Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

A data-driven biocomputing pipeline with meta-analysis on high throughput transcriptomics to identify genome-wide miRNA markers associated with type 2 diabetes

Kushan De Silva [a,*], Ryan T. Demmer [b,c], Daniel Jönsson [d,e], Aya Mousa [a], Andrew Forbes [f], Joanne Enticott [a]

[a] Monash Centre for Health Research and Implementation, School of Public Health and Preventive Medicine, Faculty of Medicine, Nursing, and Health Sciences, Monash University, Clayton, 3168, Australia
[b] Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, Minnesota, USA
[c] Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, USA
[d] Department of Clinical Sciences, Faculty of Medicine, Lund University, Malmö, 21428, Sweden
[e] Public Dental Service of Skane, Lund, 22647, Sweden
[f] Biostatistics Unit, Division of Research Methodology, School of Public Health and Preventive Medicine, Faculty of Medicine, Nursing, and Health Sciences, Monash University, Melbourne, 3004, Australia

HIGHLIGHTS

- Discovered a miRNA meta-signature associated with T2D via a data-driven pipeline.
- Validated in-silico findings against existing evidence and via downstream analyses.
- Meta-signature could help decode etiologic mechanisms and therapeutic targets of T2D.
- Broader utility of the pipeline for biomedical evidence synthesis is envisioned.

ARTICLE INFO

Keywords:
Biomarkers
Differential expression
High throughput transcriptomics
Meta-analysis
Micro-RNAs
Type 2 diabetes

ABSTRACT

Background: MicroRNAs (miRNAs) are sought-after biomarkers of complex, polygenic diseases such as type 2 diabetes (T2D). Data-driven biocomputing provides robust and novel avenues for synthesizing evidence from individual miRNA seq studies.

Objective: To identify miRNA markers associated with T2D, via a data-driven, biocomputing approach on high throughput transcriptomics.

Materials and methods: The pipeline consisted of five sequential steps using miRNA seq data retrieved from the National Center for Biotechnology Information Gene Expression Omnibus platform: systematic review; identification of differentially expressed miRNAs (DE-miRNAs); meta-analysis of DE-miRNAs; network analysis; and downstream analyses. Three normalization algorithms (trimmed mean of M-values; upper quartile; relative log expression) and two meta-analytic algorithms (robust rank aggregation; Fisher's method of *p*-value combining) were integrated into the pipeline. Network analysis was conducted on *miRNet 2.0* while enrichment and over-representation analyses were conducted on *miEAA 2.0*.

Results: A total of 1256 DE-miRNAs (821 downregulated; 435 upregulated) were identified from 5 eligible miRNA seq datasets (3 circulatory; 1 adipose; 1 pancreatic). The meta-signature comprised 9 miRNAs (*hsa-miR-15b-5p; hsa-miR-33b-5p; hsa-miR-106b-3p; hsa-miR-106b-5p; hsa-miR-146a-5p; hsa-miR-483-5p; hsa-miR-539-3p; hsa-miR-1260a; hsa-miR-4454*), identified via the two meta-analysis approaches. Two hub nodes (*hsa-miR-106b-5p; hsa-miR-15b-5p*) with above-average degree and betweenness centralities in the miRNA-gene interactions network were identified. Downstream analyses revealed 5 highly conserved- (*hsa-miR-33b-5p; hsa-miR-15b-5p; hsa-miR-106b-3p; hsa-miR-106b-5p; hsa-miR-146a-5p*) and 7 highly confident- (*hsa-miR-33b-5p; hsa-miR-15b-5p; hsa-miR-106b-3p; hsa-miR-106b-5p; hsa-miR-146a-5p; hsa-miR-483-5p; hsa-miR-539-3p*) miRNAs. A total of 288 miRNA-

* Corresponding author.
E-mail address: kushan.ranakombu@monash.edu (K. De Silva).

https://doi.org/10.1016/j.heliyon.2022.e08886
Received 27 August 2021; Received in revised form 23 November 2021; Accepted 29 January 2022

disease associations were identified, in which 3 miRNAs (*hsa-miR-15b-5p; hsa-miR-106b-3p; hsa-miR-146a-5p*) were highly enriched.

Conclusions: A meta-signature of DE-miRNAs associated with T2D was discovered via in-silico analyses and its pathobiological relevance was validated against corroborating evidence from contemporary studies and downstream analyses. The miRNA meta-signature could be useful for guiding future studies on T2D. There may also be avenues for using the pipeline more broadly for evidence synthesis on other conditions using high throughput transcriptomics.

1. Introduction

Type 2 diabetes (T2D) is a chronic, polygenic disorder of multifactorial etiology, which affected nearly 462 million individuals worldwide in 2017, equaling to 6.28% of the global population [1]. According to the projections of the International Diabetes Federation, the prevalence of T2D will increase by 25% by 2030 and 51% by 2045 [2]. The natural history of T2D entails a relatively long, early asymptomatic period characterized by subclinical disease, posing difficulties for its timely diagnosis. The magnitude of this diagnostic dilemma is exemplified by the reports that nearly half of the population with diabetes across the world remains undetected [3]. Classic hallmarks of T2D include insulin resistance primarily in skeletal muscle and adipose tissue, worsening pancreatic β-cell failure, hyperglycemia, liver adiposity, and glucose toxicity mediated chronic inflammation, inducing multiple direct and indirect complications. While genetics, environment, and gene-environment interactions all likely play key roles in the onset of T2D, its pathogenic mechanisms are not fully known.

Micro-RNAs (miRNAs) are highly sought-after as novel biomarkers of complex diseases including T2D [4], as they demonstrate a high degree of stability and reproducibility in various tissues and body fluids [5]. Besides enhancing current diagnostic efforts, deeper analysis of miRNA markers could also shed light on disease etiologies, guide drug target identification and facilitate the ultimate development of precision therapeutics [6]. These small (∼22 nucleotides), evolutionarily well-conserved, non-coding, single-stranded molecules have possible roles in multiple pathological and physiological phenomena such as cellular communications, angiogenesis, immune responses, and metastasis via gene-regulating effects on target mRNA translation [7]. Over 60% of the protein-coding genes in the human genome are reportedly regulated by miRNAs [8]. Recent studies ascribed certain miRNAs to T2D pathogenesis-related functions such as insulin action and secretion [9], insulin resistance and β-cell activity [10], glucose metabolism [11], as well as adipogenesis and obesity [12].

Data-driven research forms a pillar of precision medicine whereas multi-omics insights are fundamental to personalizing care pathways for

