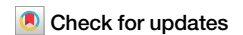


<https://doi.org/10.1038/s43856-025-00922-7>

A multi-omics integration approach relying on circulating factors does not discern subtypes of childhood type 1 diabetes



Valentina Codazzi¹, Nicola Baldoni¹, Giulia M. Scotti¹, Anna Giovenzana¹, Andrea Rigamonti¹, Giulio Frontino¹, Eugenia Bezecchi¹, Camillo Bechi Genzano¹, Alessandra Mandelli¹, Debora Carnovale¹, Ilaria Marzinotto¹, Vito Lampasona¹, Paolo Fiorina², Andrea Giustina^{1,3}, Lorenzo Piemonti^{1,3}, Manuela Battaglia⁴, Marco J. Morelli¹, Riccardo Bonfanti^{1,3} & Alessandra Petrelli^{1,5} ✉

Abstract

Background Type 1 Diabetes (T1D) exhibits considerable heterogeneity, impacting prediction, prevention, diagnosis, and treatment. Precision medicine aims to tailor treatments using ‘endotypes’—subtypes of disease with distinct pathophysiological mechanisms. However, proposed endotypes often lack mechanistic associations with clinical outcomes for accurately identifying T1D cases.

Methods This study introduces an approach leveraging the multi-omics factor analysis (MOFA) strategy, a computational method for unsupervised integration analysis, to explore endotypes. Analyzing data from 146 new-onset children with T1D (54 females, 92 males; age range 5–18 years), including circulating immunome, transcriptome, and serum metabolic hormones, we identify 12 factors explaining variability across the three data sets.

Results Here we find no associations, either direct or through clustering, between these 12 factors and clinical parameters, genetic predisposition, or disease outcome. These results suggest that a combination of clinical phenotypes might be responsible for the differences across T1D cases.

Conclusions These findings challenge the assumption that T1D heterogeneity reflects diverse developmental mechanisms. These results add to the ongoing debate on endotypes and carry important implications for clinical trial design—particularly in how treatments are evaluated for their effectiveness across broad and diverse patient populations.

Plain Language Summary

This study seeks to understand why type 1 diabetes (T1D) affects individuals differently and whether specific subtypes of the disease, known as “endotypes,” can explain this variability to inform personalized treatment strategies. We analyze data from 146 children newly diagnosed with T1D, examining immune activity, gene expression, and hormone levels. We identify 12 factors that account for some variability among patients but find no clear associations between these factors and disease outcomes or genetic risk. These findings suggest that T1D is shaped by a complex interplay of factors rather than by distinct underlying mechanisms, highlighting the need for a more refined understanding of disease heterogeneity to guide effective treatments and the design of future clinical trials.

Type 1 Diabetes (T1D) displays considerable diversity in epidemiology, etiopathogenesis, clinical course, and responses to intervention, posing challenges in prediction, prevention, diagnosis, and treatment. Precision medicine aims to deliver personalized treatments in the context of this heterogeneity, aligning with the concept of ‘endotypes’—distinct disease subtypes characterized by unique aetiopathogenesis and specific targetable interventions. The notion of endotypes, originally adopted from the asthma

field¹, was introduced by the T1D scientific community as essential for designing tailored interventions for a disease marked by considerable heterogeneity².

Major studies on large pediatric T1D cohorts utilized unsupervised clustering of clinical data, considering autoantibody status, age at onset, glucose control, and body mass index (BMI), for the identification of endotypes. While these studies revealed subtypes associated with residual

¹IRCCS San Raffaele Scientific Institute, Milan, Italy. ²University of Milan, Milan, Italy. ³Vita-Salute San Raffaele University, Milan, Italy. ⁴Katholieke Universiteit, University Hospital Gasthuisberg, Leuven, Belgium. ⁵Present address: Department of Clinical Sciences and Community Health, University of Milan and Pio Albergo Trivulzio, Milan, Italy. ✉e-mail: alessandra.petrelli@unimi.it

beta-cell function and disease progression trajectories^{3–5}, none have effectively pinpointed molecular associations with clinical traits. Recently proposed endotypes, T1DE1 and T1DE2, can be distinguished based on genetics, age at onset, type of first autoantibody, degree of immune pancreatic infiltration, and responsiveness to immunomodulators⁶. However, their conceptualization primarily relies on associations reported by various investigators linking age to specific genetics or systemic and pancreatic traits. Furthermore, this description is not exhaustive as forms of T1D with distinct pathobiological findings have been described⁷. Consequently, a comprehensive description confirming the existence of endotypes and their associations with clinical outcomes or response to targeted treatments is currently lacking.

In line with the hypothesis that endotypes inherently involve specific molecular mechanisms, our methodology employs a ‘bottom-up’ approach, starting from the integration of multi-omics data and culminating in the association with clinical phenotypes, which represent clusters of visible traits characterizing disease expression. Here, we initially apply a computational method for unsupervised integration analysis of circulating immunome, transcriptome, and serum metabolic hormone data, enabling the identification of hidden factors. These factors are defined as linear combinations of the original features from all integrated layers, aimed at disentangling shared heterogeneity. Subsequently, we evaluate associations with observable traits, including clinical parameters, genetic predisposition, and disease outcome. The findings of this study demonstrate a uniform distribution of factor scores across clinical traits, with no clear segregation of patients based on their clinical phenotypes. This suggests that the observed differences among patients are more likely attributable to a complex interplay of clinical characteristics rather than distinct pathogenic mechanisms.

