HHS Public Access

Author manuscript

Diabetologia. Author manuscript; available in PMC 2024 March 01.

Published in final edited form as:

Diabetologia. 2023 March ; 66(3): 520–534. doi:10.1007/s00125-022-05843-x.

Stratifying Risk for Onset of Type 1 Diabetes Using Islet Autoantibody Trajectory Clustering

Sejal Mistry[1], **Ramkiran Gouripeddi**[1,2], **Vandana Raman**[3], **Julio C. Facelli**[1,2,*]

[1]Department of Biomedical Informatics, University of Utah, Salt Lake City, UT, USA

[2]Clinical and Translational Science Institute, University of Utah, Salt Lake City, UT, USA

[3]Division of Pediatric Endocrinology, Department of Pediatrics, University of Utah, Salt Lake City, UT, USA

Abstract

Aims/Hypothesis: Islet autoantibodies can be detected prior to the onset of type 1 diabetes and are important tools for etiologic studies, prevention trials, and disease screening. Current risk stratification models rely on positivity status of islet autoantibodies alone, but additional autoantibody characteristics may be important for understanding disease onset. This work aimed to determine if a data-driven model incorporating characteristics of islet autoantibody development, including timing, type, and titer, could stratify risk for type 1 diabetes onset.

Methods: Data on autoantibodies against GAD (GADA), tyrosine phosphatase islet antigen-2 (IA-2A) and insulin (IAA) were obtained for 1,415 children enrolled in The Environmental Determinants of Diabetes in the Young study with at least one positive autoantibody measurement from years 1 – 12 of life. Unsupervised machine learning algorithms were trained to identify clusters of autoantibody development based on islet autoantibody timing, type, and titer. Risk for type 1 diabetes across each identified cluster was evaluated using time-to-event analysis.

Results: We identified 2 – 4 clusters in each year cohort that differed by autoantibody timing, titer, and type. During the first 3 years of life, risk for type 1 diabetes onset was driven by membership in clusters with high titers of all three autoantibodies (1-year risk: 20.87–56.25%, 5-year risk: 67.73–69.19%). Type 1 diabetes risk transitioned to type-specific titers during ages 4 – 8, as clusters with high titers of IA-2A (1-year risk: 20.88–28.93%, 5-year risk: 62.73–78.78%) showed faster progression to diabetes compared to high titers of GADA (1-year risk: 4.38–6.11%, 5-year risk: 25.06–31.44%). The importance of high GADA titers decreased during ages 9 – 12, with clusters containing high titers of IA-2A alone (1-year risk: 14.82–30.93%) or both GADA and IA-2A (1-year risk: 8.27–25.00%) demonstrating increased risk.

Conclusions/Interpretation: This unsupervised machine learning approach provides a novel tool for stratifying risk for type 1 diabetes onset using multiple autoantibody characteristics. These

*Corresponding Author: Julio C. Facelli, julio.facelli@utah.edu.

Contribution Statement: S.M and J.C.F. retrieved the data from the NIDDK Central Repository. S.M., R.G., and J.C.F. developed the analytical plan of the study. S.M. performed the preprocessing, unsupervised clustering, and statistical analyses. S.M., R.G., V.R., and J.C.F. interpreted the results, provided input in discussion, reviewed and edited the manuscript, and approved the final draft of the manuscript. J.C.F. is the guarantor of this work.

Authors' Relationships and Activities: The authors declare that there are no conflicts of interest.

findings suggest that age-dependent changes in IA-2A titers modulate risk for type 1 diabetes onset across 12 years of life. Overall, this work supports incorporation of islet autoantibody timing, type, and titer in risk stratification models for etiologic studies, prevention trials, and disease screening.

Keywords

Type 1 diabetes mellitus; islet autoantibodies; clustering; unsupervised machine learning; disease progression; risk stratification

INTRODUCTION:

Type 1 diabetes is a chronic autoimmune disease with increasing incidence worldwide [1–3]. Individuals with type 1 diabetes experience significant disease burden [4–6], making prevention a priority. Current strategies for type 1 diabetes prevention focus on modifying the pre-symptomatic disease course in individuals with significant disease risk. Pre-symptomatic type 1 diabetes is characterized by progressive immune-mediated destruction of insulin-producing pancreatic islet cells [7]. Autoantibodies against islet cells, including glutamic acid decarboxylase (GADA), tyrosine phosphatase islet antigen-2 (IA-2A), and insulin (IAA), are important biomarkers for type 1 diabetes that can be detected during this period [8, 9].

Current risk stratification models for type 1 diabetes rely on positivity status of islet autoantibodies in genetically susceptible individuals [7]; however, not all individuals that develop islet autoantibody positivity go on to develop type 1 diabetes and the time to disease onset varies considerably. Islet autoantibody characteristics, including age at appearance [10–14], the type and combinations [8, 12, 15, 16], and the titer of islet autoantibodies [17, 18], have been individually shown to stratify risk for type 1 diabetes. While these studies provide insight into islet autoantibody development, few studies have considered the longitudinal effects of these characteristics in unison due to insufficient analytical methods.

Unsupervised machine learning are data-driven methods that are well-equipped to characterize these complex interactions [19, 20]. Unsupervised learning algorithms aim to identify patterns in data without making any apriori assumptions on disease status. Previous studies leveraging unsupervised machine learning identified clusters on the basis of timing and type [21, 22] or timing and titer of islet autoantibodies [23, 24]. However, no studies have evaluated the combined effects of timing, type, and titer of islet autoantibody development on type 1 diabetes risk. Unsupervised machine learning methods may provide novel insights into the relationship between multiple autoantibody characteristics and may improve risk stratification models for disease onset.

Precise stratification of type 1 diabetes risk is needed to identify at-risk individuals, ultimately improving etiologic studies, disease screening, and enrollment in prevention studies. Therefore, this work aimed to determine if a data-driven model incorporating islet autoantibody timing, type, and titer could stratify risk for type 1 diabetes in children with high-risk HLA genotypes in The Environmental Determinants of Diabetes in the Young (TEDDY) study.

METHODS:

An overview of the methods is summarized in Fig 1. This research was approved by the University of Utah Institutional Review Board. All analyzes were performed in the HIPAA-compliant protected environment at the Center for High Performance Computing at the University of Utah. Additional details on the participants and data collection, inclusion/exclusion criteria, preprocessing, cohort extraction, identification of optimal models, and analysis of identified clusters are described in the electronic supplementary material [ESM] Methods.

