

Transcriptomics of type 1 diabetes progression: a validation study in newly diagnosed patients



Tomi Suominen,^{a,b,m} Inna Starskaia,^{a,b,m} Omid Rasool,^{a,b} Ubaid Ullah Kalim,^{a,b} Sylvaine Brugger,^c M. Loredana Marcovecchio,^c Emile Hendricks,^c Lut Overbergh,^d Mark Peakman,^{e,f} Timothy Tree,^f Søren Brunak,^g Anke M. Schulte,^h Chantal Mathieu,^d Mikael Knip,^{i,j,k} Riitta Lahesmaa,^{a,b,l,*} and Laura L. Elo,^{a,b,l,**} On behalf of the INNODIA Consortium



^aTurku Bioscience Centre, University of Turku and Åbo Akademi University, FI-20520, Turku, Finland

^bInFLAMES Research Flagship Center, University of Turku, Turku, Finland

^cDepartment of Paediatrics, University of Cambridge, Cambridge, UK

^dDepartment of Chronic Diseases and Metabolism, Endocrinology, Katholieke Universiteit Leuven, Leuven, Belgium

^eImmunology & Inflammation Research Therapeutic Area, Sanofi, MA, USA

^fDepartment of Immunobiology, King's College, London, UK

^gNovo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

^hSanofi-Aventis Deutschland GmbH, Frankfurt, Germany

ⁱPaediatric Research Centre, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

^jResearch Program for Clinical and Molecular Metabolism, Faculty of Medicine, University of Helsinki, Helsinki, Finland

^kTampere Centre for Child Health Research, Tampere University Hospital, Tampere, Finland

^lInstitute of Biomedicine, University of Turku, FI-20520, Turku, Finland

Summary

Background Type 1 diabetes is an autoimmune disease with significant long-term complications. Variability in the decline of insulin secretion after diagnosis complicates both the development of treatments and disease management. We previously reported that gene expression changes within the first year post-diagnosis were associated with C-peptide decline at two years in the first INNODIA cohort of patients with newly diagnosed type 1 diabetes. Here, we aimed to validate these findings in an independent follow-up cohort and to increase statistical power by combining the data from both cohorts.

Methods We analysed transcriptomic data from a follow-up INNODIA cohort of 168 individuals with newly diagnosed type 1 diabetes to assess whether previously identified associations with disease progression could be replicated. We then combined data from the original and follow-up cohorts for integrated analysis. Longitudinal gene expression changes during the first year after diagnosis were examined in relation to disease progression, alongside age and estimated immune cell abundances.

Findings Analysis of the follow-up cohort validated the previously observed longitudinal changes in gene expression during the first year after diagnosis. In the combined dataset, transcriptomic analysis identified a large number of genes that were differentially expressed during the first year after disease onset. More rapid disease progression was associated with younger age and a relative decrease in neutrophil abundance. In addition, changes in the expression of several genes were associated with the rate of disease progression.

Interpretation These findings support the existence of biological heterogeneity in disease progression after diagnosis of type 1 diabetes and contribute to an improved understanding of the molecular dynamics associated with disease progression. These findings may help future studies aiming to enable patient stratification and design of more targeted and personalised therapeutic approaches in type 1 diabetes.

Funding This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 115797 (INNODIA) and No 945268 (INNODIA HARVEST).

Copyright © 2026 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

*Corresponding author. Turku Bioscience Centre, University of Turku and Åbo Akademi University, FI-20520, Turku, Finland.

**Corresponding author. Turku Bioscience Centre, University of Turku and Åbo Akademi University, FI-20520, Turku, Finland.

E-mail addresses: riitta.laheesmaa@utu.fi (R. Laheesmaa), laura.elo@utu.fi (L.L. Elo).

^{††}Shared first authors.

eBioMedicine
2026;130: 106374
Published Online xxx
<https://doi.org/10.1016/j.ebiom.2026.106374>

Keywords: Type 1 diabetes; Transcriptomics; Beta cell decline; C-peptide; Disease progression

Research in context

Evidence before this study

Type 1 diabetes is characterised by autoimmune destruction of pancreatic β -cells, but the rate of post-diagnosis decline in insulin secretion varies between individuals. Circulating C-peptide reflects β -cell function, and several studies have demonstrated heterogeneity in C-peptide decline after disease onset. Transcriptomic profiling of peripheral blood has shown promise in identifying immune-related changes in type 1 diabetes, including interferon-associated signatures, yet many studies have been limited by small sample sizes, cross-sectional designs, or lack of validation. As part of the INNODIA consortium, we previously reported changes in peripheral blood gene expression during the first year after diagnosis in a cohort of 100 newly diagnosed patients, including associations with the rate of C-peptide decline at two years. However, validation in a larger cohort was required to establish the robustness and generalisability of these findings.

Added value of this study

This study provides validation of previously identified progression-associated transcriptomic changes in a follow-up cohort of 168 newly diagnosed patients with type 1 diabetes from the INNODIA study. By combining transcriptomic data from two cohorts, we substantially increased statistical power and enabled a more comprehensive characterisation

of gene expression changes occurring during the first year after disease onset. We confirmed consistent expression patterns for the top genes previously associated with disease progression and identified numerous additional differentially expressed genes, including down-regulation of interferon-related pathways. We also identified associations between more rapid disease progression and younger age, reduced relative neutrophil abundance, and changes in the expression of several specific genes. These findings strengthen the evidence that early post-diagnosis blood transcriptomic changes reflect underlying disease heterogeneity and are associated with subsequent loss of β -cell function.

Implications of all the available evidence

The combined evidence indicates that the early period after type 1 diabetes diagnosis is characterised by dynamic changes in peripheral blood gene expression that are associated with the rate of disease progression. Blood-based transcriptomic signatures, particularly when validated across independent cohorts, have the potential to support patient stratification and risk prediction in individuals with newly diagnosed type 1 diabetes. These insights may inform the design of more efficient clinical trials and could contribute to the development of personalised therapeutic approaches aimed at preserving residual β -cell function.

Introduction

Type 1 diabetes (T1D) is one of the most common chronic autoimmune diseases in children. It is characterised by immune-mediated destruction of insulin-producing pancreatic β -cells, leading to dependence on exogenous insulin. Despite advances in disease management, T1D is associated with severe long-term complications and presents a significant burden to the individuals living with T1D, their families, and health care systems.

Accumulating evidence indicates considerable heterogeneity in the rate of post-onset insulin secretion decline, which can be measured using circulating C-peptide as a marker of residual β -cell function. Heterogeneity in T1D is evident already during preclinical stages and across age at onset, immunopathology, residual insulin secretory capacity, rate of disease progression, and response to therapy, consistent with the existence of biologically distinct disease endotypes.^{1–4} This heterogeneity is also evident after diagnosis, as the rate of C-peptide decline differs substantially between individuals and is linked to age-related immune phenotypes.^{5,6} In particular, younger individuals generally show a more rapid decline in insulin secretion and

more aggressive immune profiles, whereas pancreatic studies have identified age-associated CD20Hi and CD20Lo patterns that differ in immune-cell infiltration and residual β -cell preservation.^{5–7} Preservation of C-peptide is clinically meaningful, as higher stimulated C-peptide levels are associated with lower insulin requirements, better glycaemic control, and reduced risk of hypoglycaemia and retinopathy.⁸ In addition, evidence from islet transplantation studies suggests that restoration of endogenous insulin secretion may provide broader clinical benefits beyond glycaemic control, including cardiovascular and other diabetes-related outcomes.^{9–11}

The heterogeneity complicates the understanding of disease pathogenesis and poses challenges to the development of effective therapeutic strategies.^{1,12} Improved understanding of the molecular mechanisms underlying this heterogeneity could therefore facilitate patient stratification and the development of more effective clinical trials and personalised therapeutic interventions.^{1–3}

To better understand this heterogeneity, large longitudinal studies with harmonised clinical phenotyping and biosampling are required. The INNODIA consortium

was established as a European platform for standardised recruitment and follow-up of individuals with newly diagnosed T1D and unaffected family members, with the aim of improving prediction, staging, and evaluation of disease progression.^{13,14} The INNODIA Natural History Study provides a large multicentre longitudinal cohort of individuals, including children, adolescents and adults with recent-onset T1D, enabling investigation of disease heterogeneity across age groups and over time.¹⁴

As part of the INNODIA study, we recently showed that gene expression changes in peripheral blood between baseline and one year post-diagnosis were associated with the rate of C-peptide decline at two years, when analysing samples from the cohort of first 100 patients newly diagnosed with T1D (First ND).¹⁵ These findings highlighted the potential of identifying transcriptomic changes as early indicators of the rate of disease progression. However, larger cohorts of newly diagnosed patients with T1D are essential to confirm the robustness of the findings.

In the present study, we aimed to validate and extend our earlier observations by analysing an independent follow-up cohort of 168 patients newly diagnosed with T1D from the INNODIA study (Next ND). We performed transcriptome profiling of peripheral blood samples collected at diagnosis and at 1-year follow-up and integrated these data with the data from the previously analysed First ND cohort. By combining data from both cohorts, we sought to increase the sample size and statistical power, enabling more comprehensive characterisation of gene expression changes during the first year after T1D onset and their association with the rate of disease progression, to inform future strategies for biomarker development and personalised intervention in T1D.

Methods

Ethics

The study followed the guidelines of the Declaration of Helsinki for research on human participants. The study protocol was initially approved by the London–City & East Research Ethics Committee (REC 16/LO/1750; IRAS Project ID 210497). Following translation of the participant documentation, approval was also obtained from the relevant local ethics authorities at the participating study sites. Written informed consent was obtained from all adult participants and from parents or legal guardians of minors. In addition, age-appropriate assent was obtained from underage participants, who were asked to provide their own consent once they reached the age at which this was applicable.

