Received: 6 December 2023 | Revised: 30 April 2024 | Accepted: 7 May 2024

DOI: 10.1002/dmrr.3833

RESEARCH ARTICLE

WILEY

Multi-omics analysis reveals drivers of loss of β-cell function after newly diagnosed autoimmune type 1 diabetes: An INNODIA multicenter study

Jose Juan Almagro Armenteros[1] | Caroline Brorsson[1] | Christian Holm Johansen[1] | Karina Banasik[1] | Gianluca Mazzoni[1] | Robert Moulder[2,3] | Karoliina Hirvonen[2,3] | Tomi Suomi[2,3] | Omid Rasool[2,3] | Sylvaine F. A. Bruggraber[4] | M. Loredana Marcovecchio[4] | Emile Hendricks[4] | Naba Al-Sari[5] | Ismo Mattila[5] | Cristina Legido-Quigley[5] | Tommi Suvitaival[5] | Piotr J. Chmura[1] | Mikael Knip[6,7] | Anke M. Schulte[8] | Jeong Heon Lee[9] | Guido Sebastiani[10,11] | Giuseppina Emanuela Grieco[10,11] | Laura L. Elo[2,3,12] | Simranjeet Kaur[13] | Flemming Pociot[13] | Francesco Dotta[10,14] | Tim Tree[15] | Riitta Lahesmaa[2,3] | Lut Overbergh[16] | Chantal Mathieu[16] | Mark Peakman[9,15] | Søren Brunak[1] | on behalf of the INNODIA investigators

Correspondence
Søren Brunak, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Blegdamsvej 3A, Copenhagen DK-2200, Denmark.
Email: soren.brunak@cpr.ku.dk

Funding information
Innovative Medicine Initiative 2 Joint Undertaking; European Federation of Pharmaceutical Industries and Associations; European Union's Horizon 2020 research and innovation program; Leona M. and Harry B. Helmsley Charitable Trust; JDRF

Abstract

Aims: Heterogeneity in the rate of β-cell loss in newly diagnosed type 1 diabetes patients is poorly understood and creates a barrier to designing and interpreting disease-modifying clinical trials. Integrative analyses of baseline multi-omics data obtained after the diagnosis of type 1 diabetes may provide mechanistic insight into the diverse rates of disease progression after type 1 diabetes diagnosis.

Methods: We collected samples in a pan-European consortium that enabled the concerted analysis of five different omics modalities in data from 97 newly diagnosed patients. In this study, we used Multi-Omics Factor Analysis to identify molecular signatures correlating with post-diagnosis decline in β-cell mass measured as fasting C-peptide.

Results: Two molecular signatures were significantly correlated with fasting C-peptide levels. One signature showed a correlation to neutrophil degranulation, cytokine signalling, lymphoid and non-lymphoid cell interactions and G-protein coupled receptor signalling events that were inversely associated with a rapid decline in β-cell function. The second signature was related to translation and viral

‡INNODIA: 'Innovative approaches to understanding and arresting type 1 diabetes'- List of contributors included with submission.

Jose Juan Almagro Armenteros, Caroline Brorsson and Christian Holm Johansen were the first authors.

infection was inversely associated with change in β-cell function. In addition, the immunomics data revealed a Natural Killer cell signature associated with rapid β-cell decline.

Conclusions: Features that differ between individuals with slow and rapid decline in β-cell mass could be valuable in staging and prediction of the rate of disease progression and thus enable smarter (shorter and smaller) trial designs for disease modifying therapies as well as offering biomarkers of therapeutic effect.

KEYWORDS
disease progression, multi-omics, type 1 diabetes

1 | INTRODUCTION

Type 1 diabetes is an autoimmune disease involving environmental and genetic factors that trigger immune-mediated pancreatic β-cell dysfunction and destruction that results in insulin loss and symptomatic hyperglycemia requiring lifelong insulin therapy.[1] Globally, around 1.2 million people below the age of 20 years have type 1 diabetes, with an annual increase in incidence of 3% strongly influenced by geography.[2] Insulin replacement therapy is unable to fully mimic physiological control of blood glucose and therefore, many people living with type 1 diabetes develop severe disease complications that are directly attributable to prolonged glycaemic exposure,[3] with markedly reduced life expectancy.[4] In the face of the disease burden and unmet need, international consortia are mobilising to develop disease-modifying therapies. For example, therapies that maintain even minimal residual C-peptide secretion capacity have been found to have demonstrable clinical benefit.[5]

An emerging barrier to this effort is disease heterogeneity. In particular, the rate of decline of β-cell capacity is highly variable for reasons that remain unclear. As a result, clinical trial designs for disease-modifying therapies are necessarily cumbersome, requiring large sample sizes and prolonged observation. In addition, opportunities for tailored disease-modifying therapies are limited by an unclear understanding of the factors that drive diabetes progression after diagnosis. Type 1 diabetes endotypes have previously been proposed to describe the disease heterogeneity.[6,7] Additionally, multi-omics strategies have already demonstrated promising results in modelling the heterogeneity of diabetic kidney disease onset.[8] Applying multi-omics analyses could therefore help explain aspects of the heterogeneity in progression after diagnosis and identify endotypes in a data driven manner. This knowledge could provide the attainment of improved participant inclusion in the focused designed trials to foster their success and participant benefit.

Studies with this goal conducted to date have typically been constrained by limitations to cohort size and the number of different data dimensions available for analysis.[9,10] Thus, it has not been possible to conduct studies with an emphasis on hypothesis-generating, unbiased approaches, and integration of data across pathophysiological systems. These require large-scale, inter-disciplinary research efforts in which carefully curated longitudinal clinical cohorts are aligned with multi-parametric technology platforms. Such a systems-based approach can address key questions with less bias and generate novel hypotheses on disease drivers. We used this strategy in the setting of a pan-European research consortium in which people with newly diagnosed type 1 diabetes as well as people at risk of developing type 1 diabetes (antibody positive) were enrolled into a master protocol[11] to conduct a prospective study in search of factors that associate with the rate of decline in β-cell mass/function. This endeavour was supported by the Innovative Medicines Initiative-2 Joint Undertaking, where INNODIA was created, a private-public partnership of 40 partners in 16 countries. In the natural history study, people with newly diagnosed type 1 diabetes and unaffected family members are in follow up, allowing deep clinical characterisation as well as multi-omics analysis of samples (blood, urine, stool) collected and analysed using standardized operating procedures (www.innodia.eu). Here we report the multi-omics analysis of the 'first 100' people with newly diagnosed disease. We report the existence of latent factors integrated from transcriptomic, small RNA, genomic, targeted proteomic, lipidomic, metabolomic, and immunomic-level data that show a relationship to subsequent rates of disease progression and have potential value as stratification and therapeutic target identification tools.

2 | METHODS

Here we present the methods for analysing the clinical data, integration of the multi-omics data and further analyzes of the integrated results. The materials and methods used to generate the different omics types can be found in the supplementary material.

2.1 | Subjects with type 1 diabetes

For this in depth analysis, we included the first 100 subjects with newly diagnosed (<6 weeks) type 1 diabetes enrolled in the INNODIA natural history study. Using a consecutive recruitment approach, subjects were included based on baseline omics data availability, an even gender distribution and biosample availability, positivity for at least one diabetes-related autoantibody (GADA, IA-2A, IAA, ZnT8A), and age between one and 45 years. Two subjects were excluded due

to incomplete 'omics datasets and one following the detection of a MODY gene mutation. The final analysis cohort comprised 49 male and 48 female study participants (Table 1), the average age at diabetes diagnosis of 13.2 years (SD 8.5; two-ten years $n = 38$; ten-18 years $n = 41$, 18–39 years $n = 18$), mean disease duration of 3.9 weeks (SD 1.5) and at baseline an average total daily insulin dose of 0.51 IU/kg (SD 0.27), HbA1c of 75 mmol/mol (SD 21.3), fasting C-peptide level of 269 pmol/L (IQR 25.7), fasting glucose level of 7.78 mmol/L (IQR 2.8) and BMI SDS of 0.327 units (SD 1.1)'.

C-peptide measurements were made at four visits (Figure 1A). Fasted C-peptide was calculated from measurements made prior to the mixed meal tolerance test (MMTT) at each visit. This measure was preferred to Area under the curve (AUC) C-peptide and fasted C-peptide/glucose ratio due to missingness in these variables. Fasted C-peptide was available for a total of 354 visits, while AUC C-peptide and fasted C-peptide/glucose ratio were available for 232 and 328 visits, respectively. Fasted C-peptide has also previously been suggested as a reliable substitute for AUC C-peptide.[12]

Comparison of log-transformed fasted C-peptide to log-transformed AUC C-peptide and log-transformed fasted C-peptide/glucose ratio for non-missing measurements showed a high Pearson correlation of 0.92 and 0.94, respectively (Figure S1).

C-peptide is widely accepted as a measure of endogenous insulin secretion.[13] To define the rate of C-peptide decline over time, we utilized linear mixed-effect models to fit the log-transformed fasted C-peptide from day of diagnosis to 12 months (Figure 1B). The model was fitted using subject-level random effects and the rate of C-peptide change over time. Mixed-effect models were fit using the lme4 R package[14] with an unstructured random effects variance-covariance matrix.

A total of 69 individuals completed all four visits (baseline, three, six, and 12 months), 21 individuals completed three visits, five individuals completed two visits, and two individuals completed only the baseline visit (not necessarily consecutive visits).

At each visit, HbA1c was measured and the stimulated C-peptide response was determined by MMTT from individuals of at least five years of age. The islet autoantibodies GADA, IA-2A, IAA, and ZnT8A were quantified with the use of specific radiobinding assays as described earlier.[15]

2.2 | Multi-omics data pre-processing, integration, and analysis

We normalised the counts for transcriptomics and miRNA data by variance stabilising transformation from the DESeq 2 package.[16] The transcriptomics and miRNA data sets were filtered for low counts (features with less than 10 counts in total or features with zero counts in more than 90% of the samples). The proteomics, metabolomics, lipidomics and immunomics data sets were log2 transformed. The transcriptomics, proteomics, and immunomics data were corrected for batch effects associated with dataset-specific factors (sequencing at different days or different handling of the samples) using the limma package in R. Finally, all the data sets were corrected for age. Age was log-transformed to account for the growth effect in children (one year difference in adults is not equivalent to one year difference in children). This was necessary due to the high degree of age heterogeneity present in the cohort and the association of the fasted C-Peptide slopes with age.