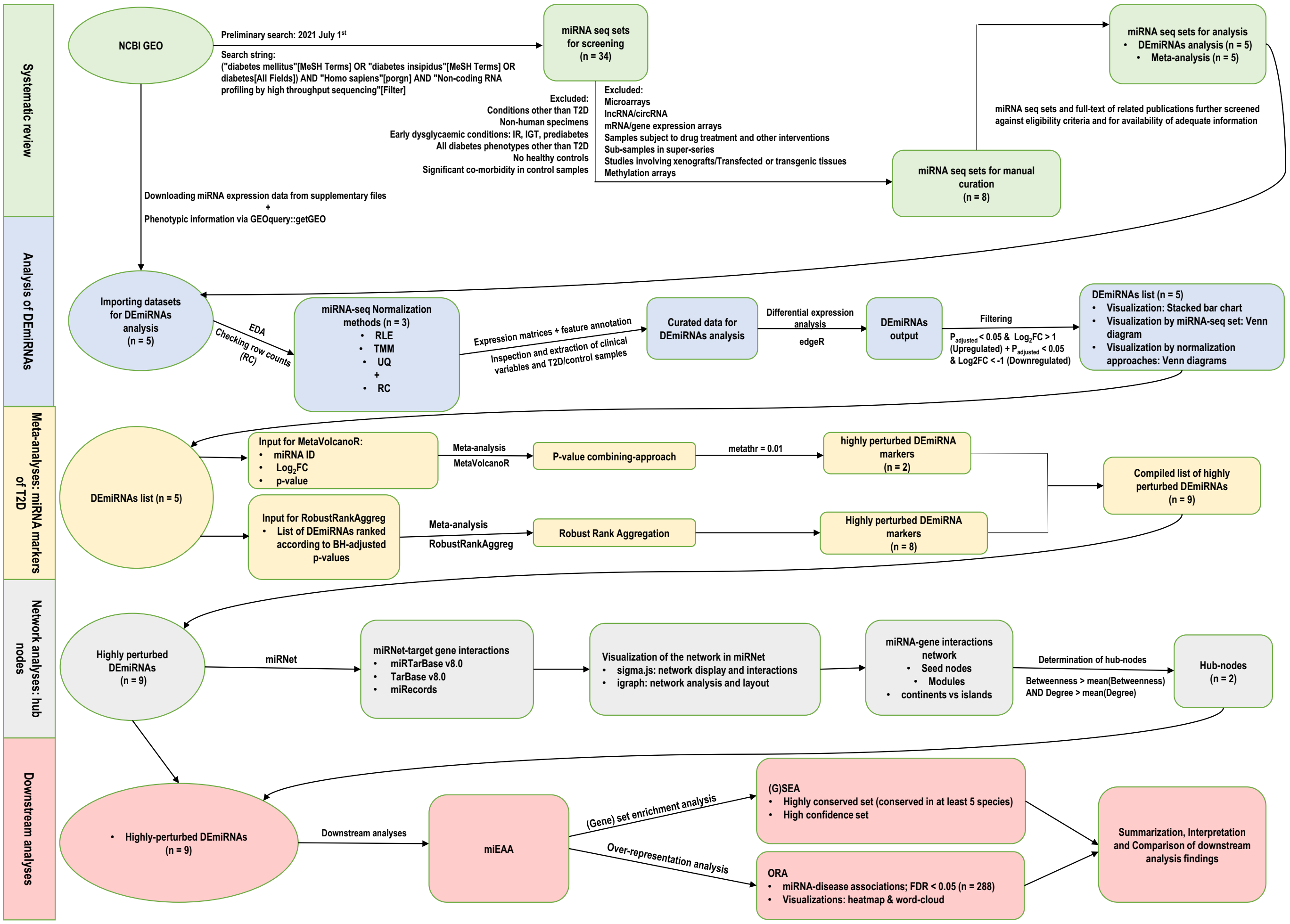
Systematic review
NCBI GEO
Preliminary search: 2021 July 1st
Search string:
("diabetes mellitus"[MeSH Terms] OR "diabetes insipidus"[MeSH Terms] OR diabetes[All Fields]) AND "Homo sapiens"[porgn] AND "Non-coding RNA profiling by high throughput sequencing"[Filter]
miRNA seq sets for screening (n = 34)
Excluded:
Conditions other than T2D
Non-human specimens
Early dysglycaemic conditions: IR, IGT, prediabetes
All diabetes phenotypes other than T2D
No healthy controls
Significant co-morbidity in control samples
Excluded:
Microarrays
lncRNA/circRNA
mRNA/gene expression arrays
Samples subject to drug treatment and other interventions
Sub-samples in super-series
Studies involving xenografts/Transfected or transgenic tissues
Methylation arrays
miRNA seq sets for manual curation (n = 8)
miRNA seq sets and full-text of related publications further screened against eligibility criteria and for availability of adequate information
miRNA seq sets for analysis
• DEmiRNAs analysis (n = 5)
• Meta-analysis (n = 5)
Downloading miRNA expression data from supplementary files
+
Phenotypic information via GEOquery::getGEO
Analysis of DEmiRNAs
Importing datasets for DEmiRNAs analysis (n = 5)
EDA
Checking row counts (RC)
miRNA-seq Normalization methods (n = 3)
• RLE
• TMM
• UQ
+
• RC
Expression matrices + feature annotation
Inspection and extraction of clinical variables and T2D/control samples
Curated data for DEmiRNAs analysis
Differential expression analysis
edgeR
DEmiRNAs output
Filtering
$P_{adjusted} < 0.05$ & $Log_2FC > 1$ (Upregulated) + $P_{adjusted} < 0.05$ & Log2FC < -1 (Downregulated)
DEmiRNAs list (n = 5)
• Visualization: Stacked bar chart
• Visualization by miRNA-seq set: Venn diagram
• Visualization by normalization approaches: Venn diagrams
Meta-analyses: miRNA markers of T2D
DEmiRNAs list (n = 5)
Input for MetaVolcanoR:
• miRNA ID
• Log_2FC
• p-value
Meta-analysis
MetaVolcanoR
P-value combining-approach
metathr = 0.01
highly perturbed DEmiRNA markers (n = 2)
Input for RobustRankAggreg
• List of DEmiRNAs ranked according to BH-adjusted p-values
Meta-analysis
RobustRankAggreg
Robust Rank Aggregation
Highly perturbed DEmiRNA markers (n = 8)
Compiled list of highly perturbed DEmiRNAs (n = 9)
Network analyses: hub nodes
Highly perturbed DEmiRNAs (n = 9)
miRNet
miRNet-target gene interactions
• miRTarBase v8.0
• TarBase v8.0
• miRecords
Visualization of the network in miRNet
• sigma.js: network display and interactions
• igraph: network analysis and layout
miRNA-gene interactions network
• Seed nodes
• Modules
• continents vs islands
Determination of hub-nodes
Betweenness > mean(Betweenness) AND Degree > mean(Degree)
Hub-nodes (n = 2)
Downstream analyses
• Highly-perturbed DEmiRNAs (n = 9)
Downstream analyses
miEAA
(Gene) set enrichment analysis
Over-representation analysis
(G)SEA
• Highly conserved set (conserved in at least 5 species)
• High confidence set
ORA
• miRNA-disease associations; FDR < 0.05 (n = 288)
• Visualizations: heatmap & word-cloud
Summarization, Interpretation and Comparison of downstream analysis findings

Figure 1. Methodological workflow illustrating the five-step biocomputing pipeline employed in the study.

Table 1. The distribution of differentially expressed miRNA identified in each RNA-seq dataset.

Dataset	Tissue type/biospecimen	# DE-miRNA[a]	# Upregulated	# Downregulated	p-value[b]
GSE139577	Circulatory	174	80	94	0.1339
GSE109266	Circulatory	838	207	631	<0.0001
GSE90028	Circulatory	172	99	73	0.0051
GSE174502	Adipose	62	45	17	<0.0001
GSE52314	Pancreatic	10	4	6	0.3833
Total[c]	–	1256	435	821	<0.0001

a: defined as miRNA with absolute $\log_2$ fold change >1 and Benjamini- Hochberg adjusted p-value < 0.05; b: computed by chi-squared test for two proportions; c: including genes commonly expressed across multiple tissues.
DE-miRNA: differentially expressed miRNA.

complex, heterogeneous diseases such as T2D. Avenues for high level evidence synthesis are offered by expanding omics data repositories, continuing advancement of data science, and the diminishing cost of next generation sequencing techniques. Such large-scale studies could also allow for the discovery of potentially novel pathobiological pathways of heterogeneous diseases such as T2D, that would not be revealed in smaller individual studies. Moreover, high throughput miRNA sequencing is increasingly used and is preferred over traditional microarrays, due to its added merits such as less technical biases, higher multiplexing power, and the ability to detect previously uncharacterized miRNAs [13]. While individual miRNA seq studies focused on T2D are steadily reported, common drawbacks inherent in such studies are noteworthy. Individual miRNA-disease association studies tend to be under-powered due to small sample sizes, suffer from non-negligible experimental noise, and are prone to yielding inconsistent findings [14]. Meta-analysis is a viable approach for alleviating the biases of individual studies and merging their findings in order to derive high-level evidence. Thus far, only a few meta-analyses of profiling studies focused on miRNA-T2D associations have been conducted, all of which resorted to pooling published summary measures [15, 16, 17]. However, data-driven meta-analysis of raw, original omics data emanating from individual studies, though computationally intensive, could yield more reliable and robust results, compared to using summary measures [18]. With the growing availability of large raw data repositories such as the National Center for Biotechnology Information Gene Expression Omnibus (NCBI GEO) [19], such novel computational analyses are now possible.

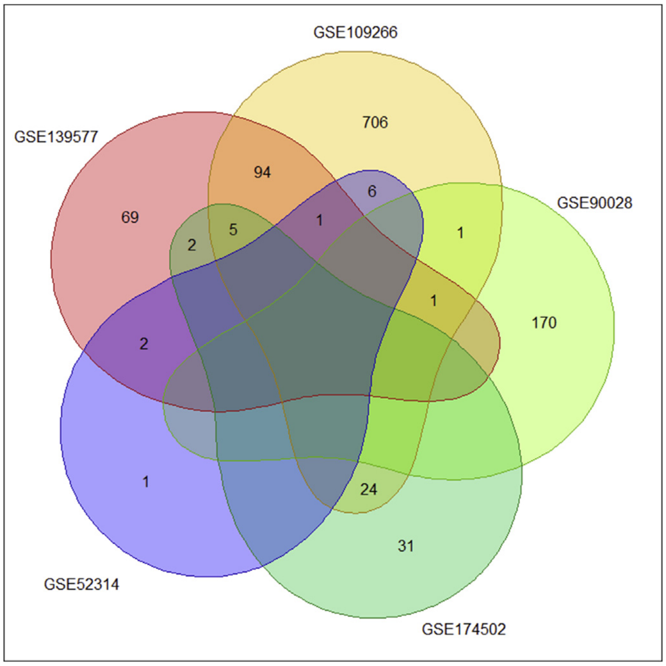

Figure 2. Venn diagram depicting the number and overlap of differentially expressed miRNA identified by each high throughput transcriptomics set. Number of shared and non-shared miRNAs across the five sets are discernible.