Software

Unsupervised clustering was performed in R (v4.0.2) [25] using kml3d (v2.4.2) [26]. All other analyses were performed with Python (v3.8.12; www.python.org): statistical analysis was performed using SciPy (v1.7.2) [27], time-to-event analysis was performed using lifelines (v0.26.4) [28], and the network diagram was generated using network (v2.8.0) [29].

Participants & Data Collection:

Data from the TEDDY study was obtained from the National Institute of Diabetes and Digestive and Kidney Disease Central Repository. The TEDDY study enrolled 8,677 children with high-risk HLA genotypes for type 1 diabetes across 6 clinical centers in the United States and Europe [30]. Participants were followed every 3 months from 3 months of age until 4 years and every 3 or 6 months until 15 years, depending on autoantibody positivity status [30, 31]. Additional descriptions of the TEDDY study can be found elsewhere [30–34] (ESM Methods).

Inclusion/Exclusion Criteria:

Participants with at least one positive GADA, IA-2A, or IAA measurement as determined by the thresholds defined by the testing laboratory for the specific autoantibody and collection assay were included for analysis (ESM Methods). Participants with too few islet autoantibody measurements, as defined by having less than 4 autoantibody measurements in total, last study visit occurring before 12 months of age, first study visit occurring after 3 months of age, or less than 50% of autoantibody measurements available, were excluded (n = 238) to limit the amount of data imputation required during analysis and remove participants with only occasional or episodic measurements. These criteria led to 1,415 participants.

Preprocessing:

GADA, IA-2A, and IAA titers were extracted for included participants. Harmonization of measurements across the type of autoantibody, collection assay, and processing laboratories was performed and log-scaled z-score of GADA, IA-2A, and IAA titers were calculated as described in previous literature (ESM Methods) [11, 23]. To address the multiplicity of measurements at the same time point due to different assays, processing laboratories, and sample retesting [31, 33], a modified measurement selection procedure was adapted from

previous literature [11, 23] and is described (ESM Methods). Missing data were linearly imputed (ESM Methods).

Year Cohort Extraction:

To address varying autoantibody trajectory lengths due to loss-to-follow-up or reaching the study endpoint, measurements for all participants were divided into 1-year cohorts from 3 months to 15 years of age as defined (ESM Methods). To assess adequacy of sample size for unsupervised clustering, the feature-to-observation ratio was calculated and cohorts were excluded if they did not meet the required sample size [35]. Sex, clinical center, high-risk HLA genotype group, and islet autoantibody positivity status were extracted as covariates, and the status and age at type 1 diabetes diagnosis were extracted as outcome measures (ESM Methods). Differences in the distribution of covariates were assessed using a X^2 test with Yates Continuity correction [27] and evaluated at the 0.05 significance level.

Unsupervised Clustering & Evaluation:

For each year cohort, unsupervised machine learning was performed to identify clusters of GADA, IA-2A, and IAA development. Non-parametric algorithms for clustering joint trajectories (kml3d) [26] were developed with pre-specified clusters ranging from 2 – 10, Euclidean distance, and the k-means++ algorithm with the centroid method. These parameters were tested across 100 different initializations to determine if clustering results were consistent across different starting conditions set by k-means ++. The Calinski Harabasz score was used to assess clustering performance and identify the optimal number of clusters [36]. The optimal numbers of clusters for each year cohort were identified as described in ESM Methods.

Analysis of Identified Clusters:

Cluster centers, defined as the arithmetic mean of all points within a cluster, and standard deviations were calculated for each cluster. The log-scaled z-score of GADA, IA-2A, and IAA autoantibody cluster centers and standard deviations were plotted for each year cohort.

Time-to-event analysis [28, 37] was used to examine risk of progression to type 1 diabetes for each cluster in each year. The period from last autoantibody measurement in each year to age at type 1 diabetes diagnosis or age at last autoantibody measurement was used as the event time. Kaplan-Meier survival estimates were generated for each cluster in each year. A log-rank test was used to compare the overall difference in survival curves between clusters each year and pairwise log-rank tests were used to compare progression of type 1 diabetes between each cluster each year. Results were evaluated at the 0.05 significance level. Adjusted p-values were calculated for pairwise log-rank tests with more than one comparison using a Benjamini-Hochberg correction for multiple comparisons. For each cluster, the 1-year and 5-year risk for type 1 diabetes were calculated with a 95% CI when applicable and the titer percentile for each autoantibody at each timepoint for each cluster were calculated.

For each cluster, the distribution of participants by sex, HLA genotype group, clinical center, and islet autoantibody positivity status was calculated. Differences in the distribution of

covariates for each cluster in each year were assessed using a X^2 test with Yates Continuity correction and evaluated at the 0.05 significance level.

To determine whether cluster membership was stable or varied across each year, the number and percentage of individuals that transitioned from a given cluster each year to a different cluster in the subsequent year were calculated. Transitions of cluster membership across Years 1 – 12 were visualized using a network diagram [29, 38] (ESM Methods).

RESULTS:

Year Cohort Characteristics:

Overall, 1,415 participants were included for analysis. The feature-to-observation ratio for assessing adequacy of sample size was calculated to be 280 (70*k, where k is the number of variables, i.e., k = 4 visits per year). All year cohorts met the sample size inclusion threshold except Years 13 (n = 136), 14 (n = 18), and 15 (n = 0). Therefore, Years 1 – 12 were included for further analysis. Year cohorts did not differ significantly by sex (p = 1.000), clinical center (p = 0.996), high-risk HLA genotypes (p = 1.000), or islet autoantibody positivity status (p = 0.719) (ESM Table 2), indicating that covariates were similarly distributed across all years.

Unsupervised Clustering:

kml3d identified 2 – 4 clusters of GADA, IA-2A, and IAA development across year cohorts. Years 1, 4, 5, 6, 7, and 8 cohorts contained 3 clusters, Years 2 and 3 cohorts contained 2 clusters, and Years 9, 10, 11, and 12 cohorts contained 4 clusters with the highest and most consistent Calinski Harabasz scores (ESM Table 3; ESM Fig. 1). These models were selected for further analysis.

Analysis of Identified Clusters:

Years were categorized into four groups based on the similarity of visualized cluster centers. The first year in each group was selected as a representative image (Fig 2–5). Cluster centers (ESM Fig. 2), time-to-event analysis (ESM Fig. 3, ESM Table 4, Table 1), and covariates (ESM Table 5) are discussed for each cluster in each year. Cluster transitions between select years are described (Fig 6, ESM Table 6).