Clinical cohort

Whole-blood samples were collected as part of the INNODIA study, following a Master Protocol,¹³ at baseline (within 6 weeks of diagnosis) and at 1 year post-diagnosis.

This study included 139 samples from the First ND cohort (47 males, 46 females), with a median age of 11.0 years at diagnosis (range 1–38 years), and 317 samples from the Next ND cohort (103 males, 65 females), with a median age of 10.5 years at diagnosis (range 1–42 years). In total, 46 and 149 individuals had samples from both baseline and 1-year follow-up visits from the First ND and Next ND cohorts, respectively.

Sex was recorded as part of the clinical metadata collected at enrolment in the INNODIA study; it was included as a covariate in the statistical analyses where appropriate.

Fasted C-peptide/glucose ratio was used as a proxy to measure disease progression, because mixed meal tolerance test (MMTT) data was not available for participants aged 5 years or younger, greatly limiting the number of participants that could be included in the analyses. C-peptide corrected for glucose has been suggested as a suitable surrogate of MMTT AUC.^{15,16} Additional clinical characteristics are provided in [Table 1](#), including HbA1c at baseline and 1-year follow-up, BMI at baseline, and autoantibody positivity at baseline.

RNA-sequencing

Prior to RNA extraction, frozen whole-blood PAXgene samples were thawed at room temperature for 2 h and subjected to RNA extraction using PAXgene Blood miRNA Kit (PreAnalytix/QIAGEN, Cat# 763134). Total RNA, including RNA longer than approximately 18 nucleotides, was purified, following the protocol supplied by the kit manufacturer. Sample concentration was measured with Nanodrop 2000 spectrophotometer and Qubit Fluorometric Quantitation (Thermo Fisher Scientific). The quality of the samples was ensured with Experion Automated Electrophoresis System (Bio Rad) and Agilent 2100 Bioanalyzer RNA Pico chip. Library preparation and sequencing were carried out at the Finnish Functional Genomics Centre (FFGC). Before starting library preparation, ERCC Spike-in Mix 1 (Invitrogen P/N 4456739) was added to 100 ng RNA according to the kit's protocol. RNA-sequencing libraries were prepared using TruSeq stranded mRNA HT kit and protocol #15031047 (Illumina). The quality and quantity of the amplified libraries were measured using Advanced Analytical Fragment Analyser (Agilent) and Qubit Fluorometric Quantitation, respectively. Pooled libraries were sequenced on an Illumina NovaSeq 6000 instrument, using 2 × 50 bp paired-end sequencing.

Preprocessing of RNA-sequencing data

RNA-sequencing reads were aligned to the human reference genome (hg38) using the Rsubread package (v.2.6.4) which was also used to generate the gene-level read counts based on RefSeq gene annotations. Three haemoglobin-related genes (HBA1, HBA2, HBB) were overrepresented in the data, accounting for around

	First ND	Next ND	Combined
Patients	93	168	261
RNA-seq profiles			
Baseline	93	158	251
1-year	46	159	205
Sex			
Female	46	65	111
Male	47	103	150
Baseline			
IAA			
Positive	70	132	202
Negative	23	35	58
n/a	0	1	1
GAD65			
Positive	73	129	202
Negative	20	38	58
n/a	0	1	1
ZnT8A			
Positive	62	119	181
Negative	31	48	79
n/a	0	1	1
IA-2A			
Positive	64	133	197
Negative	29	34	63
n/a	0	1	1
Age			
Mean	13.1	12.0	12.4
SD	8.5	7.6	7.9
BMI			
Mean	18.9	18.8	18.8
SD	3.9	3.3	3.4
HbA1c			
Mean	79.3	76.2	76.8
SD	22.1	19.3	19.9
Fasted cpep/glucose			
Mean	38.5	40.1	39.7
SD	31.8	28.6	29.3
1-year			
HbA1c			
Mean	54.7	53.4	63.7
SD	14.4	10.3	11.3
Fasted cpep/glucose			
Mean	34.5	33.9	34.0
SD	30.9	29.0	29.4

Table 1: Clinical characteristics of samples used in the study.

11.6% of all reads. These reads were filtered prior to normalisation. The filtered count data were normalised using the trimmed mean of the M-values (TMM), converted to counts per million (CPM), and $\log_2$ -transformed using the R package edgeR (v.4.6.3).¹⁷ Genes with mean logarithmic CPM <1 across the samples were removed to filter out genes with low expression. Principal component analysis (PCA) was performed as a quality control step to evaluate global transcriptomic variation, identify potential outliers, and assess possible cohort-related batch effects.

Estimation of relative cell type abundances

We quantified relative cell type abundances for seven major cell populations in whole blood using marker-based estimates, including T cells, B cells, plasma-blasts, monocytes, neutrophils, erythrocytes, and platelets. For each cell type, 5–6 cell type-specific markers were selected from literature and their cell type-specific expression was confirmed using the CZ CELLxGENE single cell resource,¹⁸ covering ~5.6 million cells from healthy human blood. Additionally, we confirmed that the markers for each cell type were highly correlated across the samples in our RNA-sequencing data. The mean value of the cell type-specific markers per sample was used as a proxy for the relative abundance of each cell type. The accuracy of the estimates was confirmed by comparing them with measured cell type proportions using samples with both RNA-sequencing and flow cytometry data and Pearson and Spearman correlation.

Statistical analysis of changes between baseline and 1-year follow-up

Linear mixed effects modelling was used to identify significant transcriptional changes between samples collected at diagnosis and 1-year follow-up. A separate model was fitted for each gene and cell type, with gene expression level or the relative cell type abundance as the dependent variable, respectively. Visit, age, sex, and sequencing pool (batch) were treated as fixed effects, while study site and individual nested within study site were included as random effects. There were no missing values in the clinical variables used as covariates in the models. The same modelling approach was applied separately to First ND and Next ND cohorts, as well as to the combined dataset. To account for variation in cellular composition in the whole blood RNA-sequencing data, we further included the estimated relative cell type abundances as additional fixed effects in the gene-wise mixed effects models for the combined dataset. The gene-wise linear mixed effects model can be written as

$$y_{ij} = \beta_0 + \beta_1 \text{Visit}_{ij} + \beta_2 \text{Age}_i + \beta_3 \text{Sex}_i + \beta_4 \text{Batch}_{ij} + \sum_k \gamma_k \text{Celltype}_{kij} + b_{0,\text{site}} + b_{0,i(\text{site})} + \varepsilon_{ij}$$

where y_{ij} denotes the expression level of a given gene for individual i at visit j . For analyses of estimated relative cell type abundances, the same model structure was used without the cell type covariates. The analyses were performed using the R package lmerTest (v.3.1-3). For the gene-wise analyses, the distribution of nominal p-values was assessed using histogram, quantile–quantile (QQ) plot, and genomic inflation factor (λ). To account for multiple testing across genes, p-values were adjusted using the Benjamini-Hochberg method. Reported p-values refer

to nominal p-values unless otherwise indicated, whereas p-values adjusted for multiple testing are reported as false discovery rate (FDR) values.

Statistical analysis of disease progression rate based on C-peptide/glucose ratios

Similar to our earlier study,¹⁵ we used the rate of decline in the fasted C-peptide/glucose ratio as a surrogate marker of disease progression, reflecting residual β -cell function. For each individual, we estimated the trajectory of the log-transformed C-peptide/glucose ratio over follow-up visits using linear regression. To ensure robust estimates, we excluded measurements collected at the 3-month visit post-diagnosis to avoid confounding by the so-called honeymoon phase when transient recovery of β -cell function may happen.^{19,20} Based on the fitted individual-specific trajectory, we then estimated the time at which the predicted logarithmic C-peptide/glucose ratio reached the prespecified threshold of zero. This estimated time was used as the event time in the Cox proportional hazards modelling in a time-to-threshold fashion. Individuals whose predicted trajectory did not reach this threshold within 30 years were censored.

To assess the association between changes in relative abundances of different cell types from baseline to the 1-year follow-up and the rate of disease progression, univariable and multivariable Cox analyses were performed, with the latter adjusting for age and sex. The proportional hazards assumption was evaluated using Schoenfeld residuals and tests for time-dependent effects for the model covariates.

To assess the association between gene expression changes and the rate of disease progression, Cox models were considered at two levels. First, gene-wise models were constructed using within-individual changes in gene expression between diagnosis and 1-year follow-up, adjusted for changes in relative cell type abundances, age at diagnosis, and sex. For each gene, statistical significance was determined using the likelihood ratio test by comparing the model including the gene to a model excluding it. This gene-wise direct change Cox model can be written in the form.

$$h_i(t) = h_0(t) \exp \left(\beta_1 \Delta y_i + \beta_2 \text{Age}_i + \beta_3 \text{Sex}_i + \sum_k \gamma_k \Delta \text{Celltype}_{ki} \right)$$

where Δy_i denotes the within-individual change in expression of a given gene for individual i and $\Delta \text{Celltype}_{ki}$ the corresponding change in the estimated relative abundance of cell type k . Second, we considered within-individual changes in gene expression residuals derived from linear mixed effects models adjusted for age, sex, batch, site, and relative cell type abundances, and utilised the reproducibility-optimisation approach

(ROTS) for the analysis. This gene-wise residual-based Cox model can be written in the form.

$$h_i(t) = h_0(t) \exp(\beta_1 \Delta r_i)$$

where Δr_i denotes the within-individual change in expression residuals of a given gene for individual i . The analyses were performed using the R packages survival (v.3.8-3) and ROTS (v.2.0.0), respectively. For the gene-wise analyses, the distribution of nominal p-values was assessed using histograms, QQ plots, and genomic inflation factors. Reported p-values refer to nominal p-values unless otherwise indicated.