Methods

Subjects and data collection

A cohort of 194 pediatric new-onset type 1 diabetes patients, hospitalized in the Pediatric Department of the IRCCS Ospedale San Raffaele (OSR), Milan (Italy), were enrolled from October 2018 to October 2022. This study includes only pediatric patients, given the significant clinical and biological differences between childhood- and adult-onset T1D⁶. By focusing on the pediatric cohort, we aimed to investigate whether underlying mechanistic drivers of disease heterogeneity could be still identified. The study was approved by the OSR Ethics Committee (protocol #TIGET004-DRI003) and written informed consent was obtained from the parents of all minor participants. Blood samples were prospectively collected between 5 and 10 days after diagnosis. Inclusion criteria for the study comprised: (1) age 5–18 years, (2) recent diagnosis of T1D, (3) the presence of at least 1 islet autoantibody. Exclusion criteria were: (1) treatment with drugs including antibiotics, steroids, non-steroidal anti-inflammatory drugs and immunosuppressive agents, (2) sign of infection (fever, cough, rhinitis) over the last two weeks prior the blood withdrawal, (3) monogenic diabetes. After applying these criteria, 174 patients were selected. Based on the availability of at least one layer of information among immunome, transcriptome and metabolic hormones data, a final cohort of $n = 146$ patients, comprising $n = 54$ female and $n = 92$ male, was included in the study. Subjects were tested for glycosylated hemoglobin (HbA1c) and islet autoantibodies to glutamic acid decarboxylase (GADA), insulin (microinsulin antibody assay, mIAA), insulinoma-associated antigen 2 (IA-2A) and zinc transporter 8 (ZnT8A) at the central laboratory of OSR. Body mass index percentiles (BMIp) were calculated referring to sex- and age-specific charts of the Italian population. Partial clinical remission was assessed as the insulin dose adjusted HbA1c (IDAA1c) ≤ 9 , where IDAA1c was calculated at a median time point of one year after the diagnosis as $\text{HbA1c (\%)} + 4 \times \text{insulin dose (U/kg per 24 h)}$ ^{8–11}.

T1D genetic risk score 2 (T1D-GRS-2)

DNA extraction from whole blood was conducted on a subset of 90 subjects using the Maxwell® RSC Blood DNA Kit and the Maxwell® RSC Instrument

from Promega. For each participant, the Type 1 Diabetes Genetic Risk Score 2 (T1D GRS-2) was calculated. This score accounts for the contribution of 67 HLA and non-HLA single nucleotide polymorphisms (SNPs) and the interactions between 18 HLA DR-DQ haplotype combinations, known to be associated with an increased risk of T1D¹². The T1D GRS-2 was obtained by weighting for the corresponding beta value the sum of each SNP allele frequency and the presence of specific HLA interactions. This method provides an improved assessment of genetic predisposition to T1D compared with the previously proposed GRS-1¹³.

Immunome

The circulating immunome was assessed on 136 subjects. Whole blood was collected in a Vacuette® blood collection tube with ACD-B anticoagulant solution (BD Biosciences). After the lysis of the red blood cells, the sample was washed and stained with specific markers. For the T regulatory cells, additional fixation and permeabilization steps were required before intracellular staining with the marker forkhead box P3 (FoxP3). Flow cytometry analysis was performed based on five panels of antibodies that comprised: T cells (CD45, CD3, CD4, CD8, CD45RA, CCR7, PD1, CD57), T&NK cells (CD45, CD3, CD4, CD8, CD27, CD28, CD56, CD69), B cells (CD45, CD19, CD27, IgD, IgM, CD24, CD38, CD21), Tregs (CD45, CD3, CD4, CD25, CD127, CD45RA, FoxP3) and DCs/monocytes (CD45, HLADR, LIN, CD11c, CD123, CD1c, CD16, CD14) (as described in ref. 14). Antibodies used for flow cytometry analysis are described in Supplementary Table 1 while gating strategy is reported in Supplementary Fig. 1. Samples were acquired using a FACSCanto-II flow cytometer (BD Bioscience) and FACSDiva v.8 software (BD Pharmingen). To assess instrument performance, CS&T Beads (BD Biosciences) and SPHERO Rainbow Beads (Spherotech Inc., Lake Forest, IL) were used for daily verification. Data analysis was assessed with unsupervised clustering using the FlowSOM algorithm¹⁵ included in the CyTOF/CATALYST pipeline (version 1.14.1), which allowed the identification of 46 distinct immune cell subsets. The list of immune cell subsets and their frequencies is provided in the Supplementary Table 2.