Three clusters were identified during the first year of life that stratified risk for type 1 diabetes (log-rank p = 4.951E-42, Fig 2). The first cluster center showed a baseline pattern with titers below the 50th percentile in all three autoantibodies (baseline; Fig 2a–c; Table 1). Survival analysis of participants in the baseline cluster revealed minimal progression to type 1 diabetes (Fig 2d), with a 1-year risk of 1.80% and a 5-year risk of 10.07% (Table 1). The second cluster center showed elevated GADA titers at 3 months with minimal elevations in IA-2A and IAA that declined over time (all declining; Fig 2a–c). The survival curve of participants in the all declining cluster was not distinguishable from the baseline cluster (pairwise log-rank adjusted p = 0.059, ESM Table 4), indicating a low risk for diabetes (1-year risk: 3.12%, 5-year: 3.12%, Table 1). Mothers of individuals in the all declining cluster were GADA positive at 0 or 9 months (ESM Results). The third cluster center

demonstrated an increase in titers of all islet autoantibodies (all increasing; Fig 2a–c), with an incline to the 99th percentile at 9 months noted in IAA (Table 1). Individuals in the all increasing cluster rapidly progressed to type 1 diabetes (Fig 2d), with a 56.25% 1-year and 68.75% 5-year risk (Table 1) and had a higher proportion of participants with the DR3/4 genotype (71.66%, ESM Table 5). The distribution of individuals who developed islet autoantibody positivity varied significantly across each cluster, with the all declining and all inclining clusters having 94.12% and 100.00% of individuals with positivity (X^2 test p = 4.729E-10, ESM Table 5). These findings suggest that islet autoantibody development among genetically susceptible individuals is characterized by increasing or decreasing patterns in all three autoantibodies, with a subset of individuals demonstrating declining patterns in islet autoantibody titers.

During years 2 – 3, two clusters of islet autoantibody development were identified that effectively stratified risk for type 1 diabetes (log-rank p = 1.132E-113 & 1.912E-136 for years 2 and 3, ESM Fig. 2b–c; ESM Fig. 3b–c). Similar to clusters identified in year 1, clusters in years 2 – 3 showed baseline patterns with titers below the 50th percentile in all autoantibodies or all increasing patterns (Table 1; Fig 3a–c). The all increasing clusters demonstrated increased risk for progression to type 1 diabetes (1-year risk: 27.96 and 20.87%, 5-year risk: 69.19 and 67.73% in years 2 and 3, Table 1; Fig 3d) compared to the baseline clusters (1-year risk: 1.21 and 0.52%, 5-year risk: 6.10 and 4.80% in years 2 and 3, Table 1; Fig 3d). Though most participants remained in the same cluster they were previously assigned to, participants in the all declining cluster in year 1 transitioned to the baseline cluster in year 2 (Fig 6; ESM Table 6). Most individuals in the all increasing clusters were islet autoantibody positive, while individuals in the baseline cluster were split (ESM Table 5). Together, these findings indicate that risk for type 1 diabetes is characterized by all increasing patterns of islet autoantibody development during the first three years of life, with few individuals returning to baseline thereafter.

During years 4 – 8, three clusters of islet autoantibody development were identified that effectively stratified risk for type 1 diabetes (log-rank p = 4.804E-150, 3.518E-146, 1.899E-91, 5.892E-82, 2.716E-57 for years 4, 5, 6, 7, and 8, ESM Fig. 2d–h, ESM Fig. 3d–h). Similar to the first three years, baseline clusters (Fig 4a–c) with minimal progression to type 1 diabetes (1-year risk: 1.03, 0.60, 0.63, 0.66, and 0.70% for years 4, 5, 6, 7, and 8, 5-year risk: 4.50, 4.34, 5.41, and 5.05% for years 4, 5, 6, and 7, Table 1; Fig 4d) were identified. Individuals previously assigned to all increasing clusters in year 3 primarily transitioned to one of two novel clusters in years 4 – 8 (Fig 6, ESM Table 6): 49.57% transitioned to a cluster with IA-2A titers above the 90th percentile (IA-2A dominant; Table 1; Fig 4b) while 29.57% transitioned to a cluster with GADA titers above the 90th percentile (GADA dominant; Table 1; Fig 4a). Though both IA-2A and GADA dominant clusters demonstrated greater risk for diabetes compared to baseline clusters (Fig 4d), individuals assigned to IA-2A dominant clusters progressed to type 1 diabetes faster than individuals assigned to the GADA dominant clusters (ESM Table 4). On average, the 1-year diabetes risk for IA-2A dominant clusters was 4.8 times higher than for GADA dominant clusters (IA-2A: 28.93, 25.49, 20.88, 24.42, and 24.08% vs. GADA: 5.93, 6.11, 4.86, 4.38, and 4.81% for IA-2A in years 4, 5, 6, 7, and 8, Table 1) and the 5-year risks were 2.4 times higher than GADA dominant cluster (IA-2A: 78.78, 74.06, 63.40, 62.73% vs. GADA:

30.45, 28.49, 25.06, 31.44% for years 4, 5, 6, and 7, Table 1). Individuals assigned to IA-2A dominant clusters in years 6 – 8 were more likely to be male and islet autoantibody positive (ESM Table 5). Together, these findings suggest that during ages 4 – 8, diabetes risk transitions to autoantibody and titer-specific clusters, with IA-2A dominant clusters having faster progression to type 1 diabetes.