Rigorous biocomputing pipelines consisting of impeccable preprocessing steps are required to gain accurate insights from miRNA seq studies. In particular, normalization strategies are critically important, as they are upstream factors that can strongly influence results [20]. While there is no gold standard for miRNA seq data normalization [21], various optimal methods have been proposed by previous studies [22]. Trimmed mean of M (TMM), upper quartile (UQ), and relative log expression (RLE) are standard normalization algorithms in an optimal miRNA seq preprocessing pipeline [23] and various combinations of these in previous studies yielded discordant results [20, 21, 22]. Meta-analytic algorithms regularly used on high throughput transcriptomics include robust rank aggregation (RRA) [24], p-value combining methods [25], and vote counting [26]. Of these, vote counting was found inadequate as a stand-alone meta-analytic strategy [27].

In this context, we aimed to identify miRNA markers of T2D via an extensive biocomputing pipeline utilizing raw miRNA seq data in the NCBI GEO repository. This consisted of five sequential steps: systematic review; identification of differentially expressed miRNAs (DE-miRNAs); meta-analysis of DE-miRNAs; network analysis; and downstream analyses. Three normalization strategies (TMM, UQ, RLE) and two meta-analytic processes (RRA, Fisher's method of p-value combining) were incorporated into the pipeline.

2. Methods

The analytic workflow is presented in Figure 1.

2.1. Systematic review

A preliminary search on the NCBI GEO platform was conducted on July 1, 2021 using the search string ("diabetes mellitus"[MeSH Terms] OR "diabetes insipidus"[MeSH Terms] OR diabetes[All Fields]) AND "Homo sapiens"[porgn] AND "Non-coding RNA profiling by high throughput sequencing"[Filter]. Publicly available data on the NCBI GEO were thus utilized in this study and no new unpublished data were included. All datasets resulting from this preliminary search were manually assessed against pre-defined eligibility criteria. We included miRNA seq sets focused on T2D with well-defined cases and controls, conducted on humans, with adequate information on tissues of origin, sample sizes, and feature annotation. Microarrays, lncRNA/circRNA, datasets with no healthy controls, samples subjected to drug treatments/other interventions, those focused on diabetes phenotypes other than T2D, non-human studies, and super-series were excluded. Datasets deemed eligible or unclear following this primary screening were included in secondary screening with downloaded raw data files and related publications.

2.2. Data retrieval and pre-processing

All eligible raw miRNA seq supplementary files were retrieved from the NCBI GEO platform while their phenotypic information was acquired using the '*getGEO*' function of the *GEOquery* R package [28]. In each miRNA seq set, multiple probes hybridized to the same miRNA were collapsed by calculating mean counts across samples. Lowly expressed probes, defined as those with a mean raw count across samples <5, were removed.

2.3. Normalization

We applied three normalization techniques (TMM, UQ, RLE) on raw miRNA seq sets, via the *edgeR* R package [29]. A brief description of each method follows.

2.3.1. Trimmed mean of M-values (TMM)

Considered as a robust technique for standardizing miRNA seq synthesis ratios using a weighted trimmed mean of the log expression ratios, this method was first proposed by Robinson and Oshlack [30]. The TMM algorithm is based on fold-change (M) and absolute expression (A) values, which are defined below.

For a specific miRNA 'a' with expression levels K_{a1} and K_{a2} under the two conditions of interest, and N_1 and N_2 denoting the total number of reads in the two libraries, respectively:

$$M = \log_2\left(\frac{K_{a1}/N_1}{K_{a2}/N_2}\right)$$

$$A = \frac{1}{2}\log_2((K_{a1} / N_1)(K_{a2} / N_2))$$

Based on the assumption that most miRNAs are not differentially expressed, the normalization factor for a given sample against a reference sample is estimated as follows:

$$\log_2(d_b TMM) = \frac{\sum_{a\in A'} \omega_{ab} M_{ab}}{\sum_{a\in A'} \omega_{ab}}$$

where,

$$M_{ab} = \log_2(K_{ab} / N_b) / (K_{ar} / N_r),$$

$$\omega_{ab} = (N_b - K_{ab}) / N_b K_{ab} + (N_r - K_{ar}) / N_r K_{ar},$$

$K_{ab},\ \ K_{ar} > 0.$

K_{ab} = *Read counts of the miRNA "a" in sample "b"*K_{ar} = *Read counts of the miRNA "a" in sample "r"*d_b = *scaling factor of the* b^{th} *sample*N_b = *Total number of reads per sample* b^{th},

N_r = *Total number of reads per reference sample* $'r'$,

A' = *miRNA set with untrimmed logfold changes & absolute expression values.*

The TMM is calculated as a weighted average after certain upper and lower percentages of both M and A values (by default, 30% of M and 5% of A, by the *edgeR* package [29]) are removed, while precision (inverse variance) weights are used to account for disparities in read counts [30]. The library whose upper quartile is closest to the mean upper quartile is chosen as the reference by *edgeR* [29].

2.3.2. Upper quartile (UQ)

In this method, following exclusion of transcripts with null raw counts in all libraries, the scaling factor is determined by the 75th percentile of reads [20], as follows.

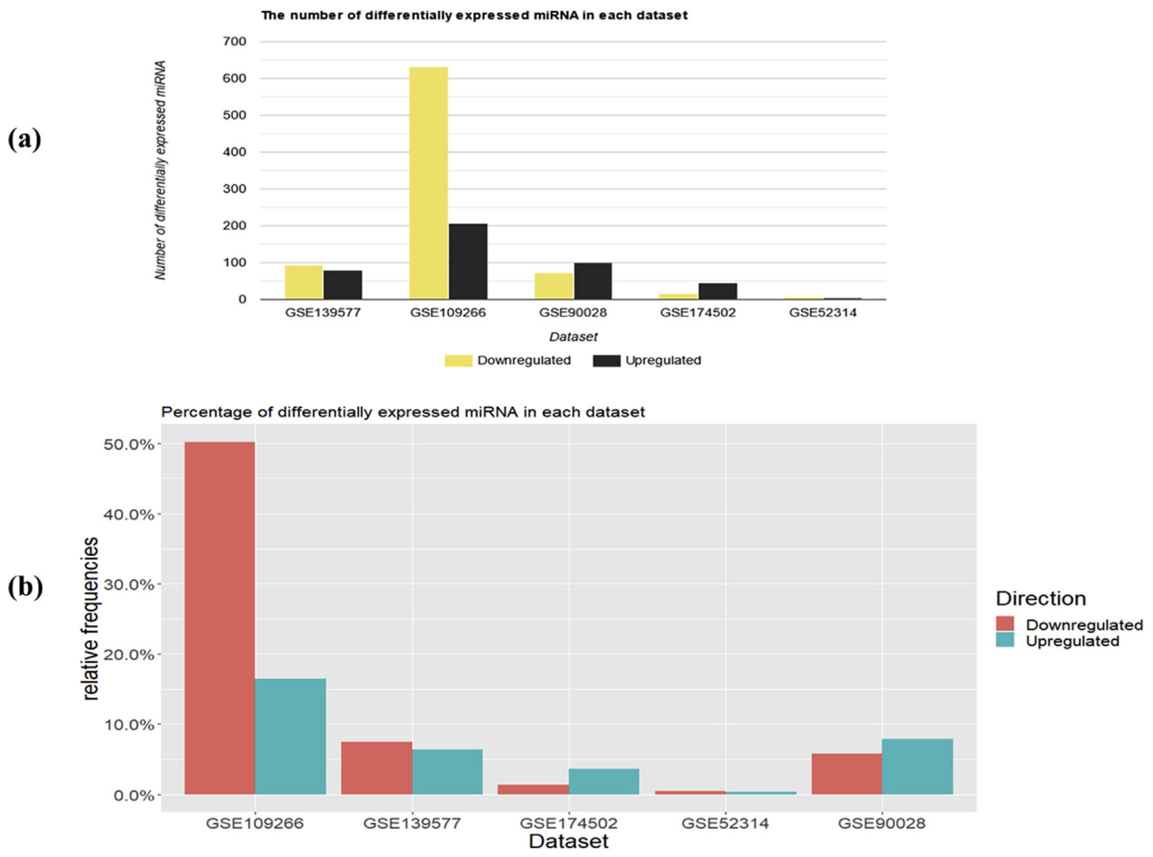

Figure 3. Stacked bar charts depicting (a) the number of up- and down-regulated miRNA in each high throughput transcriptomic set (b) the percentage of up- and down-regulated miRNA in each high throughput transcriptomic set.

$$d_b^{UQ} = Q\left(\frac{K_{ab}}{\sum_{a=1}^{A} K_{ab}}\right)$$

where,

$$Q = 3/4(N+1)^{th}$$

Scaling entails the division of raw reads by the upper quartile of counts in each sample and multiplication by the mean upper quartile across all samples, as proposed by Bullard et al [31].