Transcriptome

Bulk RNA sequencing analysis was performed on the whole blood of 103 patients. RNA was extracted from whole blood collected in in PAXgene® Blood RNA Tube (BD Biosciences) using the Maxwell® 16 LEV simplyRNA Blood Kit (Promega) and the AS2000 Maxwell® 16 Instrument (Promega). Prior to library preparation, the quality of RNA samples was assessed using the TapeStation 4100 (Agilent) to determine the RNA Integrity Number (RIN) for each sample. NGS libraries were then constructed utilizing the TruSeq Stranded mRNA (Illumina) protocol, which permits library preparation starting from 50 to 100 ng of total RNA. These libraries were subsequently barcoded, pooled, and sequenced on an Illumina NovaSeq 6000 sequencing system. Each RNA library yielded approximately 30 million reads, configured in single-read mode with a length of 100 nucleotides. Raw data underwent demultiplexing using the latest version of Bcl2Fastq (Illumina). The raw reads produced from sequencing were trimmed using Trimmomatic (version 0.39) to remove adapters and to exclude low-quality reads from the analysis. The remaining reads were then aligned to the human genome GRCh38, annotated according to Gencode basic annotations version 43, using the STAR aligner (version 2.5.3a) and assigned to the corresponding genes using featureCounts (version 1.6.4). Expressed genes were defined as those genes showing a Counts per Million (CPM) value greater than 1 in at least 16 samples. The top 5000 most variable genes were selected for the subsequent analysis.

Metabolic hormones (Meth)

The levels of 9 metabolic hormones in the serum were tested on 120 patients. The serum concentration of proinsulin (PI) was determined by enzyme-linked immunosorbent assay (ELISA) (Human Total Proinsulin Elisa kit, EMD Millipore, Merck) and BioTek Epoch Microplate Spectrophotometer.

Gastric inhibitory polypeptide (GIP), leptin and pancreatic polypeptide (PP) were measured in serum samples using the Human Metabolic Hormone Magnetic Bead Panel (EMD Millipore, Merck). Adiponectin, adipsin, resistin, lipocalin and plasminogen activator inhibitor-1 (PAI-1) were measured in serum with the Human Adipokine Magnetic Bead Panel (Millipore, Merck). Data acquisition was performed using the Bio-plex MAGPIX Multiplex Reader (BIO-RAD) and the Bio-Plex Manager MP software (BIO-RAD).

Data integration and multi-omics factor analysis

The metabolic hormones, along with the immunome and the 5000 most variable genes derived from RNA sequencing, were included in the unsupervised Multi-Omics Factor Analysis (MOFA)¹⁶. We conducted MOFA to integrate and analyze multiple omics data, aiming to understand underlying sources of variability in our data. The MOFA2 R library, version MOFA2_1.8.0, was applied for this purpose. Initially, the data were normalized and organized into a MOFA object. Subsequently, different MOFA models were trained, using the `run_mofa` function, to identify latent factors in the data, testing the impact of different numbers of factors. At the end, the model with 12 latent factor was selected. We evaluated the variance explained by each factor for each omics modality using the `plot_variance_explained` function, thus highlighting the contribution of each layer to the definition of the final model. Scatterplots on latent factor scores were then generated to visualize the distribution of samples in the MOFA factor space, coloring them based on clinical covariates such as age, sex, family history of T1D, BMIp, presence of diabetic DKA, C-peptide levels, PI/C-peptide ratio, neutrophil count, GRS-2 and IDAA1c. Finally, we identified the features that mostly contributed to the definition of the MOFA factors using the `plot_top_weights` function, applied to all the different integrated assays. The top 20 genes with the highest weights for each factor were reported, together with the top 10 features for the immunological layer, while all the 9 measured hormones were considered for the metabolic dataset. A heatmap was constructed using the complete MOFA scores matrix, representing the multidimensional integrated data across patient samples. Annotations were incorporated to delineate various clinical variables pertinent to each patient. This visualization technique enabled the elucidation of patterns in MOFA scores across patient samples, facilitating the identification of relationships between molecular profiles and clinical characteristics within the dataset.

Principal Component Analysis (PCA)

Single omics assays were initially analyzed individually, by performing a Principal Component Analysis (PCA) on each of them. For transcriptomics data, the top 5000 most variable genes were considered, while the entire molecular profile was included for the immunological and the metabolic datasets. To adjust for batch effects, the `removeBatchEffect` function from the `limma` R package was preliminary employed. This preprocessing step was applied to both transcriptomics and immunological data. PCA was performed using the `prcomp` R package, on preliminary centered and scaled data. The resulting principal components (PCs) were examined to determine the proportion of variance explained by each component. Score plots were generated to visualize the distribution of patients in the PC space. Separate plots were created for each combination of principal components (PC1 vs. PC2 and PC3 vs. PC4) to assess potential clustering patterns based on different phenotypic variables such as age, sex, family history of T1D, BMIp, presence of diabetic DKA, C-peptide levels, PI/C-peptide ratio, neutrophil count, GRS-2 and IDAA1c.

Uniform Manifold Approximation and Projection

The UMAP dimensional reduction algorithm was applied to the first 30 principal components of the PCAs of the IMM and RNA layers. The original features of MetH, which initially comprises only 9 features, were scaled and centered before applying the UMAP dimensional reduction algorithm. All the considered clinical features were then projected onto the bi-dimensional UMAP score plots obtained. UMAP R package was employed.

