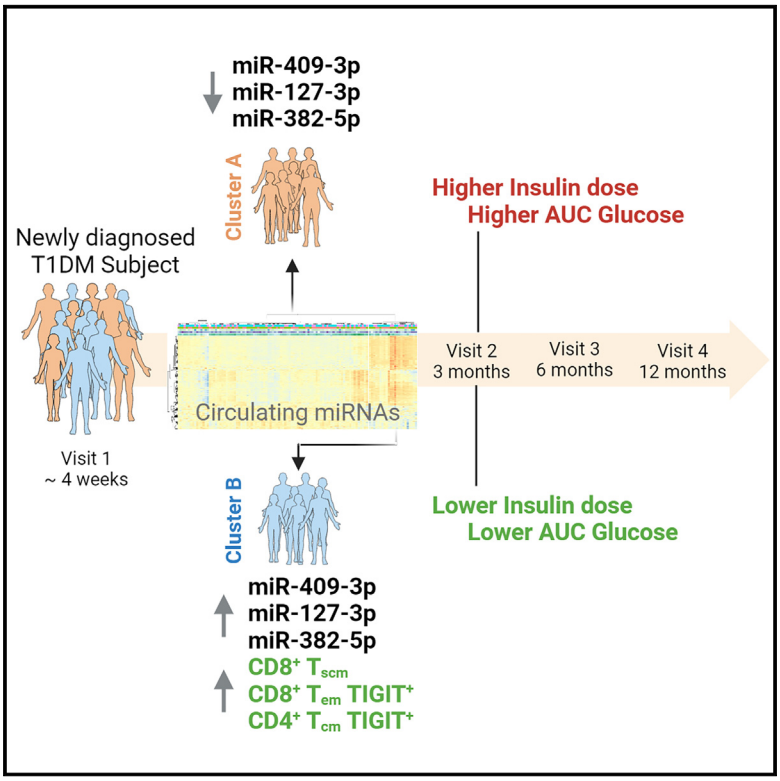
Cell Reports
Medicine

Article

A set of circulating microRNAs belonging to the 14q32 chromosome locus identifies two subgroups of individuals with recent-onset type 1 diabetes

Graphical abstract



Highlights

- Circulating miRNAs in T1DM individuals distinguish two groups: cluster A and cluster B
- miR-409-3p, miR-127-3p, and miR-382-5p are increased in cluster B T1DM individuals
- T1DM individuals in cluster B show an improved glycemic profile at follow-up
- Blood immunomics show higher frequency of memory and exhausted T cells in cluster B

Authors

Guido Sebastiani,
Giuseppina Emanuela Grieco,
Marco Bruttini, ..., Chantal Mathieu,
Francesco Dotta, INNODIA investigators

Correspondence

francesco.dotta@unisi.it

In brief

Sebastiani et al. found that circulating microRNAs, particularly those from the 14q32 locus, differentiate two distinct subgroups of individuals with type 1 diabetes mellitus, termed cluster A and cluster B. Cluster B, characterized by elevated expression of 14q32 miRNAs, exhibits better glycemic control and unique immunomics profiles, indicating potential implications for future interventional immunotherapies.

Sebastiani et al., 2024, Cell Reports Medicine *5*, 101591
June 18, 2024
https://doi.org/10.1016/j.xcrm.2024.101591

CellPress

Cell Reports Medicine

CellPress
OPEN ACCESS

Article

A set of circulating microRNAs belonging to the 14q32 chromosome locus identifies two subgroups of individuals with recent-onset type 1 diabetes

Guido Sebastiani,[1,2,12] Giuseppina Emanuela Grieco,[1,2,12] Marco Bruttini,[1,2,3] Stefano Auddino,[1,2] Alessia Mori,[1,2,3] Mattia Toniolli,[1,2] Daniela Fignani,[1,2] Giada Licata,[1,2] Elena Aiello,[1,2] Laura Nigi,[1,2] Caterina Formichi,[1,2] Juan Fernandez-Tajes,[4] Alberto Pugliese,[5,6] Carmella Evans-Molina,[7] Lut Overbergh,[8] Timothy Tree,[9] Mark Peakman,[10] Chantal Mathieu,[8] Francesco Dotta,[1,2,3,13,*] and on behalf of INNODIA investigators[11]

[1]Diabetes Unit, Department of Medicine, Surgery and Neurosciences, University of Siena, Siena, Italy
[2]Fondazione Umberto Di Mario ONLUS c/o Toscana Life Science, Siena, Italy
[3]Tuscany Centre for Precision Medicine (CReMeP), Siena, Italy
[4]INNODIA iVZW, Leuven, Belgium
[5]Diabetes Research Institute, Leonard Miller School of Medicine, University of Miami, Miami, FL, USA
[6]Department of Diabetes Immunology, Arthur Riggs Diabetes and Metabolism Research Institute, Beckman Research Institute, City of Hope, Duarte, CA, USA
[7]Center for Diabetes and Metabolic Diseases and the Wells Center for Pediatric Research, Indiana University School of Medicine, Indianapolis, IN, USA
[8]Katholieke Universiteit Leuven/Universitaire Ziekenhuizen, Leuven, Belgium
[9]Department of Immunobiology, School of Immunology and Microbial Sciences, King's College London, London, UK
[10]Immunology & Inflammation Research Therapeutic Area, Sanofi, Boston, MA, USA
[11]Further details can be found in the supplemental information
[12]These authors contributed equally
[13]Lead contact
*Correspondence: francesco.dotta@unisi.it
https://doi.org/10.1016/j.xcrm.2024.101591

SUMMARY

Circulating microRNAs (miRNAs) are linked to the onset and progression of type 1 diabetes mellitus (T1DM), thus representing potential disease biomarkers. In this study, we employed a multiplatform sequencing approach to analyze circulating miRNAs in an extended cohort of prospectively evaluated recent-onset T1DM individuals from the INNODIA consortium. Our findings reveal that a set of miRNAs located within T1DM susceptibility chromosomal locus 14q32 distinguishes two subgroups of individuals. To validate our results, we conducted additional analyses on a second cohort of T1DM individuals, confirming the identification of these subgroups, which we have named cluster A and cluster B. Remarkably, cluster B T1DM individuals, who exhibit increased expression of a set of 14q32 miRNAs, show better glycemic control and display a different blood immunomics profile. Our findings suggest that this set of circulating miRNAs can identify two different T1DM subgroups with distinct blood immunomics at baseline and clinical outcomes during follow-up.

INTRODUCTION

Type 1 diabetes mellitus (T1DM) is an autoimmune disease caused by immune-mediated destruction and dysfunction of pancreatic β cells, resulting into chronic hyperglycemia, lifelong insulin therapy, and occurrence of diabetic complications.[1] A marked disease heterogeneity leads to an incomplete understanding of T1DM pathogenesis and variable success of interventional therapies.[2] Age at diagnosis recapitulates profound differences in genetic predisposition,[3,4] islet autoantibody appearance,[5,6] clinical presentation,[7] and β cell functional decline progression,[8,9] with younger individuals showing a severe clinical presentation and a faster C-peptide decline after onset. A marked heterogeneity among individuals is also evident in terms of immune-cell infiltrates in pancreatic islets[10–12] and of circulating islet autoantibodies.[13] In light of this heterogeneity, the existence of multiple distinct subgroups/phenotypes has been hypothesized.[14–19] Overall, these studies demonstrated the existence of potentially distinct subgroups of T1DM individuals; however, it is currently unclear how the identification of these subgroups can be beneficial for a specific interventional therapy and how they can be easily identified in clinical practice. Nevertheless, the classification of T1DM individuals into disease subgroups still remains of high interest and could be beneficial for a precision-medicine approach.[20] Hence, it is imperative to find easily accessible and measurable biomarkers to detect

and further characterize putative T1DM subgroups.[21] The analysis of circulating biomarkers coupled to unsupervised, data-driven-omics methodologies may help in the unbiased stratification of T1DM individuals and, thus, in the identification of novel disease endotypes.

MicroRNAs (miRNAs) are a class of small noncoding RNAs[22,23] reported to have a critical role in the regulation of gene expression.[24] They have been associated with T1DM pathogenesis[25] by mediating function and dysfunction of pancreatic β cells[26,27] as well as immune cells.[28–31] miRNAs also represent an abundant class of circulating biomarkers.[32–34] Hence, the interception of these intercellular messages holds the potential to yield valuable insights into the status of specific diseases and facilitate the characterization of disease dysfunctions that remain incompletely understood.[35]

Numerous investigations have assessed circulating miRNAs in T1DM individuals. Currently, certain miRNAs have consistently shown repeated associations with disease onset (i.e., miR-24-3p,[36–43] miR-146a-5p,[37–39,44,45] and miR-375-3p[37,40,46–49]) or also linked to disease progression (i.e., miR-375-3p[37,50] and miR-24-3p[38,50]). However, it is important to note that many other findings have not been consistently confirmed in multiple studies. This underscores the variability in miRNA measurements, which may be attributed to the heterogeneity among cohorts of T1DM individuals, pre-analytical variables that can impact sample collection, and the performance of analytical platforms. Additionally, to date, an unsupervised and unbiased analysis of circulating miRNAs aimed at stratifying T1DM individuals into multiple subgroups based solely on their circulating expression levels has not been attempted.

In light of the limited overlap among different reports and the lack of an miRNA-based unsupervised classification of T1DM individuals, we employed two different sequencing platforms to comprehensively and unbiasedly investigate the circulating expression profile of miRNAs[51] within two large cohorts of recently diagnosed T1DM individuals (within 6 weeks from diagnosis) who were recruited and followed up as part of the INNODIA project.[52] This approach has allowed us to identify and validate two distinct subgroups of T1DM individuals characterized by different expression levels of a set of miRNAs belonging to the 14q32 chromosomal locus and reporting differences in glycemic control and peripheral blood immune-cell profiles.

RESULTS

T1DM individuals, data collection, and miRNA study design

Circulating miRNA profiling from blood-derived plasma was performed in two independent cohorts (first and second T1DM INNODIA cohorts) of individuals enrolled within 6 weeks from diagnosis of stage 3 T1DM in the European Consortium INNODIA Natural History Study (https://www.innodia.eu).[52] The study design is reported in Figure 1.

The study population of the first cohort consisted of n = 115 recently diagnosed T1DM individuals (58 female and 57 male, age 12.4 ± 7.7 years; mean duration 4.5 ± 1.5 weeks) (Table 1). T1DM individuals were followed up until 12 months post diagnosis and subjected to four visits (visit 1 [V1], baseline; V2, 3 months; V3, 6 months; V4, 12 months) (Figure 1A; Table S1A) when the main demographic data and clinical characteristics were measured and collected (Tables 1, S1A, and S2). Plasma-EDTA samples for miRNA analysis were collected at baseline visit (V1) using a standardized protocol adopted by all clinical sites.[51] Circulating miRNAs from n = 115 T1DM individuals were analyzed using two different library preparation strategies on all plasma samples, followed by short read sequencing with Illumina platforms. Specifically, we used (1) an RNA extraction-free targeted strategy adopting the HTG EdgeSeq miRNA whole transcriptome sequencing (hereafter "targeted-seq") and (2) a previously standardized untargeted approach[51] with QIAseq-miRNA/small RNA sequencing (hereafter "untargeted-seq"). Hence, each plasma sample was analyzed using both methods (Figure 1B). A subset of plasma samples from n = 6 T1DM individuals were run in duplicate for each platform, yielding a total of n = 121 samples analyzed. Notably, whole blood samples from a subset of T1DM individuals from the first cohort (67/115) were collected at V1, processed to isolate peripheral blood mononuclear cells (PBMCs), and analyzed through high-parameter flow cytometry to profile the circulating immunome.

The study population of the second cohort consisted of n = 147 T1DM individuals (55 male and 92 female; age 11,9 ± 7,9 years; mean duration 3.9 ± 1.7 weeks) (Table 1). Plasma samples of T1DM individuals from this second cohort were analyzed using the untargeted-seq approach, and results were further validated through droplet digital PCR (ddPCR) for selected miRNAs of interest (Figure 1B).

Circulating miRNA profile analysis of first-cohort T1DM individuals using a dual-sequencing platforms approach

Small RNA-sequencing (RNA-seq) technologies have enhanced the detection of miRNAs from plasma samples.[53] However, the limited RNA content in plasma, variations in RNA extraction methods, and differences in cDNA library preparation protocols and sequencing approaches have introduced biases into the analytical workflow, resulting in inconsistent findings across various studies.[54,55]

Therefore, to ensure the identification of miRNAs that were consistently detected by both targeted-seq and untargeted-seq, we performed a cross-validation of the results obtained from both platforms. This approach allowed us to establish a set of circulating miRNAs that exhibited concordant expression patterns and could therefore be effectively employed for stratifying T1DM individuals at baseline or potentially predicting disease progression during the follow-up period.

In the targeted-seq, two plasma samples failed to generate libraries, resulting into a total of n = 119/121 samples sequenced. Overall, sequencing quality metrics including Q30, total yield, and total reads passing filter met the acceptance criteria (Figures S1A and S1B). The mean total read count for each sample was 4.7 ± 0.88 × 10^6 reads (Figures S1C and S1D), while the mean read count aligned to miRNAs for each sample was 3.2 ± 0.67 × 10^6 (Figure S1E).

In the untargeted-seq, all plasma samples successfully generated cDNA libraries (n = 121/121), as shown by the correctly sized cDNA fragments analyzed by capillary electrophoresis

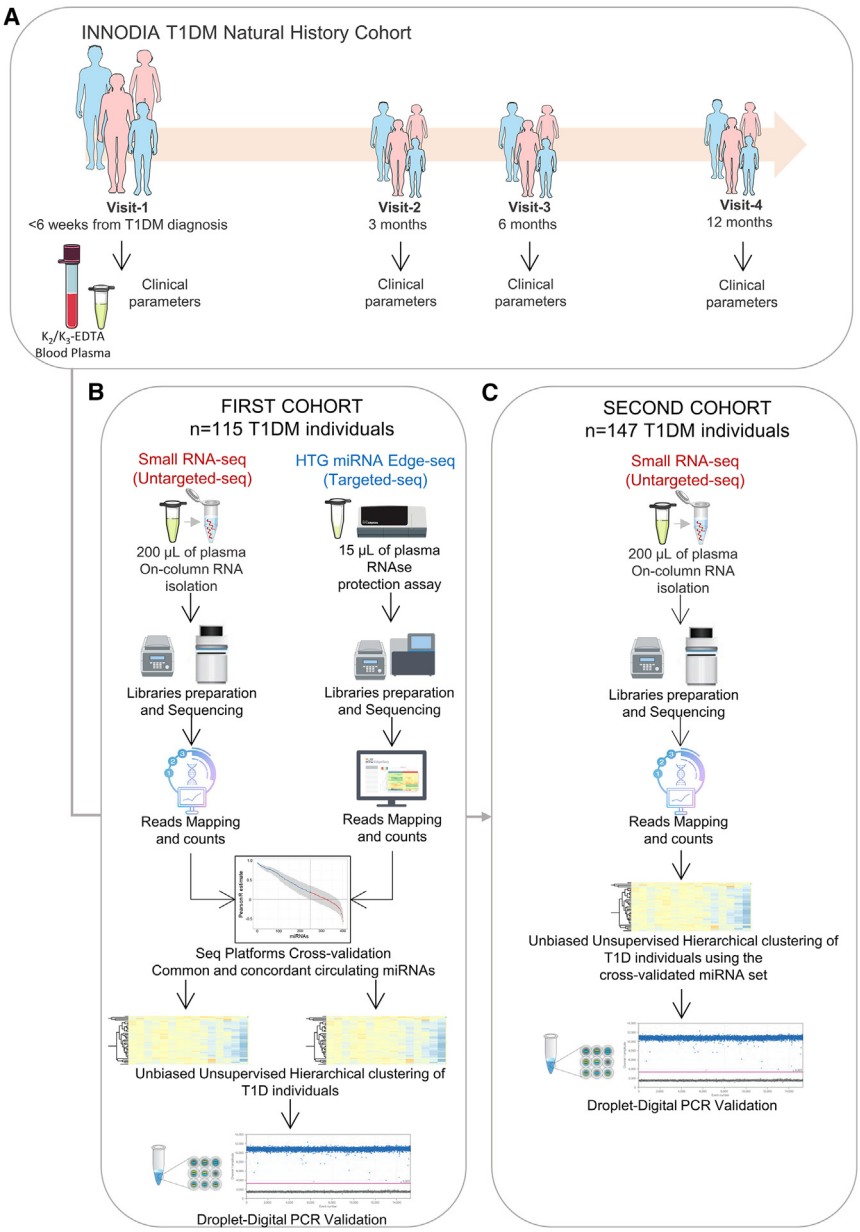


Figure 1. Schematic of the miRNA study design

Graphical description of the study design, which involved the clinical follow-up of T1DM individuals (A) and the analysis of two cohorts of individuals—first cohort (B) $n = 115$, second cohort (C) $n = 147$—using a multiplatform sequencing approach and ddPCR validation.

(Figures S2A and S2B). The quality controls of the untargeted-seq metrics returned high-quality parameters that met the acceptance criteria, including the *Phred* score (Figure S2C). The mean total read count for each sample was $7.5 \pm 2.1 \times 10^6$ reads (Figures S2D and S2E), while the miRNAs mean read count was $2.6 \pm 1.3 \times 10^6$ (Figure S2F). Overall, a total of $n = 114$ unique T1DM individuals were successfully profiled for circulating miRNAs using both sequencing approaches.

