, M.J., Zhang, W., Below, J.E., Gaulton, K.J. *et al.* (2014) Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat. Genet.*, **46**, 234–244.
31. Osborn, O. and Olefsky, J.M. (2012) The cellular and signaling networks linking the immune system and metabolism in disease. *Nat. Med.*, **18**, 363–374.
 32. Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.R., Anttila, V., Xu, H., Zang, C., Farh, K. *et al.* (2015) Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.*, **47**, 1228–1235.
 33. Chen, J.-Q., Papp, G., Szodoray, P. and Zeher, M. (2016) The role of microRNAs in the pathogenesis of autoimmune diseases. *Autoimmun. Rev.*, **15**, 1171–1180.
 34. Malmström, V., Catrina, A.I. and Klareskog, L. (2017) The immunopathogenesis of seropositive rheumatoid arthritis: from triggering to targeting. *Nat. Rev. Immunol.*, **17**, 60–75.
 35. Chen, J.-Q., Papp, G., Szodoray, P. and Zeher, M. (2016) The role of microRNAs in the pathogenesis of autoimmune diseases. *Autoimmun. Rev.*, **15**, 1171–1180.
 36. Murata, K., Yoshitomi, H., Furu, M., Ishikawa, M., Shibuya, H., Ito, H. and Matsuda, S. (2014) MicroRNA-451 Down-Regulates neutrophil Chemotaxis via p38 MAPK. *Arthritis Rheumatol.*, **66**, 549–559.
 37. Finucane, H.K., Reshef, Y.A., Anttila, V., Slowikowski, K., Gusev, A., Byrnes, A., Gazal, S., Loh, P.R., Lareau, C., Shores, N. *et al.* (2018) Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.*, **50**, 621–629.
 38. Chou, C.-H., Shrestha, S., Yang, C.-D., Chang, N.-W., Lin, Y.-L., Liao, K.-W., Huang, W.-C., Sun, T.-H., Tu, S.-J., Lee, W.-H. *et al.* (2018) miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res.*, **46**, D296–D302.