Statistics and Reproducibility

In our study, the primary objective was the unsupervised integration of multi-omics data using the MOFA (Multi-Omics Factor Analysis) model. It is important to emphasize that MOFA identifies latent factors that capture shared sources of variation across multiple biological datasets. Since the focus of our analysis was to uncover underlying data structures—rather than perform inferential hypothesis testing—we did not apply classical statistical tests. Instead, the statistical insights presented are derived directly from the model-inferred parameters automatically estimated by MOFA. To ensure full reproducibility, all code and parameter settings used for MOFA model optimization are publicly available through the official MOFA repository. The entire analytical workflow strictly adheres to the standard MOFA tutorial guidelines. Moreover, all datasets, sample sizes, and figure outputs are fully documented and consistent with the descriptions provided in the manuscript, drawing exclusively from publicly accessible data sources.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Results

Patient Cohort and Data Collection

Blood samples from pediatric new-onset T1D patients aged 5–18, admitted to IRCCS Ospedale San Raffaele, Milan (Italy) were prospectively collected from October 2018 to October 2022. After applying selection criteria (see methods section), a cohort of 146 patients was included in the study. Table 1 presents characteristics of our patient cohort, while the age distribution per year is shown in Supplementary Fig. 2.

Table 1 | Characteristics of the cohort of pediatric new-onset T1D patients

Age, median (IQR)	11.46 (9.38–13.94)	<i>n</i> = 146
Sex, F (%)	54 (36.99)	<i>n</i> = 146
Presence of T1D relatives, <i>n</i> (%)	32 (22.54)	<i>n</i> = 142
GRS-2, median (IQR)	14.42 (13.16–16.07)	<i>n</i> = 90
BMIp, median (IQR)	36.00 (6.78–73.60)	<i>n</i> = 144
Presence of DKA, <i>n</i> (%)	45 (30.82)	<i>n</i> = 146
HbA1c (mmol/mol), median (IQR)	105.00 (88.75–119.00)	<i>n</i> = 146
C-peptide (ng/ml), median (IQR)	0.45 (0.28–0.72)	<i>n</i> = 146
Anti-tTG, <i>n</i> (%)	20 (13.70)	<i>n</i> = 146
Anti-TPO, <i>n</i> (%)	19 (13.01)	<i>n</i> = 146
Neutrophil count (10 ⁹ /L), median (IQR)	2.30 (1.80–2.70)	<i>n</i> = 138
Number of AAb:		<i>n</i> = 145
1, <i>n</i> (%)	21 (14.48)	
2, <i>n</i> (%)	47 (32.41)	
3, <i>n</i> (%)	54 (37.24)	
4, <i>n</i> (%)	23 (15.86)	
Type of AAb:		<i>n</i> = 145
GADA, <i>n</i> (%)	125 (86.21)	
IA-2A, <i>n</i> (%)	99 (68.28)	
IAA, <i>n</i> (%)	38 (26.21)	
ZnT8A, <i>n</i> (%)	107 (73.79)	
IDAA1c ≤ 9, <i>n</i> (%)	41 (35.65)	<i>n</i> = 115

GRS-2 T1D-Genetic Risk Score-2, BMIp percentile of body mass index, DKA diabetic ketoacidosis, AAb autoantibodies, anti-TPO anti-thyroid peroxidase antibody, anti-tTG IgA anti-tissue transglutaminase antibodies, IDAA1c insulin- dose adjusted A1c, IQR interquartile range.

Multi-Omics Integration and Factor Analysis

RNA sequencing analysis was conducted on the whole blood of 103 patients, capturing the entire transcriptome. Subsequently, the top 5000 most variable genes were extracted and included in the multi-omics integration (RNA, 5000 genes). The circulating immunome (IMM, 46 immune cell subsets listed in Supplementary Table 2) was assessed using unsupervised flow cytometry analysis across 136 subjects, while metabolic hormones (MetH, 9 metabolites) were measured in serum samples from 120 individuals using a multiplex assay (Fig. 1). The three molecular layers were integrated using unsupervised multi-omics factor analysis (MOFA)¹⁶. This approach aims to unravel heterogeneity by relying on the identification of a set of hidden factors generated through the linear combination of original features derived from the integration of all layers of data. Moreover, MOFA is able to handle missing data in a probabilistic manner rather than using imputation¹⁷. Factors reveal the principal sources of variability across the three data sets. Subsequently, the association between the inferred factors and phenotypes was inspected, encompassing age, sex, family history of T1D, percentile of BMI (BMIp), presence of diabetic ketoacidosis (DKA), C-peptide levels, proinsulin/C-peptide ratio (PI/C-pep) and neutrophil count. Genetic predisposition, represented by the genetic risk score 2 (GRS-2)¹² and disease outcome, measured by the 12-month follow-up lower insulin- dose adjusted A1c (IDAA1c)⁸, were also considered. This downstream analysis was expected to unveil T1D endotypes as depicted in Fig. 1.

Evaluation of MOFA Models

Different MOFA models were evaluated, each characterized by a distinct number of factors. All the assessed MOFA models yielded nearly identical results, confirming the consistency of the obtained factorization. Scatter-plots displaying the distribution of the scores obtained for all the tested MOFA models are provided in Supplementary Fig. 3. The final model was selected based on the criterion that it had a minimum variance explained by any of the three layers greater than 10%. This model, meeting the specified criteria, consisted of 12 factors, with RNA accounting for the majority of variance (50.3%), IMM for 21.3%, and MetH for 12.2% (Fig. 2a). The total variance is distributed in each of the layers within single factors (Fig. 2b). The weights of the top RNA genes (Supplementary Fig. 4), immune cell subsets (Supplementary Fig. 5), and metabolic hormones (Supplementary Fig. 6) in the definition of each factor are shown.