2.3.3. Relative log expression (RLE)

This algorithm, proposed by Anders and Huber [32], is also motivated by the assumption that a majority of miRNAs are not differentially expressed. The scaling factor, under this method, can be presented as follows:

$$d_{abc}^{RLE} = \frac{K_{abc}}{median_a\left(\frac{K_{abc}}{\sqrt[BC]{\prod_{b=1}^{B}\prod_{c=1}^{C}K_{abc}}}\right)}$$

where,

K_{abc} = Observed number of reads/counts of the miRNA:

$$a \in \{1, \ldots\ldots, A\}$$

In the condition of interest:

$$b \in \{1, \ldots\ldots\ldots, B\}$$

For the biological replicate:

$$c \in \{1, \ldots\ldots\ldots, C\}$$

The median library is enumerated via the geometric mean of all samples while the median ratio of each sample to the median library is considered as the scaling factor.

2.4. Differential expression analysis

Using the *edgeR* package in *R* version 4.1.0, differential expression analysis was conducted on four matrices (matrices normalized by the three methods described above, and non-normalized, original raw counts) of each miRNA seq set. The miRNAs with absolute $\log_2$Fold Change (FC) > 1 and Benjamini-Hochberg adjusted *p*-value < 0.05 were defined as differentially expressed. For those DE-miRNAs found in multiple matrices, we selected the row containing the highest absolute $\log_2$FC. Of note, one study (GSE109266) had directly provided information on DE-miRNAs instead of raw supplementary files, which was used in the current analysis.

2.5. Meta-analysis of differentially expressed miRNA

In order to derive a genome-wide meta-signature of miRNA markers associated with T2D, we meta-analyzed the DE-miRNAs using two strategies: Fisher's *p*-value combining method via the *MetaVolcanoR* package [33] and robust rank aggregation implemented in the *RobustRankAggreg* package [24]. Brief explanations of the two algorithms follow.

2.5.1. Fisher's p-value combining method

Fisher's method combines *p*-values from a set of independent tests, based on the same null hypothesis, according to the following formula:

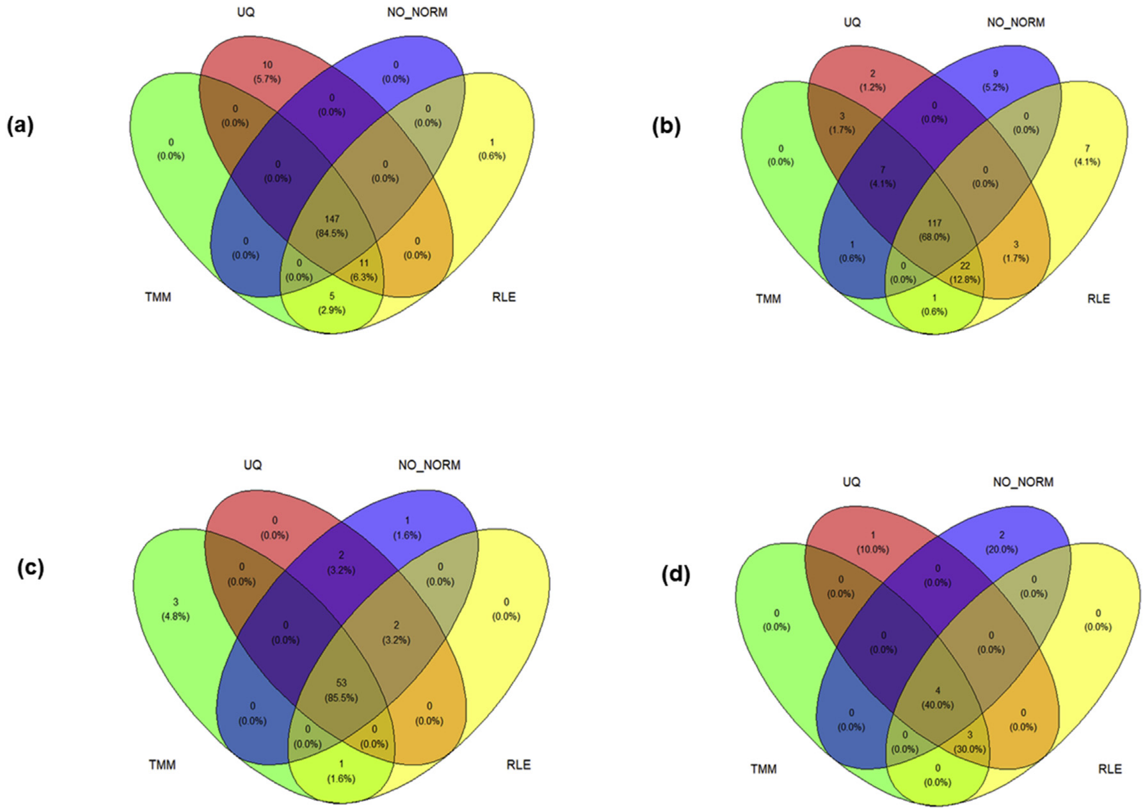

Figure 4. Distribution of differentially expressed miRNA in each dataset and the overlap across normalization methods (a) GSE139577: circulatory (b) GSE90028: circulatory (c) GSE174502: adipose (d) GSE52314: pancreatic [GSE109266 (circulatory) provided differentially expressed miRNA information directly and these were used for analyses]. NO_NORM = no normalization – raw counts used; RLE = relative log expression normalization; TMM = trimmed mean of M-values normalization; UQ = upper quartile normalization.

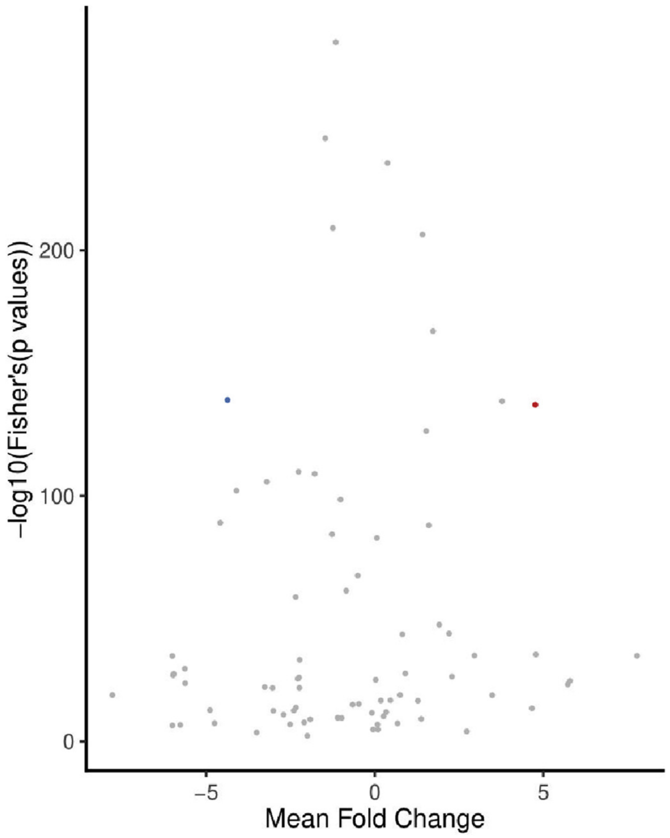

Figure 5. Volcano plot depicting Fisher's *p*-value combining meta-analysis output employed via *MetaVolcanoR*. Two highly perturbed genes were identified at metathr = 0.01 (default threshold specified by the package). Red and blue dots indicate *hsa-miR-33b-5p* (upregulated) and *hsa-miR-539-3p* (downregulated), respectively.

$$\chi^2_{2k} \sim -2\sum_{j=1}^{k} \ln(P_j),\ \text{ given }\ P \sim Unif(0,1)$$

where,

P_j = *p*-value of the jth hypothesis test

k = number of tests being pooled

Under the null hypothesis for a given test, its *p*-value conforms to a uniform distribution on the interval [0,1] and the sum of *k* independent tests attains a chi-squared distribution with 2*k* degrees of freedom. Recent reports confirmed the robustness of this method, including its ability to detect incomplete associations, outperforming other conventional meta-analysis techniques [25].