The expression scores based on the identified 47 genes were determined as the difference between the mean expression of genes with a hazard ratio <0.5 and the mean expression of genes with a hazard ratio >2 , following a similar approach as in.²¹

For Kaplan–Meier analyses, the first and third quartiles were considered. The log-rank test was used to assess the significance of differences between the groups.

Gene set enrichment analysis

Ranked gene set enrichment analysis against MSigDB (v2023.2) was performed using the R package fgsea (v1.34.2). Genes were ranked based on the formula $(1 - p) * \text{sign}(\text{coefficient})$. Here p refers to the p-value from the linear mixed effects or Cox proportional hazards model for a gene, which is multiplied by the sign of the corresponding effect for that gene. To assess whether glycolysis-related expression changes might reflect glycaemic state, the gene-wise mixed effects models for the combined dataset were additionally adjusted for HbA1c and gene set enrichment analysis was repeated using the updated gene ranking. To account for multiple testing, p-values were adjusted using the Benjamini-Hochberg method.

Role of funders

The funders had no role in the study design, data collection, analysis and interpretation of data, in the writing of the article, or in the decision to submit the paper for publication.

Results

Transcriptome changes during type 1 diabetes progression validated in the follow-up cohort

We previously identified gene expression changes during the first year after the disease onset in the First ND cohort from INNODIA.¹⁵ Here, we carried out RNA-sequencing on whole blood samples collected at diagnosis and at 1-year follow-up visit from the follow-up cohort of 168 patients newly diagnosed with T1D as part of the INNODIA study (Next ND) (Fig. 1A, Supplementary Fig. S1). We applied linear mixed

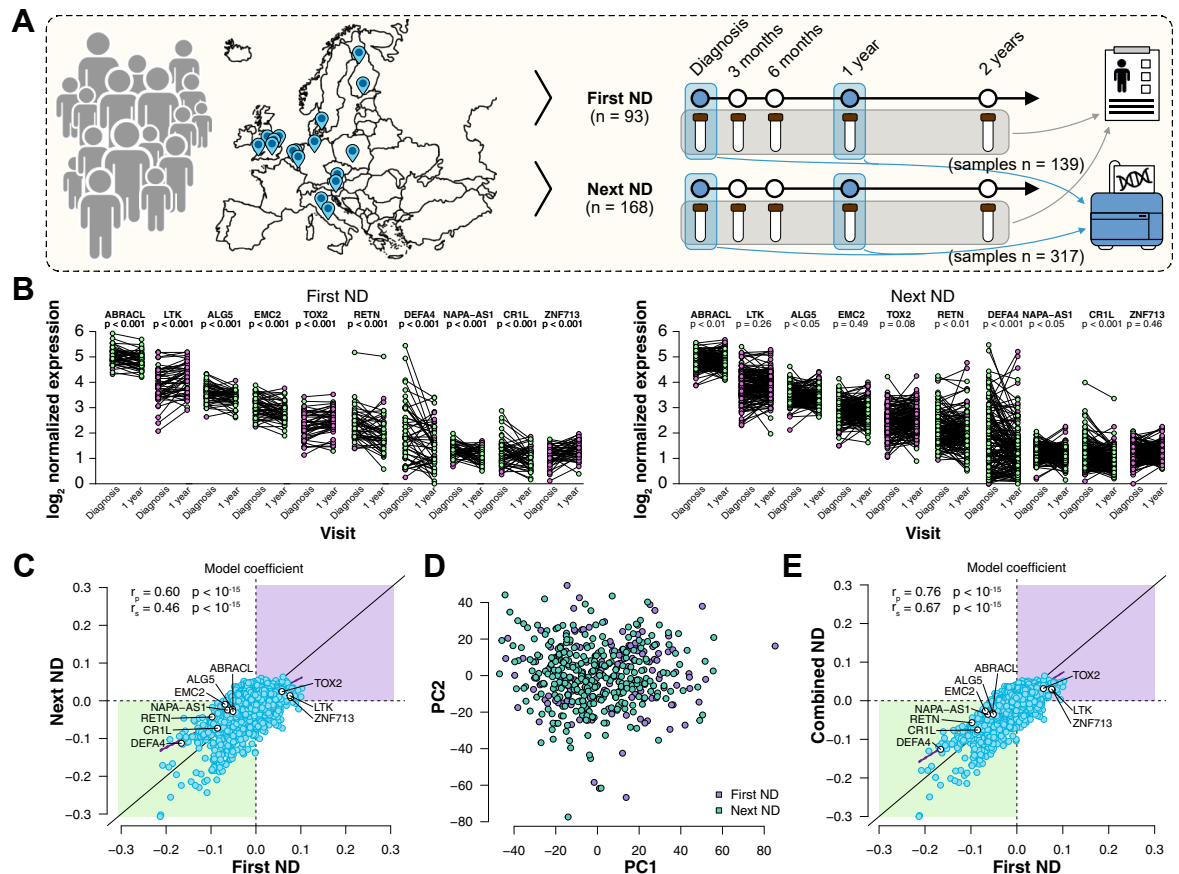


Fig. 1: Comparison of transcriptome changes during type 1 diabetes progression between the First ND (93 patients) and Next ND (168 patients) cohorts. **(A)** Whole blood samples were collected from newly diagnosed patients as part of the INNODIA study at diagnosis and 1-year follow-up (139 samples from First ND, 317 samples from Next ND cohorts), while clinical variables were available across multiple follow-up visits. **(B)** Gene expression changes between the baseline and 1-year follow-up visits across individuals for the top ten genes from the First ND cohort.¹⁵ Log-transformed normalised expression levels (counts per million) are shown for paired baseline and 1-year follow-up samples, with values from the same individual connected with lines. The p-values were obtained from linear mixed effects modelling. **(C)** Model coefficients from linear mixed effects modelling for each gene in the First ND and Next ND cohorts, comparing the baseline and 1-year follow-up visits, together with Pearson (r_p) and Spearman (r_s) correlation coefficients and their corresponding p-values. The top ten previously identified candidate genes from the First ND cohort are highlighted. **(D)** Principal component analysis of the RNA-sequencing data from the First ND and Next ND cohorts, showing no clear separation by cohort and supporting comparability of the datasets. **(E)** Model coefficients from linear mixed effects modelling for each gene in the First ND and combined cohorts, comparing the baseline and 1-year follow-up visits. The top ten previously identified candidate genes from the First ND cohort are highlighted.

effects modelling to assess the reproducibility of the results from the First ND cohort for the top ten candidates, defined in the previous study by a threshold of $p < 0.001$ and absolute model coefficient $> 0.05^{15}$: *ABRACL*, *LTK*, *ALG5*, *EMC2*, *TOX2*, *RETN*, *DEF4A*, *NAPA-AS1*, *CR1L*, and *ZNF713*. The observed changes in expression levels of these genes demonstrated consistent trends across both the First ND and Next ND cohorts, with six genes reaching statistical significance also in the Next ND cohort ($p < 0.05$, Fig. 1B). Moreover, effect sizes across all genes were highly concordant between the cohorts (Pearson correlation 0.60, $p < 10^{-15}$; Spearman correlation 0.46, $p < 10^{-15}$), supporting the robustness of the observed transcriptomic changes (Fig. 1C).

To increase the sample size and statistical power of the analysis, we combined data from the First ND and Next ND cohorts. Principal component analysis (PCA) was used as a quality control step to assess the global structure of the transcriptomic data and potential cohort-related batch effects. The PCA showed an even distribution of individuals across the principal components, with no clear clustering by cohort, supporting comparability of the datasets and their suitability for combined analysis (Fig. 1D).

Analysis of the combined dataset revealed that the differentially expressed genes between the baseline and 1-year follow-up visits were predominantly down-regulated over time, consistent with the pattern

observed in the First ND cohort (Fig. 1E). The observed changes in the expression levels of the previously identified top ten candidate genes showed again consistent trends in both the First ND and combined cohorts, and effect sizes across all genes were highly concordant (Pearson correlation 0.76, $p < 10^{-15}$; Spearman correlation 0.67, $p < 10^{-15}$; Fig. 1E).

Adjusting for the cell types revealed new interesting candidates

Due to the complex cellular heterogeneity of whole blood, we next sought to get insight into variation in cell type composition by computationally estimating the relative abundances of circulating cell populations, including T cells, B cells, plasmablasts, monocytes, neutrophils, erythrocytes, and platelets, using cell type-specific marker genes (Fig. 2A). To confirm the accuracy of the estimates, we first compared them with measured cell type proportions using samples with both RNA-sequencing and available flow cytometry data. The estimated and measured proportions showed high correlations for most cell types (Pearson correlation 0.74–0.89, $p < 0.001$; Spearman correlation 0.52–0.86, $p < 0.05$; Fig. 2B), supporting the reliability of the computational approach. Monocytes showed a more modest correlation of 0.54 ($p = 0.02$), in line with recent community benchmarking.²²

Analysis of the individual cell types revealed significant decreases in the estimated relative abundances of erythrocytes and plasmablasts from baseline to the 1-year follow-up (linear mixed effects model $p < 0.001$, adjusted for age, sex, and batch; Supplementary Table S1). As expected, the estimated relative abundances of most cell types were significantly associated with age (Supplementary Table S1).