Absence of Age-Related and Phenotypic Clustering

No clustering related to the age of the patients was observed for any of the 12 factors (Fig. 3), indicating homogeneity in the distribution of factor scores across age groups—a trait known to be critically associated with disease severity and outcome^{12–14}. The same pattern occurred for all the other variables tested (Supplementary Fig. 7), including the BMIp known to be associated with specific genetic, clinical and metabolic traits^{18,19} as well as an increased risk of developing symptomatic T1D²⁰. Additionally, no patient

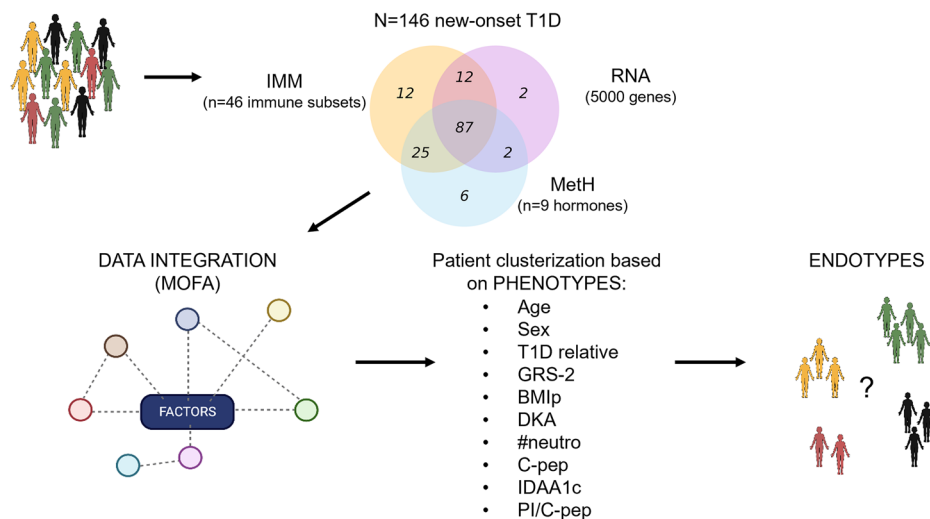
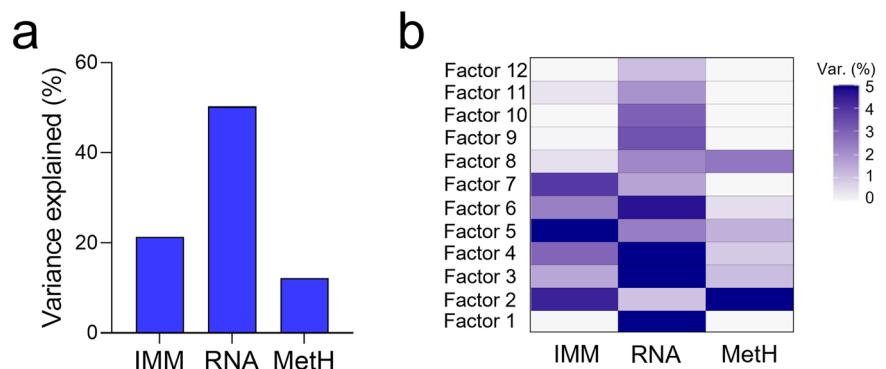


Fig. 1 | Integration of three molecular layers derived from the circulation of pediatric T1D patients and their association with clinical phenotypes. A cohort of new-onset pediatric T1D patients was prospectively enrolled at our institute, resulting in 146 individuals meeting the selection criteria. Whole transcriptome sequencing (RNA) was performed on whole blood ($n = 103$), the circulating immunome (IMM) was assessed using unsupervised flow-cytometry analysis

($n = 136$), and metabolic hormones (MetH, i.e., adiponectin, resistin, adipsin, leptin, GIP, PP, proinsulin, PAI-1, lipocalin) were measured in the serum by multiplex assay ($n = 120$). Data integration was conducted using the multi-omics factor analysis (MOFA). Factors were obtained to explain the variance of the dataset. Attempts were made to unravel T1D endotypes by performing patient clusterization based on phenotypes (clinical data, genetic predisposition, and disease outcome) and factors.

Fig. 2 | 12-factor model characteristics. **a** The percentage of variance explained by immunome (IMM), transcriptome (RNA), and metabolic hormones (MetH) using the 12-factor model. **b** In the 12-factor model, each factor explains a certain percentage of the variance across the different data modalities. This variance decomposition plot summarizes the sources of variation within the data, providing an overview of the contribution of each factor to the overall variability.