2.5.2. Robust rank aggregation

Based on the null hypothesis of unassociated inputs, RRA is a parameter-free meta-analytic algorithm which is resistant to outliers, bias, and noise.

For a normalized rank vector ν, reordered as $\nu_1 \le \ldots\ldots \le \nu_n$, $\beta_{k,n}(\nu)$ indicates the probability of $\acute{\upsilon}_k \le \nu_k$, given the rank vector $\acute{\upsilon}$ is produced by the null model of uniform distribution. The probability of $\acute{\upsilon}_k \le$ y, under the null model, can then be written as a binomial probability:

$$\beta_{k,n}(y) := \sum_{l=k}^{n} \binom{n}{l} y^l (1-y)^{n-l}$$

The meta score ρ of the rank vector ν, is expressed as the minimum of *p*-values, based on the assumption that the number of informative ranks is not known:

$$\rho(\nu) = min_{k=1,\ldots\ldots,n}\ \beta_{k,n}(\nu)$$

Details of the RRA algorithm, proposed by Kolde et al., are available elsewhere [24].

2.6. Network analysis

Network analysis of the meta-signature was carried out on the ***miRNet 2.0*** [34], which, by default, defines hub-nodes as those miRNAs with above-average degree and betweenness in the miRNA-target gene interactions network. These two measures of centrality can be defined as follows:

2.6.1. Degree centrality

$$\text{Degree centrality}(h) = \frac{d(h)}{n-1}$$

where,

d(h) = degree of the vertex ‘h’

n = sum of all the vertices in the network graph

2.6.2. Betweenness centrality

$$\text{Betweenness centrality}(h) = \frac{2}{(n-1)(n-2)} \sum_{h \neq a \neq b \in H} \frac{\varphi_{a,b}(h)}{\varphi_{a,b}}$$

Table 2. The meta-signature of micro-RNA markers associated with type 2 diabetes identified by the two meta-analytic strategies: *p*-value combining method and robust rank aggregation.

miRNA	Meta-*p* value/ρ score[a]	Expressed in	Regulation
Identified by meta-analysis with *p*-value combining method			
hsa-miR-33b-5p	6.87E-138	Adipose, Circulatory	Up
hsa-miR-539-3p	7.92E-140	Circulatory	Down
Identified by meta-analysis with robust rank aggregation			
hsa-miR-15b-5p	0.0224214885	Circulatory	Down
hsa-miR-106b-3p	0.0490074419	Circulatory	Down
hsa-miR-106b-5p	0.0260325689	Circulatory	Down
hsa-miR-146a-5p	0.0438733059	Circulatory	Down
hsa-miR-483-5p	0.0241123607	Circulatory	Down
hsa-miR-539-3p	0.0003613095	Circulatory	Down
hsa-miR-1260a	0.0224214885	Circulatory	Down
hsa-miR-4454	0.0447624691	Circulatory	Down

a: *p*-value combining meta-analysis in MetaVolcanoR computes a meta-*p* score based on meta $\log_2$ fold-change across studies while robust rank aggregation computes a ρ score based on *p*-values of DEmiRNAs in individual studies.

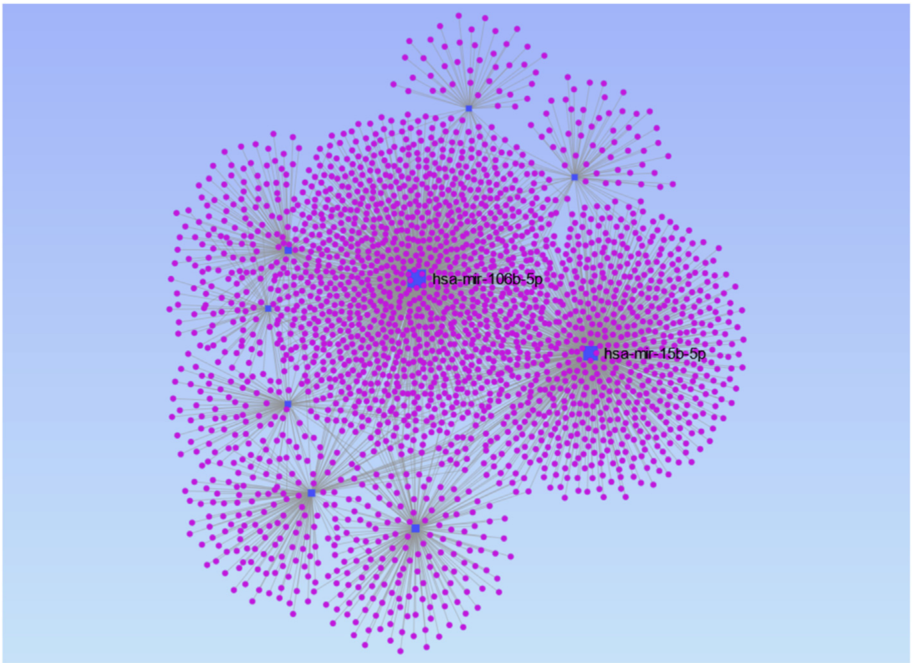

Figure 6. The miRNA-gene interactions network visualized in *miRNet 2.0*. The two hub-nodes (*hsa-miR-106b-5p* & *hsa-miR-15b-5p*) with above-average degree and betweenness are highlighted.

Table 3. Topological characteristics of the miRNA-gene interactions network.

miRNA ID	Degree	Betweenness	Hub node status
	Mean = 296	Mean = 510899.445555	
hsa-miR-15b-5p	**760**	**1322515.59471805**	**Yes**
hsa-miR-146a-5p	203	367850.990663118	No
hsa-miR-106b-5p	**1091**	**1871815.20015558**	**Yes**
hsa-miR-33b-5p	101	166566.827091811	No
hsa-miR-106b-3p	76	123910.221308144	No
hsa-miR-483-5p	143	253701.019330487	No
hsa-miR-1260a	162	288253.085747575	No
hsa-miR-4454	58	96535.6018030646	No
hsa-miR-539-3p	67	106946.459182154	No

where,

$\varphi_{a,b}$ = number of shortest paths between vertices a and b

$\varphi_{a,b}(h)$ = number of shortest paths between vertices a & b passing through vertex h

n = Total number of vertices in the network/graph

As the densely and tightly connected hub nodes in biological networks [35] tend to be critical functional regulators as well, they are sought after in precision medicine approaches [36]. We visualized the miRNA-target gene interactions network on *miRNet 2.0* and determined the network attributes including hub nodes, modules, and continents versus islands.

2.7. Downstream analyses

Functional enrichments were detected via downstream analyses on the miRNA Enrichment Analysis and Annotation Tool 2.0 (*miEAA 2.0*) [37]. Enrichment analysis entailed the detection of highly conserved- (defined as conserved in at least 5 species) and high confidence- (those miRNAs with high probability expression profiles via small read mappings) sets. Through over-representation analysis, we detected miRNA-disease associations at a false discovery rate <0.05, and visualized them on *miEAA2.0* as disease-miRNA association map and word-cloud figures.

3. Results

3.1. Systematic review

As shown in S1 Table, the preliminary search identified 34 eligible datasets, 8 of which were selected for secondary screening. A summary of these 8 datasets is provided in S2 Table. Three of these (GSE151496, GSE160308, GSE178721) had no adequate and conclusive information on the diabetes phenotype, and were therefore excluded. The final five sets selected for subsequent analysis emanated from circulatory (n = 3; GSE139577, GSE109266, GSE90028), adipose (n = 1; GSE174502), and pancreatic (n = 1; GSE52314) tissues. These comprised 113 samples in total (97 circulatory; 9 adipose; 7 pancreatic), of which 56 were T2D- and 57 were non-T2D control- samples (S2 Table).