We then applied a cross-validation strategy to obtain a reliable set of data (Figure S3A). Overall, we obtained raw data counts of 2,083 miRNAs for the targeted-seq and 2,422 miRNAs for the untargeted-seq. After low-counts filtering, a total of 892 and 753 miRNAs were retained in the targeted-seq and untargeted-seq dataset, respectively (Figure S3A). A total of 402 unique miRNAs were commonly detected in both sequencing methods. Next, we selected only those miRNAs that had a positive and significant Pearson's correlation coefficient for each corresponding miRNA pair between targeted-seq and untargeted-seq (Figures S3B and S4). By employing this strategy, we successfully obtained two datasets comprising a total of $n = 248$ plasma-derived miRNAs, each of which exhibited consistent and concordant expression patterns following cross-validation. Using the expression datasets obtained through the cross-validation approach, we checked the internal platform reproducibility by inspecting the distance matrices of the technical replicates. Analysis of replicates demonstrated a good internal reproducibility, as the sample duplicates clustered together in both sequencing platforms (Figures S3C and S5).

Coefficient of variation (CV) was calculated for each miRNA read count across all samples analyzed in both platforms. The distribution pattern of CV values showed overlap between targeted-seq (CV median: 88.1%; 95% confidence interval [CI]: 80.8%–100.1%) and untargeted-seq (CV median: 68.7%; 95% CI: 63.7%–79.6%) (Figure S3D). The comparison of miRNA expression levels ranking between targeted-seq and untargeted-seq demonstrated a significant correlation (Figure S3E). When we specifically examined the 20 most highly expressed miRNAs in plasma of T1DM patients, we observed a similar top-ranking order between the two platforms, although some exceptions were noted (Figures S3F and S3G). Overall, miRNA expression ranking observed in this study was consistent with previous findings reported in other studies.[56,57]

Circulating miRNome analysis stratifies T1DM individuals into two subgroups: Cluster A and cluster B

To identify potential subgroups within T1DM individuals, we unbiasedly conducted unsupervised miRNA hierarchical clustering analyses separately for both the targeted-seq and untargeted-seq datasets (Figures 2A and 2B). To determine the number of

Table 1. Baseline clinical characteristics of first and second cohort of T1DM individuals

Demographics	First T1DM cohort (n = 115)	Second T1DM cohort (n = 147)
	Baseline (visit 1)	Baseline (visit 1)
Age (years)	12.51 ± 7.70 [115]	12.03 ± 7.82 [147]
Gender (female/male)	58/57	55/92
BMI (kg/m^2)	23.12 ± 2.82 [16]	22.26 ± 2.99 [21]
BMI-SDS	0.15 ± 1.09 [99]	0.38 ± 1.12 [126]
Disease duration (weeks)	4.08 ± 1.53 [115]	3.91 ± 1.76 [142]
Fasting C-peptide (pmol/L)	277.64 ± 203.54 [115]	270.03 ± 194.79 [145]
HbA1c (mmol/mol)	77.45 ± 19.44 [112]	76.01 ± 19.26 [143]
Insulin dose (units/kg/day)	0.52 ± 0.27 [113]	0.59 ± 0.40 [144]
IAA (% positive)	76 [88]	77.55 [114]
IA-2A (% positive)	72.1 [83]	78.23 [115]
GADA (% positive)	76 [88]	77.55 [114]
ZnT8A (% positive)	66 [76]	70.75 [104]

Mean values ± SD are reported for continuous variables; n or n% for categorical values. Number of T1DM individuals with available measurements for each specific variable is reported in brackets. BMI-SDS is exclusively reported for T1DM individuals younger than 18 years (first cohort, n = 99; second cohort, n = 126) at the moment of diagnosis.

clusters that optimally define distinct patient subgroups, we applied the silhouette method[58] and gap statistic techniques on both platforms. These gave us the best score for k = 2 (Figure S6). Thus, by pruning the obtained hierarchical tree with k = 2, two distinct subgroups of T1DM individuals were optimally identified for both targeted-seq (Figure 2A) and untargeted-seq (Figure 2B). Principal component analysis (PCA) confirmed the observed division of T1DM individuals into two subgroups for both platforms (Figures 2C and 2D).

The two clusters, denoted as cluster A and cluster B, comprised n = 87 and n = 22 T1DM individuals, respectively, resulting into a total of n = 109/115 individuals consistently assigned to their respective cluster across both sequencing platforms (Figure 2E). The analysis of the clinical site of sample collection (Figures S7A and S7B), the total number of miRNA reads detected in each analytical platform (Figures S7C and S7D), and the hemolysis rate assessed through erythrocyte-enriched miR-451a (Figures S7E and S7F) did not exhibit significant differences between the two clusters.

Cluster B T1DM individuals of INNODIA first cohort have a reduced insulin requirement at follow-up visits

To examine the key clinical parameters associated with cluster A and cluster B, we fit a univariate logistic regression model adjusted for multiple testing (p_{adj}) at each visit, including age, gender, body mass index (BMI) and BMI standard deviation score (BMI-SDS), autoantibody titers, and selected key metabolic outcomes. Notably, such features were not associated with these subgroups at baseline (Figure 3A; Tables S1B and S3). Additionally, the frequency of T1DM individuals who showed diabetes onset with ketoacidosis (DKA) (35.5% of total T1DM individuals in the first cohort) also did not significantly differ between cluster A and cluster B individuals (36% and 46% with DKA at diagnosis, respectively) (Figure S8A; Table S1B). However, in this model, insulin autoantibody (IAA) titers at baseline (V1) showed a tendency to be associated with T1DM subgroups being the odds ratio of having a higher IAA titer lower in cluster B compared to cluster A ($\log_2$ odds ratio [OR] −0.038; 95% CI −0.009, −0.089; p = 0.05; p_{adj} = 0.4) (Figures 3A and S8B). Human leukocyte antigen (HLA) risk genotype data, available for 106/109 T1DM individuals, showed that the high-risk HLA genotype DR3-DQ2 (DRB1*03:01-DQA1*05:01-DQB1*02:01) tended to be more common in individuals from cluster A than from cluster B (A: 51.8% versus B: 28.6%, p = 0.05, χ^2 test, not adjusted); a similar trend was observed for the DR3/DR4 genotype (A: 28.2% versus B: 9.5%, p = 0.07, χ^2 test, not adjusted) (Figure S8C).

Next, we investigated whether T1DM individuals in cluster A and cluster B had different clinical characteristics during the 12 months of follow-up. We found that individuals in cluster B required a lower insulin dose per kilogram at 3 months ($\log_2$ OR −3.84; 95% CI −6.4, −1.5; p = 0.002; p_{adj} = 0.012) and at 6 months ($\log_2$ OR −2.30; 95% CI −4.5, −0.2; p = 0.036; p_{adj} = 0.21) after diagnosis (Figures 3C and 3D), while no differences were observed at 12 months after diagnosis (Figure 3D). Overall, the profile of insulin requirements of T1DM patients in cluster B during the follow-up showed a lower insulin dose per kilogram at 3 and 6 months after diagnosis compared to cluster A, while returning to similar levels at 12 months after diagnosis (Figure 3E). These results are independent of the clinical site, as no major differences were found in insulin dosing between the clinical centers involved in the study (Figures S8D–S8G). Stimulated C-peptide (mixed meal tolerance test [MMTT] area under the curve [AUC]) measured at 3, 6, and 12 months showed no significant differences between the two clusters, along with insulin dose-adjusted HbA1c (IDAA1c) and fasting glucose (Figures 3F, 3G, and 3H).

Peripheral blood immunomic profile analysis of cluster A and cluster B T1DM individuals

To look for further differences between cluster A and cluster B individuals, we evaluated the peripheral blood immunomic profile at V1 in a subset of T1DM individuals of the INNODIA first cohort (67 out of 109 subjects) belonging to cluster A and cluster B (A, n = 48; B, n = 19) (Figures 4A and 4B). High-parameter flow-cytometry analysis identified a total of n = 150 different immune-cell subpopulations, which were tested for differences between cluster A and cluster B (Figure 4C); using this approach, n = 16 immune-cell subsets showed different frequencies between cluster A and cluster B with $p < 0.05$ (Figure 4D; Table S4), while four of them resulted statistically significant after false discovery rate adjustment ($p_{adj} < 0.1$) (Figures 4E–4H; Table S4). Of note, in cluster B we observed a significant increase of CD8$^+$ T stem cell memory (T_{scm}) cells (CCR7$^+$CD95$^+$CD45RA$^+$) (Figure 4E), T_H1-T_H2-like T_{FH} memory T_{conv} cells (Figure 4F), TIGIT$^+$ central memory (T_{cm} CD45RA$^-$CCR7$^+$CD28$^+$) CD8 T cells (Figure 4G), and TIGIT$^+$ effector memory (T_{EM} CD45RA$^-$ CCR7$^-$ CD28$^-$) T_{conv} cells (Figure 4H).

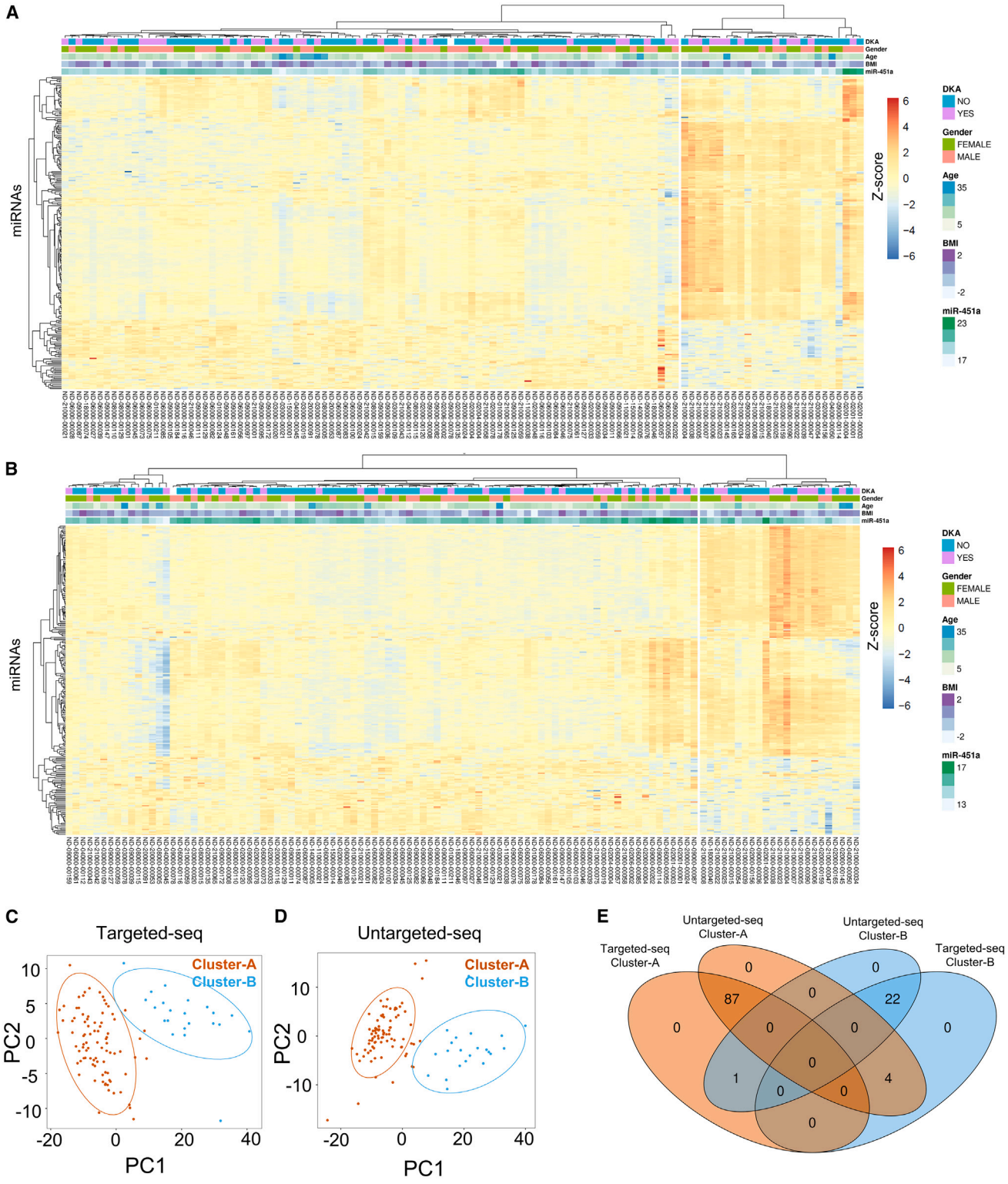


Figure 2. miRNA clustering analysis identifies two distinct T1DM subgroups

(A and B) Unsupervised hierarchical clustering analysis performed on all patients (columns) using Pearson's *R* distance on $\log_2$ normalized counts (after the addition of a pseudo-count) and complete linkage as agglomeration method in targeted-seq (A) and untargeted-seq (B) ($n = 115$ T1DM individuals). The heatmap displays clustering results, with miRNAs as rows and patients' information on diabetic ketoacidosis (DKA), gender, age, and the expression of miR-451a (an indicator of hemolysis rate). miRNA expression is represented as scaled *Z*-score values ranging from red (+6) to blue (−6).

(legend continued on next page)

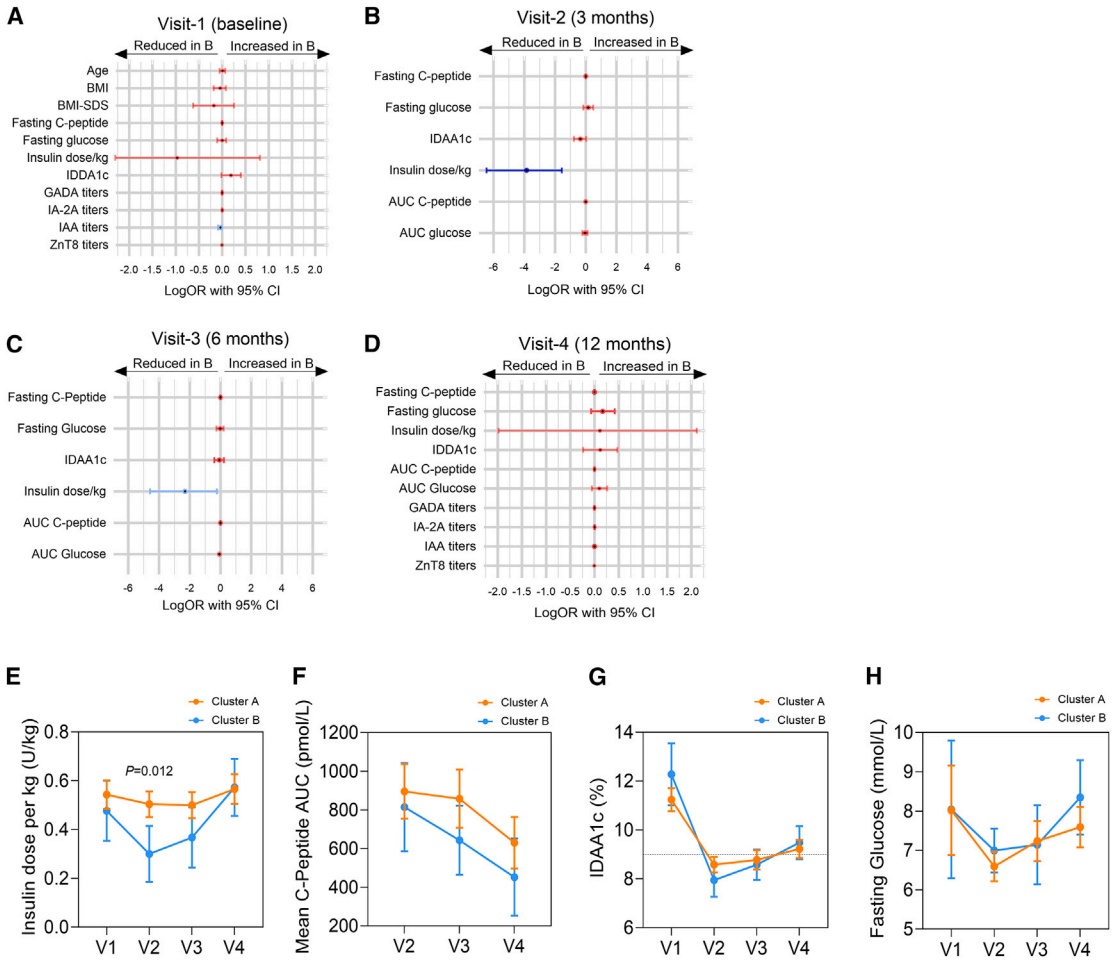


Figure 3. Clinical differences between cluster A and cluster B T1DM subjects at baseline and during follow-up

(A–D) Forest plots presenting the effect estimates and 95% confidence intervals (CIs) for cluster B across selected clinical variables collected at visit 1 (A), visit 2 (B), visit 3 (C), and visit 4 (D). The effects of cluster B are presented as log odds ratio using univariate logistic regression analysis adjusted for multiple testing. Light-blue bars indicate $p < 0.05$, while dark-blue lines indicate $p_{adj} < 0.1$.

(E) Insulin daily dose/kg profile over time from visit 1 to visit 4 in cluster A (orange) and cluster B T1DM subjects; data are reported as median and 95% CI.

(F) Area under the curve of C-peptide in mixed meal tolerance test (MMTT) performed at visit 2, visit 3, and visit 4 in cluster A and cluster B T1DM subjects.

(G) HbA1c-adjusted by insulin dose (IDAA1c) at visit 1, visit 2, visit 3, and visit 4 in cluster A and cluster B T1DM subjects.

(H) Fasting glucose at visit 1, visit 2, visit 3, and visit 4 in cluster A and cluster B T1DM subjects.