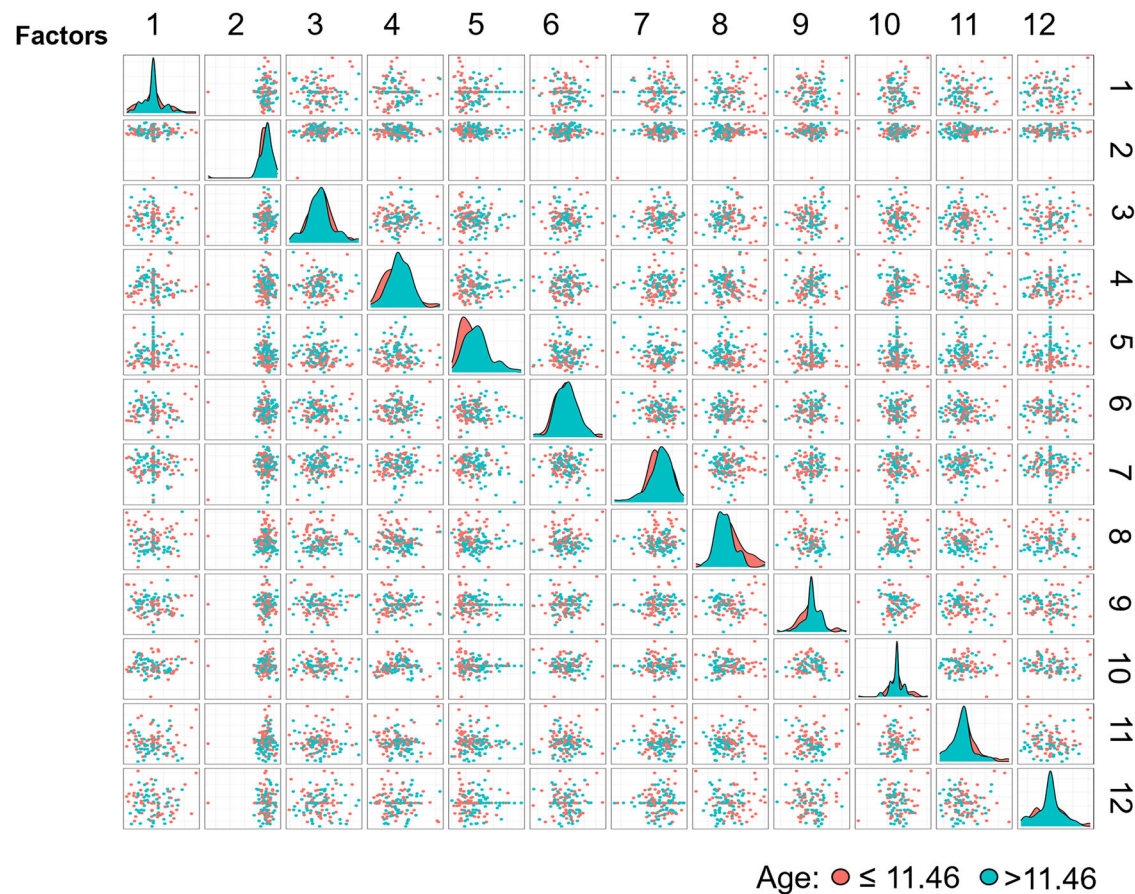


Fig. 3 | MOFA scoreplots of patient distribution based on median age groups. The scoreplots illustrate the distribution of patients in relation to the principal factors identified by multi-omics factor analysis (MOFA), with patients color-coded according to their median age groups. Each point represents an individual patient.

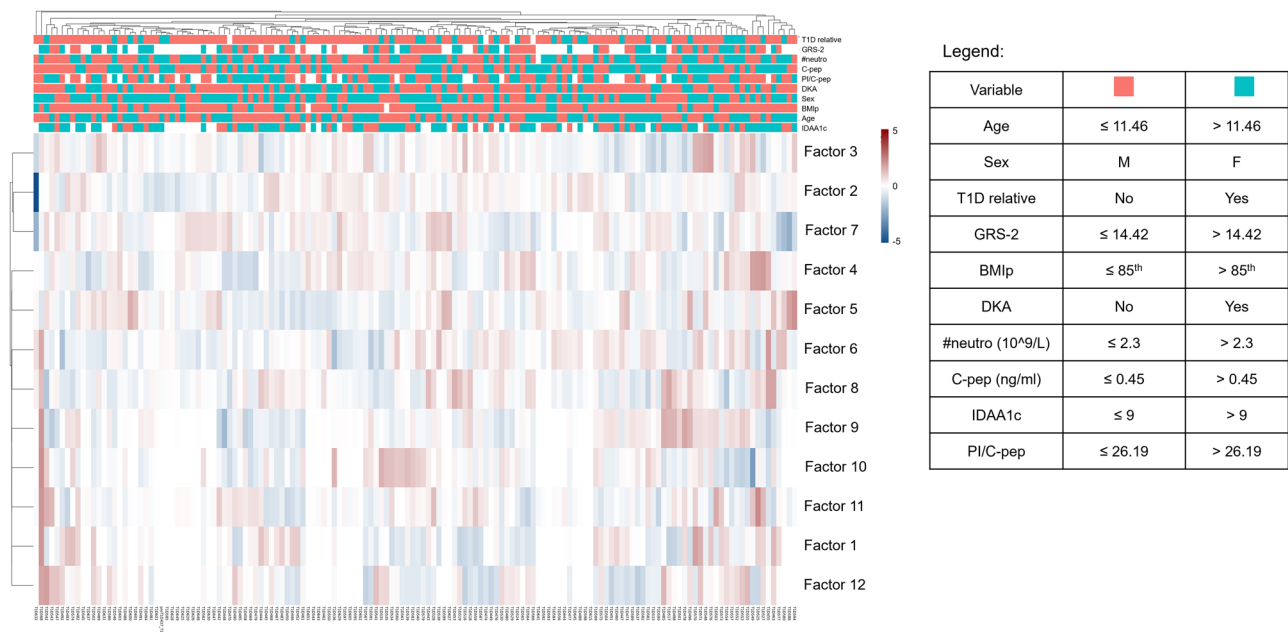


Fig. 4 | Clusterization of patients based on factors and phenotypes. Phenotypes include age, sex, family history of T1D (T1D relative), genetic predisposition (GRS-2), percentile of BMI (BMIp), presence of diabetic ketoacidosis (DKA), neutrophil count, C-peptide levels, disease outcome measured by the 12-month follow-up lower insulin- dose adjusted A1c (IDAA1c) and proinsulin/C-peptide ratio (PI/C-pep). White squares indicate that the phenotypic trait is not available for that patient. The legend shows cut-off values used to discriminate patient groups. Each variable is represented with a binary color scale: red tones indicate values equal to or lower than the reference threshold, while green tones highlight those higher.