During years 9 – 12, four clusters of islet autoantibody development were identified that effectively stratified risk for type 1 diabetes (log-rank p = 8.614E-33, 5.347E-31, 8.673E-13, and 4.024E-14 for years 9, 10, 11, and 12, ESM Fig. 2i–l, ESM Fig. 3i–l). Similar to previous years, baseline clusters (Fig 5a–c) with minimal progression to type 1 diabetes (1-year risk: 1.05, 1.39, 1.25, 0.00% in years 9, 10, 11, and 12, Table 1; Fig 5d) were identified. GADA dominant clusters and IA-2A dominant clusters were also identified in years 9 – 12 (Fig 5a–c). The importance of GADA dominance in diabetes risk diminished (1-year risk: 6.09, 3.74, 4.18, 3.85% in years 9, 10, 11, and 12, Table 1; Fig 5d) compared to IA-2A dominance, which remained equally important (1-year risk: 23.71, 21.65, 14.82, 30.93% for IA-2A in years 9, 10, 11, and 12, Table 1; Fig 5d). Though most individuals assigned to GADA or IA-2A dominant clusters in years 4 – 8 remained in the same clusters in years 9 – 12 (Fig 6, ESM Table 6), a subset of individuals transitioned to novel clusters that demonstrated elevated titers above the 90th percentile in both GADA and IA-2A (GADA and IA-2A dominant; Table 1; Fig 5a–c). Notably, the survival curves of GADA & IA-2A dominant clusters were not significantly different from IA-2A dominant clusters (pairwise log-rank adjusted p = 0.613, 0.947, 0.538, and 0.865 for years 9, 10, 11, and 12, ESM Table 4; Fig 5d). Both IA-2A dominant and GADA & IA-2A dominant clusters demonstrated faster progression to type 1 diabetes compared to baseline and GADA dominant clusters (1-year diabetes risk: 8.27, 13.5, 25.00, 23.61% for years 9, 10, 11, and 12, Table 1). All participants in IA-2A dominant clusters and IA-2A & GADA dominant clusters were islet autoantibody positive (ESM Table 5). Together, these findings indicate that IA-2A clusters play more important roles in type 1 diabetes risk stratification at older ages.

DISCUSSION:

Type 1 diabetes risk varies substantially according to islet autoantibody characteristics. A more precise stratification of type 1 diabetes risk is needed to better identify at-risk individuals for etiologic studies, disease screening, and enrollment in prevention trials. In the present study, we leveraged data-driven methods to identify clusters of GADA, IA-2A, and IAA islet autoantibody development that stratified the risk for type 1 diabetes in children with genetic susceptibility enrolled in the TEDDY study.

During the first 3 years of life, risk for type 1 diabetes was characterized by clusters with increasing titers of all three autoantibodies (Fig 2–3). Though increased diabetes risk associated with multiple islet autoantibody positivity in early life is well-established [10, 29], these findings are the first to provide insight into the joint development of GADA, IA-2A, and IAA over time. Notably, our findings suggest that increases in all three autoantibodies above the 70th percentile can be detected as early as 12 months of age and continue to increase until 3 years (Table 1). By providing a more precise estimate of

early type 1 diabetes risk and characterizing increases in multiple autoantibody titers, these findings may aid in planning type 1 diabetes prevention trials.

We also identified a subgroup of participants with a decline in GADA titers from the 99th percentile to the 50th percentile during the first year of life that likely resulted from elevated maternal autoantibodies [40] (ESM Results). Participants in this cluster demonstrated a decreased 5-year risk for type 1 diabetes compared to the baseline cluster (Table 1). This finding supports previous observations that maternal GADA may serve a protective role against type 1 diabetes onset [41] and further studies are needed to assess the role of GADA in slowing disease progression. Overall, cluster 2 reflects a heterogenous group in which autoantibodies generated from mothers with type 1 diabetes were likely transmitted to participants. These mothers are additionally likely to transmit insulin antibodies that may be generated as a consequence of maternal insulin therapy [42]. The findings from this analysis support a reduced risk for type 1 diabetes in infants born to mothers with type 1 diabetes.

Type 1 diabetes risk transitioned to type-specific titers during ages 4 – 8, as clusters with titers of IA-2A above the 98th percentile demonstrated faster progression to diabetes compared to clusters with titers of GADA above the 98th percentile (Fig 4). These findings suggest that while titers of GADA and IA-2A may both provide insights into the lifetime risk of type 1 diabetes, higher titers of IA-2A may indicate an earlier onset of type 1 diabetes. Type 1 diabetes is a highly heterogeneous disease, and recent studies have proposed autoantibody-driven subgroups of presymptomatic disease [43–45]. Our findings may reflect a novel age-related endotype driven by IA-2A dominance with faster progression to diabetes. The importance of GADA decreased during ages 9 – 12, with clusters containing IA-2A or both GADA and IA-2A demonstrating increased risk for type 1 diabetes (Fig 5). Together, these findings suggest that IA-2A plays a more important role in type 1 diabetes risk stratification later in life. Previous studies have also found that higher titers of IA-2A but not GADA increased the risk for type 1 diabetes [11, 46], and our findings add temporal context by highlighting the utility of GADA in years 4 – 8, but diminished effects after that.

This analysis detected single baseline clusters in each year cohort (Fig 2–5). We also found that individuals who did not meet the criteria for confirmed persistently positive islet autoantibody status but had at least one positive islet autoantibody were most prevalent in the baseline clusters (ESM Table 5). These findings suggest that subclinical variations in islet autoantibody titers are minimal and do not confer significant risk for future onset of type 1 diabetes.

Though HLA genotype groups significantly differed between clusters in most year cohorts, no individual genotype accounted for cluster membership (ESM Table 4). This suggests that factors beyond genotype influence autoantibody and type 1 diabetes development. Environmental exposures are implicated in the etiology of type 1 diabetes [47–49] and the clusters identified in this work may provide a useful framework to investigate etiologic factors in type 1 diabetes. We also noted a significant association between IA-2A dominant clusters in years 6 – 10 and male sex (ESM Table 5). Further research is needed to corroborate these findings in other diverse datasets and explore why males are more likely to demonstrate dominance of IA-2A in years 6 – 10 of life, especially given our finding

of increased type 1 diabetes risk in this group (Table 1). Future studies should explore underlying etiopathologic mechanisms associated with the data-driven clusters identified in this work.

Strategies to improve the prediction of risk for development of type 1 diabetes are needed to improve enrollment in prevention and etiologic studies [50]. Overall, this work offers evidence for including the timing, type, and titer of islet autoantibody measurements in risk stratification for type 1 diabetes. This is the first work to include all three autoantibody characteristics in a risk stratification model. These findings may also provide insights into optimal times and types of autoantibodies to guide screening programs during pre-symptomatic stages of type 1 diabetes. In addition, the importance of IA-2A in risk stratification after 4 years of age was established, which may guide strategies for recruitment and enrollment in prevention studies.

Strengths of these findings include a robust longitudinal cohort and a novel data-driven clustering method. The TEDDY study is the largest prospective birth cohort study that monitors longitudinal changes in islet autoantibodies [30], allowing unprecedented opportunities for big-data analytics to elucidate the pre-symptomatic features of type 1 diabetes. Though several studies have evaluated select autoantibody characteristics in type 1 diabetes risk [21–24], this is the first data-driven study that considers the combinatory role of the timing, type, and titer of islet autoantibodies in one model. The unsupervised machine learning algorithm utilized in this analysis facilitates the evaluation of joint trajectories across multiple variables [26]. We presented a year-based model of multiple islet autoantibody development that captured changes in the timing, type, and titer of autoantibody development that correlated with type 1 diabetes risk. The clusters identified in this study can serve as a computational framework for investigating other factors in pre-symptomatic type 1 diabetes development.