Data in (E)–(H) are reported as median and 95% CI. Number of individuals with available clinical parameters are reported in Table S8. p values are considered significant with $p_{adj} < 0.1$.

A set of miRNAs belonging to the chromosome locus 14q32 drives T1DM individual separation into cluster A and cluster B

To comprehensively characterize which set(s) of miRNAs were more relevant for dividing T1DM individuals into cluster A and cluster B, we performed a differential miRNA expression analysis between the two clusters. We analyzed the targeted-seq and untargeted-seq datasets separately and then validated the results by selecting only those miRNAs that were differentially expressed with $p_{adj} < 0.01$ on both platforms (Figure 5A). In total, we observed $n = 197$ differentially expressed miRNAs between cluster A and cluster B. Specifically, $n = 151$ miRNAs were upregulated and $n = 46$ miRNAs were downregulated in cluster B compared to cluster A T1DM individuals.

By ranking miRNAs from the most upregulated to the most downregulated in cluster B, we observed a significant enrichment

(C and D) Principal component analysis (PCA) of the targeted-seq dataset (C) and the untargeted-seq dataset (D), showing the elliptical grouping of samples based on miRNA expression. Targeted-seq, $n = 114$ T1DM individuals; untargeted-seq, $n = 115$ T1DM individuals.

(E) Venn diagram showing the number of T1DM individuals analyzed using both platforms and consistently identified as belonging to either cluster A or cluster B by both analytical sequencing methodologies.

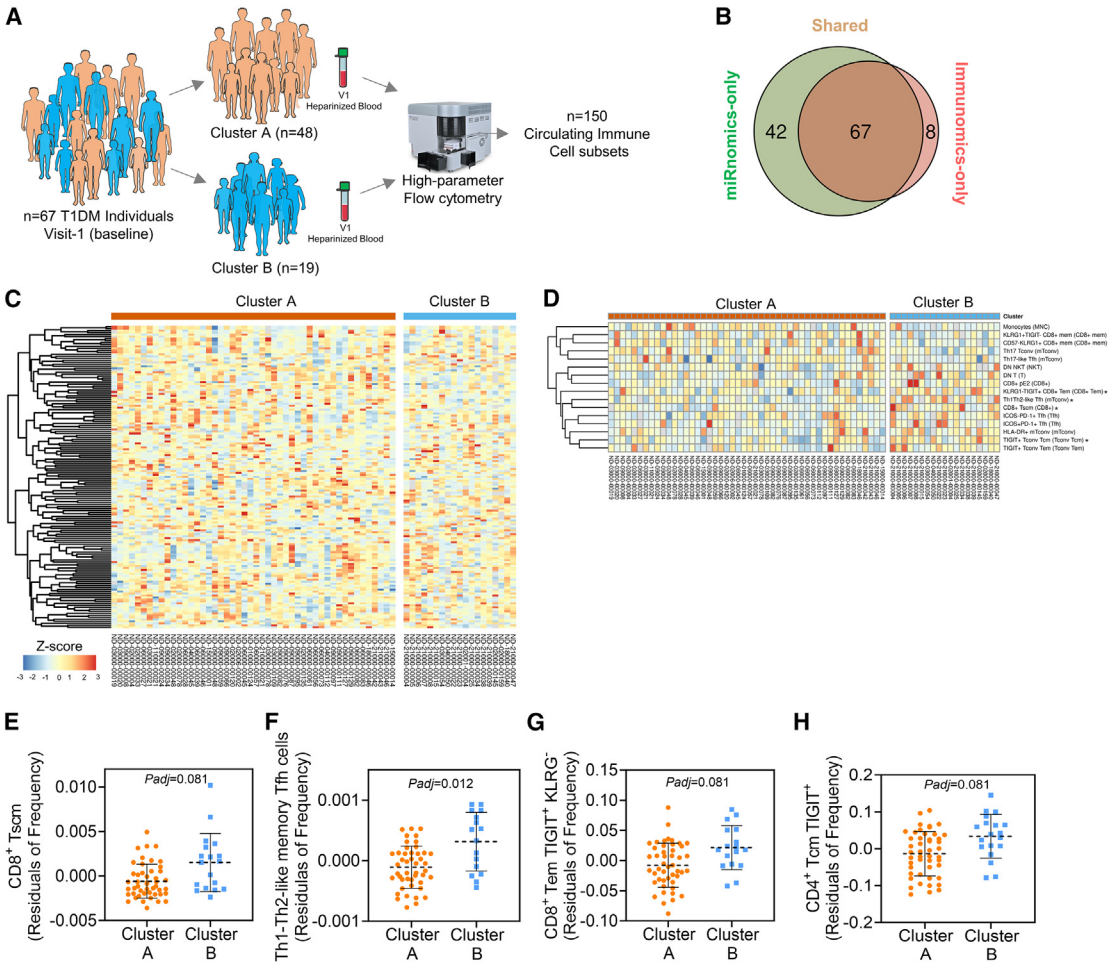


Figure 4. Blood immunomic profile at baseline visit in cluster A and cluster B T1DM individuals

(A) Schematic illustration of blood immunomics analysis of T1DM subjects at baseline.

(B) Venn diagram showing the number of T1DM subjects in the INNODIA first cohort with shared analysis of both miRNomics and immunomics. n = 67 T1DM individuals had complete data available for both miRNomics and immunomics analyses.

(C and D) (C) Heatmap and clustering analysis of immune-cell subset frequency in cluster A and cluster B T1DM subjects. The immune-cell subsets demonstrating $p < 0.05$ differential frequencies between cluster A and cluster B are further illustrated in (D) and listed in Table S4, where those with $p_{adj} < 0.1$ are further highlighted and indicated with an asterisk.

(E–H) Beta-regression models of the proportions for different immune-cell populations between clusters A and B with $p_{adj} < 0.1$. The beta regression depicts the proportions of T_{scm} $CD8^+$ cells (E) (cluster A n = 47, cluster B n = 18), T_H1-T_H2-like T_{FH} memory T_{conv} cells (F) (cluster A n = 46, cluster B n = 16), central memory $TIGIT^+KLRG^-$ CD8 T cells (G) (cluster A n = 47, cluster B n = 18), and effector memory $TIGIT^+$ T_{CONV} cells (H) (cluster A n = 47, cluster B n = 17) populations in cluster A and cluster B. p_{adj} ($p_{adj} < 0.1$) values were calculated using a beta-regression model correcting for the processing effect; mean and SD of the proportions of immune cells after residualization for the processing effect are shown.

of miRNAs originating from chromosomal locus 14q32. Specifically, miR-409-3p emerged as the most highly upregulated miRNA in cluster B (log_2 fold change [FC] = 3.99, B vs. A; $p_{adj} < 0.01$), followed by miR-127-3p (log_2 FC = 3.92, B vs. A, $p_{adj} < 0.01$) (Figure 5A). Notably, four out of the five most upregulated miRNAs in cluster B derived from the 14q32 locus, including miR-409-3p, miR-127-3p, miR-431-5p, and miR-382-5p (Figure 5A).

To investigate the presence of modules or groups of closely related miRNAs that may contribute to distinguishing the two clusters or correlate with clinical parameters at baseline or follow-up, we conducted a miRNA network analysis. This analysis was performed on both sequencing datasets, and a consensus hierarchical tree model was constructed to identify common miRNA modules across the two analytical platforms. The results of the topological overlap matrix from each dataset, along with the module assignment, are presented in Figure 5B. To identify the most interconnected and relevant miRNAs within each module, we selected the top five hub miRNAs and calculated the eigenmiRNA value as a surrogate measure of module expression (Figure 5C). Notably, we observed that the module ME#8 almost exclusively consisted of miRNAs from the 14q32 locus, which were upregulated in cluster B compared to cluster A. Among these miRNAs,

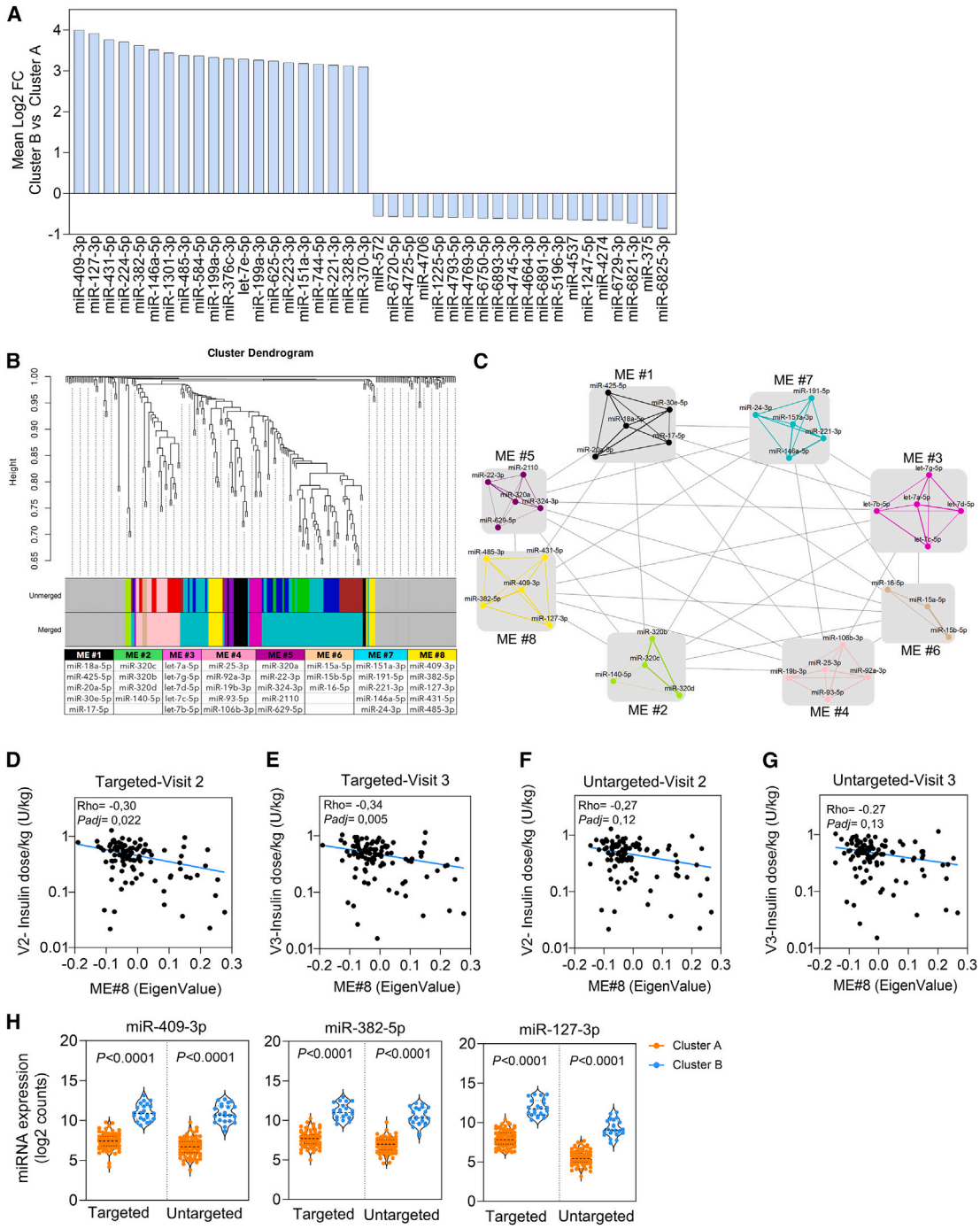


Figure 5. Relevance of miRNAs from chromosome 14q32 locus in the stratification of T1DM subjects into cluster A and cluster B

(A) Bar plot showing the top 20 differentially expressed miRNAs (up- or downregulated) in cluster B ($n = 22$) vs. cluster A ($n = 87$) T1DM subjects. Data represent the mean $\log_2$ fold-change values of cluster B vs. cluster A subjects obtained from separate analyses of targeted and untargeted datasets. Statistical analysis was performed using the Wald test (DESeq2), considering $p_{adj} \leq 0.01$.

(B) Targeted-seq and untargeted-seq consensus clustering dendrogram of miRNAs with module colors. The top five highly interconnected miRNAs assigned to each module are reported below the hierarchical tree model.

(C) Scheme illustrating modules, composition of miRNAs within each module, and strength of connections among miRNAs. Line thickness in each module indicates the level of interconnectedness among miRNAs.

(legend continued on next page)

miR-409-3p exhibited the highest level of interconnectedness (Figure 5C; Table S5).

Subsequently, we examined the correlation between the eigenmiRNA module values and available clinical parameters for each sequencing dataset. Notably, the ME#8 module exhibited a significant association with insulin dose per kilogram at follow-up visits after 3 months (V2) and 6 months (V3) in the targeted dataset, while a similar trend was observed in the untargeted one. The correlation analysis between ME#8 and clinical parameters recapitulates the previously observed differences between cluster A and cluster B T1DM individuals. Indeed, those miRNAs showing the highest connectivity in the ME#8 module (miR-409-3p, miR-382-5p, and miR-127-3p) (Table S5) differed significantly between T1DM individuals of cluster A and cluster B in both sequencing platforms (Figure 5H) and were selected for a ddPCR validation analysis.

Droplet digital PCR validation of miR-409-3p, miR-382-5p, and miR-127-3p in cluster A and cluster B T1DM individuals

The present findings underscore the discriminatory potential of miR-409-3p, miR-382-5p, and miR-127-3p in distinguishing T1DM individuals from cluster A and cluster B. Moreover, we demonstrated their high interconnectedness within a distinct network module (ME#8), which almost exclusively comprised 14q32 miRNAs. Remarkably, this interconnection is further substantiated by the discernible mid-collinearity observed among the expression of the three miRNAs in the plasma of individuals with T1DM (Figure S9A).

Co-expression and specificity analysis of the three miRNAs in multiple small RNA-seq datasets obtained from isomiRdb[59] and derived from $n = 99$ different human primary tissues and cells did not show a clear co-expression pattern (Figure S9B), suggesting that this indeed represents a peculiar characteristic of our dataset. Notably, in this analysis, miR-382-5p resulted in being specifically enriched in human primary sorted β cells (Figure S9C).

Hence, using ddPCR we analyzed miR-409-3p, miR-382-5p, and miR-127-3p in plasma samples of T1DM patients derived from cluster A and cluster B of the first cohort. The results confirmed a significantly higher expression level of miR-382-5p, miR-409-3p, and miR-127-3p in cluster B individuals compared to cluster A counterparts (Figures 6A–6C). Receiver-operating characteristic (ROC) curve analysis demonstrated the high specificity and sensitivity of these miRNAs in assigning T1DM individuals to the identified clusters. Notably, all three miRNAs exhibited a significant inverse correlation with the insulin dose per kilogram at V2 (Figure 6D) and V3 (Figure 6E).

14q32 miRNAs distinguish cluster A and cluster B T1DM individuals in the INNODIA second cohort

To further validate the identification of cluster A and cluster B subgroups in T1DM individuals, we conducted an analysis on an additional cohort consisting of $n = 147$ T1DM individuals (Tables 1, S6A, and S7).

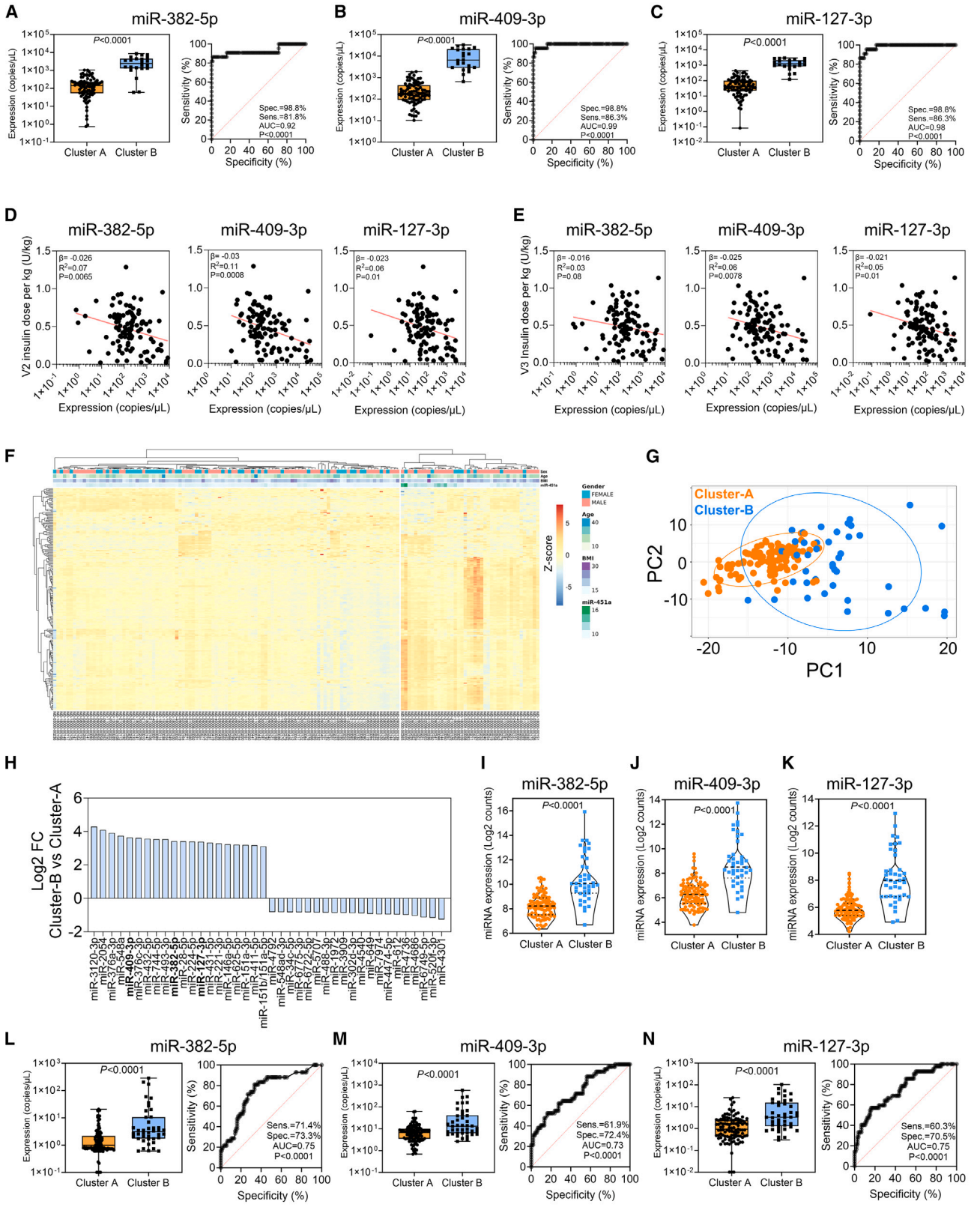
This analysis was performed using the small RNA-seq pipeline previously described (untargeted-seq) followed by ddPCR analysis. The sequencing metrics and reads counts, as shown in Figures S10A–S10D, confirmed the validity of the sequencing run also in this second cohort.