Incorporation of the relative cell type abundance estimates (T cells, B cells, plasmablasts, neutrophils, erythrocytes, platelets) into the linear mixed effects modelling of the transcriptomics data identified a total of 1083 differentially expressed genes at false discovery rate (FDR) < 0.01 between the baseline and 1-year follow-up (Supplementary Table S2), consistent with substantial within-individual transcriptomic changes during the first year after diagnosis. Of these genes, 707 (65%) were down-regulated in the 1-year follow-up and 376 (35%) were up-regulated. The most significantly down-regulated genes were RHOT1, C3orf38, GPR65, TLR1, and GPR171, while the most significantly up-regulated genes included RRP12 and FUS (Fig. 2C). Several of the differentially expressed genes have been associated with T1D in previous studies including the strong candidates PTPN22 and PTPN2,^{23,24} and also TOX2 (FDR < 0.003) identified as a top candidate in our previous study.¹⁵

Examination of the p-value histogram, QQ plot, and genomic inflation factor ($\lambda = 4.794$; Supplementary Fig. S2) showed substantial deviation from the null

distribution, which can be expected when a large proportion of genes undergo true expression changes over time.²⁵

We further performed ranked gene set enrichment analysis (GSEA) to infer biological functions of the differentially expressed genes. The analysis revealed both down- and up-regulated gene sets during the first year after the disease onset (Supplementary Table S3). Among the down-regulated gene sets, interferon alpha and gamma responses were the most prominent (Fig. 2D). The type I interferon transcriptional signature has been previously associated with the development of T1D.^{26,27} The most significantly enriched pathways among the up-regulated genes included MYC targets, myogenesis, mitotic spindle, and glycolysis, which are associated with cellular proliferation, differentiation, and metabolism.

To assess whether the glycolysis-related expression changes might reflect glycaemic state, the gene set enrichment analysis was repeated after additional adjustment for HbA1c in the mixed effects models. Notably, these same pathways remained significant, including glycolysis, although the strength of the glycolysis signal was somewhat attenuated (Supplementary Table S3).

Decreased relative abundance of neutrophils is associated with rapid C-peptide decline

We next investigated whether changes in gene expression during the first year after T1D diagnosis were associated with the rate of disease progression. As a surrogate measure of β -cell function, we used the fasted C-peptide/glucose ratio, as described earlier.¹⁵ For each individual, we estimated the trajectory of the log-transformed C-peptide/glucose ratio over time as a measure of disease progression rate and used Cox proportional hazards modelling to determine associations (Fig. 3A). Visual inspection of the individual log-transformed C-peptide/glucose ratio trajectories and residual diagnostic plot supported that a log-linear model provided a reasonable approximation for estimating the disease progression rate (Supplementary Fig. S3), consistent with previous literature.²⁸

Univariable analysis of the individual cell types suggested a lower relative abundance of neutrophils and a higher relative abundance of T cells at the 1-year follow-up, compared to baseline, to be associated with a more rapid disease progression ($p < 0.05$; Supplementary Table S4). After adjusting for age and sex, a lower relative abundance of neutrophils at the 1-year follow-up remained associated with more rapid disease progression ($p = 0.05$), whereas the other major cell types did not show significant associations ($p > 0.1$; Supplementary Table S4). Kaplan–Meier analysis of the first and third quartiles of the within-individual changes in the relative neutrophil abundances supported a highly significant association with the rate of disease progression (log-rank test $p < 0.01$; Fig. 3B). Consistent with previous

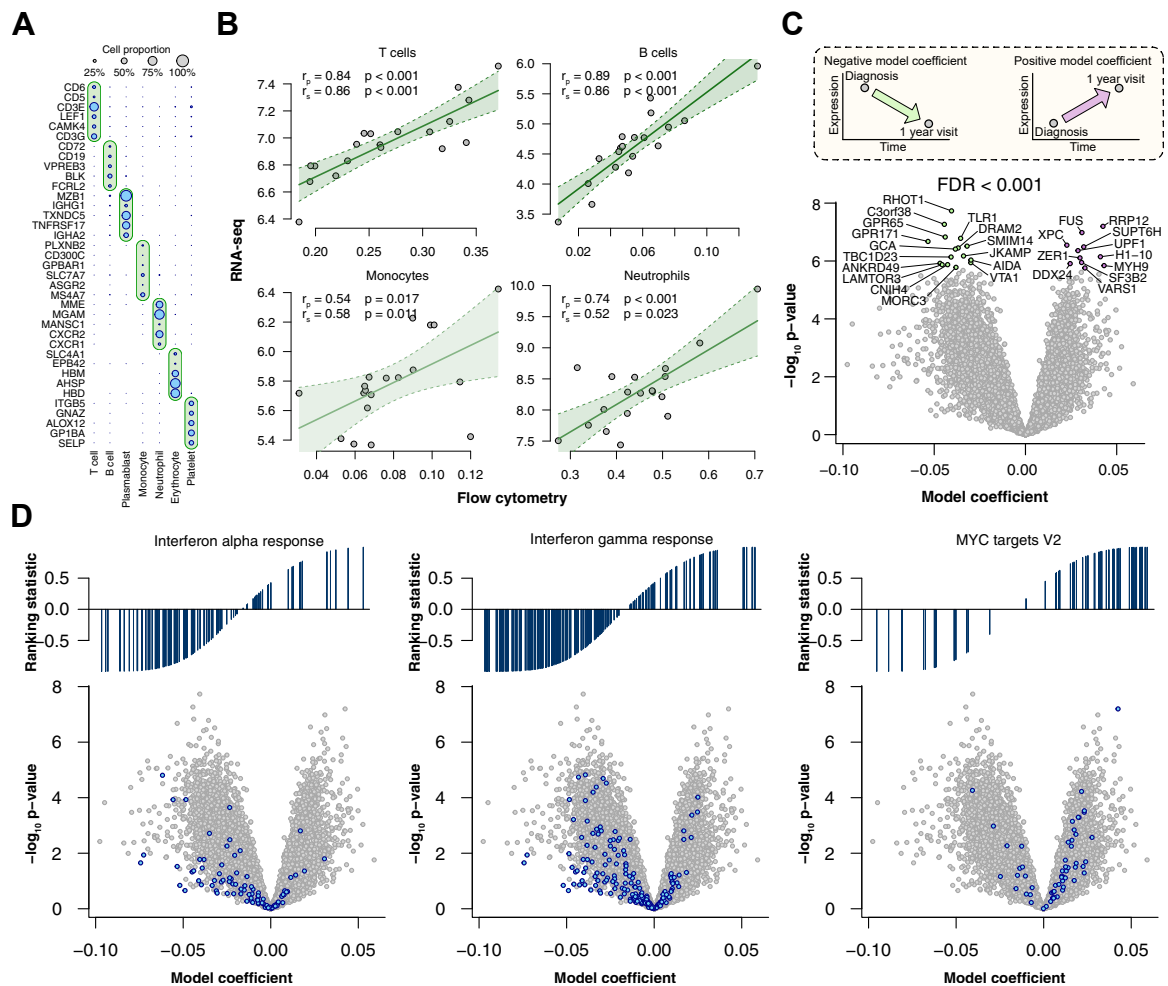


Fig. 2: Linear mixed effects modelling while accounting for relative cell type abundances. **(A)** Proportion of cells expressing cell type specific marker genes used for estimating the relative cell type abundances, according to CZ CELLxGENE single cell resource.¹⁸ **(B)** Pearson (r_p) and Spearman (r_s) correlations between RNA-sequencing based estimates of relative cell type abundances and flow cytometry based cell type proportions across samples with both measurements available. **(C)** Volcano plot showing the model coefficient from linear mixed effects modelling for each gene, comparing the baseline and 1-year follow-up visits, visualised against their $-\log_{10}$ nominal p-value. Genes with false discovery rate (FDR) < 0.001 are highlighted. **(D)** Selected top pathways from ranked gene set enrichment analysis. Genes in the pathway are highlighted.

studies,^{29–32} younger age was highly significantly associated with more rapid disease progression ($p < 10^{-7}$; Fig. 3C and Supplementary Table S4). Assessment of the proportional hazards assumption indicated evidence of non-proportionality for age (Supplementary Fig. S4). However, since the deviation appeared gradual rather than suggesting a clear alternative model structure based on stratification, age was retained as an important adjustment covariate, while its hazard ratio should be interpreted cautiously as an average effect over follow-up.

Gene expression changes associated with C-peptide decline

To study gene expression changes associated with the rate of disease progression in a gene-wise manner, we

applied Cox proportional hazards models using two complementary approaches. First, we considered within-individual gene expression changes directly while including within-individual changes in the estimated relative cell type abundances in the Cox model to account for compositional cell type changes over time. Second, we considered within-individual changes in gene expression residuals derived from linear mixed effects models adjusted for age, sex, batch, site, and relative cell type abundances using a reproducibility-optimised approach.³³ Examination of the p-value histograms, QQ plots, and genomic inflation factors ($\lambda = 0.816$ for the direct change approach and 0.853 for the residual-based approach; Supplementary Fig. S2) did not indicate substantial deviation from the null