clustering based on phenotypes was identified, as clinical parameters, disease outcome, and genetic predisposition were unable to stratify patients according to factors (Fig. 4). This finding suggests that the molecular variances observed within our patient population could be attributed to a combination of clinical traits.

Principal Component and UMAP Analyses

In line with this evidence, a principal component analysis (PCA) depicting patient distribution across individual layers of information—RNA (Supplementary Figs. 8 and 9), IMM (Supplementary Figs. 10 and 11), and MetH (Supplementary Figs. 12 and 13)—based on clinical phenotypes revealed no segregation once again. Similarly, the Uniform Manifold Approximation and Projection (UMAP) for each layer—RNA (Supplementary Fig. 14), IMM (Supplementary Fig. 15), and MetH (Supplementary Fig. 16)—did not reveal clustering of patients based on clinical variables.

Discussion

Unlike conventional ‘top-down’ approaches that primarily seek endotypes based on clinical phenotypes, our methodology introduces a different paradigm by starting from the biological foundations of T1D. This strategy prioritizes the integration of mechanistic insights at the molecular level before associating them with clinical phenotypes, offering a unique perspective to investigate potential mechanisms driving T1D. Although in a different context the MOFA-based approach successfully identified patterns associated with a certain disease state²¹, it did not result in the identification of endotypes in pediatric new-onset T1D. Several factors may have contributed to this outcome: the modest sample size may have limited the detection of subtle variations, the focus on circulating factors rather than pancreatic traits might have overlooked crucial indicators, and challenges in sample collection resulted in the absence of children under five and a lower prevalence of children aged 5–6 years, potentially skewing the analysis. Therefore, we cannot exclude the possibility that this cohort may not adequately represent the broader T1D population, as unintentional selection bias may not have led to the sufficient inclusion of subjects with diverse endotypes. Future research should aim to address these limitations by incorporating larger cohorts, including younger children as well as adult individuals, and conducting direct assessments of pancreatic tissue. These efforts are crucial for achieving a more comprehensive understanding of the underlying mechanisms driving T1D.

Conclusion

This study, encompassing 146 pediatric new-onset T1D patients, integrated transcriptomic, immunomic, and metabolic hormone profiles to investigate molecular endotypes. Using MOFA, we identified 12 factors representing primary variability across data layers. However, these factors lacked associations with clinical phenotypes, disease outcomes, or genetic predisposition. The absence of patient clustering based on clinical traits highlights a remarkable homogeneity in molecular profiles, suggesting that T1D clinical heterogeneity may arise from non-specific variability rather than fundamentally different mechanisms. While this report does not rule out the existence of distinct endotypes at the level of disease initiation, it underscores a convergence in immune profiles following the initial triggering event. These findings substantially contribute to the ongoing discussion regarding the existence of T1D endotypes, highlighting the disease’s complexity and implying that molecular differences may not align with traditional clinical phenotypes.

Data availability

The research data associated with this paper, including gene expression data, frequencies of circulating immune cell subsets, levels of serum metabolic hormones, and MOFA scores, from which Figs. 1–4 are derived, are accessible at DOI:10.6084/m9.figshare.25623762²² without any access restriction. RNA-sequencing FastQ data have been deposited in the Gene Expression Omnibus (GEO) database and are publicly accessible under the

accession code GSE287275. The source data underlying the graphs and charts presented in Figs. 1–4 and in the Supplemental Material are available via Figshare at the link <https://figshare.com/s/4324fbd6cd9d779a6c24>²². All raw numerical data used to generate the main figures are provided in this repository.

Code availability

All code used for the MOFA analysis—developed and provided by the MOFA team—is openly accessible. To facilitate easy and direct access for readers, we include links to both the official MOFA webpage and the relevant code repositories. This ensures continuous access to the most up-to-date tools and resources used in our analysis. The MOFA code and tutorials are available at: <https://biofam.github.io/MOFA2/tutorials.html>; https://raw.githubusercontent.com/bioFAM/MOFA2_tutorials/master/R_tutorials/getting_started_R.html; https://raw.githubusercontent.com/bioFAM/MOFA2_tutorials/master/R_tutorials/downstream_analysis.html.