In the first cohort, we have successfully identified a robust set of 248 miRNAs using a cross-validation approach involving two distinct platforms. To verify the reproducibility of our findings, we applied the same rigorous filtering process to this second cohort. We were able to detect a total of 226 out of the initial 248 miRNAs in this independent dataset. To assess the reliability of our small RNA-seq method, we conducted correlation analyses on both internal replicates (Pearson's $R > 0.77$, $p < 2.2 \times 10^{-16}$) (Figure S10E) and inter-cohort replicates (Pearson's $R > 0.70$, $p < 2.2 \times 10^{-16}$) (Figure S10F); these analyses yielded positive results, further validating the robustness of our approach.

Next, we analyzed this dataset by applying a hierarchical clustering analysis pipeline as in the first cohort. Interestingly, we observed a distribution of T1DM individuals into two clusters (Figure 6F), in line with the findings from the initial cohort and resembling the previously identified cluster A and cluster B subgroups (cluster A, $n = 105$ T1DM individuals; cluster B, $n = 42$ T1DM individuals). Additionally, PCA endorsed the classification of T1DM individuals into distinct groups (Figure 6G) and was in line with the results obtained in the first cohort (Figures S11A and S11B), supporting the reliability of miRNAs in classifying T1DM individuals into these two clusters. Differential expression analysis revealed an enrichment of 14q32 miRNAs that were upregulated in cluster B compared to cluster A (Figures 6H–6K). The upregulation of miR-409-3p, miR-382-5p, and miR-127-3p in cluster B compared to cluster A was also confirmed using ddPCR, providing additional support to the analysis of the sequencing dataset (Figures 6L and 6M).

Next, we investigated the clinical differences between cluster A and cluster B T1DM individuals in this second cohort. In line with first-cohort findings, we did not observe differences in age, gender, BMI, number of autoantibodies, glutamic acid decarboxylase autoantibody (GADA), insulinoma-associated autoantibody (IA-2A), and ZnT8 autoantibody (ZnT8A) titers, and other metabolic outcomes (Table S6B). We did not find any significant differences in terms of IAA titers at baseline (Figures S12A and S12B) or insulin dose at V2 (Figure S12C), although a trend was evident for both parameters and in line with observations made in the first cohort (Figures S12A and S12C). Notably, in this second cohort, we observed a significant reduction in MMTT AUC glucose at the follow-up visit V2 in cluster B T1DM individuals (cluster A: 13.5 ± 3.3 mmol/L; cluster B: 11.5 ± 2.9 mmol/L; $p_{adj} = 0.036$), suggesting a better glycemic control compared to cluster A T1DM individuals at follow-up (Figure S12C).

(D–G) Scatterplots depicting relevant correlations between eigenvalues of the ME#8 module and insulin dose/kg at visit 2 (V2) and visit 3 (V3) for each T1DM subject in the targeted (D and E) and untargeted (F and G) datasets. Spearman's rho test ($p_{adj} \leq 0.1$) was performed, reporting for each graph rho and p values. (H) Comparison of expression levels of miR-409-3p, miR-382-5p, and miR-127-3p in cluster A vs. cluster B T1DM subjects in targeted-seq and untargeted-seq. Values are presented as $\log_2$ read counts. Statistical analysis was conducted using the Wald test (DESeq2) with p_{adj}.



(legend on next page)

In summary, these findings demonstrate that miR-409-3p, miR-382-5p, and/or miR-127-3p can be utilized to stratify newly diagnosed T1DM individuals into two distinct subgroups characterized by different glycemic control at follow-up visits after diagnosis.

DISCUSSION

Numerous studies have demonstrated the association between circulating miRNAs and T1DM, including its onset, progression, and the decline of β cell function.[36–44,46–49,60–63] Within this field, there is growing appreciation that circulating miRNAs may serve as biomarkers for T1DM. However, none of the previous studies have taken an unbiased approach to utilize circulating microRNAs for stratifying individuals with T1DM shortly after disease onset. Notably, identification of disease subgroups associated with specific phenotypes may shed light on the heterogeneity of T1DM and aid in the stratification of individuals and their assignment to specific interventional immunotherapy.[64]

In the current study, we sought to characterize circulating miRNAs in a large cohort of T1DM individuals using a multiplatform sequencing approach coupled with the analysis of two cohorts of T1DM individuals belonging to the European INNODIA Study Consortium.[52] It is important to highlight that within the consortium, plasma samples were collected and processed uniformly and consistently, following a standardized operating procedure.[51,65] As a result, the present study has a significant advantage in terms of sample collection and initial screening of miRNAs. Moreover, in the first-cohort analysis two different sequencing approaches were utilized, further strengthening the robustness of the findings.

The set of selected circulating miRNAs in individuals with T1DM underwent an unbiased and unsupervised clustering approach. This analytical approach was utilized in other contexts[66–69] and demonstrated preferable results compared to a plethora of other machine-learning approaches where dichotomous/multichotomous outcomes are necessary for selecting predictive features. This unbiased and outcomes-free methodology appears as an optimal choice for exploratory and hypothesis-free studies. In the present work, results of the hierarchical clustering methodology were also confirmed using PCA stratification, both in the first and in the second cohort.

Using this approach, we have selected three miRNAs (miR-409-3p, miR-382-5p, and miR-127-3p) that can differentiate two distinct groups of T1DM individuals (cluster A and cluster B). The mid to high collinearity observed for the three miRNAs of interest supports the potential utility of exclusively utilizing one of them to assign an individual to a specific T1DM cluster. Moreover, the selected miRNAs were validated, in both cohorts, through ddPCR. Due to the absolute quantification nature of ddPCR, data were not subjected to normalization[70]; hence, absolute copies for each miRNA are reported per microliter of plasma. This ensures the independence of ddPCR miRNA expression data from normalization strategies.

At V1, we observed that cluster B individuals displayed lower IAA titers and a lower prevalence of the high-risk T1DM genotype HLA-DR3. These findings, while not reaching significance after correction for multiple comparisons in both cohorts, exhibited a consistent trend. Given the substantial heterogeneity among individuals with T1DM, this outcome is noteworthy and merits dedicated attention in subsequent T1DM cohort studies to enhance the robustness and generalizability of the present observations. Indeed, the appearance of IAA as the first autoantibody was associated with an increased risk of developing future multiple autoantibodies and of T1DM onset[71] and a rapid decline of β cell function after onset.[72] These findings suggest a more severe phenotype of T1DM in cluster A vs. cluster B individuals despite the absence of age differences at diagnosis between the two subgroups, which was reported to be associated with disease severity. This hypothesis is further corroborated

Figure 6. Droplet digital PCR validation and independent miRNA clustering confirmation

(A–C) Stem-loop reverse transcriptase and TaqMan-based droplet digital PCR analysis of circulating miR-382-5p (A), miR-409-3p (B), and miR-127-3p (C) in cluster A ($n = 87$) and cluster B ($n = 22$) T1DM individuals of the first cohort. Logistic regression and receiver-operating characteristic (ROC) curve analysis for each of the three miRNAs are also presented. Values are reported as miRNA copies per microliter of plasma and shown on a $\log_{10}$ scale. Statistical analysis was performed using the non-parametric Mann-Whitney U test ($p < 0.05$). Logistic regression and ROC curves provide information on specificity, sensitivity, area under the curve, and corresponding p values ($p < 0.05$).

(D and E) Simple linear regression analyses between miR-382-5p, miR-409-3p, and miR-127-3p and visit 2 (V2) insulin daily dose/kg (D) and visit 3 (V3) insulin daily dose/kg (E). Linear regression analyses report slope (β) values, R^2, and corresponding p values ($p < 0.05$).

(F) Unsupervised hierarchical clustering analysis performed on all patients ($n = 147$) (in columns) using Pearson's R distance on $\log_2$ normalized counts (after the addition of a pseudo-count) and complete-linkage agglomeration method. The heatmap displays clustering results, with miRNAs as rows and patients' information on diabetic ketoacidosis (DKA), gender, age, and the expression of miR-451a. miRNA expression is represented as scaled Z-score values ranging from red (+6) to blue (−6).

(G) Principal component analysis (PCA) of the untargeted-seq dataset of the second cohort, showing grouping of samples based on miRNA expression (orange dots: cluster A individuals, $n = 105$; blue dots: cluster B individuals, $n = 42$).

(H) Bar plot showing the top 20 significantly upregulated or downregulated miRNAs in cluster B vs. cluster A T1DM individuals. Data represent the average $\log_2$ fold-change values of cluster B vs. cluster A individuals obtained from analysis of the untargeted datasets. Statistical analysis was performed using the Wald test (DESeq2), considering $p_{adj} < 0.01$. miR-409-3p, miR-382-5p, and miR-127-3p are reported in bold.

(I–K) Comparison of expression levels of miR-382-5p (I), miR-409-3p (J), and miR-127-3p (K) in cluster A vs. cluster B T1DM individuals in untargeted-seq of second-cohort samples. Values are presented as $\log_2$ values of normalized read counts.

(L–N) Stem-loop reverse transcriptase and TaqMan-based droplet digital PCR analysis of circulating miR-382-5p (L), miR-409-3p (M), and miR-127-3p (N) in cluster A and cluster B T1DM individuals of the second cohort. Logistic regression and ROC curve analysis for each of the three miRNAs are also presented. Values are reported as miRNA copies per microliter of plasma and shown on a $\log_{10}$ scale. Statistical analysis was performed using the non-parametric Mann-Whitney U test ($p < 0.05$). Logistic regression and ROC curves provide information on specificity, sensitivity, area under the curve, and corresponding p values ($p < 0.05$).

by clinical observations made at V2 and V3, where a better glycemic profile was observed in individuals in cluster B compared to cluster A, underscoring a less severe phenotype of cluster B subjects during follow-up. In both T1DM cohorts, we did not detect any differences in β cell functional profiles between cluster A and cluster B at baseline or follow-up visits. These observations indicate the possible presence of variations in insulin sensitivity between the two clusters of individuals. However, it is important to note that, in T1DM cohorts of the INNODIA consortium, direct measurements of insulin sensitivity were not performed at diagnosis, and baseline (V1) and follow-up (V2, V3, and V4) visits were influenced by administration of insulin therapy, impairing a reliable insulin-sensitivity measurement.

To gain further insights into specific differences between cluster A and cluster B, we had the valuable opportunity to investigate peripheral blood immune-cell profiles. Cluster B individuals showed increased frequency of $CD8^+$ and $CD4^+$ T cell subpopulations exhibiting a partially exhausted phenotype (i.e., increased frequency of $TIGIT^+$ central memory $CD8^+$ T cells and $TIGIT^+$ effector memory T_{conv} cells) and increased frequency of $CD8^+$ T_{scm} cells and T_H1-T_H2-like T_{FH} memory T_{conv} cells. The role of inhibitory receptors is well known, with TIGIT expression representing a hallmark of T cell exhaustion.[73,74] It is noteworthy that partial exhaustion and central memory phenotype of $CD8^+$ T cells define responders in anti-CD3 clinical trials.[75–79] In addition, $CD8^+$ T cell exhaustion also characterizes T1DM individuals experiencing a slow progression of the disease after onset.[80] On the other hand, the role of T_H1-T_H2-like T_{FH} memory T_{conv} cells and of $CD8^+$ T_{scm} cells in autoimmune diseases is still to be fully deciphered. Overall, while T_H1-T_H2-like T_{FH} memory T_{conv} cells have not yet been characterized, T_{scm} cells are known to play a crucial role in promoting antitumor and immune reconstitution because of their enhanced stem cell-like self-renewal capacity, which can serve as a reservoir of effector T cells. Of note, circulating T_{scm} cells showed increased frequency in T1DM individuals, thus potentially promoting autoimmunity.[81,82] Hence, the precise role of these immune-cell subsets in cluster B T1DM individuals should be further analyzed.

In both T1DM cohorts, levels of circulating miR-409-3p, miR-382-5p, and miR-127-3p were significantly elevated in plasma of individuals belonging to cluster B compared to cluster A. These miRNAs are all part of the 14q32 chromosomal locus.[83] Intriguingly, miRNA network analysis showed that these three miRNAs were the most interconnected among others within a specific module (ME#8), hosting 13 out of 14 miRNAs whose genes are located in the 14q32 genomic region. These findings highlight the significance of this locus, which has been previously indicated as a susceptibility region for T1DM.[84] Furthermore, network analysis revealed the presence of additional interesting modules. For example, module ME#3 comprised miRNAs from the let-7 family, which have been linked to microvascular complications in diabetes[85,86]; ME#4 and ME#7 modules contained miRNAs associated with T1D, inflammation, and β cell function, such as miR-93-5p,[87] miR-25-3p,[36] and miR-106b-3p[88] in ME#4 module and miR-151a-3p,[89] miR-24-3p,[36–43] and miR-146a-5p[89–93] in ME#7 module. Among these, miR-25-3p displayed the highest degree of interconnectedness within ME#4, while miR-151a-3p held this distinction in ME#7. Interestingly, both miRNAs have previously been linked to metabolic impairment and β cell function in T1DM.[36,89] Overall, these findings provide additional support to the validity and robustness of our analysis.

Chromosome locus 14q32 hosts the largest polycistronic miRNA clusters in mammals. In humans, it contains 54 miRNA genes, organized in different subclusters or interspersed within this region, and included between DLK1 and DIO3 genes.[94] miR-127-3p, miR-409-3p, and miR-382-5p derived from two different clusters, both residing in the 14q32.2 region: the miR-127/miR-136 cluster (including miR-127a-3p) and the miR-379/miR-410 cluster (including miR-409-3p and miR-382-5p). It is noteworthy that even within the same cluster, each individual miRNA may undergo distinct transcriptional regulation. For instance, miR-127 and miR-433, which are part of the same cluster, are transcribed from independent promoters,[95] implying an independent regulation of gene expression. Additionally, despite residing in the same cluster and genomic locus and potentially being subject to similar signal alterations, it appears that their functions may be entirely divergent and strictly dependent on the context.[96] Consistent with this, their tissues/cells expression patterns, analyzed across multiple small RNA-seq datasets from 99 distinct human primary cells and tissues in a physiological state (Figures S10B and S10C), do not align with their observed collinearity in T1DM plasma samples of our cohort. This misalignment suggests that their cellular origins may be different, but their expression could be influenced by similar pathological mechanisms, hence providing a high co-expression signature in plasma of T1DM individuals.

Concerning the functions of the three miRNAs, they have been documented as regulators of several functions. Genetic ablation of the miR-379/miR-410 cluster in mice resulted in impaired glucose homeostasis through the alteration of gluconeogenesis and glycogenolysis.[97] Particularly, In support of their role in metabolism, several other studies showed that miRNAs from the 14q32.2 region, including miR-409 and miR-127, exhibit high expression in human pancreatic islets,[59] where they are enriched in β cells rather than α cells.[98–100] In addition, it has been shown that these miRNAs are differentially expressed in islets of type 2 diabetic donors,[99] thus highlighting their role in glucose metabolism.

The specific association between 14q32 miRNAs and their involvement in T1DM pathogenesis or progression has been previously suggested. First, 14q32 miRNAs are expressed in β cells where they have been reported to regulate multiple functions through the modulation of a set of target genes encoding specific T1DM autoantigens[101]; hence, their downregulation may potentially increase the visibility of β cells to the immune system in T1DM. Second, a specific single-nucleotide polymorphism has been previously identified as associated with T1DM in a genome-wide association study.[84] Third, we previously demonstrated that miR-409-3p expression is decreased in the plasma of diabetic NOD mice in comparison with nondiabetic NOD mice, and this reduction is mirrored in the pancreas-infiltrating lymphocytes. Importantly, miR-409-3p was reduced in plasma of recent onset (<2 years from diagnosis) T1DM individuals compared to nondiabetic controls.[30] Interestingly, we found a

similar reduction in miR-409-3p expression in cluster A T1DM individuals compared to cluster B. This reduction in miR-409-3p expression is in line with the more severe phenotype observed in cluster A individuals, both at baseline and during follow-up visits, thus potentially depicting a protective role for this miRNA in the context of T1DM. Additional studies are needed to fully understand the molecular mechanisms encompassing miRNA-mediated communication among different cell types.[102]

In summary, the results of this study provide evidence supporting the use of miR-409-3p, miR-382-5p, and/or miR-127-3p as effective markers for stratifying newly diagnosed individuals with T1DM into two distinct subgroups that display different immune-related characteristics at baseline and different levels of glycemic control over time. Hence, this stratification can be taken into consideration to verify the responsiveness of T1DM individuals of cluster A and cluster B to multiple immunotherapies, thus implementing a tailored precision-medicine approach to treat T1DM.