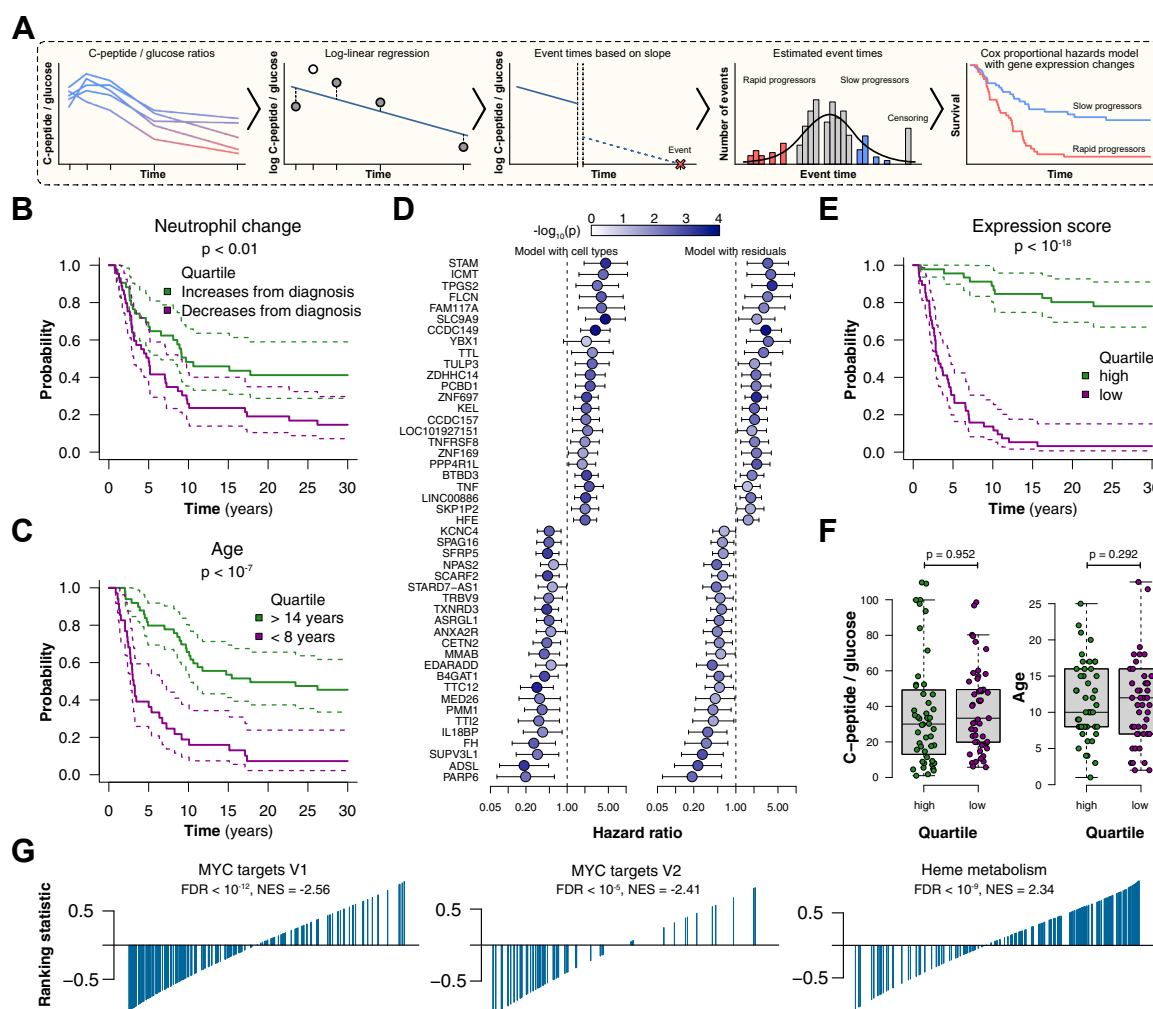


Fig. 3: Analysis of gene expression changes associated with rate of disease progression. **(A)** Schematic overview of the analysis approach, using the trajectory of the fasted C-peptide/glucose ratio over time as a surrogate for the rate of disease progression and Cox proportional hazards modelling to determine associations. **(B–C)** Kaplan–Meier curves illustrating the association of decreased abundance of neutrophils ($n = 49$ in both groups) and younger age ($n = 49$ for >14 years, $n = 47$ for <8 years) with more rapid disease progression, respectively. **(D)** Genes whose expression changes from baseline to 1-year follow-up were associated with the rate of disease progression at $p < 0.01$ and hazard ratio (HR) > 2 or < 0.5 with at least one of the two complementary approaches and supportive evidence from the other ($p < 0.1$). The model with cell types (left) refers to Cox model using gene expression changes directly, while adjusting for within-individual changes in estimated relative cell type abundances. The model with residuals (right) refers to reproducibility-optimised Cox model using gene expression residuals from linear mixed effects models adjusted for age, sex, batch, site, and relative cell type abundances. Hazard ratios with 95% confidence intervals are shown for each gene. **(E)** Kaplan–Meier curves of the first and third quartiles of expression scores determined based on the identified 47 genes, showing their association with rapid versus slow progression ($n = 49$ in both groups). **(F)** Box plots of baseline C-peptide/glucose ratios and ages in these two groups based on the first and third quartiles of expression scores ($n = 49$ in both groups). Between-group comparisons were performed using the Wilcoxon rank sum test. **(G)** Selected top pathways from ranked gene set enrichment analysis.

distribution, consistent with the more modest effects expected in this setting.

While none of the genes reached statistical significance after correcting for multiple testing ($FDR < 0.05$) with either approach, several showed consistent associations with both approaches, supporting the

robustness of these findings. We therefore focused particularly on 47 genes that were identified at nominal $p < 0.01$ and hazard ratio (HR) > 2 or < 0.5 with at least one approach and had supportive evidence with the other ($p < 0.1$). Among these, 24 genes (51%) had higher relative expression associated with rapid

progression and 23 genes (49%) lower relative expression associated with rapid progression at the 1-year follow-up compared to baseline (Supplementary Table S5, Fig. 3D).

Notably, Kaplan–Meier analysis of the first and third quartiles of expression scores determined based on these 47 genes confirmed their highly significant association with the rate of disease progression (log-rank test $p < 10^{-18}$; Fig. 3E). Interestingly, individuals in the first and third quartiles did not show significant differences in their baseline age nor their baseline C-peptide/glucose ratios (Wilcoxon rank sum test $p > 0.2$, Fig. 3F). Overall, these results suggest that gene expression changes during the first year after diagnosis may capture molecular signals linked to β -cell decline in addition to compositional changes in circulating cell populations.

Gene set enrichment analysis suggested higher relative expression of haem metabolism and TGF β signalling, and lower relative expression of MYC targets to be significantly associated with more rapid disease progression concordantly with both approaches (FDR < 0.05 ; Fig. 3G, Supplementary Table S6).

Discussion

In this study, we investigated gene expression changes associated with T1D progression by analysing RNA-sequencing data from a follow-up cohort of 168 newly diagnosed patients as part of the INNODIA study. The analysis confirmed consistent patterns in the expression of the top ten previously identified genes between baseline and the 1-year follow-up across both cohorts. Further integrated analysis across both cohorts revealed numerous other changes in gene expression during the first year after the disease onset, including down-regulation of genes related to interferon responses. Importantly, we also identified several genes associated with the rate of T1D progression, as well as a significant association of rapid disease progression with decreased relative abundance of neutrophils. Overall, our findings contribute to understanding of gene expression changes during the first year after the diagnosis and their implications for disease progression and treatment.

We observed a decrease of erythrocytes in patients recently diagnosed with T1D during the first year following disease onset. Previous research has reported altered erythrocyte metabolism in patients with T1D,³⁴ and hyperglycemia may also contribute to shortened erythrocyte lifespan, potentially explaining the reduced number of the cells. In addition, we detected decreased plasmablast levels in newly diagnosed patients over the same follow-up period. Interestingly, a previous study demonstrated elevated plasmablast frequencies in patients with new-onset T1D compared to non-diabetic and individuals with long-term T1D.³⁵ The higher

level of plasmablasts in recent onset patients may reflect an active autoimmune response at diagnosis that subsequently diminishes as the disease progresses, leading to lower plasmablast levels during follow-up.

Interestingly, among the genes whose expression level changed during the one follow-up year, we identified several differentially expressed genes that have previously been implicated in T1D pathogenesis, including TYK2, TLR8, PTPN2, PTPN22, ITGB7, all of which harbour SNPs associated with the disease.

We observed that MYC targets and glycolysis-related genes were significantly upregulated in patients with new onset T1D during the one-year follow-up. Notably, several key glycolytic enzymes, including enolase 1 (ENO1), phosphofructokinase (PFKM), pyruvate kinase (PKM), and glucose-6-phosphate isomerase (GPI), were among the upregulated transcripts in immune cells of patients with T1D. These findings may reflect altered metabolic activity during the early post-diagnosis period; however, they should be interpreted cautiously. In the context of recent-onset T1D, changes in metabolic pathway-related gene expression may reflect not only disease progression, but also glycaemic state at sampling, exogenous insulin exposure, and changes in blood cell composition over time. Notably, the pathway-level findings, including glycolysis, remained significant after additional adjustment for HbA1c, supporting that this signal was not explained solely by glycaemic control. In addition, because whole blood comprises multiple immune cell populations with distinct metabolic programmes, pathway enrichment related to glycolysis cannot be directly interpreted as evidence of mechanisms driving β -cell decline. Rather, these findings indicate transcriptomic changes associated with the first year after diagnosis and require a further investigation in more targeted settings.

Next, we examined whether gene expression changes observed during the first year following T1D diagnosis were associated with the rate of disease progression. We found the decrease in neutrophil proportions to be significantly associated with rapid C-peptide loss, which is consistent with our previous study.¹⁵ Previous studies have shown lower levels of neutrophils in patients newly diagnosed with T1D, as well in autoantibody-positive individuals compared to non-diabetic controls.^{36,37} A further study confirmed neutropenia in presymptomatic individuals and patients with T1D and demonstrated accumulation of neutrophils in the pancreas in patients with clinical disease and islet autoantibody-positive at-risk subjects.³⁸ However, the present study does not allow conclusions about causality, and the observed decrease in circulating neutrophils could either contribute to or result from the disease progression. In particular, it may reflect systemic metabolic dysregulation, insulin deficiency, altered immune cell trafficking, or other processes accompanying early T1D.