The omics data sets were integrated using the general framework in the Multi-Omics Factor Analysis (MOFA) package from 2018.[17] MOFA performs a dimensionality reduction of the omics data into a lower-dimensional latent space (Figure 1C). The latent factors generated by MOFA capture sources of global variability across the different omics data sets. Each factor has an underlying weight for every feature, which can be used to annotate the factors in terms of omics features, yielding a specific molecular signature for each factor. MOFA was run with default parameters and 20 latent factors. The model was initialised with different random seeds yielding similar results, generally only altering the number assigned to each factor associated with the C-peptide slopes.

Latent factors were associated with the estimated C-peptide slopes using the Spearman correlation. Other covariates were also analysed such as age at baseline and C-peptide at baseline. The association of the latent factors with the progression groups was calculated using the Kruskal-Wallis test and each group was compared using a Mann-Whitney *U* test.

Gene Set Enrichment Analysis (GSEA)[18] was performed to characterise the genes with the largest weight in the latent factors. This analysis was performed using the MOFA GSEA function, which utilises a modified version of the principal component gene set enrichment scheme (PCGSE).[19] The Reactome database was used as the gene set annotation for this analysis. The GSEA was performed separately for genes with a positive and negative weight in each latent factor. This was done to avoid combining genes that are upregulated (positive weight) and downregulated (negative weight) in the latent factor as these two groups of genes might be involved in different biological pathways. The top 15 significant pathways for each latent factor (positive and negative weights) were selected and grouped by biological pathway.

Differential gene expression of individual genes was determined by the DESeq 2 package.[16] Covariates for the batch variable (different runs) and age were included in the model. Age was encoded as a categorical variable defined in three groups: less than ten years, between ten and 18 years, and more than 18 years. Differential gene expression was assessed between the rapid and slow groups, rapid and increasing and slow and increasing, respectively.

For biological network analysis, two types of interaction networks were constructed. One was compiled from the STRING[20] database to study protein-protein interactions or associations only. Another network was compiled from the mirTarBase[21] (miRNA-gene), which was combined with the STRING network to study protein-protein-miRNA interactions or associations. The STRING database was filtered for high-confidence interactions (combined score above 0.7) and miRNA-gene interactions had to be reported by at least two publications and two non-high-throughput methods. We decided to

TABLE 1 Clinical and demographic data for the type 1 diabetes cohort across progression groups.

	Rapid ($N = 32$)	Slow ($N = 31$)	Increasing ($N = 32$)	p-value	Overall ($N = 95$)
Sex					
Female	20 (62.5%)	15 (48.3%)	13 (40.6%)	0.211	48 (49.5%)
Male	12 (37.5%)	16 (51.6%)	19 (59.4%)	(0.159)	49 (50.5%)
Age (years)					
Mean (SD)	11.2 (9.17)	12.9 (7.94)	15.6 (7.94)	0.0125	13.2 (8.49)
Median (IQR) [min, max]	9.60 (8.97) [2.01, 38.1]	10.9 (8.68) [2.08, 36.4]	13.4 (5.23) [7.57, 38.8]	(<0.001)	11.8 (8.08) [2.01, 38.8]
Age intervals					
<10	16 (50.0%)	13 (41.9%)	7 (21.9%)	0.13	38 (39.2%)
>10–18	11 (34.4%)	12 (38.7%)	18 (56.3%)	(0.021)	41 (42.3%)
>18	5 (15.6%)	6 (19.4%)	7 (21.9%)		18 (18.6%)
Ketoacidosis at diagnosis					
Yes	10 (31.3%)	11 (35.5%)	10 (31.2%)	0.981	32 (33.0%))
No	22 (68.7%)	19 (61.3%)	22 (68.8%)	(0.991)	64 (66.0%)
Missing	0 (0%)	1 (3.2%)	0 (0%)		1 (1.0%)
diagnosis To baseline (weeks)					
Mean (SD)	3.88 (1.52)	3.80 (1.66)	4.10 (1.43)	0.843	3.93 (1.53)
Median (IQR). [Min, max]	4.40 (2.50) [0.900, 6.40]	4.35 (2.35) [0.900, 6.30]	4.40 (1.70) [0.700, 6.10]	(0.639)	4.40 (2.20) [0.700, 6.40]
BMI SDS at baseline					
Mean (SD)	0.0719 (1.05)	0.397 (1.02)	0.511 (1.20)	0.153	0.327 (1.10)
Median (IQR) [min, max]	0.0800 (1.19). [−2.32, 2.47]	0.350 (1.60) [−1.56, 2.72]	0.610 (1.35) [−2.00, 2.28]	(0.105)	0.260 (1.57). [−2.32, 2.72]
Missing	1 (3.1%)	0 (0%)	0 (0%)		1 (1.0%)
BMI SDS at 12 months					
Mean (SD)	0.234 (0.923)	0.309 (0.991)	0.549 (1.20)	0.416	0.366 (1.04)
Median (IQR) [min, max]	0.165 (0.937) [−1.79, 2.33]	0.49 (1.44) [−1.57, 2.03]	0.505 (1.162) [−1.98, 2.56]	(0.242)	0.4 (1.43) [−1.98, 2.56]
Missing	3 (9.4%)	4 (12.9%)	2 (6.3%)		10 (10.3%)
Glucose at baseline (mmol/L)					
Mean (SD)	8.43 (5.57)	8.42 (5.33)	6.45 (2.20)	0.22	7.78 (4.69)
Median (IQR) [min, max]	7.10 (2.70) [3.70, 31.5]	6.70 (2.93) [3.60, 26.9]	6.10 (2.25) [3.70, 13.8]	(0.085)	6.40 (2.80) [3.60, 31.5]
Glucose at 12 months (mmol/L)					
Mean (SD)	7.56 (2.27)	7.57 (2.97)	8.15 (1.75)	0.655	7.79 (2.33)
Median (IQR) [min, max]	7.35 (4.15). [4.2, 11.4]	6.90 (3.40) [4.10, 18.6]	8.24 (2.90) [4.90, 11.0]	(0.261)	7.80 (3.30) [4.10, 18.6]
Missing	8 (25.0%)	6 (19.4%)	2 (6.25%)		18 (18.6%)
Daily insulin dose pr kg at baseline (IU/kg)					
Mean (SD)	0.544 (0.279)	0.498 (0.247)	0.498 (0.281)	0.819	0.514 (0.268)
Median (IQR). [Min, max]	0.545 (0.291) [0.136, 1.46]	0.468 (0.243) [0.108, 1.20]	0.500 (0.429) [0.0359, 1.01]	(0.616)	0.500 (0.354) [0.0359, 1.46]
Missing	1 (3.1%)	1 (3.2%)	0 (0%)		2 (2.1%)

TABLE 1 (Continued)

	Rapid ($N = 32$)	Slow ($N = 31$)	Increasing ($N = 32$)	p-value	Overall ($N = 95$)
Daily insulin dose per kg at 12 months (IU/kg)					
Mean (SD)	0.700 (0.320)	0.517 (0.249)	0.486 (0.211)	0.005	0.567 (0.277)
Median (IQR) [min, max]	0.640 (0.320) [0.024, 1.41]	0.466 (0.280) [0.125, 1.41]	0.48 (0.255) [0.067, 0.91]	(<0.001)	0.561 (0.312) [0.024, 1.41]
Missing	6 (18.8%)	6 (19.4%)	3 (9.4%)		16 (16.5%)
HbA1c at baseline (mmol/mol)					
Mean (SD)	72.2 (21.4)	74.9 (24.2)	78.0 (18.1)	0.699	75.0 (21.3)
Median (IQR) [min, max]	75.0 (27.3) [8.70, 119]	73.0 (29.5) [13.4, 130]	80.3 (22.4) [50.0, 130]	(0.332)	77.5 (25.7) [8.70, 130]
Missing	1 (3.1%)	1 (3.2%)	1 (3.1%)		3 (3.1%)
HbA1c at 12 months (mmol/mol)					
Mean (SD)	55.7 (15.9)	53.2 (12.7)	50.6 (8.89)	0.427	53.3 (12.96)
Median (IQR) [min, max]	51.9 (9.75) [40.0, 130]	52.0 (13.3) [37.0, 99.0]	50.0 (10.6) [34.0, 71.0]	(0.110)	51.0 (12.1) [34.0, 130]
Missing	2 (6.3%)	5 (16.1%)	4 (12.5%)		13 (13.4%)
C-peptide estimated slopes					
Mean (SD)	−0.425 (0.173)	−0.125 (0.050)	0.046 (0.062)	<0.001	−0.171 (0.225)
Median (IQR) [min, max]	−0.412 (0.273) [−0.866, −0.209]	−0.134 (0.083) [−0.205, −0.0481]	0.047 (0.101) [−0.047, 0.185]	(<0.001)	−0.135 (0.264) [−0.866, 0.185]
C-peptide at baseline (pmol/L)					
Mean (SD)	235 (183)	248 (185)	324 (263)	0.296	269 (215)
Median (IQR) [min, max]	165 (228) [26.1, 809]	240 (204) [15.0, 986]	264 (268) [25.8, 1290]	(0.07)	224 (239) [15.0, 1290]
C-peptide at 12 months (pmol/L)					
Mean (SD)	64.7 (59.8)	194 (137)	432 (206)	<0.001	234 (214)
Median (IQR) [min, max]	41.6 (83.5) [11.7, 222]	161 (136) [45.8, 635]	433 (246) [96.7, 896]	(<0.001)	163 (283) [11.7, 896]
Missing	6 (18.8%)	8 (25.8%)	5 (15.6%)		21 (21.6%)
GADA at baseline					
Negative	5 (15.2%)	11 (34.4%)	8 (25.0%)	0.199	24 (24.7%)
Positive	28 (84.8%)	21 (65.6%)	24 (75.0%)	(0.204)	73 (75.3%)
IA-2A at baseline					
Negative	9 (27.3%)	12 (37.5%)	9 (28.1%)	0.615	30 (30.9%)
Positive	24 (72.7%)	20 (62.5%)	23 (71.9%)	(0.495)	67 (69.1%)
IAA at baseline					
Negative	4 (12.1%)	10 (31.3%)	10 (31.3%)	0.118	24 (24.7%)
Positive	29 (87.9%)	22 (68.8%)	22 (68.8%)	(0.490)	73 (75.3%)
ZnT8A at baseline					
Negative	12 (36.4%)	10 (31.3%)	14 (43.8%)	0.582	36 (37.1%)
Positive	21 (63.6%)	22 (68.8%)	18 (56.3%)	(0.481)	61 (62.9%)
Detectable autoantibodies at baseline					
Mean (SD)	3.09 (0.947)	2.66 (0.902)	2.72 (0.851)	0.0961	2.82 (0.913)
Median (IQR) [min, max]	3.00 (2.00) [1.00, 4.00]	3.00 (1.00) [1.00, 4.00]	3.00 (1.00) [1.00, 4.00]	(0.014)	3.00 (2.00) [1.00, 4.00]