Limitations of the study

We acknowledge limitations of this study. First, the sample size used for the association study between T1DM clusters (A or B) and peripheral blood immunomics was relatively small, consisting of 67 individuals, while a large number of immune-cell subpopulations (n = 150) were analyzed. Consequently, we did perform multiple testing corrections setting the p_{adj} value at 0.1, being an exploratory study.

Second, in the second cohort we acknowledge a gender imbalance, females being more prevalent than males. Several analyses of this cohort are still pending because the INNODIA study is ongoing.

Finally, although circulating miRNAs were validated using ddPCR in two different T1DM cohorts, comprising a total of 256 T1DM individuals analyzed, their diagnostic applicability is still limited and requires future studies to determine specific cut-off values and undergo extensive validation in multiple cohorts and real-life analyses.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - T1DM individuals of the INNODIA first cohort
  - T1DM individuals of the INNODIA second cohort
- METHOD DETAILS
  - Blood samples and plasma processing
  - HTG EdgeSeq miRNA whole transcriptome assay (targeted-seq)
  - QIAseq small RNA sequencing (untargeted-seq)
  - Primary analysis of miRNAs expression
  - Circulating miRNAs unsupervised hierarchical clustering analysis
  - HLA genotyping
  - PBMC (cryopreserved) multi-dimensional flow cytometry (MultiFACS) immunomics
  - MiRNAs differential expression analysis
  - Weighted miRNA correlation network analysis (WMCNA) for the identification of miRNA modules
  - Droplet digital PCR (ddPCR)
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xcrm.2024.101591.

ACKNOWLEDGMENTS

This work is supported by the Innovative Medicines Initiative 2 (IMI2) Joint Undertaking under grant agreements no. 115797-INNODIA and 945268-INNODIA HARVEST. This joint undertaking receives support from the Union's Horizon 2020 research and innovation program and EFPIA, JDRF, and The Leona M. and Harry B. Helmsley Charitable Trust. G.S. is supported by the University of Siena within *F-CUR* funding program grant no. 2268-2022-SG-PSR2021-FCUR_001. F.D. was supported by the Italian Ministry of University and Research (2017KAM2R5_003). This work is also supported by the EU within the Italian Ministry of University and Research PNRR "National Center for Gene Therapy and Drugs based on RNA Technology" (project no. CN00000041 CN3 Spoke #5 "Inflammatory and Infectious Diseases"). F.D. was also supported by the Italian Ministry of University and Research with the project PNC 0000001 D3 4 Health, the National Plan for Complementary Investments to the NRRP, funded by the European Union - NextGenerationEU. This work is also supported by JDRF and The Leona M. and Harry B. Helmsley Charitable Trust for the project "Collaborative Effort to Identify and Validate miRNAs as Biomarkers of T1D."

AUTHOR CONTRIBUTIONS

Conceptualization, G.S. and F.D.; methodology, G.S., G.E.G., and A.M.; investigation, G.S., G.E.G., A.M., D.F., G.L., E.A., L.N., C.F., and T.T.; formal analysis, G.S., M.B., S.A., M.T., and J.F.-T.; data curation, M.B., S.A., M.T., J.F.-T., G.E.G., and G.S.; writing – original draft, G.S., G.E.G., and F.D.; writing – review & editing, C.M., L.O., M.P., A.P., C.E.-M., and F.D.; funding acquisition, F.D. and G.S.; supervision, F.D. and C.M.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 19, 2023
Revised: February 2, 2024
Accepted: May 13, 2024
Published: June 4, 2024

REFERENCES

1. American Diabetes Association Professional Practice Committee (2022). 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes—2022. Diabetes Care *45*, S17–S38. https://doi.org/10.2337/dc22-S002.
2. Ilonen, J., Lempainen, J., and Veijola, R. (2019). The heterogeneous pathogenesis of type 1 diabetes mellitus. Nat. Rev. Endocrinol. *15*, 635–650. https://doi.org/10.1038/s41574-019-0254-y.
3. Krischer, J.P., Liu, X., Lernmark, Å., Hagopian, W.A., Rewers, M.J., She, J.-X., Toppari, J., Ziegler, A.-G., and Akolkar, B.; TEDDY Study Group (2017). The influence of type 1 diabetes genetic susceptibility regions, age, sex, and family history on the progression from multiple autoantibodies to type 1 diabetes: A TEDDY study report. Diabetes *66*, 3122–3129. https://doi.org/10.2337/db17-0261.
4. Inshaw, J.R.J., Cutler, A.J., Crouch, D.J.M., Wicker, L.S., and Todd, J.A. (2020). Genetic Variants Predisposing Most Strongly to Type 1 Diabetes

Diagnosed Under Age 7 Years Lie Near Candidate Genes That Function in the Immune System and in Pancreatic β-Cells. Diabetes Care *43*, 169–177. https://doi.org/10.2337/dc19-0803.

5. Ziegler, A.G., and Bonifacio, E.; BABYDIAB-BABYDIET Study Group (2012). Age-related islet autoantibody incidence in offspring of patients with type 1 diabetes. Diabetologia *55*, 1937–1943. https://doi.org/10.1007/s00125-012-2472-x.
6. Krischer, J.P., Lynch, K.F., Lernmark, Å., Hagopian, W.A., Rewers, M.J., She, J.-X., Toppari, J., Ziegler, A.-G., and Akolkar, B.; TEDDY Study Group (2017). Genetic and Environmental Interactions Modify the Risk of Diabetes-Related Autoimmunity by 6 Years of Age: The TEDDY Study. Diabetes Care *40*, 1194–1202. https://doi.org/10.2337/dc17-0238.
7. Leete, P., Mallone, R., Richardson, S.J., Sosenko, J.M., Redondo, M.J., and Evans-Molina, C. (2018). The effect of age on the progression and severity of type 1 diabetes: potential effects on disease mechanisms. Curr. Diab. Rep. *18*, 115. https://doi.org/10.1007/s11892-018-1083-4.
8. Greenbaum, C.J., Beam, C.A., Boulware, D., Gitelman, S.E., Gottlieb, P.A., Herold, K.C., Lachin, J.M., McGee, P., Palmer, J.P., Pescovitz, M.D., et al. (2012). Fall in C-peptide during first 2 years from diagnosis: evidence of at least two distinct phases from composite Type 1 Diabetes TrialNet data. Diabetes *61*, 2066–2073. https://doi.org/10.2337/db11-1538.
9. Hao, W., Gitelman, S., DiMeglio, L.A., Boulware, D., and Greenbaum, C.J.; Type 1 Diabetes TrialNet Study Group (2016). Fall in C-Peptide During First 4 Years From Diagnosis of Type 1 Diabetes: Variable Relation to Age, HbA1c, and Insulin Dose. Diabetes Care *39*, 1664–1670. https://doi.org/10.2337/dc16-0360.
10. Campbell-Thompson, M., Fu, A., Kaddis, J.S., Wasserfall, C., Schatz, D.A., Pugliese, A., and Atkinson, M.A. (2016). Insulitis and β-Cell Mass in the Natural History of Type 1 Diabetes. Diabetes *65*, 719–731. https://doi.org/10.2337/db15-0779.
11. Apaolaza, P.S., Balcacean, D., Zapardiel-Gonzalo, J., and Rodriguez-Calvo, T. (2023). The extent and magnitude of islet T cell infiltration as powerful tools to define the progression to type 1 diabetes. Diabetologia *66*, 1129–1141. https://doi.org/10.1007/s00125-023-05888-6.
12. Nigi, L., Maccora, C., Dotta, F., and Sebastiani, G. (2020). From immunohistological to anatomical alterations of human pancreas in type 1 diabetes: New concepts on the stage. Diabetes. Metab. Res. Rev. *36*, e3264. https://doi.org/10.1002/dmrr.3264.
13. Ross, C., Ward, Z.J., Gomber, A., Owais, M., Yeh, J.M., Reddy, C.-L., and Atun, R. (2022). The prevalence of islet autoantibodies in children and adolescents with type 1 diabetes mellitus: A global scoping review. Front. Endocrinol. *13*, 815703. https://doi.org/10.3389/fendo.2022.815703.
14. Battaglia, M., Ahmed, S., Anderson, M.S., Atkinson, M.A., Becker, D., Bingley, P.J., Bosi, E., Brusko, T.M., DiMeglio, L.A., Evans-Molina, C., et al. (2020). Introducing the endotype concept to address the challenge of disease heterogeneity in type 1 diabetes. Diabetes Care *43*, 5–12. https://doi.org/10.2337/dc19-0880.
15. Leete, P., Willcox, A., Krogvold, L., Dahl-Jørgensen, K., Foulis, A.K., Richardson, S.J., and Morgan, N.G. (2016). Differential Insulitic Profiles Determine the Extent of β-Cell Destruction and the Age at Onset of Type 1 Diabetes. Diabetes *65*, 1362–1369. https://doi.org/10.2337/db15-1615.
16. Arif, S., Leete, P., Nguyen, V., Marks, K., Nor, N.M., Estorninho, M., Kronenberg-Versteeg, D., Bingley, P.J., Todd, J.A., Guy, C., et al. (2014). Blood and islet phenotypes indicate immunological heterogeneity in type 1 diabetes. Diabetes *63*, 3835–3845. https://doi.org/10.2337/db14-0365.
17. Leete, P., Oram, R.A., McDonald, T.J., Shields, B.M., Ziller, C., TIGI study team; Hattersley, A.T., Richardson, S.J., and Morgan, N.G. (2020). Studies of insulin and proinsulin in pancreas and serum support the existence of aetiopathological endotypes of type 1 diabetes associated with age at diagnosis. Diabetologia *63*, 1258–1267. https://doi.org/10.1007/s00125-020-05115-6.
18. Krischer, J.P., Lynch, K.F., Schatz, D.A., Ilonen, J., Lernmark, Å., Hagopian, W.A., Rewers, M.J., She, J.-X., Simell, O.G., Toppari, J., et al. (2015). The 6 year incidence of diabetes-associated autoantibodies in genetically at-risk children: the TEDDY study. Diabetologia *58*, 980–987. https://doi.org/10.1007/s00125-015-3514-y.
19. Vehik, K., Bonifacio, E., Lernmark, Å., Yu, L., Williams, A., Schatz, D., Rewers, M., She, J.-X., Toppari, J., Hagopian, W., et al. (2020). Hierarchical order of distinct autoantibody spreading and progression to type 1 diabetes in the TEDDY study. Diabetes Care *43*, 2066–2073. https://doi.org/10.2337/dc19-2547.
20. Deligne, C., You, S., and Mallone, R. (2022). Personalized immunotherapies for type 1 diabetes: who, what, when, and how? J. Pers. Med. *12*, 542. https://doi.org/10.3390/jpm12040542.
21. Mathieu, C., Lahesmaa, R., Bonifacio, E., Achenbach, P., and Tree, T. (2018). Immunological biomarkers for the development and progression of type 1 diabetes. Diabetologia *61*, 2252–2258. https://doi.org/10.1007/s00125-018-4726-8.
22. Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell *75*, 843–854. https://doi.org/10.1016/0092-8674(93)90529-y.
23. Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell *75*, 855–862. https://doi.org/10.1016/0092-8674(93)90530-4.
24. Flynt, A.S., and Lai, E.C. (2008). Biological principles of microRNA-mediated regulation: shared themes amid diversity. Nat. Rev. Genet. *9*, 831–842. https://doi.org/10.1038/nrg2455.
25. Ventriglia, G., Nigi, L., Sebastiani, G., and Dotta, F. (2015). MicroRNAs: Novel Players in the Dialogue between Pancreatic Islets and Immune System in Autoimmune Diabetes. BioMed Res. Int. *2015*, 749734. https://doi.org/10.1155/2015/749734.
26. Roggli, E., Gattesco, S., Caille, D., Briet, C., Boitard, C., Meda, P., and Regazzi, R. (2012). Changes in microRNA expression contribute to pancreatic β-cell dysfunction in prediabetic NOD mice. Diabetes *61*, 1742–1751. https://doi.org/10.2337/db11-1086.
27. Grieco, F.A., Sebastiani, G., Juan-Mateu, J., Villate, O., Marroqui, L., Ladrière, L., Tugay, K., Regazzi, R., Bugliani, M., Marchetti, P., et al. (2017). MicroRNAs miR-23a-3p, miR-23b-3p, and miR-149-5p Regulate the Expression of Proapoptotic BH3-Only Proteins DP5 and PUMA in Human Pancreatic β-Cells. Diabetes *66*, 100–112. https://doi.org/10.2337/db16-0592.
28. Sebastiani, G., Ventriglia, G., Stabilini, A., Socci, C., Morsiani, C., Laurenzi, A., Nigi, L., Formichi, C., Mfarrej, B., Petrelli, A., et al. (2017). Regulatory T-cells from pancreatic lymphnodes of patients with type-1 diabetes express increased levels of microRNA miR-125a-5p that limits CCR2 expression. Sci. Rep. *7*, 6897. https://doi.org/10.1038/s41598-017-07172-1.
29. Guay, C., Kruit, J.K., Rome, S., Menoud, V., Mulder, N.L., Jurdzinski, A., Mancarella, F., Sebastiani, G., Donda, A., Gonzalez, B.J., et al. (2019). Lymphocyte-Derived Exosomal MicroRNAs Promote Pancreatic β Cell Death and May Contribute to Type 1 Diabetes Development. Cell Metab. *29*, 348–361.e6. https://doi.org/10.1016/j.cmet.2018.09.011.
30. Ventriglia, G., Mancarella, F., Sebastiani, G., Cook, D.P., Mallone, R., Mathieu, C., Gysemans, C., and Dotta, F. (2020). miR-409-3p is reduced in plasma and islet immune infiltrates of NOD diabetic mice and is differentially expressed in people with type 1 diabetes. Diabetologia *63*, 124–136. https://doi.org/10.1007/s00125-019-05026-1.
31. Sebastiani, G., Grieco, F.A., Spagnuolo, I., Galleri, L., Cataldo, D., and Dotta, F. (2011). Increased expression of microRNA miR-326 in type 1 diabetic patients with ongoing islet autoimmunity. Diabetes. Metab. Res. Rev. *27*, 862–866. https://doi.org/10.1002/dmrr.1262.

32. Lawrie, C.H., Gal, S., Dunlop, H.M., Pushkaran, B., Liggins, A.P., Pulford, K., Banham, A.H., Pezzella, F., Boultwood, J., Wainscoat, J.S., et al. (2008). Detection of elevated levels of tumour-associated microRNAs in serum of patients with diffuse large B-cell lymphoma. Br. J. Haematol. *141*, 672–675. https://doi.org/10.1111/j.1365-2141.2008.07077.x.

33. Chim, S.S.C., Shing, T.K.F., Hung, E.C.W., Leung, T.-Y., Lau, T.-K., Chiu, R.W.K., and Lo, Y.M.D. (2008). Detection and characterization of placental microRNAs in maternal plasma. Clin. Chem. *54*, 482–490. https://doi.org/10.1373/clinchem.2007.097972.

34. Thomou, T., Mori, M.A., Dreyfuss, J.M., Konishi, M., Sakaguchi, M., Wolfrum, C., Rao, T.N., Winnay, J.N., Garcia-Martin, R., Grinspoon, S.K., et al. (2017). Adipose-derived circulating miRNAs regulate gene expression in other tissues. Nature *542*, 450–455. https://doi.org/10.1038/nature21365.

35. Sebastiani, G., Nigi, L., Grieco, G.E., Mancarella, F., Ventriglia, G., and Dotta, F. (2017). Circulating microRNAs and diabetes mellitus: a novel tool for disease prediction, diagnosis, and staging? J. Endocrinol. Invest. *40*, 591–610. https://doi.org/10.1007/s40618-017-0611-4.

36. Nielsen, L.B., Wang, C., Sørensen, K., Bang-Berthelsen, C.H., Hansen, L., Andersen, M.-L.M., Hougaard, P., Juul, A., Zhang, C.-Y., Pociot, F., and Mortensen, H.B. (2012). Circulating levels of microRNA from children with newly diagnosed type 1 diabetes and healthy controls: evidence that miR-25 associates to residual beta-cell function and glycaemic control during disease progression. Exp. Diabetes Res. *2012*, 896362. https://doi.org/10.1155/2012/896362.

37. Samandari, N., Mirza, A.H., Nielsen, L.B., Kaur, S., Hougaard, P., Fredheim, S., Mortensen, H.B., and Pociot, F. (2017). Circulating microRNA levels predict residual beta cell function and glycaemic control in children with type 1 diabetes mellitus. Diabetologia *60*, 354–363. https://doi.org/10.1007/s00125-016-4156-4.

38. Garavelli, S., Bruzzaniti, S., Tagliabue, E., Di Silvestre, D., Prattichizzo, F., Mozzillo, E., Fattorusso, V., La Sala, L., Ceriello, A., Puca, A.A., et al. (2020). Plasma circulating miR-23∼27∼24 clusters correlate with the immunometabolic derangement and predict C-peptide loss in children with type 1 diabetes. Diabetologia *63*, 2699–2712. https://doi.org/10.1007/s00125-020-05237-x.