Notably, a decrease in circulating neutrophils has also been found to correlate with an increased risk of progression to T1D among autoantibody-positive at-risk individuals.^{37,38} Additionally, an inverse correlation between peripheral neutrophil levels and the titres of autoantibodies against GAD, IA2, and ZnT8 in individuals with T1D and latent autoimmune diabetes in adults (LADA) has been reported.³⁹ Furthermore, individuals who tested positive for all three autoantibodies demonstrated the most significant decrease in neutrophil levels.³⁹ The results of our study are in line with previous observations on the important role of neutrophils in T1D and showed that reduction in peripheral neutrophils was associated with rapid loss of β -cell function in patients with clinical disease. Interestingly, a recent study by Sassi et al. identified a neutrophil-enriched blood gene signature associated with teplizumab non-response in both human cohorts (stage 2 and new-onset stage 3) and nonobese diabetic (NOD) mice.⁴⁰

Another aspect of neutrophil involvement in T1D pathogenesis is changes in neutrophil function upon diabetes microenvironment exposure. Interestingly, hyperglycemia affects the neutrophil metabolism, shifting towards a more pro-inflammatory signature and leading to oxidative stress and neutrophil extracellular trap (NET) formation, thereby potentially exacerbating inflammation in T1D. On the other hand, essential functions such as phagocytosis and the oxidative burst, which are critical for effective bacterial clearance, are compromised under conditions of metabolic dysregulation.^{41,42} This might be reflected in the increased incidence of bacterial infections observed in patients with T1D compared to unaffected controls.^{43,44}

Previous studies have associated more rapid post-diagnosis decline in β -cell function with higher B cell and lower neutrophil immune signatures.⁵ In our study, the association between lower neutrophil abundance and more rapid disease progression was consistent with these earlier observations. However, we did not observe the association of higher B cell gene expression in the present analysis. This difference may reflect variation in study population, including cohort age composition, as well as differences in study design and analyses. In particular, Dufort et al. reported that higher B cell gene expression was associated with more rapid disease progression in younger individuals, but with slower disease progression with older individuals,⁵ while our analyses considered all individuals together. Our findings strengthen previous observation by showing that neutrophil-related changes remain associated with the disease progression also in an independent cohort analysed using a different approach.

Many previous studies have reported a strong correlation between age at diagnosis and the loss of C-peptide, reflecting a more rapid β -cell decline in younger individuals.^{29–32,45} The findings of our study

align with these earlier observations, indicating that children under the age of 8 years show a significantly more rapid progression compared to those older than 14 years. Moreover, an accumulating body of evidence suggests the existence of age-related phenotypes and potentially distinct mechanisms underlying the pathogenesis of T1D in children diagnosed before the age of 7 years compared to those diagnosed after the age of 13 years.^{7,46–48}

Further, our analysis identified 47 genes that showed consistent association with the rate of disease progression. Among these genes, we would like to highlight *TNF*, *TNFRSF8*, *HFE*, and *STAM*, all of which had higher relative expression in rapid progressors. *TNF* encodes a pro-inflammatory cytokine, whereas *TNFRSF8* is another TNF superfamily member, encoding CD30, which is expressed on activated T and B cells, NK cells, and monocytes. CD30 has been shown to be upregulated in multiple immune-mediated inflammatory diseases⁴⁹ and a relative expansion of a CD30⁺ T cell subpopulation was observed in the peripheral blood of subjects with newly diagnosed T1D.⁵⁰ Thus, the elevated expression of *TNF* and *TNFRSF8* in rapid progressors may suggest increased inflammatory response in the pancreas.

HFE encodes a protein that regulates iron homeostasis. Beyond iron metabolism, altered *HFE* has been shown to modulate type I interferon responses during viral infection.⁵¹ Additionally, *HFE* variants have been associated with increased risk of late-onset T1D⁵² and maternal *HFE* genotypes have been associated with increased risk of T1D in the offspring.⁵² Increased expression of the signal transducing adaptor molecule *STAM* in regulatory T cells has previously been associated with progressively worse future C-peptide decline in patients with newly diagnosed T1D.⁵³

In contrast, genes with lower relative expression in rapid progressors may represent anti-inflammatory or protective pathways. Among these, *IL18BP* encodes an inhibitor of the pro-inflammatory cytokine IL18, and overexpression of *IL18BP* has been observed to delay hyperglycemia in mice.⁵⁴ *SFRP5* has been shown to have anti-inflammatory properties and to support glucose homeostasis in type 2 diabetes.⁵⁵ Fumarate hydratase (FH), which encodes the enzyme involved in the fumarate metabolism pathway, has been identified as a pivotal regulator of cytokine production. In macrophages, the inhibition of FH results in a reduction of interleukin-10 (IL-10) secretion and an associated increase in *TNF* levels.⁵⁶ Additionally, the loss of FH contributes to an elevation in interferon β -levels.⁵⁶ Furthermore, lower levels of FH have been observed in the immune cells of patients diagnosed with systemic lupus erythematosus (SLE)⁵⁶ and multiple sclerosis.⁵⁷

One of the main strengths of the present study is that the key findings were evaluated in an independent

follow-up cohort. This demonstrates that the observed gene expression changes after diagnosis are reproducible across independent datasets and are not limited to a single cohort. This is particularly important in longitudinal whole-blood transcriptomic studies, where biological heterogeneity between individuals and technical variability can be substantial. This also helps distinguish robust biological signals from cohort-specific findings. In addition, the combined analysis further increased statistical power to characterise progression-associated transcriptomic changes.

Although our findings were validated in an independent follow-up cohort and supported by the increased sample size of the combined analysis, further validation in larger and more diverse cohorts will still be needed, and this study has some other limitations. More detailed metabolic and clinical variables, including detailed MMTT measures, were not available from all study participants, which did not allow additional correlation analyses to be performed with sufficient statistical power. In addition, C-peptide-based measures reflect residual β -cell function rather than exact β -cell mass. Thus, the fasted C-peptide/glucose ratio used here should be interpreted as a functional proxy of disease progression, not as a direct measure of pancreatic β -cell mass.

Because the analyses were based on whole-blood transcriptomic data, mechanistic interpretation should be cautious, as the observed pathway⁵⁶ changes may reflect a mixture of signals from different blood cell populations. An important limitation of whole-blood transcriptomic studies is that differences in bulk gene expression may partly reflect differences in cell type composition. We addressed this by adjusting for estimated relative abundances of major blood cell populations derived through *in silico* deconvolution, after confirming good agreement with measured cell type proportions from flow cytometry. However, since these estimates are derived from the same bulk transcriptomic data, they are not fully independent of the downstream gene expression values. Although methods for inferring cell type-specific expression or differential expression from bulk data exist, their performance depends on the abundance of the cell populations considered and the specificity of their marker genes.^{22,58,59} We therefore considered adjustment for estimated relative cell type abundances to be the most robust and interpretable approach for the present study, while interpreting the findings at the bulk whole-blood level. Finally, the association between lower peripheral neutrophil abundance and more rapid progression is observational and does not establish causality. Reduced neutrophil levels in blood could reflect either a process contributing to disease progression or a consequence of the metabolic and/or immunological changes occurring after diagnosis.

In conclusion, we identified gene expression changes in circulating immune cells of T1D patients

during the first year after the diagnosis. Importantly, our results suggest that, in addition to their compositional changes, gene expression changes during the first year of the clinical disease can capture the rate of disease progression and might be helpful in future studies aimed at patient stratification and developing new therapeutics.

Contributors

INNODIA and INNODIA HARVEST, LO, MP, Søb, AMS, CM provided expertise and facilitated data collection and curation. TS, IS, MK, RL and LLE designed the study. TS, IS, OR, UKK, SyB, LM, EH, TT collected data and performed analyses. TS, IS, OR, RL, LLE wrote the manuscript. TS, IS, OR verified the underlying data. All authors revised the manuscript and approved the final version.

Data sharing statement

The generated data is person-sensitive and access can be provided by application to the INNODIA Data Access Committee.

Declaration of interests

MLM has received grants from Breakthrough T1D. MLM reports consulting fees, lecture honoraria, and support for attending the ISPAD Conference 2025 from Sanofi. MLM reports board participation for SAB Bio and a leadership role as Vice-President of INNODIA iVZW. Søb reports lecture honoraria from Sanofi A/S. Søb also held stock in Novo Nordisk A/S, Lundbeck A/S, and Novonovon A/S. CM reports research support paid to KU Leuven by Dexcom, Abbott, Novo Nordisk, and Sanofi. CM also reports that KU Leuven received financial compensation for lectures and other activities from Eli Lilly, Vertex, Roche, Dexcom, Abbott, Medtronic, Novo Nordisk, Insulet, and Sanofi. CM reports that KU Leuven received financial compensation for board participation from Bayer, Biomea Fusion, Boehringer Ingelheim, Eli Lilly, Abbott, Insulet, Medtronic, Novartis, Novo Nordisk, Roche, SAB Bio, Sanofi, and Vertex. CM holds a leadership role as President of the European Diabetes Forum.

Acknowledgements

We are grateful to the staff of the University of Cambridge Department of Paediatrics laboratory, particularly Diane Wingate, for their contributions to the management of the samples. We are grateful to the people living with T1D and their families for their participation in the study and the clinical personnel for excellent work with the families, and collection of the samples for the study. We thank Marjo Hakkarainen and Sarita Heinonen (Turku Bioscience Centre, University of Turku, Finland) for their excellent technical help. We acknowledge Turku Bioscience Centre's core facility, Finnish Functional Genomics Centre (FFGC) supported by Biocenter Finland, for their assistance. This work has been performed using the Danish National Life Science Supercomputing Center, Computerome. The authors also wish to acknowledge CSC—IT Center for Science, Finland, and ELIXIR Finland for computational resources.