Note: p-values were calculated using the Kruskal-Wallis test for associations with progression groups. p-values for associations with estimated C-peptide slopes were calculated with Kruskal-Wallis and Spearman's rank correlation coefficient for categorical and continuous data, respectively.

focus on genes and miRNAs only because these omics types generally have greater weights in the most relevant latent factors.

The association between the biological networks and the latent factors was performed using the PCSF graph optimization approach. This method allows us to interpret the biological landscape of the interaction network based on the importance/weight of each gene/miRNA in the latent factor. The output is a subnetwork that captures interactions between the genes/miRNAs with a higher importance in the latent factor. In order to select a subset of genes and miRNAs to construct the network, only genes with normalised absolute weights three-fold higher than expected by chance and miRNAs with normalised absolute weights two-fold higher than expected by chance were selected. Grid-search was performed to select the best parameters based on the network that had a high number of genes/miRNA from the latent factor while keeping the number of genes/miRNA not observed in the latent factor low ($\mu = 0.005$, $\omega = 1$, $\beta = 5000$). The final network was constructed using 20 runs with noise to edge costs ($r = 0.1$) that were combined and clustered using the edge-betweenness algorithm. Gene Set Enrichment Analysis was performed on the clusters obtained from the network.

3 | RESULTS

3.1 | C-peptide decline over time

The individual rates of C-peptide decline were calculated as the slope of fasted C-peptide change over 12 months as described in Methods (Supplementary Figure S2).

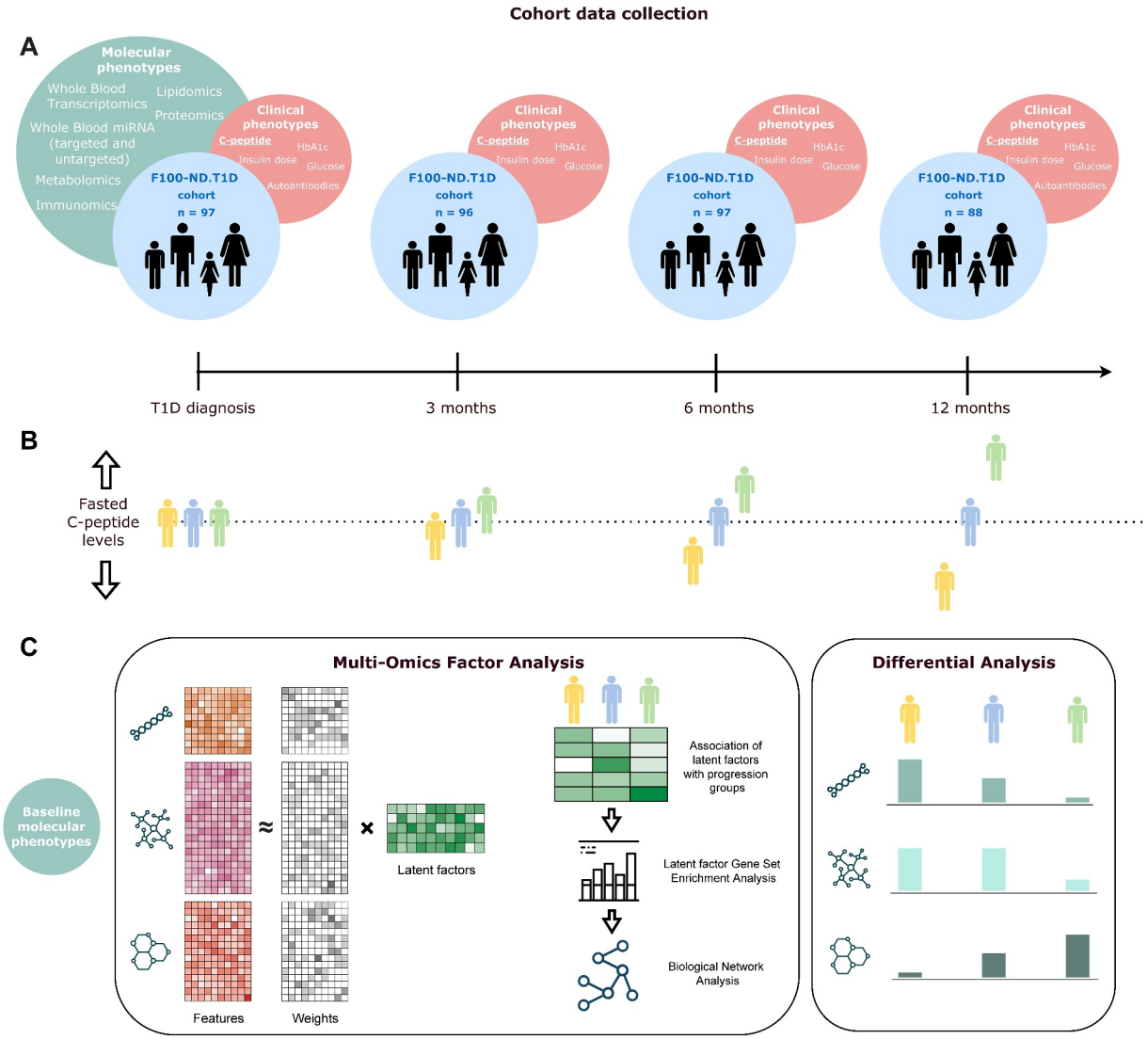

FIGURE 1 Cohort data and analysis overview. (A) The cohort consists of 97 people with newly diagnosed type 1 diabetes. Multi-omics data were collected at baseline (within six weeks after diagnosis of type 1 diabetes) and clinical data were collected at baseline and at three, six, and 12 months (B) Participants were divided into three groups based on their change of insulin secretion levels (fasted C-peptide measurements) from baseline to 12 months (C) Multi-Omics Factor Analysis was performed to obtain an integrated signature across omics data types followed by differential expression analysis for each omics data type independently.

An overall trend of C-peptide decline over time was observed for the entire cohort (*p*-value 0.0001) (Supplementary Figure S3A, Supplementary Figure S4A). C-peptide slopes were used to divide patients into terciles (equal-sized), classified as rapid, slow, and increasing progression groups, yielding three groups with distinct C-peptide progression patterns (Supplementary Figure S3B, Supplementary Figure S4B). Two individuals had only one available C-peptide measurement and were not assigned a group as C-peptide decline could not be determined. These patients were included in the MOFA model as their omics data are still valuable for constraining the integration of the omics data into factors. All three progression groups had a similar estimated baseline C-peptide (intersect) (*p*-value 0.296), while C-peptide slopes differed significantly (*p*-value <0.0001). Time from diagnosis to baseline and frequency of ketoacidosis at diagnosis showed no significant differences (*p*-value 0.843 and 0.98). Furthermore, there was a significant association between C-peptide slopes and the number of autoimmune antibodies at baseline (*p*-value 0.014). Estimated baseline C-peptide and glucose levels were borderline significant (*p*-value 0.07 and 0.085). For clinical features 12 months after baseline, only C-peptide and average insulin dose proved to be significantly associated. BMI, HbA1c and glucose and insulin dose showed no differences between the progression groups. These findings were consistent for both associations with patient groups and C-peptide slopes. The clinical features for each of the progression groups are shown in Table 1.

We further analysed the relationship between age, divided into children (<10), adolescents (10–18) and adults (>18), and C-peptide levels (Supplementary Figure S3C, Supplementary Figure S4C). Having three intervals coincides with shifts in the general incidence of diabetes, although other thresholds could have been used. The analysis showed that the participants in the <10 years group had a significantly lower baseline C-peptide level compared to both 10–18 years and >18 years groups (*p*-values <0.0001). In addition, the 10–18 years group had reduced but non-significant baseline C-peptide levels compared to >18 years age groups (*p*-value 0.27).

Changes in C-peptide over time showed no significant differences in the interactions between time and the age groups. The 10–18 years group had a non-significant, yet more negative trend compared to the <10 years and >18 years groups (*p*-values: 0.25 and 0.40, respectively), while the <10 years and >18 years groups were similar (*p*-value: 0.90). These findings indicate a degree of association between C-peptide levels and age among children, and as a result, age was included after log transformation as a covariate in our models to correct for potential effects on the relationship between C-peptide slopes and omics.

Evaluating the association of sex with the C-peptide change over time (Supplementary Figure S3D, Supplementary Figure S4D), we found no significant association with baseline C-peptide (*p*-value 0.64) or slope (*p*-value 0.16).