39. Grieco, G.E., Cataldo, D., Ceccarelli, E., Nigi, L., Catalano, G., Brusco, N., Mancarella, F., Ventriglia, G., Fondelli, C., Guarino, E., et al. (2018). Serum Levels of miR-148a and miR-21-5p Are Increased in Type 1 Diabetic Patients and Correlated with Markers of Bone Strength and Metabolism. Noncoding RNA *4*, 37. https://doi.org/10.3390/ncrna4040037.

40. Seyhan, A.A., Nunez Lopez, Y.O., Xie, H., Yi, F., Mathews, C., Pasarica, M., and Pratley, R.E. (2016). Pancreas-enriched miRNAs are altered in the circulation of subjects with diabetes: a pilot cross-sectional study. Sci. Rep. *6*, 31479. https://doi.org/10.1038/srep31479.

41. Erener, S., Marwaha, A., Tan, R., Panagiotopoulos, C., and Kieffer, T.J. (2017). Profiling of circulating microRNAs in children with recent onset of type 1 diabetes. JCI Insight *2*, e89656. https://doi.org/10.1172/jci.insight.89656.

42. Åkerman, L., Casas, R., Ludvigsson, J., Tavira, B., and Skoglund, C. (2018). Serum miRNA levels are related to glucose homeostasis and islet autoantibodies in children with high risk for type 1 diabetes. PLoS One *13*, e0191067. https://doi.org/10.1371/journal.pone.0191067.

43. Małachowska, B., Wyka, K., Nowicka, Z., Bartłomiejczyk, M.A., Młynarski, W., and Fendler, W. (2018). Temporal dynamics of serum let-7g expression mirror the decline of residual beta-cell function in longitudinal observation of children with type 1 diabetes. Pediatr. Diabetes *19*, 1407–1415. https://doi.org/10.1111/pedi.12783.

44. Assmann, T.S., Recamonde-Mendoza, M., Puñales, M., Tschiedel, B., Canani, L.H., and Crispim, D. (2018). MicroRNA expression profile in plasma from type 1 diabetic patients: Case-control study and bioinformatic analysis. Diabetes Res. Clin. Pract. *141*, 35–46. https://doi.org/10.1016/j.diabres.2018.03.044.

45. Liu, Y., Ma, M., Yu, J., Ping, F., Zhang, H., Li, W., Xu, L., and Li, Y. (2019). Decreased Serum microRNA-21, microRNA-25, microRNA-146a, and microRNA-181a in Autoimmune Diabetes: Potential Biomarkers for Diagnosis and Possible Involvement in Pathogenesis. Int. J. Endocrinol. *2019*, 8406438. https://doi.org/10.1155/2019/8406438.

46. Latreille, M., Herrmanns, K., Renwick, N., Tuschl, T., Malecki, M.T., McCarthy, M.I., Owen, K.R., Rülicke, T., and Stoffel, M. (2015). miR-375 gene dosage in pancreatic β-cells: implications for regulation of β-cell mass and biomarker development. J. Mol. Med. *93*, 1159–1169. https://doi.org/10.1007/s00109-015-1296-9.

47. Marchand, L., Jalabert, A., Meugnier, E., Van den Hende, K., Fabien, N., Nicolino, M., Madec, A.-M., Thivolet, C., and Rome, S. (2016). miRNA-375 a Sensor of Glucotoxicity Is Altered in the Serum of Children with Newly Diagnosed Type 1 Diabetes. J. Diabetes Res. *2016*, 1869082. https://doi.org/10.1155/2016/1869082.

48. Lakhter, A.J., Pratt, R.E., Moore, R.E., Doucette, K.K., Maier, B.F., DiMeglio, L.A., and Sims, E.K. (2018). Beta cell extracellular vesicle miR-21-5p cargo is increased in response to inflammatory cytokines and serves as a biomarker of type 1 diabetes. Diabetologia *61*, 1124–1134. https://doi.org/10.1007/s00125-018-4559-5.

49. Bertoccini, L., Sentinelli, F., Incani, M., Bailetti, D., Cimini, F.A., Barchetta, I., Lenzi, A., Cavallo, M.G., Cossu, E., and Baroni, M.G. (2019). Circulating miRNA-375 levels are increased in autoantibodies-positive first-degree relatives of type 1 diabetes patients. Acta Diabetol. *56*, 707–710. https://doi.org/10.1007/s00592-019-01297-7.

50. Januszewski, A.S., Cho, Y.H., Joglekar, M.V., Farr, R.J., Scott, E.S., Wong, W.K.M., Carroll, L.M., Loh, Y.W., Benitez-Aguirre, P.Z., Keech, A.C., et al. (2021). Insulin micro-secretion in Type 1 diabetes and related microRNA profiles. Sci. Rep. *11*, 11727. https://doi.org/10.1038/s41598-021-90856-6.

51. Grieco, G.E., Sebastiani, G., Fignani, D., Brusco, N., Nigi, L., Formichi, C., Licata, G., Bruttini, M., D'Aurizio, R., Mathieu, C., et al. (2021). Protocol to analyze circulating small non-coding RNAs by high-throughput RNA sequencing from human plasma samples. STAR Protoc. *2*, 100606. https://doi.org/10.1016/j.xpro.2021.100606.

52. Dunger, D.B., Bruggraber, S.F.A., Mander, A.P., Marcovecchio, M.L., Tree, T., Chmura, P.J., Knip, M., Schulte, A.M., and Mathieu, C.; INNODIA consortium (2022). INNODIA Master Protocol for the evaluation of investigational medicinal products in children, adolescents and adults with newly diagnosed type 1 diabetes. Trials *23*, 414. https://doi.org/10.1186/s13063-022-06259-z.

53. Chugh, P., and Dittmer, D.P. (2012). Potential pitfalls in microRNA profiling. Wiley Interdiscip. Rev. RNA *3*, 601–616. https://doi.org/10.1002/wrna.1120.

54. Androvic, P., Benesova, S., Rohlova, E., Kubista, M., and Valihrach, L. (2022). Small RNA-Sequencing for Analysis of Circulating miRNAs: Benchmark Study. J. Mol. Diagn. *24*, 386–394. https://doi.org/10.1016/j.jmoldx.2021.12.006.

55. Raabe, C.A., Tang, T.-H., Brosius, J., and Rozhdestvensky, T.S. (2014). Biases in small RNA deep sequencing data. Nucleic Acids Res. *42*, 1414–1426. https://doi.org/10.1093/nar/gkt1021.

56. Coenen-Stass, A.M.L., Magen, I., Brooks, T., Ben-Dov, I.Z., Greensmith, L., Hornstein, E., and Fratta, P. (2018). Evaluation of methodologies for microRNA biomarker detection by next generation sequencing. RNA Biol. *15*, 1133–1145. https://doi.org/10.1080/15476286.2018.1514236.

57. Wong, R.K.Y., MacMahon, M., Woodside, J.V., and Simpson, D.A. (2019). A comparison of RNA extraction and sequencing protocols for detection of small RNAs in plasma. BMC Genom. *20*, 446. https://doi.org/10.1186/s12864-019-5826-7.

58. Rousseeuw, P.J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. *20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7.

59. Aparicio-Puerta, E., Hirsch, P., Schmartz, G.P., Fehlmann, T., Keller, V., Engel, A., Kern, F., Hackenberg, M., and Keller, A. (2023). isomiRdb: microRNA

expression at isoform resolution. Nucleic Acids Res. *51*, D179–D185. https://doi.org/10.1093/nar/gkac884.

60. Erener, S., Mojibian, M., Fox, J.K., Denroche, H.C., and Kieffer, T.J. (2013). Circulating miR-375 as a biomarker of β-cell death and diabetes in mice. Endocrinology *154*, 603–608. https://doi.org/10.1210/en.2012-1744.
61. Snowhite, I.V., Allende, G., Sosenko, J., Pastori, R.L., Messinger Cayetano, S., and Pugliese, A. (2017). Association of serum microRNAs with islet autoimmunity, disease progression and metabolic impairment in relatives at risk of type 1 diabetes. Diabetologia *60*, 1409–1422. https://doi.org/10.1007/s00125-017-4294-3.
62. Snowhite, I., Pastori, R., Sosenko, J., Messinger Cayetano, S., and Pugliese, A. (2021). Baseline assessment of circulating microRNAs near diagnosis of type 1 diabetes predicts future stimulated insulin secretion. Diabetes *70*, 638–651. https://doi.org/10.2337/db20-0817.
63. Samandari, N., Mirza, A.H., Kaur, S., Hougaard, P., Nielsen, L.B., Fredheim, S., Mortensen, H.B., and Pociot, F. (2018). Influence of Disease Duration on Circulating Levels of miRNAs in Children and Adolescents with New Onset Type 1 Diabetes. Noncoding RNA *4*, 35. https://doi.org/10.3390/ncrna4040035.
64. den Hollander, N.H.M., and Roep, B.O. (2022). From disease and patient heterogeneity to precision medicine in type 1 diabetes. Front. Med. *9*, 932086. https://doi.org/10.3389/fmed.2022.932086.
65. Grieco, G.E., Besharat, Z.M., Licata, G., Fignani, D., Brusco, N., Nigi, L., Formichi, C., Po, A., Sabato, C., Dardano, A., et al. (2022). Circulating microRNAs as clinically useful biomarkers for Type 2 Diabetes Mellitus: miRNomics from bench to bedside. Transl. Res. *247*, 137–157. https://doi.org/10.1016/j.trsl.2022.03.008.
66. Li, S., Dragan, I., Fung, C.H., Kuznetsov, D., Hansen, M.K., Beulens, J.W., 't Hart, L.M., Slieker, R.C., Donnelly, L.A., Gerl, M.J., et al. (2022). Novel subgroups of type 2 diabetes based on multi-Omics profiling: an IMI-RHAPSODY Study. Preprint at medRxiv. https://doi.org/10.1101/2022.09.03.22279563.
67. Xie, Q.Y., Almudevar, A., Whitney-Miller, C.L., Barry, C.T., and McCall, M.N. (2016). A microRNA biomarker of hepatocellular carcinoma recurrence following liver transplantation accounting for within-patient heterogeneity. BMC Med. Genomics *9*, 18. https://doi.org/10.1186/s12920-016-0179-4.
68. Kruger, D.T., Beelen, K.J., Opdam, M., Sanders, J., van der Noort, V., Boven, E., and Linn, S.C. (2018). Hierarchical clustering of activated proteins in the PI3K and MAPK pathways in ER-positive, HER2-negative breast cancer with potential therapeutic consequences. Br. J. Cancer *119*, 832–839. https://doi.org/10.1038/s41416-018-0221-8.
69. Van Laar, R., Latif, B., King, S., Love, C., Taubenheim, N., Kalansooriya, E., Wang, W., Saad, M., Winship, I., Azzi, A., et al. (2023). Validation of a microRNA liquid biopsy assay for diagnosis and risk stratification of invasive cutaneous melanoma. Br. J. Dermatol. *189*, 292–301. https://doi.org/10.1093/bjd/ljad137.
70. Ferracin, M., and Negrini, M. (2018). Quantification of circulating microRNAs by droplet digital PCR. Methods Mol. Biol. *1768*, 445–457. https://doi.org/10.1007/978-1-4939-7778-9_25.
71. Giannopoulou, E.Z., Winkler, C., Chmiel, R., Matzke, C., Scholz, M., Beyerlein, A., Achenbach, P., Bonifacio, E., and Ziegler, A.-G. (2015). Islet autoantibody phenotypes and incidence in children at increased risk for type 1 diabetes. Diabetologia *58*, 2317–2323. https://doi.org/10.1007/s00125-015-3672-y.
72. Krischer, J.P., Liu, X., Vehik, K., Akolkar, B., Hagopian, W.A., Rewers, M.J., She, J.-X., Toppari, J., Ziegler, A.-G., and Lernmark, Å.; TEDDY Study Group (2019). Predicting Islet Cell Autoimmunity and Type 1 Diabetes: An 8-Year TEDDY Study Progress Report. Diabetes Care *42*, 1051–1060. https://doi.org/10.2337/dc18-2282.
73. Johnston, R.J., Comps-Agrar, L., Hackney, J., Yu, X., Huseni, M., Yang, Y., Park, S., Javinal, V., Chiu, H., Irving, B., et al. (2014). The immunoreceptor TIGIT regulates antitumor and antiviral CD8(+) T cell effector function. Cancer Cell *26*, 923–937. https://doi.org/10.1016/j.ccell.2014.10.018.
74. Chew, G.M., Fujita, T., Webb, G.M., Burwitz, B.J., Wu, H.L., Reed, J.S., Hammond, K.B., Clayton, K.L., Ishii, N., Abdel-Mohsen, M., et al. (2016). TIGIT Marks Exhausted T Cells, Correlates with Disease Progression, and Serves as a Target for Immune Restoration in HIV and SIV Infection. PLoS Pathog. *12*, e1005349. https://doi.org/10.1371/journal.ppat.1005349.
75. Long, S.A., Thorpe, J., DeBerg, H.A., Gersuk, V., Eddy, J., Harris, K.M., Ehlers, M., Herold, K.C., Nepom, G.T., and Linsley, P.S. (2016). Partial exhaustion of CD8 T cells and clinical response to teplizumab in new-onset type 1 diabetes. Sci. Immunol. *1*, eaai7793. https://doi.org/10.1126/sciimmunol.aai7793.
76. Tooley, J.E., Vudattu, N., Choi, J., Cotsapas, C., Devine, L., Raddassi, K., Ehlers, M.R., McNamara, J.G., Harris, K.M., Kanaparthi, S., et al. (2016). Changes in T-cell subsets identify responders to FcR-nonbinding anti-CD3 mAb (teplizumab) in patients with type 1 diabetes. Eur. J. Immunol. *46*, 230–241. https://doi.org/10.1002/eji.201545708.
77. Herold, K.C., Gitelman, S.E., Willi, S.M., Gottlieb, P.A., Waldron-Lynch, F., Devine, L., Sherr, J., Rosenthal, S.M., Adi, S., Jalaludin, M.Y., et al. (2013). Teplizumab treatment may improve C-peptide responses in participants with type 1 diabetes after the new-onset period: a randomised controlled trial. Diabetologia *56*, 391–400. https://doi.org/10.1007/s00125-012-2753-4.
78. Herold, K.C., Bundy, B.N., Long, S.A., Bluestone, J.A., DiMeglio, L.A., Dufort, M.J., Gitelman, S.E., Gottlieb, P.A., Krischer, J.P., Linsley, P.S., et al. (2019). An Anti-CD3 Antibody, Teplizumab, in Relatives at Risk for Type 1 Diabetes. N. Engl. J. Med. *381*, 603–613. https://doi.org/10.1056/NEJMoa1902226.
79. Diggins, K.E., Serti, E., Muir, V., Rosasco, M., Lu, T., Balmas, E., Nepom, G., Long, S.A., and Linsley, P.S. (2021). Exhausted-like CD8+ T cell phenotypes linked to C-peptide preservation in alefacept-treated T1D subjects. JCI Insight *6*, e142680. https://doi.org/10.1172/jci.insight.142680.
80. Wiedeman, A.E., Muir, V.S., Rosasco, M.G., DeBerg, H.A., Presnell, S., Haas, B., Dufort, M.J., Speake, C., Greenbaum, C.J., Serti, E., et al. (2020). Autoreactive CD8+ T cell exhaustion distinguishes subjects with slow type 1 diabetes progression. J. Clin. Invest. *130*, 480–490. https://doi.org/10.1172/JCI126595.
81. Vignali, D., Cantarelli, E., Bordignon, C., Canu, A., Citro, A., Annoni, A., Piemonti, L., and Monti, P. (2018). Detection and characterization of CD8+ autoreactive memory stem T cells in patients with type 1 diabetes. Diabetes *67*, 936–945. https://doi.org/10.2337/db17-1390.
82. Fazeli, P., Talepoor, A.G., Faghih, Z., Gholijani, N., Ataollahi, M.R., Ali-Hassanzadeh, M., Moravej, H., and Kalantar, K. (2022). The frequency of CD4+ and CD8+ circulating T stem cell memory in type 1 diabetes. Immun. Inflamm. Dis. *10*, e715. https://doi.org/10.1002/iid3.715.
83. Benetatos, L., Vartholomatos, G., and Hatzimichael, E. (2011). MEG3 imprinted gene contribution in tumorigenesis. Int. J. Cancer *129*, 773–779. https://doi.org/10.1002/ijc.26052.
84. Wallace, C., Smyth, D.J., Maisuria-Armer, M., Walker, N.M., Todd, J.A., and Clayton, D.G. (2010). The imprinted DLK1-MEG3 gene region on chromosome 14q32.2 alters susceptibility to type 1 diabetes. Nat. Genet. *42*, 68–71. https://doi.org/10.1038/ng.493.
85. Zhou, Q., Frost, R.J.A., Anderson, C., Zhao, F., Ma, J., Yu, B., and Wang, S. (2017). let-7 Contributes to Diabetic Retinopathy but Represses Pathological Ocular Angiogenesis. Mol. Cell Biol. *37*, e00001-17. https://doi.org/10.1128/MCB.00001-17.
86. Brennan, E., Wang, B., McClelland, A., Mohan, M., Marai, M., Beuscart, O., Derouiche, S., Gray, S., Pickering, R., Tikellis, C., et al. (2017). Protective Effect of let-7 miRNA Family in Regulating Inflammation in Diabetes-Associated Atherosclerosis. Diabetes *66*, 2266–2277. https://doi.org/10.2337/db16-1405.
87. Salas-Pérez, F., Codner, E., Valencia, E., Pizarro, C., Carrasco, E., and Pérez-Bravo, F. (2013). MicroRNAs miR-21a and miR-93 are down

regulated in peripheral blood mononuclear cells (PBMCs) from patients with type 1 diabetes. Immunobiology *218*, 733–737. https://doi.org/10.1016/j.imbio.2012.08.276.