Funding: This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 115797 (INNODIA) and No 945268 (INNODIA HARVEST). This Joint Undertaking receives support from the Union's Horizon 2020 research and innovation programme, 'EFPIA', 'JDRF' and 'The Leona M. and Harry B. Helmsley Charitable Trust'.

LLE reports grants from the European Union's Horizon 2020 research and innovation programme (955321), Research Council of Finland (329278, 341342, 364700), Sigrid Jusélius Foundation, and Cancer Foundation Finland during the conduct of the study. R.L. received funding from the Research Council of Finland (292335, 294337, 314444, 329277, 331793, 335435, 335690, 369450), the Breakthrough T1D, the Novo Nordisk Foundation the Finnish Diabetes Foundation, the Sigrid Jusélius Foundation the Jane and Aatos Erkko Foundation, and the Fanconi Cancer Foundation. I.S. was supported by the Breakthrough T1D, Diabetes Wellness Suomi, Turku Doctoral

Programme of Molecular Medicine (TuDMM) and Finnish Diabetes Research Foundation.

Our research is also supported by the University of Turku Graduate School (UTUGS) and by the InFLAMES Flagship Programme of the Research Council of Finland (decision number: 337530).

The funders had no role in the study design, data collection, analysis and interpretation of data, in the writing of the article, or in the decision to submit the paper for publication.

Members of the INNODIA and INNODIA HARVEST Consortia:

Mathieu C, Gillard P, Casteels K, Overbergh L (KU Leuven, Belgium), Dunger D, Wallace C, Evans M, Thankamony A, Hendriks E, Bruggaber S, Marcovecchio L (University of Cambridge, United Kingdom), Peakman M, Tree T (King's College London, United Kingdom), Morgan NG, Richardson S, Oram R (University of Exeter, United Kingdom), Todd JA, Wicker L (University of Oxford, United Kingdom), Mander A, Dayan C, Alhadi Ali M (Cardiff University, United Kingdom), Pieber T (Medical University of Graz, Austria), Eizirik DL, Cnop M (Université Libre de Bruxelles, Belgium), Brunak S (University of Copenhagen, Denmark), Pociot F, Johannesen J, Rossing P, Legido Quigley C (Herlev University Hospital, Region Hovedstaden, Denmark), Mallone R, Scharfmann R, Boitard C (Cochin Institute Paris, France), Knip M, Otonkoski T (University of Helsinki, Finland), Veijola R (University of Oulu, Finland), Laheesmaa R, Oresic M, Toppari J (University of Turku, Finland), Danne T (Children's and Youth Hospital Hannover, Germany), Ziegler AG, Achenbach P, Rodriguez-Calvo T (Helmholtz Zentrum Muenchen, Germany), Solimena M, Bonifacio EE, Speier S (TU Dresden, Germany), Holl R (University of Ulm, Germany), Dotta F (University of Siena, Italy), Chiarelli F (University of Chieti, Italy), Marchetti P (University of Pisa, Italy), Bosi E (University Vita-Salute San Raffaele, Italy), Cianfarani S, Ciampalini P (Bambino Gesù Children's Hospital, Italy), De Beaufort C (Centre Hospitalier de Luxembourg, Luxembourg), Dahl-Jørgensen K, Skriverhaug T, Joner G, Krogvold L (Oslo University Hospital, Norway), Jarosz-Chobot P (Medical University of Silesia, Poland), Battelino T (University of Ljubljana, Slovenia), Thorens B (University of Lausanne, Switzerland), Gotthardt M (Radboud University Medical Center, The Netherlands), Roep BO, Nikolic T, Zaldumbide A (Leiden University Medical Center, The Netherlands), Lernmark A, Lundgren M, Giordano G, Tajes JF (Lund University, Sweden), Costecalde G (Univercell-Biosolutions, France), Strube T, Schulte AM, Nitsche A (Sanofi, Germany), Peakman M, Vela J (Sanofi, United States), Von Herrath M, Wesley J (Novo Nordisk, Denmark), Napolitano-Rosen A (GlaxoSmithKline, United Kingdom), Thomas M, Schlott N (Eli Lilly, United Kingdom), Goldfine A, Waldron-Lynch F, Komp J, Vedala A, Hartmann N, Nicolas G (Novartis Pharma AG, Switzerland), van Rangelbergh J, Bovy N (Imcyse SA, Belgium), Dutta S, Soderberg J, Ahmed S, Martin F, Latres E (Juvenile Diabetes Research Foundation, United States), Agiostratidou G, Korolova A (The Leona M. and Harry B. Helmsley Charitable Trust, United States).

Associated clinical sites:

Willemsen R (Barts Health NHS Trust, United Kingdom), Smith A (Northampton General Hospital NHS Trust, United Kingdom), Anand B (West Suffolk NHS FT, United Kingdom), Datta V (Norfolk & Norwich University NHS FT, United Kingdom), Puthi V (North West Anglia NHS FT, United Kingdom), Zac-Varghese S (East & North Hertfordshire NHS Trust, United Kingdom), Dias R (Birmingham Women's and Children's NHS FT, United Kingdom), Sundaram P (University Hospitals of Leicester NHS Trust, United Kingdom), Vaidya B (Royal Devon & Exeter NHS FT, United Kingdom), Patterson C (NHS Fife, United Kingdom), Owen K (Oxford University Hospitals NHS FT, United Kingdom), Dayan C (Cardiff & Vale University Health Board, United Kingdom), Piel B (Queen Elizabeth Hospital, King's Lynn FT, United Kingdom), Heller S (Sheffield Teaching Hospitals NHS FT, United Kingdom), Randall T, Gazis T (Nottingham University Hospitals NHS Trust, United Kingdom), Bismuth Reisman E, Carel J-C (Hospital Robert Debre, France), Riveline J-P, Gautier J-F (Hospital Lariboisiere, France), Andreelli F (Hospital Lapitie-Salpetriere, France), Travert F (Hospital Bichat Claude Bernard, France), Cosson E (Hospital Jean-Verdier, France), Penfornis A, Petit C (Centre Hospitalier

Sud-Francilien, France), Feve B (Hospital St Antoine, France), Lucidarme N (Hospital Jean-Verdier Pédiatrie, France), Cosson E (Hospital Avicenne, France), Beressi J-P (Hospital Andre Mignot, France), Ajzenman C (Hospital Andre Mignot Pédiatrie, France), Radu A (Hospital European Georges-Pompidou, France), Greteau-Hamoumou S (Hospital Louis Mourier, France), Bibal C (Hospital Kremlin Bicetre, France), Meissner T (Universitätsklinikum der Heinrich-Heine-Universität Düsseldorf, Germany), Heidtmann B (Katholisches Kinderkrankenhaus Wilhelmstift, Germany), Toni S (AOU Meyer, Italy), Rami-Merhar B (Medical University of Vienna, Austria), Eeckhout B, Peene B, Vantongerloo N (Algemeen Ziekenhuis Geel Sint-Dimpna Geel, Belgium), Maes T, Gommers L (Imeldziekenhuis Bonheiden, Belgium).

Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.ebiom.2026.106374>.

References

- 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;43:5–12.
- Redondo MJ, Morgan NG. Heterogeneity and endotypes in type 1 diabetes mellitus. *Nat Rev Endocrinol*. 2023;19:542–554.
- Fiorina P, Pozzilli P. Unveiling a novel type 1 diabetes endotype: opportunities for intervention. *Diabetes Metab Res Rev*. 2022;38:e3536.
- Galderisi A, Bonet J, Ismail HM, et al. Metabolic phenotype of stage 1 and stage 2 type 1 diabetes using modeling of β cell function. *J Clin Endocrinol Metab*. 2025;110:3168–3178.
- Dufort MJ, Greenbaum CJ, Speake C, Linsley PS. Cell type-specific immune phenotypes predict loss of insulin secretion in new-onset type 1 diabetes. *JCI Insight*. 2019;4:e125556. <https://doi.org/10.1172/jci.insight.125556>.
- Arif S, Leete P, Nguyen V, et al. Blood and islet phenotypes indicate immunological heterogeneity in type 1 diabetes. *Diabetes*. 2014;63:3835–3845.
- Leete P, Willcox A, Krogvold L, et al. Differential insulinitic profiles determine the extent of β -cell destruction and the age at onset of type 1 diabetes. *Diabetes*. 2016;65:1362–1369.
- Lachin JM, McGee P, Palmer JP, DCCT/EDIC Research Group. Impact of C-peptide preservation on metabolic and clinical outcomes in the diabetes control and complications trial. *Diabetes*. 2014;63:739–748.
- Gariani K, Peloso A, Haidar F, et al. Impact of islet transplantation on type 1 diabetes-related complication: a systematic review. *Transpl Int*. 2025;38:15091.
- Fiorina P, Gremizzi C, Maffi P, et al. Islet transplantation is associated with an improvement of cardiovascular function in type 1 diabetic kidney transplant patients. *Diabetes Care*. 2005;28:1358–1365.
- Perrier Q, Jambon-Barbara C, Kessler L, et al. Impact of islet transplantation on diabetes complications and mortality in patients living with type 1 diabetes. *Diabetes Care*. 2025;48:1007–1015.
- Ben Nasr M, D'Addio F, Uselli V, Tezza S, Abdi R, Fiorina P. The rise, fall, and resurgence of immunotherapy in type 1 diabetes. *Pharmacol Res*. 2015;98:31–38.
- Dunger DB, Bruggaber 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;23:414.
- Marcovecchio ML, Hendriks AEJ, Delfin C, et al. The INNODIA type 1 diabetes natural history study: a European cohort of newly diagnosed children, adolescents and adults. *Diabetologia*. 2024;67:995–1008.
- Suomi T, Starskaia I, Kalim UU, et al. Gene expression signature predicts rate of type 1 diabetes progression. *eBioMedicine*. 2023;92:104625.
- Ruan Y, Willemsen RH, Wilinska ME, Tauschmann M, Dunger DB, Hovorka R. Mixed-meal tolerance test to assess residual beta-cell secretion: beyond the area-under-curve of plasma C-peptide concentration. *Pediatr Diabetes*. 2019;20:282–285.
- Robinson MD, McCarthy DJ, Smyth GK. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26:139–140.