3.2 | Multi-omics integration analysis

The multi-omics data set overview is shown in Figure 2A. Missing values, which are seen predominantly in the immunomics data set, are disregarded by MOFA and do not affect the decomposition of the data into latent factors. After training MOFA on the multi-omics data set (Figure 2B), the latent factors that capture most of the variance across participants were represented by the mRNA and miRNA data. MOFA captures latent factors with common variance across the different omics types, even though certain data types appear to be responsible of most of the captured variance (Factors 1 to 7). This indicates that a certain degree of heterogeneity exists across omics types, making the integration more challenging.

Importantly, however, latent factors 15 and 18 were significantly associated (*p*-values <0.1 adjusted by Benjamini-Hochberg) with C-peptide slopes (Figure 2D) but not with age or baseline C-peptide. The amount of variance captured by latent factors 15 and 18 is 2.62% and 1.84%, respectively, indicating that the decline of C-peptide over time is not among the major sources of variance across the participants but is sufficiently strong to be captured by this method. The differences in the strength of associations for latent factors 15 and 18 with the progression groups and C-peptide slopes indicated that latent factor 15 captures a non-linear association with the progression groups (Figure 2C), and for latent factor 18 a linear association with the rate of C-peptide decline (Figure 2E). As the two factors correlate with C-peptide decline, they may contain molecular signatures useful for explaining the differences in disease progression between patients. Therefore, we continued a thorough scrutiny of these factors.

3.3 | Differential gene expression analysis

Differentially expressed genes (DEGs) were identified between the different progression groups, with batch and age groups used as covariates (Figure 3). *p*-values were adjusted for multiple testing using the Benjamini-Hochberg procedure and genes with an adjusted *p*-value <0.1 were reported as differentially expressed genes (DEG). Figure 3 shows the volcano plots of the different comparisons together with the genes belonging to latent factors 15 and 18. A total of 339 DEGs were observed comparing the rapid decline group and the increasing group (Figure 3A, Supplementary Table S1), 33 DEGs between the rapid and slow decline groups (Figure 3B, Supplementary Table S2), and 1205 DEGs between the slow decline group and the increasing group (Figure 3C, Supplementary Table S3). Additionally, differential gene expression was performed for the C-peptide slopes, hence studying the linear change in gene expression with respect to the rate of decline of C-peptide over time. Here we found 484 DEGs (Figure 3D, Supplementary Table S4).

More DEGs were observed when comparing the slow and increasing progression groups than when comparing the rapid and increasing progression groups, with little overlap between the

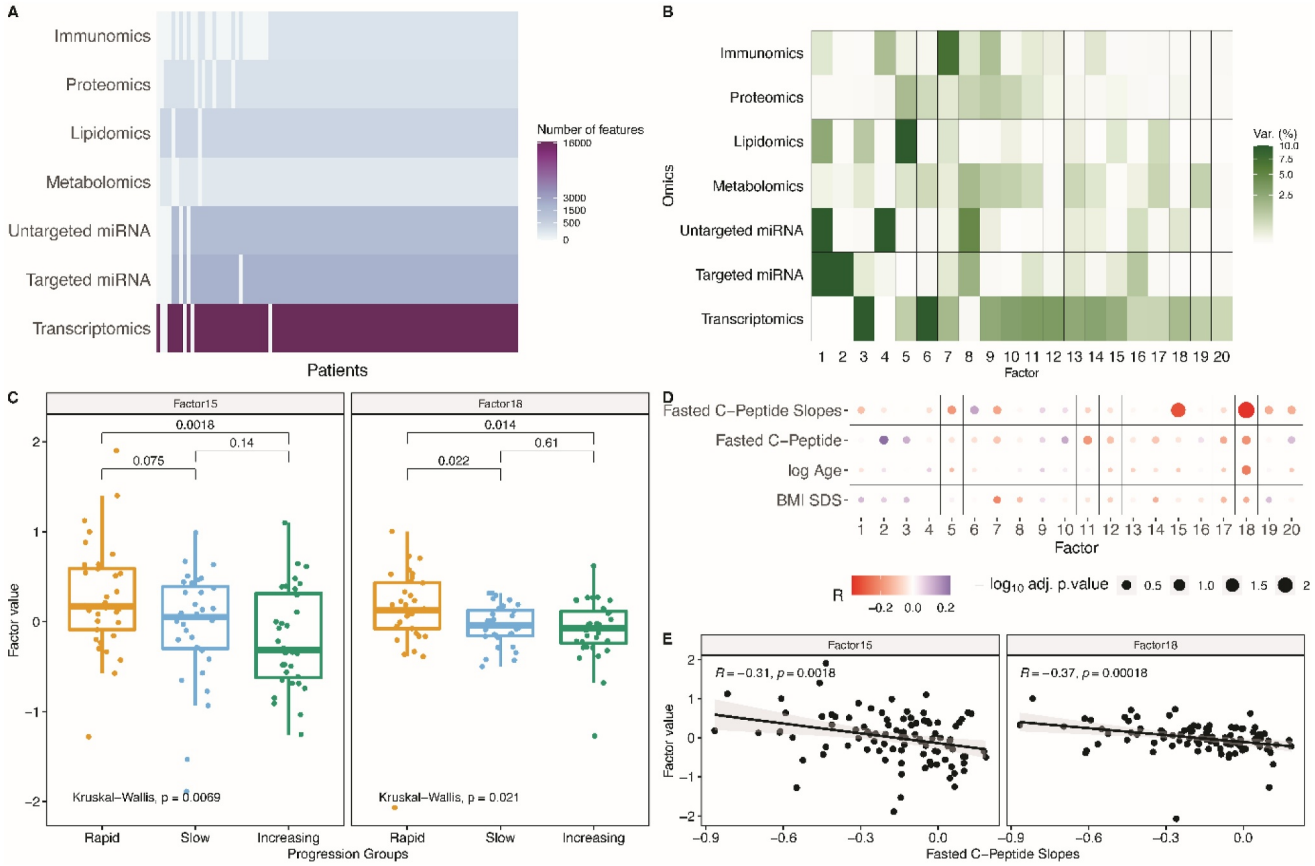

FIGURE 2 (A) Overview of the multi-omics data sets, describing the number of features per data set and the level of missing data (white). (B) Latent factors obtained from MOFA, the colour scale represents the variance captured by each of the latent factors indicating the level of integration of the data types for each factor. (C) Association of latent factors 15 and 18 values with the different progression groups. (D) Spearman correlation of latent factors with the fasted C-peptide slopes, baseline fasted C-Peptide, Age (log scale), and BMI-SDS. *p*-values were adjusted by Benjamini-Hochberg. E, Spearman correlation of fasted C-peptide slopes against latent factor values (15 and 18).

significant top-ranking genes. Additionally, very few genes were differentially expressed between the rapid and slow progression groups. Altogether, these results might indicate that even though the rapid and slow progression groups are more similar, the set of DEGs between these two groups and the increasing group are not the same. To further confirm this, two additional differential expression analyses were performed. One between the rapid-slow groups combined versus the increasing group and another between the slow-increasing groups combined versus the rapid. The analyses showed that the rapid-slow versus increasing comparison produced 1804 DEGs (Supplementary Table S5), while the slow-increasing versus rapid comparison produced 18 DEGs (Supplementary Table S6). This indicates that the increasing progression group is much more dissimilar in its blood sample expression profile compared to the other two groups at the early stage of type 1 diabetes manifestation. Therefore, the underlying biological processes involved in the developing disease progression may vary already early in the disease manifestation when comparing individuals experiencing loss of β-cell function and those experiencing an improvement in β-cell function.

The similarity between DEGs found by the continuous change in C-peptide levels ($n = 484$) and DEGs between the rapid and increasing group ($n = 339$) showed an overlap of 209 genes. The rapid-slow versus increasing DEGs ($n = 1804$) had a bigger overlap with the continuous C-peptide levels where 313 of the same DEGs were found. In all cases, the DEGs had the same sign of their log2 fold changes for both analyses. Therefore, most DEGs were observed for the continuous change were also found when investigating progression groups.

Nonetheless, we believe that the linear association of blood gene expression at baseline with the C-peptide slopes is more informative regarding the disease progression. We observe that change in β-cell function follows a gradient; therefore, by categorising participants into groups, we lose the resolution that the C-peptide slopes provide.

With respect to the C-peptide slopes, the significant DEGs (Figure 3D) also found among latent factor 15 or 18 were mostly associated with the immune system. These genes included, FOXP1, a transcription factor associated with LPS exposure to neutrophils[22] and hepatic homoeostasis of glucose in mice,[23] Notch1, associated with a decrease in insulin secretion and β-cell mass, and FADS1, associated with fasting glucose in non-diabetics.[24] Lastly, we found that some T-cell receptor genes (TRAV29DV5, TRAC) and immunoglobulin-like receptors (LILRB1 and SIRPG) were also differentially expressed.