There are some limitations to this study. Only children with genetic susceptibility to type 1 diabetes were enrolled in the TEDDY study [30], limiting the generalizability of these findings. Autoantibody titers were measured using different assays across two different laboratory sites and the visit schedule changed after 4 years of age based on autoantibody status. Though analytical measures were taken to address these limitations (ESM Methods), residual measurement bias may remain and our findings may be skewed towards individuals who developed eventual islet autoantibody positivity. Future studies should evaluate these findings in more consistently collected datasets.

In this study, we included participants with at least one autoantibody positivity to limit bias towards baseline clusters and obtain a sufficient sample size for analysis. We also excluded individuals with few autoantibody measurements, but some individuals with episodic or occasional measurements may remain. Though we were able to identify distinct clusters, the number of participants assigned to baseline clusters was much larger than in non-baseline clusters. We also divided participants into 1-year cohorts to account for loss to follow-up or reaching the study endpoint. Future studies in larger cohorts should seek to evaluate clusters among participants with persistently positive islet autoantibodies using alternate temporal intervals.

The unsupervised clustering method used in this analysis was limited to temporal islet autoantibody measurements. Future studies should seek to use more complex machine learning methods to evaluate other characteristics, including other novel islet autoantigens [51], other genetic risk markers, environmental exposures, dietary intake, and socioeconomic factors.

In conclusion, this study leveraged data-driven techniques to assess the role of multiple islet autoantibody characteristics in type 1 diabetes risk. These findings highlight the importance of islet autoantibody timing, type, and titer in type 1 diabetes risk stratification. The identification of age-dependent percentiles for each autoantibody and associated 1-year and 5-year risk for type 1 diabetes onset may help to improve screening and prediction strategies for prevention studies. Future work should validate these findings in independent cohorts with diverse populations across longer temporalities and further characterize phenotypic features.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS:

This research was performed under the auspices of the TEDDY study group, a collaborative clinical study sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases (NIAID), National Institute of Child Health and Human Development (NICHD), National Institute of Environmental Health Sciences (NIEHS), Juvenile Diabetes Research Foundation (JDRF), and Centers for Disease Control and Prevention (CDC). The TEDDY study was conducted by the TEDDY study group Investigators. The data from the TEDDY study reported here were supplied by the NIDDK Central Repository. This manuscript was not prepared in collaboration with Investigators of the TEDDY study and does not necessarily reflect the opinions or views of the TEDDY study, the NIDDK Central Repository, for the NIDDK.

Funding:

This work was supported by the National Library of Medicine (T15LM007124), the NIDDK (F30DK134113), and the National Center for Advancing Translational Sciences (UL1TR002538). The study sponsors/funders were not involved in the design of the study; the collection, analysis, and interpretation of data; writing the report; and did not impose any restrictions regarding publication of the report.

Data Availability:

Data described in this manuscript will be made available through the NIDDK Central Repository at https://repository.niddk.nih.gov/home/.

ABBREVIATIONS:

GADA	Autoantibodies against glutamic acid decarboxylase
IA-2A	Autoantibodies against tyrosine phosphatase islet antigen-2
IAA	Autoantibodies against insulin
TEDDY	The Environmental Determinants of Diabetes in the Young
kml3d	Non-parametric algorithm for clustering joint trajectories

REFERENCES

1. Karvonen M, Viik-Kajander M, Moltchanova E, Libman I, LaPorte R, Tuomilehto J (2000) Incidence of childhood type 1 diabetes worldwide. Diabetes Mondiale (DiaMond) Project Group. Diabetes Care 23(10):1516–1526. 10.2337/diacare.23.10.1516 [PubMed: 11023146]
2. Patterson CC, Dahlquist GG, Gyürüs E, Green A, Soltész G (2009) Incidence trends for childhood type 1 diabetes in Europe during 1989–2003 and predicted new cases 2005–20: a multicentre prospective registration study. Lancet 373(9680):2027–2033. 10.1016/s0140-6736(09)60568-7 [PubMed: 19481249]
3. Mayer-Davis EJ, Lawrence JM, Dabelea D et al. (2017) Incidence trends of type 1 and type 2 diabetes among youths, 2002–2012. N Engl J Med 376:1419–1429. 10.1056/NEJMoa1610187 [PubMed: 28402773]
4. Fisher L, Polonsky WH, Hessler DM et al. (2015) Understanding the sources of diabetes distress in adults with type 1 diabetes. J Diabetes Complicat 29(4):572–577. 10.1016/j.jdiacomp.2015.01.012
5. Sussman M, Benner J, Haller MJ, Rewers M, Griffiths R (2020) Estimated lifetime economic burden of type 1 diabetes. Diabetes Technol Ther 22(2):121–130. 10.1089/dia.2019.0398 [PubMed: 31886730]
6. Hamman RF, Bell RA, Dabelea D et al. (2014) The search for diabetes in youth study: rationale, findings, and future directions. Diabetes Care 37(12):3336–3344. 10.2337/dc14-0574 [PubMed: 25414389]
7. Insel RA, Dunne JL, Atkinson MA et al. (2015) Staging presymptomatic type 1 diabetes: a scientific statement of JDRF, the Endocrine Society, and the American Diabetes Association. Diabetes Care 38(10):1964–1974. 10.2337/dc15-1419 [PubMed: 26404926]
8. Achenbach P, Warncke K, Reiter J et al. (2004) Stratification of type 1 diabetes risk on the basis of islet autoantibody characteristics. Diabetes 53(2):384–392. 10.2337/diabetes.53.2.384 [PubMed: 14747289]
9. Wenzlau JM, Juhl K, Yu L et al. (2007) The cation efflux transporter ZnT8 (Slc30A8) is a major autoantigen in human type 1 diabetes. Proc Natl Acad Sci U S A 104(43):17040–17045. 10.1073/pnas.0705894104 [PubMed: 17942684]
10. Ziegler AG, Rewers M, Simell O et al. (2013) Seroconversion to multiple islet autoantibodies and risk of progression to diabetes in children. Jama 309(23):2473–2479. 10.1001/jama.2013.6285 [PubMed: 23780460]
11. Steck AK, Vehik K, Bonifacio E et al. (2015) Predictors of progression from the appearance of islet autoantibodies to early childhood diabetes: the environmental determinants of diabetes in the young (TEDDY). Diabetes Care 38(5):808–813. 10.2337/dc14-2426 [PubMed: 25665818]
12. Anand V, Li Y, Liu B et al. (2021) Islet autoimmunity and HLA markers of presymptomatic and clinical type 1 diabetes: joint analyses of prospective cohort studies in Finland, Germany, Sweden, and the U.S. Diabetes Care 44(10):2269–2276. 10.2337/dc20-1836 [PubMed: 34162665]
13. Ziegler AG, Hummel M, Schenker M, Bonifacio E (1999) Autoantibody appearance and risk for development of childhood diabetes in offspring of parents with type 1 diabetes: the 2-year analysis of the German BABYDIAB Study. Diabetes 48(3):460–468. 10.2337/diabetes.48.3.460 [PubMed: 10078544]
14. Krischer JP, Lynch KF, Schatz DA et al. (2015) The 6 year incidence of diabetes-associated autoantibodies in genetically at-risk children: the TEDDY study. Diabetologia 58(5):980–987. 10.1007/s00125-015-3514-y [PubMed: 25660258]
15. Orban T, Sosenko JM, Cuthbertson D et al. (2009) Pancreatic islet autoantibodies as predictors of type 1 diabetes in the diabetes prevention trial-type 1. Diabetes Care 32(12):2269–2274. 10.2337/dc09-0934 [PubMed: 19741189]
16. Barker JM, Barriga KJ, Yu L et al. (2004) Prediction of autoantibody positivity and progression to type 1 diabetes: Diabetes Autoimmunity Study in the Young (DAISY). J Clin Endocrinol Metab 89(8):3896–3902. 10.1210/jc.2003-031887 [PubMed: 15292324]
17. Parikka V, Näntö-Salonen K, Saarinen M et al. (2012) Early seroconversion and rapidly increasing autoantibody concentrations predict prepubertal manifestation of type 1 diabetes in children at genetic risk. Diabetologia 55(7):1926–1936. 10.1007/s00125-012-2523-3 [PubMed: 22441569]