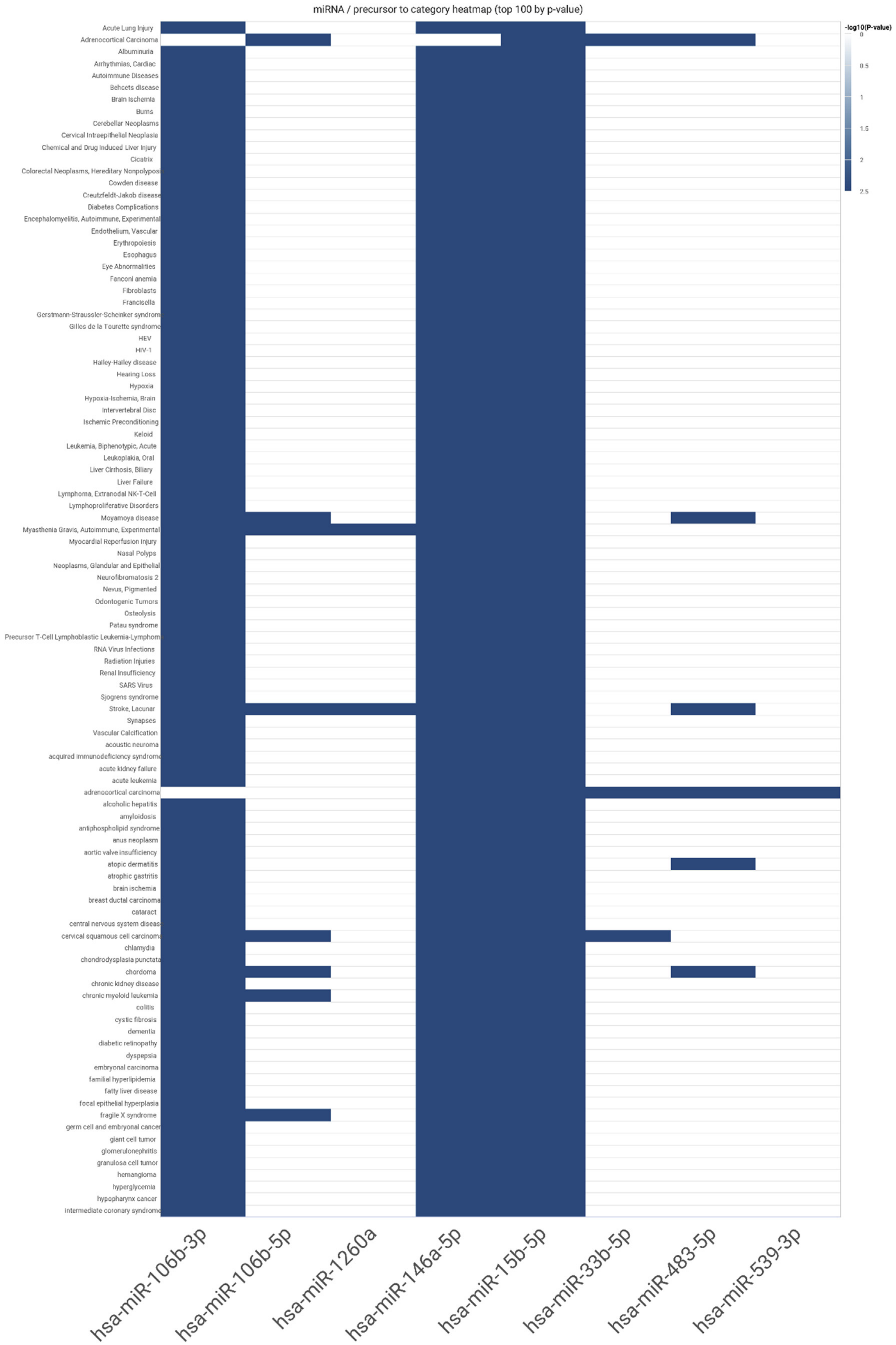

Figure 7. The default disease-miRNA associations map illustrating the top 100 associations, produced by miEAA 2.0. Three miRNAs (hsa-miR-15b-5p, hsa-miR-106b-3p, hsa-miR-146a-5p) are highly over-represented.

3.2. DE-miRNAs

As shown in Table 1, 1256 DE-miRNAs in total were identified across the five datasets. The number of DE-miRNAs in circulatory, adipose, and pancreatic tissues were 1184, 62, and 10, respectively. The majority of DE-miRNAs (977/1256) were exclusive to each dataset (GSE109266 = 706; GSE90028 = 170; GSE139577 = 69; GSE174502 = 31; GSE52314 = 1) (Figure 2). The dataset containing the greatest number of DE-miRNAs was GSE109266, while the total number of down-regulated miRNAs were significantly larger than the sum of up-regulated miRNAs (821 down-regulated and 435 up-regulated; $p < 0.0001$) (Table 1; Figure 3). The proportions of miRNAs identified as differentially expressed by all three normalization algorithms as well as via raw counts in the four datasets: GSE139577; GSE90028; GSE174502; and GSE 52314 were 147/174; 117/172; 53/62; and 4/10, respectively (Figure 4).

3.3. Meta-analyses

The meta-signature consisted of two miRNA markers identified by Fisher's *p*-value combining method (*hsa-miR-33b-5p, hsa-miR-539-3p*) (Figure 5) and eight miRNA markers elucidated by RRA (*hsa-miR-15b-5p, hsa-miR-106b-3p, hsa-miR-106b-5p, hsa-miR-146a-5p, hsa-miR-483-5p, hsa-miR-539-3p, hsa-miR-1260a, hsa-miR-4454*) (Table 2). One miRNA marker, i.e. *hsa-miR-539-3p* was identified by both methods, resulting in a meta-signature of 9 miRNAs. Of the 9 miRNAs, 8 were down-regulated while a single miRNA (*hsa-miR-33b-5p*) was up-regulated. Also, *hsa-miR-33b-5p* was differentially expressed in both circulatory and adipose tissues, while all other miRNAs constituting the meta-signature were exclusively found in circulatory tissues (Table 2).

3.4. Network analyses

The miRNA-gene interactions network derived by *miRNet 2.0* is presented in Figure 6, while the details of the network are shown in S3 Table. Two miRNAs with above average betweenness and degree centralities, namely, *hsa-miR-15b-5p* and *hsa-miR-106b-5p* were demarcated as hub nodes (Table 3, Figure 6). Topologically, the network constituted of nine modules, each clustered around a specific miRNA. Those two modules around the two hub nodes formed large and dense sub-networks characterized as continents while the remainder formed smaller sub-networks stipulated as islands (Figure 6).

3.5. Downstream analyses

Functional enrichment analyses determined five miRNA markers (*hsa-miR-33b-5p; hsa-miR-15b-5p; hsa-miR-106b-3p; hsa-miR-106b-5p; hsa-miR-146a-5p*) as highly conserved while seven miRNAs (*hsa-miR-33b-5p; hsa-miR-15b-5p; hsa-miR-106b-3p; hsa-miR-106b-5p; hsa-miR-146a-5p; hsa-miR-483-5p; hsa-miR-539-3p*) constituted the highly confident set (S4 Table). Over-representation analysis identified 288 significant miRNA-disease associations in total, including hyperglycemia, diabetes complications, diabetic retinopathy and many other chronic cardiometabolic diseases and neoplastic conditions, the details of which are presented in S5 Table. As shown in Figure 7, three miRNAs (*hsa-miR-15b-5p, hsa-miR-106b-3p, hsa-miR-146a-5p*) were highly enriched across a majority of miRNA-disease associations. The top 100 enriched diseases associated with the 9 miRNA markers according to Benjamini-Hochberg adjusted *p*-values are visualized as a word-cloud in S1 Figure.

4. Discussion

In this in-silico analysis, we identified a genome-wide meta-signature of nine miRNA markers associated with T2D via a data-driven, biocomputing workflow. Biological plausibility of the findings was validated through downstream analyses and in the backdrop of contemporary literature. We envisage that the proposed biocomputing workflow would be widely adoptable for studies using data-driven approaches on high throughput transcriptomics for biomarker discovery, including systemic or tissue non-specific meta-signatures.

A recent high throughput plasma sequencing study revealed the down-regulation of three of the circulating miRNAs present in the identified meta-signature (*hsa-miR-15b-5p, hsa-miR-106b-3p, hsa-miR-106b-5p*), in association with incident T2D [38]. Multiple studies attest to the upregulation of *hsa-miR-33b-5p* in circulatory and adipose tissues in the presence of dysglycemia. It has been reported that increased *hsa-miR-33b-5p* levels associate with both insulin resistance and adiposity [39], and leads to impaired insulin signaling as well as reduced fatty acid oxidation [40]. Moreover, upregulation of *hsa-miR-33b-5p* in cultured adipocytes inhibited GLUT4 – a key gene involved in insulin-mediated glucose transport and homeostasis [41]. Conversely, the inhibition of miR-33 was found to be able to overcome the harmful effects of diabetes on atherosclerosis [42]. Lower circulating levels of *hsa-miR-146a* were associated with both T2D and prediabetes, playing a potential role in inflammation, as per a case-control study [43]. Another experimental study revealed that *hsa-miR-146a-5p* is downregulated in human aortic endothelial cells in the presence of high glucose, acting as a mediator of high-glucose induced endothelial inflammation [44]. A nested case-control study found significantly decreased *hsa-miR-483-5p* levels in the plasma of individuals with prediabetes which indicated its potential utility as a diagnostic predictor/circulating biomarker of β-cell function [45]. Significant downregulation of *hsa-miR-539-5p* was previously detected in the circulation of individuals with uncontrolled diabetes [46] while another study discovered its implications on both diabetes and the heart [47]. Down-regulation of both *hsa-miR-1260a* [48,49] and *hsa-miR-4454* [49] in the circulation of individuals with T2D was also previously reported. Reduced circulating *hsa-miR-4454* levels are associated with early onset obesity as well [50]: a condition that could quadruple the risk of subsequent development of T2D [51]. Concordant findings from contemporary studies with respect to the meta-signature are summarized in S6 Table.