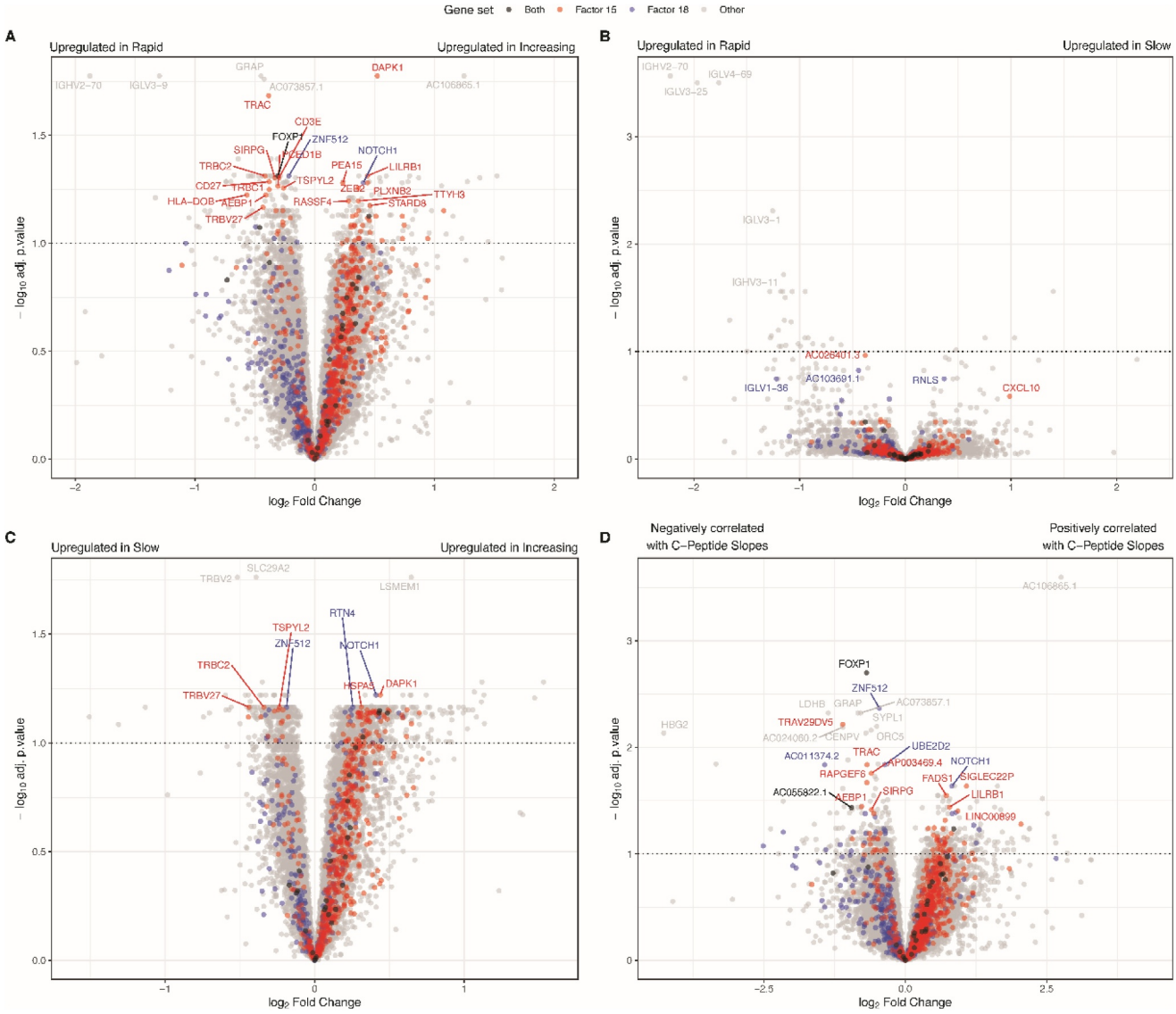

FIGURE 3 Volcano plot of differential gene expression between progression groups. The colour indicates gene membership to either or both associated latent factors. (A) DGE between rapid and increasing progression groups (339 genes are differentially expressed). (B) DGE between rapid and slow progression groups (33 genes are differentially expressed). (C) DGE between slow and increasing progression groups (1206 genes are differentially expressed). (D) DGE for C-peptide slopes (484 genes are differentially expressed).

3.4 | Annotation of latent factors

To examine the biological pathways of the integrated representation in the two most relevant latent factors, we used GSEA. This analysis was divided into genes positively correlated with the factor score (positive weights) and genes negatively correlated with the factor value (negative weights). Figure 4 displays the top 15 significant pathways in each of the two associated latent factors (p-values <0.1 adjusted by Benjamini-Hochberg). Negatively regulated genes in latent factor 15 are enriched in the immune system and signalling by G protein-coupled receptors (GPCR) pathways. Of specific interest are pathways associated with innate immunity, such as neutrophil degranulation, with high expression of granule proteins (e.g. CTSs, MPO) pointing to the presence of activated or degranulated neutrophils; and platelet activation, signalling and aggregation (the latter not shown). Also, several pathways pointing to cytokine signalling and interleukin (e.g. IL-1β) signalling emerge. Regarding the role of GPCRs, several pathways are associated with latent factor 15, such as signalling by GPCRs, downstream signalling and GPCR ligand binding. Positively regulated genes in latent factor 15 are also enriched in immune system pathways, with again an important contribution of the innate immune system. Although here it seems that the increase is mainly attributed to resting neutrophils, with for instance high LY96 expression. Collectively, these data show that GSEA pathways in innate immunity are mainly associated with the activation status of the neutrophils, with a shift in the balance of resting neutrophils versus activated or degranulated neutrophils.

Negatively regulated genes in latent factor 18 are enriched in influenza infection pathways and mRNA translation pathways, suggesting that viral mRNA replication and translation by host cell

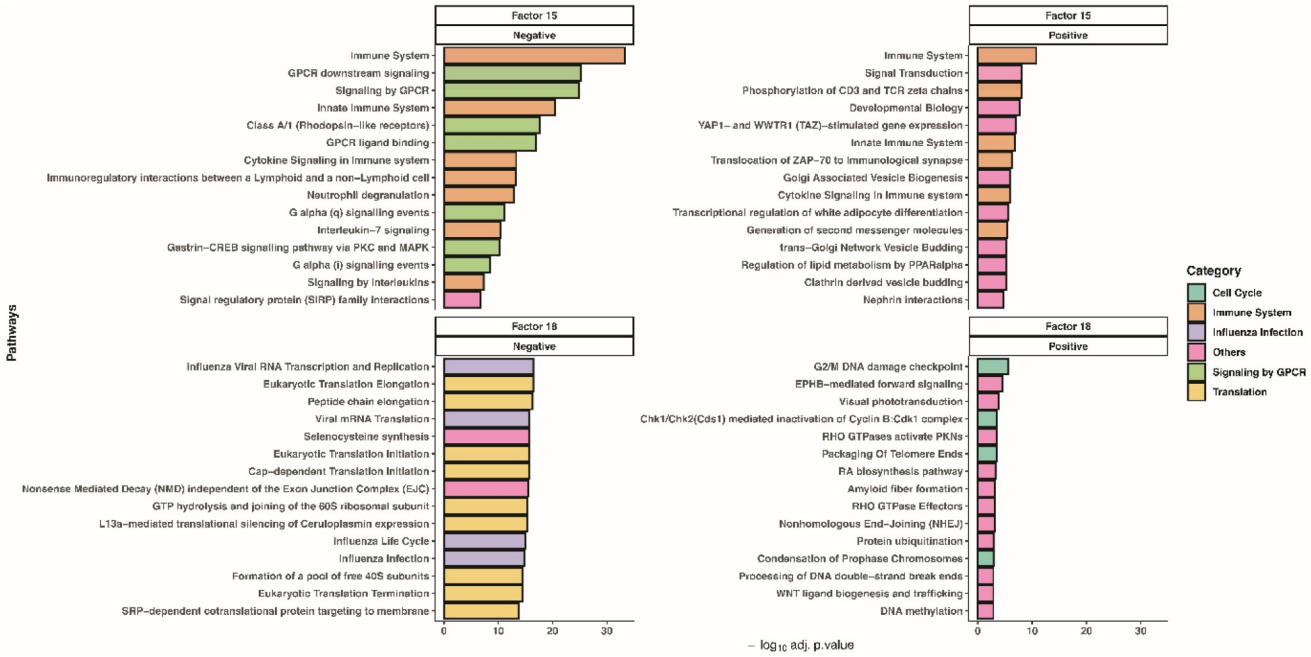

FIGURE 4 GSEA was performed separately for genes with positive weights in the latent factor (correlating negatively with C-peptide slopes) and genes with negative weights in the latent factor (correlating positively with C-peptide slopes). Pathways are coloured depending on the main pathway they belong to according to Reactome. *p*-values were adjusted using Benjamini-Hochberg.

machinery are major pathological features of this association. Positively regulated genes in latent factor 18 did not show a particular pattern of enrichment.

Furthermore, the relation of the latent factors' genes to previous type 1 diabetes publications was studied. Using the Open Targets database,[25] a total of 174 genes out of the 668 top genes (three-fold higher than expected by chance) in the latent factor 15 have been previously associated with type 1 diabetes (*p*-value 0.03). On the other hand, the overlap between the top genes in the latent factor 18 and associated disease genes was not significant. These results indicate that latent factor 15 captures a set of genes composed of known disease targets (e.g. INSR, NDUFA4, CTSH) as well as potential new candidate genes already detectable in blood in the early phase of type 1 diabetes.

3.5 | Interpretation of biological networks

Biological networks were constructed for latent factors 15 and 18 separately, based on protein-protein or protein-protein-miRNA interactions (Supplementary Figures S5–S8). As the two network types yielded similar results, we selected the protein-protein-miRNA networks as the focus of our interpretations. The latent factor 15 network revealed a diverse set of biological functions (Supplementary Figure S7), some of which overlapped with the pathways shown in Figure 4. Immune system responses, signalling by different receptors and lipid metabolism are (widely) represented in these clusters. Notably there is an enrichment of lipid metabolism pathways as the lipidomic data is also influenced latent factor 15. Some of the genes that were significantly associated with the C-peptide slopes also appeared in several of the clusters, which further validated the biological processes captured by the latent factor. The latent factor 18 network was smaller and had more loosely defined clusters (Supplementary Figure S8). Nonetheless, it captured a similar set of biological functions compared to the latent factor 15 network. Eukaryotic/viral mRNA translation is the main difference between the two networks, which appear as the biggest and more interconnected cluster of the latent factor 18 network. In this case, only one of the genes in this network was significantly associated with the C-peptide slopes.