18. Steck AK, Johnson K, Barriga KJ et al. (2011) Age of islet autoantibody appearance and mean levels of insulin, but not GAD or IA-2 autoantibodies, predict age of diagnosis of type 1 diabetes: diabetes autoimmunity study in the young. Diabetes Care 34(6):1397–1399. 10.2337/dc10-2088 [PubMed: 21562325]
19. Beam AL, Kohane IS (2018) Big data and machine learning in health care. JAMA 319(13):1317–1318. 10.1001/jama.2017.18391 [PubMed: 29532063]
20. Deo RC (2015) Machine learning in medicine. Circulation 132(20):1920–1930. 10.1161/CIRCULATIONAHA.115.001593 [PubMed: 26572668]
21. Endesfelder D, Hagen M, Winkler C et al. (2016) A novel approach for the analysis of longitudinal profiles reveals delayed progression to type 1 diabetes in a subgroup of multiple-islet-autoantibody-positive children. Diabetologia 59(10):2172–2180. 10.1007/s00125-016-4050-0 [PubMed: 27400691]
22. Endesfelder D, Zu Castell W, Bonifacio E et al. (2019) Time-resolved autoantibody profiling facilitates stratification of preclinical type 1 diabetes in children. Diabetes 68(1):119–130. 10.2337/db18-0594 [PubMed: 30305370]
23. Köhler M, Beyerlein A, Vehik K et al. (2017) Joint modeling of longitudinal autoantibody patterns and progression to type 1 diabetes: results from the TEDDY study. Acta Diabetol 54(11):1009–1017. 10.1007/s00592-017-1033-7 [PubMed: 28856522]
24. Ng K, Stavropoulos H, Anand V et al. (2022) Islet autoantibody type-specific titer thresholds improve stratification of risk of progression to type 1 diabetes in children. Diabetes Care 45(1):160–168. 10.2337/dc21-0878 [PubMed: 34758977]
25. R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 10.1186/1752-0509-7-125
26. Genolini C, Pingault JB, Driss T et al. (2013) KmL3D: a non-parametric algorithm for clustering joint trajectories. Comput Methods Prog Biomed 109(1):104–111. 10.1016/j.cmpb.2012.08.016
27. Virtanen P, Gommers R, Oliphant TE et al. (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 17(3):261–272. 10.1038/s41592-019-0686-2 [PubMed: 32015543]
28. Davidson-Pilon C (2019) lifelines: survival analysis in python. J Open Source Softw 4(40):1317. 10.21105/joss.01317
29. Hagberg A, Swart P, Chult SD (2008) Exploring network structure, dynamics, and function using NetworkX. In. Los Alamos National Lab.(LANL), Los Alamos, NM (United States)
30. TEDDY Study Group (2008) The environmental determinants of diabetes in the young (TEDDY) study. Ann N Y Acad Sci 1150:1–13. 10.1196/annals.1447.062
31. The TEDDY Study Group (2007) The environmental determinants of diabetes in the young (TEDDY) study: study design. Pediatr Diabetes 8(5):286–298. 10.1111/j.1399-5448.2007.00269.x [PubMed: 17850472]
32. Hagopian WA, Erlich H, Lernmark A et al. (2011) The Environmental Determinants of Diabetes in the Young (TEDDY): genetic criteria and international diabetes risk screening of 421 000 infants. Pediatr Diabetes 12(8):733–743. 10.1111/j.1399-5448.2011.00774.x [PubMed: 21564455]
33. Bonifacio E, Yu L, Williams AK et al. (2010) Harmonization of glutamic acid decarboxylase and islet antigen-2 autoantibody assays for national institute of diabetes and digestive and kidney diseases consortia. J Clin Endocrinol Metab 95(7):3360–3367. 10.1210/jc.2010-0293 [PubMed: 20444913]
34. The Expert Committee on the Diagnosis and Classification of Diabetes Mellitus (1997) Report of the expert committee on the diagnosis and classification of diabetes mellitus. Diabetes Care 20(7):1183–1197. 10.2337/diacare.20.7.1183 [PubMed: 9203460]
35. Dolnicar S, Grün B, Leisch F, Schmidt K (2014) Required sample sizes for data-driven market segmentation analyses in tourism. J Travel Res 53(3):296–306. 10.1177/0047287513496475
36. Caliński T, Harabasz J (1974) A dendrite method for cluster analysis. Commun Stat Theory Methods 3(1):1–27. 10.1080/03610927408827101
37. Kleinbaum DG, Klein M (2012) Survival analysis : a self-learning text. 3rd edn. Springer, New York; London, DOI: 10.1007/978-1-4419-6646-9