88. Giannella, A., Castelblanco, E., Zambon, C.F., Basso, D., Hernandez, M., Ortega, E., Alonso, N., Mauricio, D., Avogaro, A., Ceolotto, G., and Vigili de Kreutzenberg, S. (2023). Circulating small noncoding RNA profiling as a potential biomarker of atherosclerotic plaque composition in type 1 diabetes. Diabetes Care *46*, 551–560. https://doi.org/10.2337/dc22-1441.
89. Hezova, R., Slaby, O., Faltejskova, P., Mikulkova, Z., Buresova, I., Raja, K.R.M., Hodek, J., Ovesna, J., and Michalek, J. (2010). microRNA-342, microRNA-191 and microRNA-510 are differentially expressed in T regulatory cells of type 1 diabetic patients. Cell. Immunol. *260*, 70–74. https://doi.org/10.1016/j.cellimm.2009.10.012.
90. Yang, M., Ye, L., Wang, B., Gao, J., Liu, R., Hong, J., Wang, W., Gu, W., and Ning, G. (2015). Decreased miR-146 expression in peripheral blood mononuclear cells is correlated with ongoing islet autoimmunity in type 1 diabetes patients. J. Diabetes *7*, 158–165. https://doi.org/10.1111/1753-0407.12163.
91. Assmann, T.S., Duarte, G.C.K., Brondani, L.A., de Freitas, P.H.O., Martins, É.M., Canani, L.H., and Crispim, D. (2017). Polymorphisms in genes encoding miR-155 and miR-146a are associated with protection to type 1 diabetes mellitus. Acta Diabetol. *54*, 433–441. https://doi.org/10.1007/s00592-016-0961-y.
92. Wang, G., Gu, Y., Xu, N., Zhang, M., and Yang, T. (2018). Decreased expression of miR-150, miR146a and miR424 in type 1 diabetic patients: Association with ongoing islet autoimmunity. Biochem. Biophys. Res. Commun. *498*, 382–387. https://doi.org/10.1016/j.bbrc.2017.06.196.
93. Syed, F., Krishnan, P., Chang, G., Langlais, S.R., Hati, S., Yamada, K., Lam, A.K., Talware, S., Liu, X., Sardar, R., et al. (2023). β Cell microRNAs Function as Molecular Hubs of Type 1 Diabetes Pathogenesis and as Biomarkers of Diabetes Risk. Preprint at bioRxiv. https://doi.org/10.1101/2023.06.15.545170.
94. Cavaillé, J., Seitz, H., Paulsen, M., Ferguson-Smith, A.C., and Bachellerie, J.-P. (2002). Identification of tandemly-repeated C/D snoRNA genes at the imprinted human 14q32 domain reminiscent of those at the Prader-Willi/Angelman syndrome region. Hum. Mol. Genet. *11*, 1527–1538. https://doi.org/10.1093/hmg/11.13.1527.
95. Song, G., and Wang, L. (2008). MiR-433 and miR-127 arise from independent overlapping primary transcripts encoded by the miR-433-127 locus. PLoS One *3*, e3574. https://doi.org/10.1371/journal.pone.0003574.
96. Benetatos, L., Hatzimichael, E., Londin, E., Vartholomatos, G., Loher, P., Rigoutsos, I., and Briasoulis, E. (2013). The microRNAs within the DLK1-DIO3 genomic region: involvement in disease pathogenesis. Cell. Mol. Life Sci. *70*, 795–814. https://doi.org/10.1007/s00018-012-1080-8.
97. Labialle, S., Marty, V., Bortolin-Cavaillé, M.-L., Hoareau-Osman, M., Pradère, J.-P., Valet, P., Martin, P.G.P., and Cavaillé, J. (2014). The miR-379/miR-410 cluster at the imprinted Dlk1-Dio3 domain controls neonatal metabolic adaptation. EMBO J. *33*, 2216–2230. https://doi.org/10.15252/embj.201387038.
98. van de Bunt, M., Gaulton, K.J., Parts, L., Moran, I., Johnson, P.R., Lindgren, C.M., Ferrer, J., Gloyn, A.L., and McCarthy, M.I. (2013). The miRNA profile of human pancreatic islets and beta-cells and relationship to type 2 diabetes pathogenesis. PLoS One *8*, e55272. https://doi.org/10.1371/journal.pone.0055272.
99. Kameswaran, V., Golson, M.L., Ramos-Rodríguez, M., Ou, K., Wang, Y.J., Zhang, J., Pasquali, L., and Kaestner, K.H. (2018). The Dysregulation of the DLK1-MEG3 Locus in Islets From Patients With Type 2 Diabetes Is Mimicked by Targeted Epimutation of Its Promoter With TALE-DNMT Constructs. Diabetes *67*, 1807–1815. https://doi.org/10.2337/db17-0682.
100. Klein, D., Misawa, R., Bravo-Egana, V., Vargas, N., Rosero, S., Piroso, J., Ichii, H., Umland, O., Zhijie, J., Tsinoremas, N., et al. (2013). MicroRNA expression in alpha and beta cells of human pancreatic islets. PLoS One *8*, e55064. https://doi.org/10.1371/journal.pone.0055064.
101. Abuhatzira, L., Xu, H., Tahhan, G., Boulougoura, A., Schäffer, A.A., and Notkins, A.L. (2015). Multiple microRNAs within the 14q32 cluster target the mRNAs of major type 1 diabetes autoantigens IA-2, IA-2β, and GAD65. FASEB J. *29*, 4374–4383. https://doi.org/10.1096/fj.15-273649.
102. Grieco, G.E., Fignani, D., Formichi, C., Nigi, L., Licata, G., Maccora, C., Brusco, N., Sebastiani, G., and Dotta, F. (2021). Extracellular Vesicles in Immune System Regulation and Type 1 Diabetes: Cell-to-Cell Communication Mediators, Disease Biomarkers, and Promising Therapeutic Tools. Front. Immunol. *12*, 682948. https://doi.org/10.3389/fimmu.2021.682948.
103. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.
104. Cribari-Neto, F., and Zeileis, A. (2010). Beta Regression in R. J. Stat. Softw. *34*. https://doi.org/10.18637/jss.v034.i02.
105. Smithson, M., and Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. Psychol. Methods *11*, 54–71. https://doi.org/10.1037/1082-989X.11.1.54.
106. Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559. https://doi.org/10.1186/1471-2105-9-559.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Human plasma samples	INNODIA consortium project	https://www.innodia.eu/
Chemicals, peptides, and recombinant proteins		
Ethanol, 200 proof	Sigma Aldrich	SHBL.2095/SHBK9944-5
Library Dilution Buffer	HTG	307900/308171
10 mM Tris pH 8.5 Elution Buffer	QIAGEN	163034340
2N NaOH	HTG	308142
2N HCl	HTG	307876
Nuclease-Free Water (10 × 50 mL)	QIAGEN	129114
Beta-Mercaptoethanol	Sigma-Aldrich	M7522
Tris EDTA buffer, for molecular biology, DNAse, RNAse, Proteasefree ready to use, pH 8.0	Acros Organics	AC327345000
Critical commercial assays		
Plasma/Serum RNA Purification Mini Kit	NORGEN	55000
QIAseq miRNA Library Kit (96)	QIAGEN	331505
QIAseq miRNA 96 Index IL (96)	QIAGEN	331565
PhiX Control v3	ILLUMINA	FC-110-3001
NovaSeq 6000 SP Reagent Kit (100 cycles)	ILLUMINA	20027464
NovaSeq XP 2-Lane Kit	ILLUMINA	20021664
Thermo Qubit DSDNA HS Assay	LIFE TECHNOLOGIES	Q32854
Kit per DNA ad elevata intensità	AGILENT	5067–4626
NovaSeq 6000 S1 Reagent Kit v1.5 (100 cycles)	ILLUMINA	20028319
HTG Lysis Buffer	HTG	307914
MDx Proteinase K	Invitrogen	2036707
HTG Plasma Lysis Buffer	HTG	308251
EdgeSeq Drawer Reagents	HTG	308312
Human Brain RNA	Ambion	2077661-A-2
Hemo KlenTaq enzyme	New England Biolabs	10009977
Deoxynucleotide (dNTP) Solution Set	New England Biolabs	10014675
5X Hemo KlenTaq Reaction Buffer	New England Biolabs	0031801
PCR, H_20	HTG	308124
HTG Sequencing Tag Pack	HTG	308065
MIDx Ampure XP	HTG	308183
Bulk, Trizma Hydrochloride Solution (pH 8_0) Dilution, HTG EdgeSeq	HTG	308062
MDX, I-tm (Md. BIO)	Lonza	RNBH 1710/RNBH3183
KAPA Biosystems q PCR Kit	KAPA Biosystem	005098
MDx 12.5 PM PhiX	HTG	308076
HTI (Hybridization Buffer)	ILLUMINA	20373892/20384289/20391454
Nextseq 500/550 Buffer cartridge v2	ILLUMINA	20399134/20383558
NextSeq 500/550 High Output Reagent Cartridge v2	ILLUMINA	20396441
NextSeq 500/550 High Output now cell Cartridge v2	ILLUMINA	20387266/20398769
ILM Seq Read 1 Primer Mix Assy	HTG	308246
ILM Seq Index Primer Mix Assy	HTG	308245
TAQMAN PREAMP MASTER MIX (2X)	INVITROGEN	4488593

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
TAQMAN(R) MICRORNA RT KIT	INVITROGEN	4366597
TAQMAN MICRORNA ASSAY	INVITROGEN	cat.#4440887 ID 002332
TAQMAN MICRORNA ASSAY	INVITROGEN	cat.#4440887 ID 000452
TAQMAN MICRORNA ASSAY	INVITROGEN	cat.#4440887 ID 000572
TAQMAN MICRORNA ASSAY	INVITROGEN	cat.#4440887 ID 000564
ddPCR EvaGreen Supermix	BIORAD	1864034
ddPCR Smx for Probes (no dUTP), 5 × 1mL	BIORAD	1863024
DG8 Cartridges and Gaskets	BIORAD	1864007
ddPCR Plates 96-Well, Semi-Skirted (25 plates)	BIORAD	12001925
PIERCEABLE FOIL HEAT SEAL (no.100)	BIORAD	1814040
Droplet Gen Oil for Probes, 10 × 7 mL	BIORAD	1863005
Deposited data		
Small RNA seq First Cohort	GEO	GSE265980
HTG-Edge seq miRNA First Cohort	GEO	GSE265976
Small RNA seq Second Cohort	GEO	GSE265981
Software and algorithms		
Graphpad Prism v10.0	GraphPad Software (Boston, MA)	https://www.graphpad.com/
2100 Expert Software (vB.02.11)	Agilent	N/A; RRID:SCR_014466
Illumina Experiment Manager Software (v1.19.1)	ILLUMINA	N/A; RRID:SCR_021202
BaseSpace v. 1.1.0.64	ILLUMINA	N/A; RRID:SCR_011881
GeneGlobe Data Analysis Center	QIAGEN	N/A
ExpressionSuite Software (v1.3)	THERMO FISHER	N/A; RRID:SCR_021095
R/Bioconductor	N/A	N/A
Other		
Vacutainer Safety-Lok butterflies	BD Biosciences	367282
NovaSeq Xp Flow Cell Dock	ILLUMINA	20021663
Vacutainer K2 EDTA	BD Biosciences	368861
DNA LoBind Tubes 0.5 mL	Eppendorf	0030 108.035
DNA LoBind Tubes 1.5 mL	Eppendorf	0030 108.071
DNA LoBind Tubes 2.0 mL	Eppendorf	0030 108.058
96 Fast PCR Plate half skirt	SARSTEDT	72.1981.202
MicroAmp 96 Optical Adhesive Film 100 cover	Thermo Fisher	4311971
Strip 8 flat caps (Xtra-Clear)	STARLAB	I1400-0900
StarTub Reagent Reservoir (PVC)	STARLAB	E2310-1000
96 PCR Plate without skirt	SARSTEDT	72.1978.202
SimpliAmp Thermal Cycler	LIFE TECHNOLOGIES	A24811
Qubit Assay Tube Set	Thermo Fisher	Q32856
Centrifuge MiniSpin	Eppendorf	5452000010
Ministar centrifuge	VWR	521–2319
NovaSeq 6000 System	ILLUMINA	20012850
DynaMag-96 Side Magnet	INVITROGEN	12331D
DynaMag-2 Magnet	INVITROGEN	12321D
QUBIT 3.0 spectrofluorometer	INVITROGEN	N/A
2100 Bioanalyzer Instrument	Agilent	N/A
IKA MS3 S36 Agilent vortex IKA	Agilent	N/A
HTG EdgeSeq Processor	HTG	N/A
HTG EdgeSeq Processor	HTG	N/A
SimpliAmp Thermal Cycler	LIFE TECHNOLOGIES	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
StepOnePlus	APPLIED BIOSYSTEM	N/A
NextSeq	ILLUMINA	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Francesco Dotta (francesco.dotta@unisi.it).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The generated data regarding demographic and clinical parameters are person-sensitive, and access can be provided by application to the INNODIA Data Access Committee.
- Data: Circulating Small RNA-seq data from T1D individuals of the first and second cohort are available in Gene Expression Omnibus (GEO) with the following IDs: GEO: GSE265980 (First cohort untargeted-seq); GEO: GSE265976 (First cohort Targeted-seq); GEO: GSE265981 (Second cohort Untargeted-seq).
- Code: This study does not generate a custom code.
- Any additional information required to reanalyze the data reported in this work is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

T1DM individuals of the INNODIA first cohort

For circulating small RNA sequencing analysis, an initial cohort composed of 115 individuals with newly diagnosed (<6 weeks, 4.5 ± 1.5 weeks) type 1 diabetes were enrolled in INNODIA natural history study. T1DM individuals [all positive for at least one autoantibody (GADA, IA-2A, ZnT8A) and aged between 1 and 45 years] enrolled in this study were selected based on sample availability and even gender distribution (sex: 58F/57M; age at diagnosis: 12,4 ± 7,7 years) (complete clinical characteristics in Tables 1 and S1). T1DM individuals were followed-up to 24 months with programmed visits at 3- (V2), 6- (V3), 12- (V4) and 24-month (V5) after diagnosis. For circulating small RNAs study, we considered visits up to 12-month for the statistical association analysis with clinical parameters due to the uncompleteness at V5 at the moment of small RNA measurement.

Plasma samples for small RNAs sequencing were collected at baseline (V1, <6 weeks from diagnosis) through a standardised protocol[51] adopted by all clinical sites involved in the multicentric consortium.

T1DM individuals of the INNODIA second cohort

An independent cohort composed by n = 147 newly diagnosed (V1<6 weeks: 3.9 ± 1.8 weeks) T1DM individuals [sex: 92/55 (M/F); age at diagnosis: 11,89 ± 7,86] was also enrolled in INNODIA consortium to perform the same analyses described for the first cohort. All individuals were positive for at least one autoantibody (GADA, IA-2A, ZnT8A) and aged between 1 and 45 years. All individuals were followed-up with programmed visits at 3- (V2), 6- (V3) and 12-month (V4) after diagnosis (Tables 1 and S5). Plasma samples at baseline visit-1 (V1) were collected following the SOP reported above.

Study approval and ethics

The study followed the guidelines of the Declaration of Helsinki for research on human individuals, and the study was approved by the local ethical committees of each participating clinical sites (see Tables S2 and S7). Written informed consent was obtained from all participants before their participation in this study.

METHOD DETAILS

Blood samples and plasma processing

Blood samples were collected in K_2 or K_3-EDTA 3.5 mL tubes, inverted ten times and stored upright at room temperature (15°C–25°C), and processed within 2 h from blood draw. An initial centrifuge was performed at 1800Xg for 10 min at 15°C-25°C to separate blood cells from plasma. Then, plasma was collected in smaller tubes (i.e., 2 mL – Sterile, apyrogen and nuclease-free) avoiding touching the white blood cells interphase (leaving 2–3 mm of plasma layer over the leukocytes) and further centrifuged at 1200Xg

for 20 min at 10°C to remove contaminant cells and platelets. Multiple 200 μL plasma aliquots (where possible, 5 aliquots in nuclease-free tubes) were then stored at −80°C and then transferred to a central biobank located in Cambridge (UK) until final transfer to the analytical laboratory.

HTG EdgeSeq miRNA whole transcriptome assay (targeted-seq)

HTG EdgeSeq miRNA whole transcriptome assay method is an RNA extraction-free approach, and exploit quantitative nuclease protection assay (qNPA) chemistry with a subsequent Next Generation Sequencing (NGS) platform to allow semi-quantitative analysis of a panel of $n = 2102$ targeted miRNAs (including $n = 13$ housekeepings, $n = 5$ negative process controls, $n = 1$ positive process control and $n = 2083$ targeted miRNAs) from 15 μL of plasma. In PCR-based library preparation, each sample is used as a template for PCR reactions for specially designed primers (tags), which share common sequences complementary to 5′-end and 3′-end "wing" sequences of the probes and common adapters required for cluster generation on NGS platform (Illumina NextSeq550). Libraries were prepared and cleaned-up with HTG EdgeSeq AMPure cleanup of Illumina Sequencing Libraries. Following libraries preparation, their concentration has been evaluated through HTG EdgeSeq KAPA Library quantification, and each library has been normalized and pooled using HTG EdgeSeq RUO library calculator. Then, pooled libraries were denatured in 0.2 N NaOH and sequenced (final concentration 4 pM) onto Illumina NextSeq550 platform (High Output kit v2 cat. FC-404-2005). Data were returned from the sequencer as demultiplexed FASTQ files. Resulting reads were aligned referring to miRbase v20 using HTG Parser software.