3.6 | Immunomics signature

Immune cell populations were identified based on standard markers and analysed for their association with the C-peptide slopes. Natural Killer (NK) cells were found to be significantly associated after *p*-value adjustment (Figure 5A), with higher levels of NK cells observed in people with slow disease progression (Figure 5B). Examination of the relationship between C-peptide slope and NK cell frequency in individual progression groups indicated that the strongest correlation was observed among rapid progressors (Figure 5C). Unsupervised analysis of the immunomics data revealed distinct clusters among the progression groups. Figure 5D shows a FlowSOM colour-density map of CD16 expression levels with node sizes representing the frequency of cells in each cluster. Meta-cluster 19 (MC-19) was assigned as the primary NK (CD56loCD16+) subset based on lineage marker expression and was significantly more abundant in the

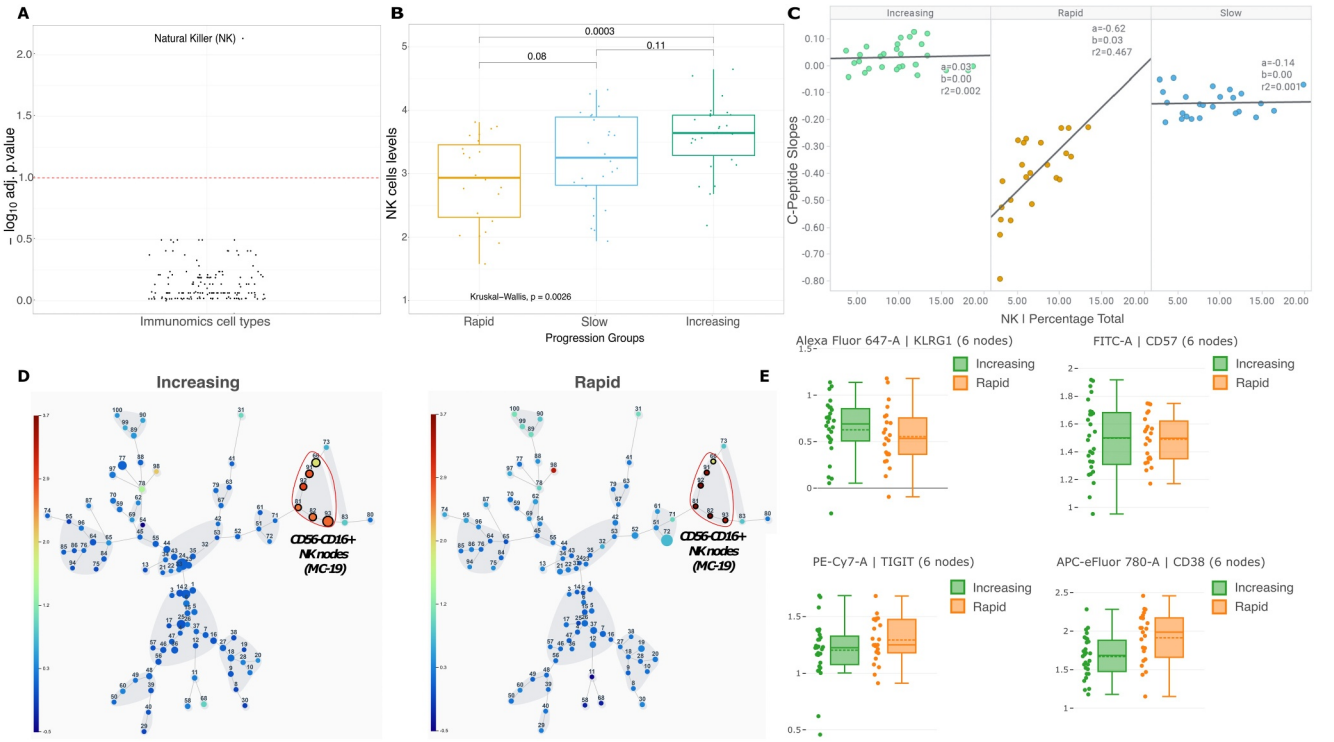

FIGURE 5 Immunomics association with C-peptide slopes. (A) Benjamini-Hochberg adjusted *p*-value for the linear association of immune cells abundances with the C-peptide slopes (correcting for batch effects and age groups). (B) NK cell levels for the different progression groups (as a frequency of total live mononuclear cells). (C) Spearman correlation of fasted C-peptide slopes vs. NK cell frequency for the different progression groups. (D) FlowSOM unbiased cluster analysis on live $CD45^+$ PBMCs of Donor D07 (Increasing, left) and K40 (Rapid, right) highlighting Metacluster-19 (red-circle) as a primary NK ($CD56^{lo}CD16^+$) subset. Colour-density and circle size were overlaid as CD16 intensity and number of events, respectively. (E) Bar charts representing NK marker expression on the selected clusters in each progression group.

increasing versus rapid progression groups (8.8 vs. 6.3% *p*-value = 0.013). Examination of markers of NK cell activation and differentiation (KLRG1, TIGIT, CD38, and CD57) in the different progression groups showed higher expression of KLRG1 in the increasing group but lower levels of CD38.

Comparison of the genes associated with latent factor 15 and the leucocyte gene signature matrix LM22[26] revealed an overlap of 52 genes and enrichment in immune cell-specific genes (*p*-value 0.0001). The genes were representative of the following five main groups: T-cell specific, macrophage M1 specific, monocyte specific, neutrophil specific, and eosinophil specific (Supplementary Figure S9). In this way, in addition to capturing immune-related pathways, the latent factor 15 broadly represented immune cell-specific genes.

3.7 | Other 'omics associations with disease progression

The association of factors derived from running MOFA without the transcriptomics data did not yield any significant associations with the progression groups or the C-peptide slopes. The variance of the miRNA and lipidomics captured by latent factor 15 might indicate that these two omics data sets are only informative of the disease progression in combination with transcriptomics. This shows how the integration of both lipidomics and transcriptomics data may aid in discovering enrichment in specific pathways. Similarly, the incorporation of transcriptomics and miRNA data allowed the discovery of enrichment in viral infection pathways.

4 | DISCUSSION

In this study, we identified two latent factors associated with β-cell decline. These factors were predominantly influenced by transcriptomics with secondary contributions of miRNA and lipidomics, respectively. Latent factor 15 revealed an enrichment of immune system pathways, the most significant being associated with neutrophil degranulation, cytokine signalling, and immunoregulatory interactions between lymphoid and non-lymphoid cells. Moreover, there were multiple pathways associated with GPCR signalling events. More detailed GSEA revealed that disease progression (C-peptide slopes) was associated with an altered balance between resting and activated/degranulated neutrophils. This is in keeping with previous studies that showed a temporary decline in the number of circulating neutrophils in people with newly diagnosed type 1 diabetes compared with healthy controls, as well as high circulating

levels of neutrophil extracellular traps (NETs).[27–29] The previous detection of neutrophils and NETs in the pancreas of deceased subjects affected by type 1 diabetes, and a correlation between circulating neutrophil numbers and C-peptide levels in pre-symptomatic subjects (non-diabetic, at-risk) has implicated that activated neutrophils play a pathogenic role in type 1 diabetes.[30] Our data add significantly to this hypothesis since we showed in longitudinal follow-up that the neutrophil profile at diagnosis is associated with the rate of disease progression.

Of further interest, GSEA shows platelet activation to be a feature of latent factor 15, linking our findings to the recent demonstration of a role for activated platelets in the formation of platelet-neutrophil aggregates, which are increased in the circulation of subjects during the development of type 1 diabetes.[31] It is tempting to speculate that activation of GPCR pathways (also associated with latent factor 15) may play a role in these events since it is a response to a variety of stimuli (chemokines, cytokines, complement fragments) and can trigger neutrophil degranulation.[32]

In contrast, latent factor 18, which was also associated with β-cell decline, is characterised by features of enrichment of viral mRNA translation and subsequent translation by the host cell machinery. There is a considerable body of literature that associates viral infections with early events in type 1 diabetes development as well as peri-diagnosis events.[33] As a result, virus infection has often been cited as an autoimmunity-triggering event as well as a disease-precipitating event. Our findings in the context of the present study design are entirely consistent with the latter hypothesis, which could be addressed in follow-up viromic studies targeted to samples in which both the relevant viral mRNA translation signals and negative slope of C-peptide decline are prominent.

Two recent single-omics papers have been published on the INNODIA cohort[34,35] with a considerable overlap of patients to our study. There are small differences in the inclusion criteria of patients. There are substantial differences in methodology that complicate a direct comparison. We use gene expression levels at baseline that do most often not carry the same information as changes in gene transcription over time (which is used in the single-omics studies). The temporal expression data can be strongly influenced by for example, treatment. We focus on signatures available already at the baseline level, as this would have a great impact on treatment strategies and is easier to incorporate in trial designs and stratification of patients. The proteomics study[34] found 12 significant associations between peptide expression changes in the initial 12 months after diagnosis and C-peptide glucose ratio slopes, with expression changes in GPX3 being indicative of future C-peptide changes. Because of the cohort and methodological differences, a direct comparison is not possible. Instead, we ranked peptides from the proteomics study by their maximum importance in factor 15 or 18 and saw an enrichment of signature genes in the most important genes with GPX3 having the 6th highest loading (Supplementary Fig. 10). The transcriptomics study[35] also investigated changes in gene expression in the 12 months after diagnosis and associations to C-peptide glucose ratio slopes and identified 392 significant genes and of those we had 360 included in the multi-omics model. We only found a small overlap of 8 genes between DEGs in the multi-omics and transcriptomics studies. The lack of overlap could be explained by the different temporal and non-temporal data. Genes that are significantly different at baseline are not necessarily the same as those found significant in a temporal single-omics analysis, and vice versa. However, both the transcriptomics study and the multi-omics analysis highlight pathways from the immune system and especially neutrophils as a driving signal for C-peptide loss. Since the untargeted transcriptomics data may have identified different but correlated gene expressions, looking at the pathways involved may be a more relevant and mechanistic comparison of the single-omics and multi-omics results.