38. Pedregosa F, Varoquaux G, Gramfort A et al. (2011) Scikit-learn: machine learning in python. J Mach Learn Res 12:2825–2830
39. Ziegler AG, Bonifacio E (2012) Age-related islet autoantibody incidence in offspring of patients with type 1 diabetes. Diabetologia 55(7):1937–1943. 10.1007/s00125-012-2472-x [PubMed: 22289814]
40. Naserke EH, Bonifacio E, Ziegler A-G (2001) Prevalence, characteristics and diabetes risk associated with transient maternally acquired islet antibodies and persistent islet antibodies in offspring of parents with type 1 diabetes. J Clin Endocrinol Metab 86(10):4826–4833. 10.1210/jcem.86.10.7931 [PubMed: 11600549]
41. Ziegler A-G, Bonifacio E (2020) Why is the presence of autoantibodies against GAD associated with a relatively slow progression to clinical diabetes? Diabetologia 63(8):1665–1666. 10.1007/s00125-020-05175-8 [PubMed: 32451571]
42. Koczwara K, Bonifacio E, Ziegler AG (2004) Transmission of maternal islet antibodies and risk of autoimmune diabetes in offspring of mothers with type 1 diabetes. Diabetes 53(1):1–4. 10.2337/diabetes.53.1.1 [PubMed: 14693690]
43. Parviainen A, Härkönen T, Ilonen J, But A, Knip M, Register tFPD (2022) Heterogeneity of type 1 diabetes at diagnosis supports existence of age-related endotypes. Diabetes Care 45(4):871–879. 10.2337/dc21-1251 [PubMed: 35147706]
44. Achenbach P, Hummel M, Thümer L, Boerschmann H, Höfelmann D, Ziegler AG (2013) Characteristics of rapid vs slow progression to type 1 diabetes in multiple islet autoantibody-positive children. Diabetologia 56(7):1615–1622. 10.1007/s00125-013-2896-y [PubMed: 23539116]
45. Giannopoulou EZ, Winkler C, Chmiel R et al. (2015) Islet autoantibody phenotypes and incidence in children at increased risk for type 1 diabetes. Diabetologia 58(10):2317–2323. 10.1007/s00125-015-3672-y [PubMed: 26138334]
46. Bauer W, Veijola R, Lempainen J et al. (2019) Age at seroconversion, HLA genotype, and specificity of autoantibodies in progression of islet autoimmunity in childhood. J Clin Endocrinol Metab 104(10):4521–4530. 10.1210/jc.2019-00421 [PubMed: 31120497]
47. Rewers M, Ludvigsson J (2016) Environmental risk factors for type 1 diabetes. Lancet 387(10035):2340–2348. 10.1016/s0140-6736(16)30507-4 [PubMed: 27302273]
48. Hummel S, Ziegler AG (2011) Early determinants of type 1 diabetes: experience from the BABYDIAB and BABYDIET studies. Am J Clin Nutr 94(suppl_6):1821S–1823S. 10.3945/ajcn.110.000646 [PubMed: 21633073]
49. Knip M, Veijola R, Virtanen SM, Hyöty H, Vaarala O, Åkerblom HK (2005) Environmental triggers and determinants of type 1 diabetes. Diabetes 54(suppl_2):S125–S136. 10.2337/diabetes.54.suppl_2.S125 [PubMed: 16306330]
50. Wherrett DK (2021) Improving prediction of risk for the development of type 1 diabetes-insights from populations at high risk. Diabetes Care 44(10):2189–2191. 10.2337/dci21-0018 [PubMed: 34548281]
51. Purcell AW, Sechi S, DiLorenzo TP (2019) The evolving landscape of autoantigen discovery and characterization in type 1 diabetes. Diabetes 68(5):879–886. 10.2337/dbi18-0066 [PubMed: 31010879]

RESEARCH IN CONTEXT:

What is already known about this subject?

- Islet autoantibodies can be detected prior to type 1 diabetes onset.
- Current risk stratification models rely on positivity status of islet autoantibodies, but additional autoantibody characteristics may be important for understanding disease onset.
- No studies have examined the joint role of islet autoantibody timing, type, and titer in type 1 diabetes onset due to insufficient analytical methods, but machine learning approaches may overcome these limitations.

What is the key question?

- Can a data-driven model incorporating islet autoantibody timing, type, and titer stratify risk for type 1 diabetes?

What are the new findings?

- Overall, we identified 2 – 4 clusters in each year cohort that differed by autoantibody timing, titer, and type.
- During the first 3 years of life, risk for type 1 diabetes was driven by membership in clusters with high titers of all three autoantibodies.
- Type 1 diabetes risk transitioned to type-specific clusters during ages 4–8, with high titers of IA-2A showing faster progression to disease compared to GADA that continued during ages 9–12.

How might this impact clinical practice in the foreseeable future?

- This data-driven approach provides a novel tool for stratifying type 1 diabetes risk using multiple autoantibody characteristics and supports the role for incorporation of islet autoantibody timing, type, and titer in risk stratification models for etiologic studies, prevention trials, and disease screening.