- 18 CZI Cell Science Program, Abdulla S, Aevertmann B, et al. CZ CELLxGENE discover: a single-cell data platform for scalable exploration, analysis and modeling of aggregated data. *Nucleic Acids Res.* 2025;53:D886–D900.
- 19 Abdul-Rasoul M, Habib H, Al-Khouly M. “The honeymoon phase” in children with type 1 diabetes mellitus: frequency, duration, and influential factors. *Pediatr Diabetes.* 2006;7:101–107.
- 20 Böber E, Dündar B, Büyükgebiz A. Partial remission phase and metabolic control in type 1 diabetes mellitus in children and adolescents. *J Pediatr Endocrinol Metab.* 2001;14:435–441.
- 21 Seyednasrollah F, Rantanen K, Jaakkola P, Elo LL. ROTS: reproducible RNA-Seq biomarker detector-prognostic markers for clear cell renal cell cancer. *Nucleic Acids Res.* 2016;44:e1.
- 22 White BS, de Reyniès A, Newman AM, et al. Community assessment of methods to deconvolve cellular composition from bulk gene expression. *Nat Commun.* 2024;15:7362.
- 23 Bottini N, Musumeci L, Alonso A, et al. A functional variant of lymphoid tyrosine phosphatase is associated with type 1 diabetes. *Nat Genet.* 2004;36:337–338.
- 24 Barrett JC, Clayton DG, Concannon P, et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Genet.* 2009;41:703–707.
- 25 Storey JD, Tibshirani R. Statistical significance for genome-wide studies. *Proc Natl Acad Sci U S A.* 2003;100:9440–9445.
- 26 Ferreira RC, Guo H, Coulson RMR, et al. A type 1 interferon transcriptional signature precedes autoimmunity in children genetically at risk for type 1 diabetes. *Diabetes.* 2014;63:2538–2550.
- 27 Kallionpää H, Elo LL, Laajala E, et al. Innate immune activity is detected prior to seroconversion in children with HLA-conferred type 1 diabetes susceptibility. *Diabetes.* 2014;63:2402–2414.
- 28 Shields BM, McDonald TJ, Oram R, et al. C-Peptide decline in type 1 diabetes has two phases: an initial exponential fall and a subsequent stable phase. *Diabetes Care.* 2018;41:1486–1492.
- 29 Petrone A, Galgani A, Spoletoni M, et al. Residual insulin secretion at diagnosis of type 1 diabetes is independently associated with both, age of onset and HLA genotype. *Diabetes Metab Res Rev.* 2005;21:271–275.
- 30 Ludvigsson J, Carlsson A, Deli A, et al. Decline of C-peptide during the first year after diagnosis of Type 1 diabetes in children and adolescents. *Diabetes Res Clin Pract.* 2013;100:203–209.
- 31 Barker A, Lauria A, Schloot N, et al. Age-dependent decline of β -cell function in type 1 diabetes after diagnosis: a multi-centre longitudinal study. *Diabetes Obes Metab.* 2014;16:262–267.
- 32 Harsunen M, Haukka J, Harjutsalo V, et al. Residual insulin secretion in individuals with type 1 diabetes in Finland: longitudinal and cross-sectional analyses. *Lancet Diabetes Endocrinol.* 2023;11:465–473.
- 33 Suomi T, Seyednasrollah F, Jaakkola MK, Faux T, Elo LL. ROTS: an R package for reproducibility-optimized statistical testing. *PLoS Comput Biol.* 2017;13:e1005562.
- 34 Herance JR, Ciudin A, Lamas-Domingo R, et al. The footprint of type 1 diabetes on red blood cells: a metabolomic and lipidomic study. *J Clin Med.* 2023;12:556. <https://doi.org/10.3390/jcm12020556>.
- 35 Ling Q, Shen L, Zhang W, et al. Increased plasmablasts enhance T cell-mediated beta cell destruction and promote the development of type 1 diabetes. *Mol Med.* 2022;28:18.
- 36 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.* 2013;45:467–470.
- 37 Valle A, Giamporcaro GM, Scavini M, et al. Reduction of circulating neutrophils precedes and accompanies type 1 diabetes. *Diabetes.* 2013;62:2072–2077.
- 38 Vecchio F, Lo Buono N, Stabilini A, et al. Abnormal neutrophil signature in the blood and pancreas of presymptomatic and symptomatic type 1 diabetes. *JCI Insight.* 2018;3:e122146. <https://doi.org/10.1172/jci.insight.122146>.
- 39 Huang J, Xiao Y, Zheng P, et al. Distinct neutrophil counts and functions in newly diagnosed type 1 diabetes, latent autoimmune diabetes in adults, and type 2 diabetes. *Diabetes Metab Res Rev.* 2019;35:e3064.
- 40 Sassi G, Lemaître P, Calvo LF, et al. Neutrophil-enriched gene signature correlates with teplizumab therapy resistance in different stages of type 1 diabetes. *J Clin Invest.* 2025;135:e176403. <https://doi.org/10.1172/JCI176403>.
- 41 Marthoff W, Stein M, Schleinkofer L, Federlin K. Evidence of ex vivo and in vitro impaired neutrophil oxidative burst and phagocytic capacity in type 1 diabetes mellitus. *Diabetes Res Clin Pract.* 1993;19:183–188.
- 42 Bilgic S, Aktas E, Salman F, et al. Intracytoplasmic cytokine levels and neutrophil functions in early clinical stage of type 1 diabetes. *Diabetes Res Clin Pract.* 2008;79:31–36.
- 43 Shah BR, Hux JE. Quantifying the risk of infectious diseases for people with diabetes. *Diabetes Care.* 2003;26:510–513.
- 44 Simonsen JR, Harjutsalo V, Järvinen A, et al. Bacterial infections in patients with type 1 diabetes: a 14-year follow-up study. *BMJ Open Diabetes Res Care.* 2015;3:e000067.
- 45 Antman KH, Nadler L, Mark EJ, Montella DL, Kirkpatrick P, Halpern J. Primary Kaposi’s sarcoma of the lung in an immunocompetent 32-year-old heterosexual white man. *Cancer.* 1984;54:1696–1698.
- 46 Leete P, Oram RA, McDonald TJ, 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.* 2020;63:1258–1267.
- 47 Parviainen A, Härkönen T, Ilonen J, But A, Knip M, Finnish Pediatric Diabetes Register. Heterogeneity of type 1 diabetes at diagnosis supports existence of age-related endotypes. *Diabetes Care.* 2022;45:871–879.
- 48 Zhou Q, Zheng X, Ma C, Gu W. Phenotypic spectrum at diagnosis of age-related endotypes of type 1 diabetes mellitus: a cross-sectional study in China. *J Diabetes.* 2025;17:e70111.
- 49 Veerasubramanian PK, Wynn TA, Quan J, Karlsson FJ. Targeting TNF/TNFR superfamilies in immune-mediated inflammatory diseases. *J Exp Med.* 2024;221:e20240806. <https://doi.org/10.1084/jem.20240806>.
- 50 Baker C, Chang L, Elsegood KA, et al. Activated T cell subsets in human type 1 diabetes: evidence for expansion of the DR+ CD30+ subpopulation in new-onset disease. *Clin Exp Immunol.* 2007;147:472–482.
- 51 Liu J, Wu X, Wang H, et al. HFE inhibits type 1 IFNs signaling by targeting the SQSTM1-mediated MAVS autophagic degradation. *Autophagy.* 2021;17:1962–1977.
- 52 Størdal K, McArdle HJ, Hayes H, et al. Prenatal iron exposure and childhood type 1 diabetes. *Sci Rep.* 2018;8:9067.
- 53 Pesenacker AM, Chen V, Gillies J, et al. Treg gene signatures predict and measure type 1 diabetes trajectory. *JCI Insight.* 2019;4:e123879. <https://doi.org/10.1172/jci.insight.123879>.
- 54 Lewis EC, Dinarello CA. Responses of IL-18- and IL-18 receptor-deficient pancreatic islets with convergence of positive and negative signals for the IL-18 receptor. *Proc Natl Acad Sci U S A.* 2006;103:16852–16857.
- 55 Liu L-B, Chen X-D, Zhou X-Y, Zhu Q. The Wnt antagonist and secreted frizzled-related protein 5: implications on lipid metabolism, inflammation, and type 2 diabetes mellitus. *Biosci Rep.* 2018;38. <https://doi.org/10.1042/BSR20180011>.
- 56 Hooftman A, Peace CG, Ryan DG, et al. Macrophage fumarate hydratase restrains mtRNA-mediated interferon production. *Nature.* 2023;615:490–498.
- 57 Sarkar P, Redondo J, Hares K, et al. Reduced expression of mitochondrial fumarate hydratase in progressive multiple sclerosis contributes to impaired in vitro mesenchymal stromal cell-mediated neuroprotection. *Mult Scler.* 2022;28:1179–1188.
- 58 Jaakkola MK, Elo LL. Computational deconvolution to estimate cell type-specific gene expression from bulk data. *NAR Genom Bioinform.* 2021;3:lqaa110.
- 59 Jaakkola MK, Elo LL. Estimating cell type-specific differential expression using deconvolution. *Brief Bioinform.* 2022;23:bbab433. <https://doi.org/10.1093/bib/bbab433>.