The main difference between transcriptomics and the other omics data was that it was determined from whole blood. The remaining omics were collected either from serum, plasma, or PBMCs. Even though some analytes may give similar results, samples obtained from the same medium are more easily comparable. Therefore, the disparity between omics that we observe in the latent factors and in the linear association of each omics data set might be caused by the source medium. This could be considered both a drawback and an advantage. On one hand, it is undesirable that this disparity exists because correlated analytes across omics cannot be studied. This makes the data integration challenging as we cannot observe the joint effect of multiple omics nor validate whether the analytes associated with the disease progression in one data type can also be observed in a different one. On the other hand, the heterogeneity across omics can be seen as complementary information. Each omics data set captures a different source of variation, thus providing additional information not captured by the other omics types. Based on the current data, we cannot conclude whether the plasma and serum omics were not associated with the disease progression due to the medium or the analytes themselves.

Molecular and cellular signatures of adaptive immune responses were by and large not observed to be associated with β-cell decline in our study, which might at first sight be considered a surprise, given the strong credentials, at genomic, pathological, and mechanistic levels, for type 1 diabetes being the archetype of an organ-specific autoimmune disease. However, it is entirely plausible that the detection of such associations is challenging in whole blood or whole mononuclear cell analyses, both because the disease-relevant, β-cell antigen-specific lymphocytes are rare, and even more importantly, because they may be sequestered at inflammatory sites. Certainly, smaller scale studies focused on using appropriately sensitive technologies have identified that the activation and differentiation state of circulating antigen-specific cytotoxic T lymphocytes, for example, correlate with changes in β-cell function after the diagnosis of type 1 diabetes.[10] Among the lymphocyte studies presented here, our observation of a prominent NK cell signal related to rapid β-cell decline is of considerable interest. NK cells have features of both innate and adaptive immune cells and play a key role in anti-viral responses. Both pro-inflammatory and regulatory functions have been ascribed to these cells, and functional subtypes can be partially differentiated by surface markers. Previous studies examining the

frequency of NK cells in individuals affected by type 1 diabetes have consistently reported lower circulating levels of both proinflammatory and regulatory NK cells when compared to aged-matched non-diabetic subjects,[36–38] potentially reflecting homing to inflammatory sites in the pancreas. Consistent with this, we observed lower circulating levels of NK cells (both effector and regulatory subtypes according to surface markers) in the rapid decline group. Of interest, reduced circulating NK cell levels are also associated with viral infection, linking this observation to the viral signatures already described. Future functional studies will be required to explore the pathological implications of these findings since the immune phenotyping performed here was limited to the expression of CD38 (NK cell activation) and KLRG1 (an inhibitory receptor associated with an exhausted phenotype).

Key strengths of the study include (i) the setting of a large longitudinal natural history study conducted across multiple European sites according to standardized clinical and laboratory protocols; (ii) access through the INNODIA network to highly specialised, systems-based technology platforms for parallel multi-parametric analysis; (iii) leverage of new tools in integrated MOFA to discover signatures that are significantly associated with rate of disease progression for the year following diagnosis. This period of the disease is important since it represents the phase during which disease-modifying immunotherapies are typically trialed for their effect on arresting β-cell decline. Factors identifiable at baseline that are associated with faster β-cell decline could be used to conduct shorter and smaller trials (e.g. by enrollment of a rapid-decline group), an important step towards de-risking the investment needed to bring disease modifying strategies into clinical use.

Our INNODIA natural history collection involves individuals from the ages of one up to 45 years, and thus includes the whole lifespan of people living with type 1 diabetes. Although this may render interpretation of findings in this clearly heterogenic disease more complex, it may add to the identification of common factors that drive disease, irrespective of age. We used fasting C-peptide as read-out for β-cell function rather than stimulated C-peptide due to missingness, but we demonstrated similar trends in decline of function using this simple parameter. Although collected throughout Europe, the population studied is almost completely white Caucasian, thus limiting the generalisability of the findings to a global population, where type 1 diabetes is becoming more prevalent in non-Caucasian people. Confirmation of our observations will therefore be needed in more diverse cohorts. Our work reports on a small cohort of just under 100 people with newly diagnosed disease, and the power to detect existing associations could be limited, notably when the population is further divided into progression groups. However, we demonstrate that even in such small numbers, using deeply phenotyped individuals and standardized operating procedures, application of systems biology techniques can lead to significant associations with top ranking pathways consistent with single-omics studies. Here we believe that the collaboration between academics, foundations, industry and people affected by type 1 diabetes and their families within INNODIA was a unique driver. We created strict protocols for follow up where people could be convinced to participate with support of materials created by the PAC (People with diabetes Advisory Committee), we set up a highly standardized sample collection (e.g. even standardising the pipet tips for miRNA sample collection) and applied homogeneous laboratory procedures. Importantly, we brought all data into a GDPR-conforming centralised database, allowing clean data collection and high-quality data for input into the analysis.

In summary, the presented study addressed the drivers of disease heterogeneity in type 1 diabetes by leveraging opportunities presented by a highly integrated clinical network featuring embedded research platforms with the capability to generate large systems-level datasets. One of the two factors identified showed correlation to neutrophil degranulation, cytokine signalling, lymphoid and non-lymphoid cell interactions and G-protein coupled receptor signalling events, while the second signature, pathways related to translation and viral infection, were inversely associated with change in β-cell function. The derived latent factors were used to identify specific signatures, which were further investigated for biomarker opportunities and mechanistic pathways that correlate with β-cell decline. Further investigation and validation of these multi-omics signatures could aid in identifying rapid and slow progressors around diagnosis, which can be utilised for designing better stratified trials of future disease-modifying therapies.

AUTHOR CONTRIBUTIONS

JJAA, CB, SFAB, DBD, PJC, MK, AMS, GS, LLE, FD, TT, RL, LO, CM, MP, SB designed the study. JJAA, CB, CHJ, KB, RM, NA-S, IM, CL-Q, TS, JHL, GEG, MKH, TS, OR, PJC, JHL, GS, TT collected data and performed analyses. JJAA, CHJ, TT, LO, CM, MP, SB wrote the manuscript. All authors revised the manuscript for crucial content and approved the final version. All authors had full access to all the data and had final responsibility for the decision to submit for publication. A complete list of the partners of the INNODIA consortium is included separately.

AFFILIATIONS

[1]Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

[2]Turku Bioscience Centre, University of Turku and Åbo Akademi University, Turku, Finland

[3]InFLAMES Research Flagship Center, University of Turku, Turku, Finland

[4]Department of Paediatrics, University of Cambridge, Cambridge, England

[5]Steno Diabetes Center Copenhagen, Systems Medicine, Herlev, Denmark

[6]Research Program for Clinical and Molecular Metabolism, Faculty of Medicine, University of Helsinki, Helsinki, Finland

[7]Pediatric Research Center, New Children's Hospital, Helsinki University Hospital, Helsinki, Finland

[8]Bayer Pharmaceuticals, Berlin, Germany

[9]Immunology & Inflammation Research Therapeutic Area, Sanofi, Massachusetts, USA

[10]Department of Medicine, Surgery and Neuroscience, Università degli Studi di Siena, Siena, Italy

[11]Fondazione Umberto di Mario, ONLUS – Toscana Life Sciences, Siena, Italy

[12]Institute of Biomedicine, University of Turku, Turku, Finland

[13]Steno Diabetes Center Copenhagen, Herlev University Hospital, Herlev, Denmark

[14]Tuscany Centre for Precision Medicine (CReMeP), Siena, Italy

[15]Department of Immunobiology, King's College, London, UK

[16]Department of Chronic Diseases and Metabolism, Endocrinology, Katholieke Universiteit Leuven, Leuven, Belgium

ACKNOWLEDGEMENTS

We gratefully acknowledge all participants of the INNODIA natural history study, which built the basis for the work presented here. Without the outstanding engagement of the people with type diabetes and their relatives and friends participating in the INNODIA clinical efforts, this data and new knowledge gain would not have been possible. The authors also want to thank all the clinical personnel for their dedication in the participant recruitment, characterization, sample collection and preparation that was at the basis of this analysis. We dedicate this work to the late Professor David Dunger, who inspired INNODIA and laid the basis of this analysis. This work is funded by the Innovative Medicine Initiative 2 Joint Undertaking (IMI2 JU) under grant agreement N° 115797 (INNODIA) and N° 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'.

CONFLICT OF INTEREST STATEMENT

CM serves or has served on the advisory panel for Novo Nordisk, Sanofi, Merck Sharp and Dohme Ltd., Eli Lilly and Company, Novartis, AstraZeneca, Boehringer Ingelheim, Roche, Medtronic, ActoBio Therapeutics, Pfizer, Imcyse, Insulet, Zealand Pharma, Avotres, Mannkind, Sandoz and Vertex. Financial compensation for these activities has been received by KU Leuven; KU Leuven has received research support for CM from Medtronic, Imcyse, Novo Nordisk, Sanofi and ActoBio Therapeutics; CM serves or has served on the speakers bureau for Novo Nordisk, Sanofi, Eli Lilly and Company, Boehringer Ingelheim, Astra Zeneca and Novartis. Financial compensation for these activities has been received by KU Leuven. S.Br. reports having received funding from INNODIA (grant agreement No 115797), having ownerships in Intomics A/S, Hoba Therapeutics Aps, Novo Nordisk A/S, Lundbeck A/S, ALK A/S and managing board memberships in Proscion A/S and Intomics A/S. MK reports ownership and managing board membership in Vactech Oy. CLQ serves on advisory boards of Fondation Alzheimer's and Institute Pasteur, and has an affiliation with Novo Nordisk. LO reports having received funding from INNODIA and INNODIA Harvest. CHJ reports having received funding from INNODIA (grant agreement No 115797).

CODE AVAILABILITY

The primary software used for this work was Multi-Omics Factor Analysis (MOFA). The code for this software is available at https://github.com/bioFAM/MOFA2.

ETHICS STATEMENT

The INNODIA study protocol was approved by the London – City & East Research Ethics Committee on 28 October 2016 (REC 16/LO/1750) IRAS Project ID 210497. Subsequently, after translation of the participants' documentation, approval was obtained from local Ethic authorities throughout the entire INNODIA clinical network.