QIAseq small RNA sequencing (untargeted-seq)

Total RNA extraction was performed from 200 μL of plasma through Serum/Plasma Norgen kit (cat. 55000, Thorold, ON L2V 4Y6, Canada). Small RNA-derived cDNA libraries were prepared using QiaSeq miRNA library kit (cat. 331505, Qiagen). QIASeq strategy assign Unique Molecular Index (bound to reverse transcription primers) during reverse transcription step to every mature miRNA molecule, to enable unbiased and accurate miRNome-wide quantification of mature miRNAs by NGS. Then, libraries quality control (QC) was performed quantifying their concentration through QUBIT 3.0 spectrofluorometer (Qubit dsDNA HS Assay Kit, cat. Q32854, Thermofisher Scientific) and assessing their quality using capillary electrophoresis in Bioanalyzer 2100 (Agilent High Sensitivity DNA kit cat. 5067–4626, Thermofisher Scientific). High quality of libraries was evaluated considering electropherograms showing a peak comprised between 175 and 185 bp. Following QC, all libraries were normalized until 2 nM and pooled, denatured in 0.2 N NaOH and further sequenced (final concentration 175 pM) on Illumina NovaSeq 6000 platform [NovaSeq 6000 SP Reagent Kit (100 cycles) cat. 20027464, NovaSeq XP 2-Lane Kit cat. 20021664, Illumina] using the XP protocol applying 75 × 1 single reads.

Data were returned from BaseSpace Sequence Hub as demultiplexed FASTQ files. Resulting raw reads were deduplicated by leveraging Unique Molecular Identifiers (UMIs) present in the library, then mapped to miRbase v21 and piRNABank using QIAseq miRNA Quantification V1 Legacy pipeline from QIAGEN GeneGlobe Data Analysis Center portal (https://geneglobe.qiagen.com/us/analyze). Briefly, resulting reads were mapped referring to miRbase v21 and piRNABank using QIAGEN Gene Globe data analysis center software, which identified a wide repertoire of small RNA species e.g., piRNA (PIWI interacting RNA), tRFs (tRNA fragments), rRNA (ribosomal RNA), miRNA (microRNA).

All these procedures (samples collection and time, RNA extraction, Small RNAs library preparation and sequencing) were also conducted on second cohort as already described for first cohort, with minor modifications. In details, Small RNAs libraries were barcoded with unique dual indexes (UDI) (cat.# 331615 and cat.# 331625).

Primary analysis of miRNAs expression

Reads assigned to miRNAs were standardized into Counts Per Million (CPM) and filtered through *edgeR* package of R (*BioConductor*), maintaining only those miRNAs expressed in at least 70% of individuals with at least 15 CPM for HTG EdgeSeq and 10 CPM for QIAseq. Following low counts filtering, Median of Ratios normalization was performed through *DESeq2*[103] package of R (*BioConductor*) and normalised counts were used for subsequent analyses. Consistently detected miRNAs in both sequencing platforms were selected, keeping only those having a positive Pearson correlation estimate between the two approaches ($R > 0$ and p value < 0.05).

Circulating miRNAs unsupervised hierarchical clustering analysis

Unsupervised hierarchical clustering was independently applied to both miRNAs expression datasets. The analyses were performed on log2 transformed data (after the addition of a pseudo-count), with *hclust* function of *stats* R package (complete-linkage agglomeration method and Pearson's distance as distance metric). In order to determine the optimal cutting threshold silhouette method was applied.

Then, both dendrograms were split in two branches (*cutree* function of *stats* package of R with $K = 2$) and named according to their size (cluster A the major and cluster B the minor). Only patients consistently identified as members of the same cluster in both expression datasets were kept.

The association among patients' clinical data and their branch membership was evaluated with logistic regression (glm function from *stats* package of R) and p-values ≤ 0.05 were considered significant. The direct association between individual miRNA

expression (log2-transformed, adding a pseudo-count) and each numerical clinical parameter was evaluated using Spearman's correlation test. The same analysis was performed on Small RNAs sequencing dataset, obtained from the second cohort, by considering the same miRNAs used for the analyses in the first cohort.

HLA genotyping

HLA typing was performed at v1 for $n = 107$ out of $n = 109$ T1DM individuals classified as cluster A or cluster B individuals through AXIOM Genotyping Array. HLA aplotype prevalence differences between cluster A and cluster B was assessed using chi-square test, considering significant p values <0.05.

PBMC (cryopreserved) multi-dimensional flow cytometry (Multi-FACS) immunomics

Immunomics profile of peripheral blood from $n = 67$ out of 115 individuals was investigated at baseline through Cytek Aurora flow cytofluorometer.

Samples were processed in five batches (between 12 and 16 samples per batch, consisting of a mixture of samples from each of the five INNODIA immune laboratories) together with two unrelated control samples in each batch using 0.54–2.8 × 10^6 PBMCs per sample. PBMCs were first stained using Live/dead blue for 15 min at room temperature, washed with FACS buffer (PBS with 0.2% BSA and 2mM EDTA), and incubated with Fc receptor blocker (TruStain FcX Fc; BioLegend) for 10 min at room temperature. Without wash, samples were stained in a 37°C waterbath for 15 min using mastermix 1 (containing antibodies against CXCR3. CD117, CD294/CRTH2, and CD161). Samples were further stained in waterbath for 15 min using mastermix 2 (containing antibodies against CXCR5, ICOS, CCR7, and CCR6), followed by 30 min at room temperature using mastermix 3. Finally, samples were washed using FACS buffer, then fixed and resuspended in PBS containing 1% paraformaldehyde (Alfa Aesar). Single color controls were made using PBMC for all colors except for CD294, CD117, CD161, and TCRgd where BD mouse or rat comp beads were used instead due to low cell expression. Single color controls were subjected to the same buffer and fixed as the multi-colour stained samples. SpectroFlo QC beads were run daily and single color controls were acquired in the reference library, which was subsequently used for live unmixing during sample acquisition on a Cytek Aurora cytometer. Flow data were analyzed using FlowJo software using the Boolean gating scheme shown in Figure S13.

In order two find differences in the proportion of immune cell populations between the two cluster of patients (cluster A, cluster B), a Beta regression analysis was performed, using the *betareg* function (version 3.1.4).[104] From the 109 patients assigned to the same cluster by the two different sequencing platforms, 67 were also present in the immunomics cohort. Among these overlapping patients, $n = 48$ belongs to cluster A and $n = 19$ to cluster B. As first step of the analysis, proportions of immune cell population were transformed according to Smithson & Verkuilen[105] to rescale the dependent variable in the interval (0.1), avoiding values of 0.1. Indeed, beta regression model cannot deal with values of 0 and 1. The Beta regression model was corrected for the time between blood draw and PBMCs isolation (same day vs. overnight). Before estimating the differential proportions of immune cell populations between the two clusters of patients, an influent points detection analysis was performed to remove outliers. For each beta regression model, Cook's distance of the samples was computed. Samples with Cook's distance 5-fold higher than the average Cook's distance of the samples in the model were marked as influent points and removed from the analysis. After this procedure, the final beta regression models were fitted and immune cell populations with an adjusted p-value (p adj) related to the cluster <0.1 were detected as significantly different in proportions between the two groups of patients. For representative purpose only, immune cell proportions were residualized for the time between blood draw and PBMCs isolation. For each immune cell population, a linear model was fit with the proportion as dependent variable and the time between blood draw and PBMCs isolation as independent variable. The residuals of the models represent the information on immune cell proportions that is not explained from the timing of PBMC's isolation.

MiRNAs differential expression analysis

Normalised reads of the two sequencing datasets were used to detect any differentially expressed miRNAs between cluster A and cluster B groups in both sequencing platforms, accounting for age and gender as covariates with *DESeq2* package of R (*BioConductor*) using Wald test and Benjamini-Hoechberg adjusted p-value (p_{adj}) < 0.01 was considered as significant.

Weighted miRNA correlation network analysis (WMCNA) for the identification of miRNA modules

Co-expression modules from the two different sequencing platforms were identified using the *WGCNA*[106] package. Normalized expression values from the 248 common correlated miRNAs were transformed in log2 scale for the analysis (after the addition of one pseudo-count). Similarities between nodes were computed using the biweight midcorrelation, setting the max p-outliers parameter at 0.1 and using a weighted signed network. The following step was the identification of the Beta parameter (for both platforms) to compute adjacencies between nodes, by applying the approximate scale-free topology criterion. This criterion assumes that few highly connected nodes (hubs) link the rest of the less connected nodes to the system. Given the power-law distribution of the connectivity (sum of the adjacencies of a node with all the other nodes of the system), the goodness of the scale free-topology assumption for Beta values was measured through the R^2 of the model regressing the log10 of probability of the connectivity and the log10 of the connectivity. High values of the R^2 of the model are related to a straight line fitting the model, suggesting the assumption of the scale free topology. Moreover, the slope of the model should be close to −1. Candidate values of Beta ranging from 5 to 25 were

manually inspected to choose the optimal ones for both platforms. Regarding the HTG-Seq platform, the scale free topology was met at a value of Beta equal to 24, which is the first value with an R^2 close to 0.8 and with a slope of −0.82. On the other hand, regarding the Untargeted platform, the first value of Beta with an R^2 of the model of at least 0.8 was 11. However, the slope of the regression model for this value was very far from −1 (−0.40), suggesting that the number of nodes with a high connectivity does not decay as expected. Moreover, two values of Beta were very different for the two sequencing strategies (11 and 24) resulting in two networks with very different architectures. For this reason, the optimal value of Beta in the untargeted platform was determined as the first value of Beta with at least an R^2 of the model of 0.8 and a regression slope of at least −0.7. The first value which satisfies these two criteria for the untargeted platform was Beta equal to 20. Once estimated the Beta parameters, similarities matrices were converted into adjacencies matrices by elevating similarities at the corresponding Beta value. The next step was the identification of the Topological Overlap Matrix (TOM) for both sequencing platforms from the adjacency matrices. The topological overlap of two nodes is a measure of similarity which defines how well the two nodes are interconnected. At this point, the information from the two sequencing platforms, managed in a separate way during the previous phases of network construction, was merged in a consensus TOM. For the estimation of the consensus TOM (TOMcons), the two TOMs were first scaled at the 95th quantile. The scaling step is crucial, because the consensus TOM was estimated as the minimum-wise component between the two TOMs. The consensus TOM, which contains the minimum-wise information about nodes interconnectivity from both platforms, was then transformed into a dissimilarity matrix (1-TOMcons). The *hclust* algorithm (hierarchical clustering), with average agglomeration method, was used to detect the modules, using the dissimilarity TOM as distance matrix. Minimum module size equal to 3 and deep split equal to 2 were set as parameters for the deepsplit function to cut the dendrogram for modules identification. At this point, each miRNA was assigned to a module, with the gray one representing un-assigned miRNAs. The last step was the merging of very similar modules, to limit the redundancy in the information held. Modules EigenMiRNAs (MEs) were computed as the first principal component of the miRNAs expression values (normalized and log2 scaled) present in the module for both sequencing platforms. Similarity between MEs was computed with Pearson's correlation coefficient, and then the dissimilarity was estimated and used for the hierarchical clustering of MEs. A cut height of 0.1 was used to merge closely related module. The final output of the *WGCNA* algorithm was a list of labels which identifies each miRNA as belonging to a specific module (after merging of closely related ones). Each module was summarized by an ME, computed as previously stated.

Once defined the modules through WMCNA, the aim of the analysis was the identification of a subset of the most representative miRNAs ($n = 5$) for each module. The centrality (hubness) of a miRNA within its module was defined using the intramodular connectivity as metric. The intramodular connectivity is the sum of the adjacencies of a miRNA with all the other miRNAs present in the module. The higher the intramodular connectivity, the higher the centrality of the node in the module. However, given the different values of Beta for the two sequencing platforms in the module's estimation step (24 HTG-Seq, 20 Untargeted), the intramodular connectivity must be normalized in order to be comparable between them. Thus, for each module and sequencing platform, the intramodular connectivity of the node was divided by the maximum value of its own module. Then, the normalized intramodular connectivity of the nodes among sequencing platforms were summed up, and miRNAs were ranked based on this value. The 5 miRNAs with the highest values were defined as the most representatives for the module. Modules eigenMiRNAs of these subsets of most representative miRNAs were then computed for both sequencing platforms and correlated with clinical parameters using Spearman's Rho correlation. Correlations with a $p_{adj} \leq 0.1$ were considered as significant.

Droplet digital PCR (ddPCR)

Validation of selected miR-409-3p, miR-127-3p and miR-382-5p, identified through both differential expression analysis and WMCNA, was performed through Custom Taqman reverse transcription and subsequent droplet digital PCR (ddPCR) detection.

In details, their expression was analyzed in all plasma samples of first and second T1DM cohort using TaqMan miRNA assay primers (Life technologies, CA, USA) through a standardised protocol. RNA (the same used for small RNA sequencing) was reverse transcribed employing Custom RT primers pool and preamplified using Custom Preamp primers pool. Briefly, 5 μL each RT or TM primer was diluted in a total volume of 500 μL Tris-EDTA 1X and used for RT or preamplification reaction. Then, 3 μL of RNA were added to 6 μL of custom primers pool, 0.30 μL 100 mM dNTPs, 3 μL of 50 U/μL Multiscribe RT, 1.50 μL 10× RT Buffer, 0.19 μL 20 U/μL RNase Inhibitor and 1.01 μL H_2O. The reaction product was incubated at 16°C for 30 min, 42°C for 30 min and then at 85°C for 5 min. Afterward, the synthesised cDNA was preamplified using Custom Preamp primer pool; the reaction included: 2.5 μL of cDNA from each sample, 12.5 μL 2× TaqMan Preamp Master Mix, 3.75 μL 10× Custom Preamp primers and 6.75 μL H_2O. The reaction was incubated at 95°C for 10 min, at 55°C for 2 min and at 72°C for 2 min, then for 12 cycles at 95°C for 15 s and 60°C for 4 min and, finally, at 99°C for 10 min. Then, droplet digital PCR was performed on a BioRad QX200 system using a Probes assay (BioRad, Mississauga, ON, Canada). Each PCR reaction contained 11 μL of QX200 super mix, 1.1 μL of each 20X TaqMan assay, 5.9 μL of H_2O and 4 μL of template cDNA in a final volume of 22 μL. The PCR reactions were mixed, centrifuged briefly and 20 μL transferred into the sample well of a DG8 cartridge. After adding 70 μL of QX200 droplet generation oil into the oil wells, the cartridge was covered using a DG8 gasket, and droplets were generated using the QX200 droplet generator. Droplets were carefully transferred into PCR plates using a multi-channel pipette and the plate sealed using PCR plate heat seal foil and the PX1 PCR plate sealer. PCR was performed in a SimpliAmp thermal cycler (Life technologies, CA, USA). The PCR protocol was 95°C for 10 min; 40 cycles of: 95°C for 30 s, optimal annealing temperature (56°C for miR-409-3p and miR-382-5p; 54°C for miR-127-3p; 98°C for 10 min; 4°C for 30 min. PCR plates were transferred into a QX200 droplet reader to count positive and negative droplets. Thresholds

to separate positive from negative droplets were set manually for each miRNA using the histogram function and reads analyzed using QuantaSoft Analysis Pro software (Version 1.2, BioRad, Mississauga, ON, Canada).

QUANTIFICATION AND STATISTICAL ANALYSIS

Sample size for circulating miRNAs analysis were determined according to our experience from previous works. Mann Whitney U test was performed between two groups when the variables did not follow a Gaussian distribution. Coefficient of Variation of miRNAs expression in Targeted- and Untargeted-seq was calculated on read counts using GraphPad Prism 10.0.

Differences in clinical parameters between individuals belonging to cluster A and cluster B were determined with the univariate logistic regression using the glm function from the *stats* package in R software. Data were modeled using the clusters as dependent variable and the clinical parameter as independent variable. Multiple testing correction of *p*-values (p_{adj}) associated to clinical parameters was performed independently for each visit using the Benjamini-Hochberg (BH) method. Clinical parameters with a $p_{adj} \leq 0.1$ were considered as significantly different between the clusters. Correlation analyses were performed using Spearman Rho Test or Pearson's R test. *p*-values from correlation analyses were corrected for multiple testing using the Benjamini-Hochberg (BH) method independently for each visit.

Collinearities of miR-409-3p, miR-127-3p, and miR-382-5p (ddPCR) were assessed using a logistic regression model with the cluster as the dependent variable. The collinearities of the three miRNAs in predicting the cluster were evaluated with the variance inflation factor (VIF). VIF values >5 were determined to indicate high collinearity.

The specificity of the three miRNAs previously mentioned was assessed using IsomiRdb datasets. These datasets were downloaded and re-analyzed to validate miRNAs expression in 99 cell types. Sample metadata and miRNAs expression (Reads Per Million) files were retrieved from IsomiRdb[59] and re-analysed using a multi-step filtering approach.

Firstly, metadata information was used to select sequencing data derived from healthy donors and primary cell studies. Initially, only sources derived from the Sequence Read Archive (SRA) [105] in which cell origin is present were retained. From the remaining dataset, external IDs were used to collect study titles and attributes from NCBI (https://www.ncbi.nlm.nih.gov/). Additional filters were applied to these attributes to produce the final metadata containing 396 small RNA sequencing runs derived from 99 cell types. After these filtering steps, the mean expression of each miRNA among cell types was estimated, and Z-scores were calculated.

Principal Component Analysis (PCA) was performed using the prcomp function from the stats package in R. Normalized expression data were scaled either by log2 or Z-scores transformation for the analyses. Statistical analyses were performed using R project (version 4.2.2) or GraphPad Prism 10.0.