Systems biology insights on miRNA-T2D associations, rendered by network analyses are noteworthy. Since the two hub nodes (*hsa-miR-106b-5p & hsa-miR-15b-5p*) which formed epicenters of the two most densely and extensively connected sub-networks might have crucial functional roles as well [35, 36], further research on their roles on T2D pathogenesis is warranted. Another insightful observation was the disproportionately high representation of three markers (*hsa-miR-15b-5p, hsa-miR-106b-3p, hsa-miR-146a-5p*) within the miRNA-disease associations network, which suggested that they likely have key roles in the development of not only T2D but also other cardiometabolic diseases and co-morbidities via multiple cross-talk and pleiotropic mechanisms. The highly conserved- and confident sets of miRNAs, as well as the large number of miRNA-disease associations identified by downstream analyses, provide important guidance and directions on the miRNA markers of T2D to inform future studies. Therefore, future studies focusing on the functions of the miRNAs constituting the meta-signature, including their roles in pathobiological pathways of T2D and as potential biomarkers of T2D-associated comorbidity, would be informative. Longitudinal studies exploring the mRNA and protein expression of identified miRNAs are also recommended as they might assist us in determining whether plausible targets of these miRNAs predict the onset of incident T2D.

The predominance of circulatory miRNAs in the meta-signature could have resulted from the nature of the origins of data sources. A vast majority of miRNA seq data emanated from circulatory tissues, while only a few DE-miRNAs were uncovered in adipose and pancreatic tissues. It should be noted, however, that the circulating miRNAs are often preferred as minimally-invasive biomarkers. They have clear merits including remarkable stability, relatively easy detectability, high sensitivity, and the ability to provide mechanistic insights via their dynamic expression patterns under different pathological or physiological conditions [52]. The meta-signature derived by current analysis may therefore have greater clinical value as a potential diagnostic tool of T2D.

Nevertheless, the lack of miRNA seq data from other tissue types was a limitation, which may need to be attended to in the quest for unravelling the complex pathobiology of T2D. The divergent nature of datasets emanating from different tissues and producing vastly different amounts of DE-miRNAs, may have influenced the meta-analytic outputs. Consequently, meta-analyses may have lacked sufficient sensitivity to identify all miRNA markers associated with T2D, in the current study.

Given the lack of consensus on gold standard algorithms for normalizing and meta-analyzing high throughput transcriptomics, it would be prudent to apply a nuanced mix of these techniques, as was done in the current analysis, instead of resorting to a random, single method. While both contemporary evidence and downstream analysis outputs corroborated our findings, further in-vitro and in-vivo experimental studies are recommended to validate the meta-signature of T2D derived via in-silico analyses in the present study. As the miRNA seq databases are gradually getting scaled up, we underscore the potential value of data-driven meta-analyses of high throughput transcriptomics which have hitherto been utilized only sparsely, compared to more abundant meta-analyses of the aggregate measures reported in published studies. Lastly, the proposed biocomputing pipeline could be widely adopted as a robust evidence synthesis strategy in data-driven, high throughput transcriptomics studies focused on various other conditions.

5. Conclusions

In this study, a meta-signature of DE-miRNAs associated with T2D was discovered via a data-driven, biocomputing pipeline on high throughput transcriptomics. In-silico findings were validated against corroboratory evidence from contemporary studies and via downstream analyses. The miRNA meta-signature could be useful for guiding future studies such as those aimed at unravelling pathological mechanisms and effective therapeutics of T2D. There may also be avenues for using the proposed pipeline more broadly for evidence synthesis on other conditions using high throughput transcriptomics data.

Declarations

Author contribution statement

Kushan De Silva: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Ryan T. Demmer: Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper.

Daniel Jönsson, Aya Mousa: Analyzed and interpreted the data; Wrote the paper.

Andrew Forbes: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Joanne Enticott: Conceived and designed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Funding statement

This work was supported by a PhD scholarship funded by the Australian Government under Research Training Program (RTP).

Data availability statement

Data associated with this study has been deposited at the National Center for Biotechnology Information Gene Expression Omnibus (NCBI GEO) portal: https://www.ncbi.nlm.nih.gov/geo/

Declaration of interests statement

The authors declare no conflict of interest.

Additional information

Supplementary content related to this article has been published online at https://doi.org/10.1016/j.heliyon.2022.e08886.

References

[1] M.A.B. Khan, M.J. Hashim, J.K. King, R.D. Govender, H. Mustafa, J. Al Kaabi, Epidemiology of type 2 diabetes - global burden of disease and forecasted trends, J. Epidemiol. Glob. Health 10 (1) (2020) 107–111.

[2] P. Saeedi, I. Petersohn, P. Salpea, B. Malanda, S. Karuranga, N. Unwin, et al., Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: results from the international diabetes federation diabetes atlas, 9th edition, Diabetes Res. Clin. Pract. 157 (2019) 107843.

[3] C.J. Weale, D.M. Matshazi, S.F.G. Davids, S. Raghubeer, R.T. Erasmus, A.P. Kengne, et al., MicroRNAs-1299, -126-3p and -30e-3p as potential diagnostic biomarkers for prediabetes, Diagnostics (Basel) 11 (6) (2021) 949.

[4] X. Chen, D. Xie, Q. Zhao, Z.H. You, MicroRNAs and complex diseases: from experimental results to computational models, Briefings Bioinf. 20 (2) (2019) 515–539.

[5] O. Faruq, A. Vecchione, microRNA: diagnostic perspective, Front. Med. 2 (2015) 51.

[6] J. Hanna, G.S. Hossain, J. Kocerha, The potential for microRNA therapeutics and clinical research, Front. Genet. 10 (2019) 478.

[7] C. Catalanotto, C. Cogoni, G. Zardo, MicroRNA in control of gene expression: an overview of nuclear functions, Int. J. Mol. Sci. 17 (10) (2016) 1712.

[8] F. Zhang, D. Wang, The pattern of microRNA binding site distribution, Genes (Basel) 8 (11) (2017) 296.

[9] S. Calderari, M.R. Diawara, A. Garaud, D. Gauguier, Biological roles of microRNAs in the control of insulin secretion and action, Physiol. Genom. 49 (1) (2017) 1–10.

[10] P. Kaur, S. Kotru, S. Singh, B.S. Behera, A. Munshi, Role of miRNAs in the pathogenesis of T2DM, insulin secretion, insulin resistance, and β cell dysfunction: the story so far, J. Physiol. Biochem. 76 (4) (2020) 485–502.

[11] P. Agbu, R.W. Carthew, MicroRNA-mediated regulation of glucose and lipid metabolism, Nat. Rev. Mol. Cell Biol. 22 (6) (2021) 425–438.

[12] M. Zaiou, H. El Amri, A. Bakillah, The clinical potential of adipogenesis and obesity-related microRNAs, Nutr. Metabol. Cardiovasc. Dis. 28 (2) (2018) 91–111.

[13] E. Andrés-León, R. Núñez-Torres, A.M. Rojas, miARma-Seq: a comprehensive tool for miRNA, mRNA and circRNA analysis, Sci. Rep. 6 (2016) 25749.

[14] E. Pala, T. Denkçeken, Differentially expressed circulating miRNAs in postmenopausal osteoporosis: a meta-analysis, Biosci. Rep. 39 (5) (2019), BSR20190667.

[15] H. Zhu, S.W. Leung, Identification of microRNA biomarkers in type 2 diabetes: a meta-analysis of controlled profiling studies, Diabetologia 58 (5) (2015) 900–911.

[16] Y.Z. Liang, J.J. Li, H.B. Xiao, Y. He, L. Zhang, Y.X. Yan, Identification of stress-related microRNA biomarkers in type 2 diabetes mellitus: a systematic review and meta-analysis, J. Diabetes 12 (9) (2020) 633–644.

[17] M. Gholami, S. Asgarbeik, F. Razi, E.N. Esfahani, M. Zoughi, A. Vahidi, et al., Association of microRNA gene polymorphisms with Type 2 diabetes mellitus: a systematic review and meta-analysis, J. Res. Med. Sci. 25 (2020) 56.