DATA AVAILABILITY STATEMENT

The data generated and analysed is person-sensitive as it can be used to identify people based on their sequence variation and can be accessed only in secure environments. Access to data can be provided by application to the INNODIA Data Access Committee by emailing Professor Lut Overbergh (lutgart.overbergh@kuleuven.be). Processed results of GSEA analysis are available as supplementary material.

ORCID

Christian Holm Johansen https://orcid.org/0000-0001-8665-2111
Robert Moulder https://orcid.org/0000-0003-4742-0566
Karoliina Hirvonen https://orcid.org/0000-0003-1575-9725
Emile Hendricks https://orcid.org/0000-0002-0795-1832
Guido Sebastiani https://orcid.org/0000-0003-4374-8564
Søren Brunak https://orcid.org/0000-0003-0316-5866

TRANSPARENT PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1002/dmrr.3833.

REFERENCES

1. Warshauer JT, Bluestone JA, Anderson MS. New frontiers in the treatment of type 1 diabetes. *Cell Metab*. 2020;31(1):46-61. https://doi.org/10.1016/j.cmet.2019.11.017
2. International Diabetes Federation. IDF Diabetes Atlas. 10th ed.; 2021. Published online.
3. Miller RG, Orchard TJ. Understanding metabolic memory: a tale of two studies. *Diabetes*. 2020;69(3):291-299. https://doi.org/10.2337/db19-0514
4. Rawshani A, Sattar N, Franzén S, et al. Excess mortality and cardiovascular disease in young adults with type 1 diabetes in relation to age at onset: a nationwide, register-based cohort study. *Lancet*. 2018;392(10146):477-486. https://doi.org/10.1016/S0140-6736(18)31506-X
5. Tatovic D, Dayan CM. Replacing insulin with immunotherapy: time for a paradigm change in Type 1 diabetes. *Diabet Med J Br Diabet Assoc*. 2021;38(12):e14696. https://doi.org/10.1111/dme.14696
6. Battaglia M, Ahmed S, Anderson MS, et al. Introducing the endotype concept to address the challenge of disease heterogeneity in type 1 diabetes. *Diabetes Care*. 2020;43(1):5-12. https://doi.org/10.2337/dc19-0880
7. Weston CS, Boehm BO, Pozzilli P. Type 1 diabetes: a new vision of the disease based on endotypes. *Diabetes Metab Res Rev*. 2024;40(2):e3770. https://doi.org/10.1002/dmrr.3770
8. Limonte CP, Valo E, Montemayor D, et al. A targeted multiomics approach to identify biomarkers associated with rapid eGFR decline in type 1 diabetes. *Am J Nephrol*. 2020;51(10):839-848. https://doi.org/10.1159/000510830

9. Speake C, Skinner SO, Berel D, et al. A composite immune signature parallels disease progression across T1D subjects. *JCI Insight*. 2019; 4(23). https://doi.org/10.1172/jci.insight.126917
10. Yeo L, Woodwyk A, Sood S, et al. Autoreactive T effector memory differentiation mirrors β cell function in type 1 diabetes. *J Clin Invest*. 2018;128(8):3460-3474. https://doi.org/10.1172/JCI120555
11. Dunger DB, Bruggraber SFA, Mander AP, et al. INNODIA Master Protocol for the evaluation of investigational medicinal products in children, adolescents and adults with newly diagnosed type 1 diabetes. *Trials*. 2022;23(1):414. https://doi.org/10.1186/s13063-022-06259-z
12. Besser REJ, Shields BM, Casas R, Hattersley AT, Ludvigsson J. Lessons from the mixed-meal tolerance test. *Diabetes Care*. 2013;36(2): 195-201. https://doi.org/10.2337/dc12-0836
13. Maddaloni E, Bolli GB, Frier BM, et al. C-peptide determination in the diagnosis of type of diabetes and its management: a clinical perspective. *Diabetes Obes Metab*. 2022;24(10):1912-1926. https://doi.org/10.1111/dom.14785
14. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J Stat Softw*. 2015;67:1-48. https://doi.org/10.18637/jss.v067.i01
15. Knip M, Virtanen SM, Seppä K, et al. Dietary intervention in infancy and later signs of beta-cell autoimmunity. *N Engl J Med*. 2010; 363(20):1900-1908. https://doi.org/10.1056/NEJMoa1004809
16. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014; 15(12):550. https://doi.org/10.1186/s13059-014-0550-8
17. Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets *Mol Syst Biol*. 2023;Accessed November 30, 2023. https://www.embopress.org/doi/full/10.15252/msb.20178124
18. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles | *Proc Natl Acad Sci USA*. 2023;Accessed November 30, 2023. https://www.pnas.org/doi/full/10.1073/pnas.0506580102
19. Frost HR, Li Z, Moore JH. Principal component gene set enrichment (PCGSE). *BioData Min*. 2015;8(1):25. https://doi.org/10.1186/s13040-015-0059-z
20. Szklarczyk D, Gable AL, Nastou KC, et al. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res*. 2021;49(D1):D605-D612. https://doi.org/10.1093/nar/gkaa1074
21. Huang HY, Lin YCD, Cui S, et al. miRTarBase update 2022: an informative resource for experimentally validated miRNA–target interactions. *Nucleic Acids Res*. 2022;50(D1):D222-D230. https://doi.org/10.1093/nar/gkab1079
22. Ismailova A, Salehi-Tabar R, Dimitrov V, Memari B, Barbier C, White JH. Identification of a forkhead box protein transcriptional network induced in human neutrophils in response to inflammatory stimuli. *Front Immunol*. 2023;14:1123344. https://doi.org/10.3389/fimmu.2023.1123344
23. Zou Y, Gong N, Cui Y, et al. Forkhead box P1 (FOXP1) transcription factor regulates hepatic glucose homeostasis. *J Biol Chem*. 2015;290(51):30607-30615. https://doi.org/10.1074/jbc.M115.681627
24. Dupuis J, Langenberg C, Prokopenko I, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet*. 2010;42(2):105-116. https://doi.org/10.1038/ng.520
25. Koscielny G, An P, Carvalho-Silva D, et al. Open Targets: a platform for therapeutic target identification and validation. *Nucleic Acids Res*. 2017;45(Database issue):D985-D994. https://doi.org/10.1093/nar/gkw1055
26. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling tumor infiltrating immune cells with CIBERSORT. In: vonStechow L, ed. *Cancer Systems Biology: Methods and Protocols. Methods in Molecular Biology*. Springer; 2018:243-259. https://doi.org/10.1007/978-1-4939-7493-1_12
27. Harsunen MH, Puff R, D'Orlando O, et al. Reduced blood leukocyte and neutrophil numbers in the pathogenesis of type 1 diabetes. *Horm Metab Res Horm Stoffwechselforschung Horm Metab*. 2013;45(6):467-470. https://doi.org/10.1055/s-0032-1331226
28. Valle A, Giamporcaro GM, Scavini M, et al. Reduction of circulating neutrophils precedes and accompanies type 1 diabetes. *Diabetes*. 2013;62(6):2072-2077. https://doi.org/10.2337/db12-1345
29. Wang Y, Xiao Y, Zhong L, et al. Increased neutrophil elastase and proteinase 3 and augmented NETosis are closely associated with β-cell autoimmunity in patients with type 1 diabetes. *Diabetes*. 2014; 63(12):4239-4248. https://doi.org/10.2337/db14-0480
30. Vecchio F, Buono NL, Stabilini A, et al. Abnormal neutrophil signature in the blood and pancreas of presymptomatic and symptomatic type 1 diabetes. *JCI Insight*. 2018;3(18). https://doi.org/10.1172/jci.insight.122146
31. Popp SK, Vecchio F, Brown DJ, et al. Circulating platelet-neutrophil aggregates characterize the development of type 1 diabetes in humans and NOD mice. *JCI Insight*. 2022;7(2). https://doi.org/10.1172/jci.insight.153993
32. *Roles of Neutrophil Granule Proteins in Orchestrating Inflammation and Immunity - Othman - 2022 - the FEBS Journal* - Wiley Online Library. Accessed November 30, 2023. https://febs.onlinelibrary.wiley.com/doi/full/10.1111/febs.15803
33. Vehik K, Lynch KF, Wong MC, et al. Prospective virome analyses in young children at increased genetic risk for type 1 diabetes. *Nat Med*. 2019;25(12):1865-1872. https://doi.org/10.1038/s41591-019-0667-0
34. Moulder R, Välikangas T, Hirvonen MK, et al. Targeted serum proteomics of longitudinal samples from newly diagnosed youth with type 1 diabetes distinguishes markers of disease and C-peptide trajectory. *Diabetologia*. 2023;66(11):1983-1996. https://doi.org/10.1007/s00125-023-05974-9
35. Suomi T, Starskaia I, Kalim UU, et al. Gene expression signature predicts rate of type 1 diabetes progression. *EBioMedicine*. 2023;92: 104625. https://doi.org/10.1016/j.ebiom.2023.104625
36. Wilson RG, Anderson J, Shenton BK, White MD, Taylor RM, Proud G. Natural killer cells in insulin dependent diabetes mellitus. *Br Med J Clin Res Ed*. 1986;293(6541):244. https://doi.org/10.1136/bmj.293.6541.244
37. Fitas AL, Martins C, Borrego LM, et al. Immune cell and cytokine patterns in children with type 1 diabetes mellitus undergoing a remission phase: a longitudinal study. *Pediatr Diabetes*. 2018;19(5): 963-971. https://doi.org/10.1111/pedi.12671
38. Oras A, Peet A, Giese T, Tillmann V, Uibo R. A study of 51 subtypes of peripheral blood immune cells in newly diagnosed young type 1 diabetes patients. *Clin Exp Immunol*. 2019;198(1):57-70. https://doi.org/10.1111/cei.13332

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Armenteros JJA, Brorsson C, Johansen CH, et al. Multi-omics analysis reveals drivers of loss of β-cell function after newly diagnosed autoimmune type 1 diabetes: An INNODIA multicenter study. *Diabetes Metab Res Rev*. 2024;e3833. https://doi.org/10.1002/dmrr.3833