Received: 9 April 2024; Accepted: 15 May 2025;

Published online: 27 May 2025

References

- Kuruvilla, M. E., Lee, F. E. & Lee, G. B. Understanding asthma phenotypes, endotypes, and mechanisms of disease. *Clin. Rev. Allergy Immunol.* **56**, 219–233 (2019).
- Battaglia, M. et al. Introducing the endotype concept to address the challenge of disease heterogeneity in type 1 diabetes. *Diab. Care* **43**, 5–12 (2020).
- Endesfelder, D. et al. Time-resolved autoantibody profiling facilitates stratification of preclinical type 1 diabetes in children. *Diabetes* **68**, 119–130 (2019).
- Achenbach, P. et al. A classification and regression tree analysis identifies subgroups of childhood type 1 diabetes. *EBioMedicine* **82**, 104118 (2022).
- Parviainen, A. et al. Heterogeneity of type 1 diabetes at diagnosis supports existence of age-related endotypes. *Diab. Care* **45**, 871–879 (2022).
- Redondo, M. J. & Morgan, N. G. Heterogeneity and endotypes in type 1 diabetes mellitus. *Nat. Rev. Endocrinol.* **19**, 542–554 (2023).
- Aida, K. et al. Distinct Inflammatory Changes of the Pancreas of Slowly Progressive Insulin-dependent (Type 1) Diabetes. *Pancreas* **47**, 1101–1109 (2018).
- Mortensen, H. B. et al. New definition for the partial remission period in children and adolescents with type 1 diabetes. *Diab. Care* **32**, 1384–1390 (2009).
- Krischer, J. P. et al. Characteristics of children diagnosed with type 1 diabetes before vs after 6 years of age in the TEDDY cohort study. *Diabetologia* **64**, 2247–2257 (2021).
- Leete, P. et al. Differential insulinitic profiles determine the extent of beta-cell destruction and the age at onset of type 1 diabetes. *Diabetes* **65**, 1362–1369 (2016).
- Leete, P. et al. 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 (2020).
- Sharp, S. A. et al. Development and standardization of an improved type 1 diabetes genetic risk score for use in newborn screening and incident diagnosis. *Diab. Care* **42**, 200–207 (2019).
- Oram, R. A. et al. A type 1 diabetes genetic risk score can aid discrimination between type 1 and type 2 diabetes in young adults. *Diab. Care* **39**, 337–344 (2016).
- Bechi Genzano, C. et al. Combined unsupervised and semi-automated supervised analysis of flow cytometry data reveals cellular fingerprint associated with newly diagnosed pediatric type 1 diabetes. *Front Immunol.* **13**, 1026416 (2022).

15. Van Gassen, S. et al. FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data. *Cytom. A* **87**, 636–645 (2015).
16. Argelaguet, R. et al. Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets. *Mol. Syst. Biol.* **14**, e8124 (2018).
17. Athieniti, E. & Spyrou, G. M. A guide to multi-omics data collection and integration for translational medicine. *Comput Struct. Biotechnol. J.* **21**, 134–149 (2023).
18. Redondo, M. J. et al. Association of TCF7L2 variation with single islet autoantibody expression in children with type 1 diabetes. *BMJ Open Diab. Res. Care* **2**, e000008 (2014).
19. Redondo, M. J. et al. Single islet autoantibody at diagnosis of clinical type 1 diabetes is associated with older age and insulin resistance. *J. Clin. Endocrinol. Metab.* **105**, 1629–1640 (2020).
20. Petrelli, A. et al. HOMA-IR and the Matsuda Index as predictors of progression to type 1 diabetes in autoantibody-positive relatives. *Diabetologia* **67**, 290–300 (2024).
21. Arian, M. et al. Metaproteogenomic analysis of saliva samples from Parkinson's disease patients with cognitive impairment. *NPJ Biofilms Microbiomes* **9**, 86 (2023).
22. Petrelli, A. et al. Source data for: A multi-omics integration approach relying on circulating factors does not discern subtypes of childhood type 1 diabetes. Figshare. Dataset. <https://doi.org/10.6084/m9.figshare.25623762>

Acknowledgements

A.P. is supported by a Juvenile Diabetes Research Foundation Advanced Postdoctoral Fellowship (3-APF-2019-744-AN). We express their gratitude to Angela Stabilini and Francesca Ragogna for their assistance with database management, as well as doctors Elisa Morotti, Valeria Castorani, Valeria Favalli, and Marianna Rita Stancampiano for their contributions to sample collection. Special thanks are extended to Raniero Chimienti for providing bioinformatic insights during the initial stages of the analysis. We also acknowledge the children with diabetes and their families whose participation made this study possible.

Author contributions

V.C. performed sample collection, literature search, data interpretation, drafted and revised the final version of the article. N.B. performed data analysis and contributed to data interpretation and revised the final version of the article. G.M.S., M.J.M., E.B., C.B.G. contributed to data analysis. D.C. and A.M. contributed to sample collection. A.R. and G.F. contributed to subject enrollment and reviewed the final version of the manuscript. A.G. contributed to data collection, data analysis and data interpretation. I.M. and V.L. performed genetic analysis and reviewed the final version of the manuscript. P.F., A.G., L.P., M.B. contributed to data interpretation and

reviewed the final version of the manuscript. R.B. contributed to subject enrollment and to data interpretation, reviewed the final version of the manuscript. A.P. designed and coordinated the study, contributed to data interpretation, drafted and edited the manuscript. All authors have read and approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s43856-025-00922-7>.

Correspondence and requests for materials should be addressed to Alessandra Petrelli.

Peer review information *Communications Medicine* thanks Noel Morgan, Mitsunori Ogihara and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025