[18] A. Gholaminejad, H. Abdul Tehrani, M. Gholami Fesharaki, Identification of candidate microRNA biomarkers in diabetic nephropathy: a meta-analysis of profiling studies, J. Nephrol. 31 (6) (2018) 813–831.

[19] T. Barrett, S.E. Wilhite, P. Ledoux, C. Evangelista, I.F. Kim, M. Tomashevsky, et al., NCBI GEO: archive for functional genomics data sets–update, Nucleic Acids Res. 41 (Database issue) (2013) D991–D995.

[20] J. Zyprych-Walczak, A. Szabelska, L. Handschuh, K. Górczak, K. Klamecka, M. Figlerowicz, et al., The impact of normalization methods on RNA-seq data analysis, BioMed Res. Int. 2015 (2015) 621690.

[21] E. Maza, In papyro comparison of TMM (edgeR), RLE (DESeq2), and MRN normalization methods for a simple two-conditions-without-replicates RNA-seq experimental design, Front. Genet. 7 (2016) 164.

[22] F. Abbas-Aghababazadeh, Q. Li, B.L. Fridley, Comparison of normalization approaches for gene expression studies completed with high-throughput sequencing, PLoS One 13 (10) (2018), e0206312.

[23] S. Tam, M.S. Tsao, J.D. McPherson, Optimization of miRNA-seq data preprocessing, Briefings Bioinf. 16 (6) (2015) 950–963.

[24] R. Kolde, S. Laur, P. Adler, J. Vilo, Robust rank aggregation for gene list integration and meta-analysis, Bioinformatics 28 (4) (2012) 573–580.

[25] S. Yoon, B. Baik, T. Park, D. Nam, Powerful p-value combination methods to detect incomplete association, Sci. Rep. 11 (1) (2021) 6980.

[26] B.A. Rikke, M.W. Wynes, L.M. Rozeboom, A.E. Barón, F.R. Hirsch, Independent validation test of the vote-counting strategy used to rank biomarkers from published studies, Biomarkers Med. 9 (8) (2015) 751–761.

[27] J. Koricheva, J. Gurevitch, Place of meta-analysis among other methods of research synthesis, in: Handbook of Meta-Analysis in Ecology and Evolution, 2013, pp. 3–13.

[28] S. Davis, P.S. Meltzer, GEOquery: a bridge between the gene expression Omnibus (GEO) and BioConductor, Bioinformatics 23 (14) (2007) 1846–1847.

[29] M.D. Robinson, D.J. McCarthy, G.K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data, Bioinformatics 26 (1) (2010) 139–140.

[30] M.D. Robinson, A. Oshlack, A scaling normalization method for differential expression analysis of RNA-seq data, Genome Biol. 11 (3) (2010) R25.

[31] J.H. Bullard, E. Purdom, K.D. Hansen, S. Dudoit, Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments, BMC Bioinf. 11 (2010) 94.

[32] S. Anders, W. Huber, Differential expression analysis for sequence count data, Genome Biol. 11 (10) (2010) R106.

[33] C. Prada, D. Lima, H. Nakaya, MetaVolcanoR: Gene Expression Meta-Analysis Visualization Tool, R package version 1.6.0, 2021.

[34] L. Chang, G. Zhou, O. Soufan, J. Xia, miRNet 2.0: network-based visual analytics for miRNA functional analysis and systems biology, Nucleic Acids Res. 48 (W1) (2020) W244–W251.

[35] E. Buchberger, A. Bilen, S. Ayaz, D. Salamanca, C. Matas de Las Heras, A. Niksic, et al., Variation in pleiotropic hub gene expression is associated with interspecific differences in head shape and eye size in Drosophila, Mol. Biol. Evol. (2021) msaa335.

[36] H. Liu, Y. Qu, H. Zhou, Z. Zheng, J. Zhao, J. Zhang, Bioinformatic analysis of potential hub genes in gastric adenocarcinoma, Sci. Prog. 104 (1) (2021), 368504211004260.

[37] F. Kern, T. Fehlmann, J. Solomon, L. Schwed, N. Grammes, C. Backes, et al., miEAA 2.0: integrating multi-species microRNA enrichment analysis and workflow management systems, Nucleic Acids Res. 48 (W1) (2020) W521–W528.

[38] P.L. Wander, D.A. Enquobahrie, T.K. Bammler, S. Srinouanprachanh, J. MacDonald, S.E. Kahn, D. Leonetti, W.Y. Fujimoto, E.J. Boyko, Short Report: circulating microRNAs are associated with incident diabetes over 10 years in Japanese Americans, Sci. Rep. 10 (1) (2020) 6509.

[39] F.I. Corona-Meraz, M. Vázquez-Del Mercado, F.J. Ortega, S.L. Ruiz-Quezada, M.O. Guzmán-Ornelas, R.E. Navarro-Hernández, Ageing influences the relationship of circulating miR-33a and miR- 33b levels with insulin resistance and adiposity, Diabetes Vasc. Dis. Res. 16 (3) (2019) 244–253.

[40] A. Dávalos, L. Goedeke, P. Smibert, C.M. Ramírez, N.P. Warrier, U. Andreo, et al., miR-33a/b contribute to the regulation of fatty acid metabolism and insulin signaling, Proc. Natl. Acad. Sci. U. S. A. 108 (22) (2011) 9232–9237.

[41] Y. Yang, H. Jiang, L. Xiao, X. Yang, MicroRNA-33b-5p is overexpressed and inhibits GLUT4 by targeting HMGA2 in polycystic ovarian syndrome: an in vivo and in vitro study, Oncol. Rep. 39 (6) (2018) 3073–3085.

[42] E. Distel, T.J. Barrett, K. Chung, N.M. Girgis, S. Parathath, C.C. Essau, et al., miR33 inhibition overcomes deleterious effects of diabetes mellitus on atherosclerosis plaque regression in mice, Circ. Res. 115 (9) (2014) 759–769.

[43] F. Zeinali, S.M. Aghaei Zarch, A. Jahan-Mihan, S.M. Kalantar, M.Y. Vahidi Mehrjardi, H. Fallahzadeh, et al., Circulating microRNA-122, microRNA-126-3p and microRNA-146a are associated with inflammation in patients with pre-diabetes and type 2 diabetes mellitus: a case control study, PLoS One 16 (6) (2021), e0251697.

[44] W.Y. Lo, C.T. Peng, H.J. Wang, MicroRNA-146a-5p mediates high glucose-induced endothelial inflammation via targeting interleukin-1 receptor-associated kinase 1 expression, Front. Physiol. 8 (2017) 551.

[45] K.J. Belongie, E. Ferrannini, K. Johnson, P. Andrade-Gordon, M.K. Hansen, J.R. Petrie, Identification of novel biomarkers to monitor β-cell function and enable early detection of type 2 diabetes risk, PLoS One 12 (8) (2017), e0182932.

[46] T.G. Nührenberg, M. Cederqvist, F. Marini, C. Stratz, B.A. Grüning, D. Trenk, et al., Uncontrolled diabetes mellitus has No major influence on the platelet transcriptome, BioMed Res. Int. 2018 (2018) 8989252.

[47] Q.A. Hathaway, M.V. Pinti, A.J. Durr, S. Waris, D.L. Shepherd, J.M. Hollander, Regulating microRNA expression: at the heart of diabetes mellitus and the mitochondrion, Am. J. Physiol. Heart Circ. Physiol. 314 (2) (2018) H293–H310.

[48] G. Sebastiani, C. Guay, M. Latreille, Circulating noncoding RNAs as candidate biomarkers of endocrine and metabolic diseases, Internet J. Endocrinol. 2018 (2018) 9514927.

[49] W. Tao, X. Dong, G. Kong, P. Fang, X. Huang, P. Bo, Elevated circulating hsa-miR-106b, hsa-miR-26a, and hsa-miR-29b in type 2 diabetes mellitus with diarrhea-predominant irritable bowel syndrome, Gastroenterol. Res. Pract. 2016 (2016) 9256209.

[50] S. Ouyang, R. Tang, Z. Liu, F. Ma, Y. Li, J. Wu, Characterization and predicted role of microRNA expression profiles associated with early childhood obesity, Mol. Med. Rep. 16 (4) (2017) 3799–3806.

[51] A. Abbasi, D. Juszczyk, C.H.M. van Jaarsveld, M.C. Gulliford, Body mass index and incident type 1 and type 2 diabetes in children and young adults: a retrospective cohort study, J. Endocr. Soc. 1 (5) (2017) 524–537.

[52] H. Wang, R. Peng, J. Wang, Z. Qin, L. Xue, Circulating microRNAs as potential cancer biomarkers: the advantage and disadvantage, Clin. Epigenet. 10 (2018) 59.