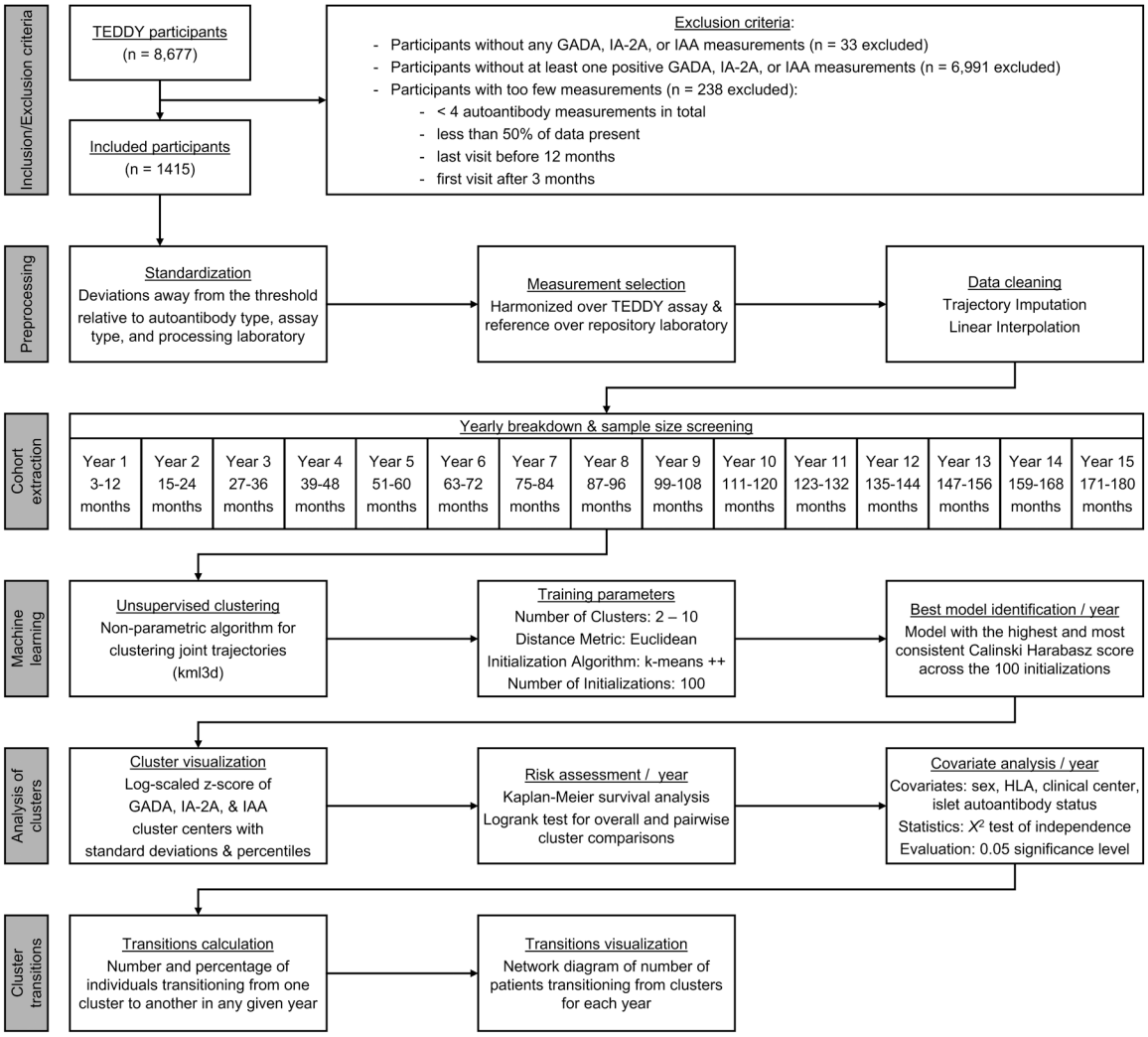
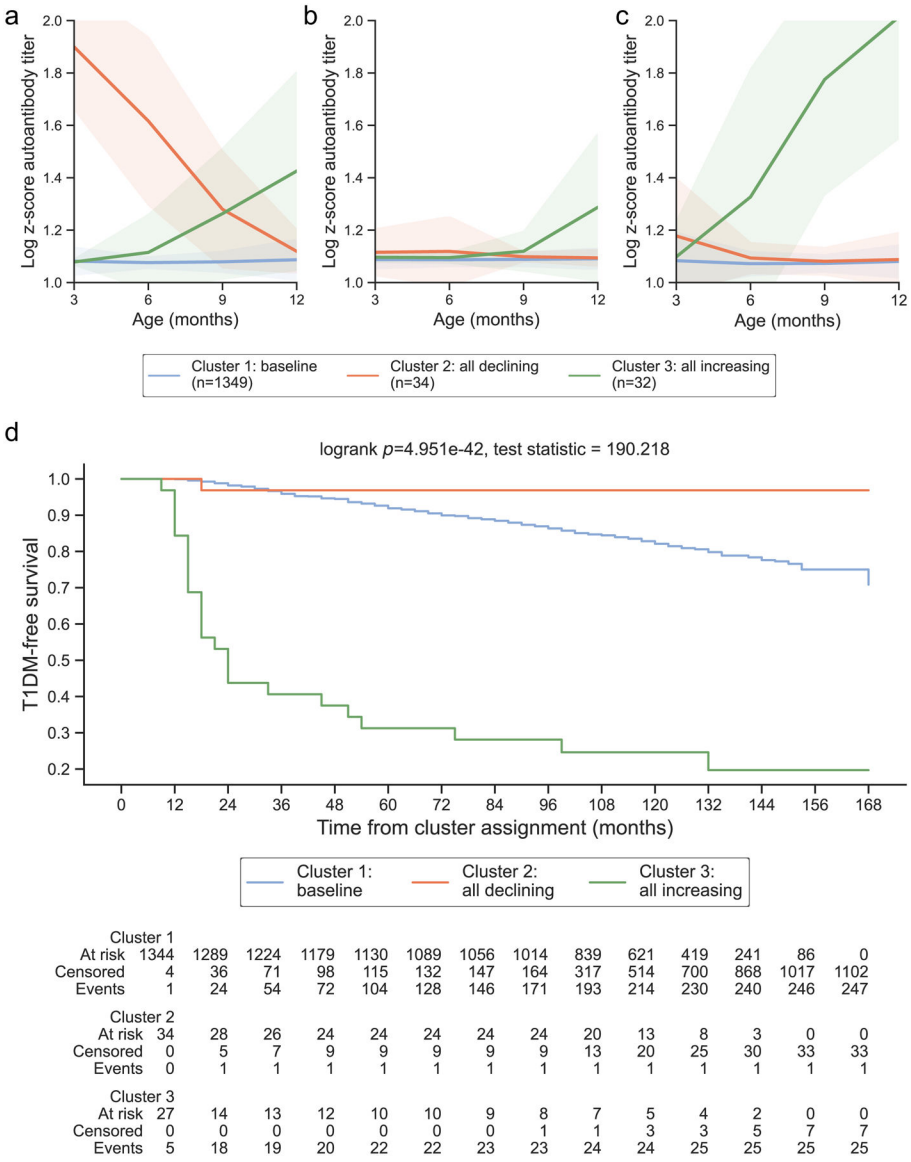


Figure 1: Methods workflow:

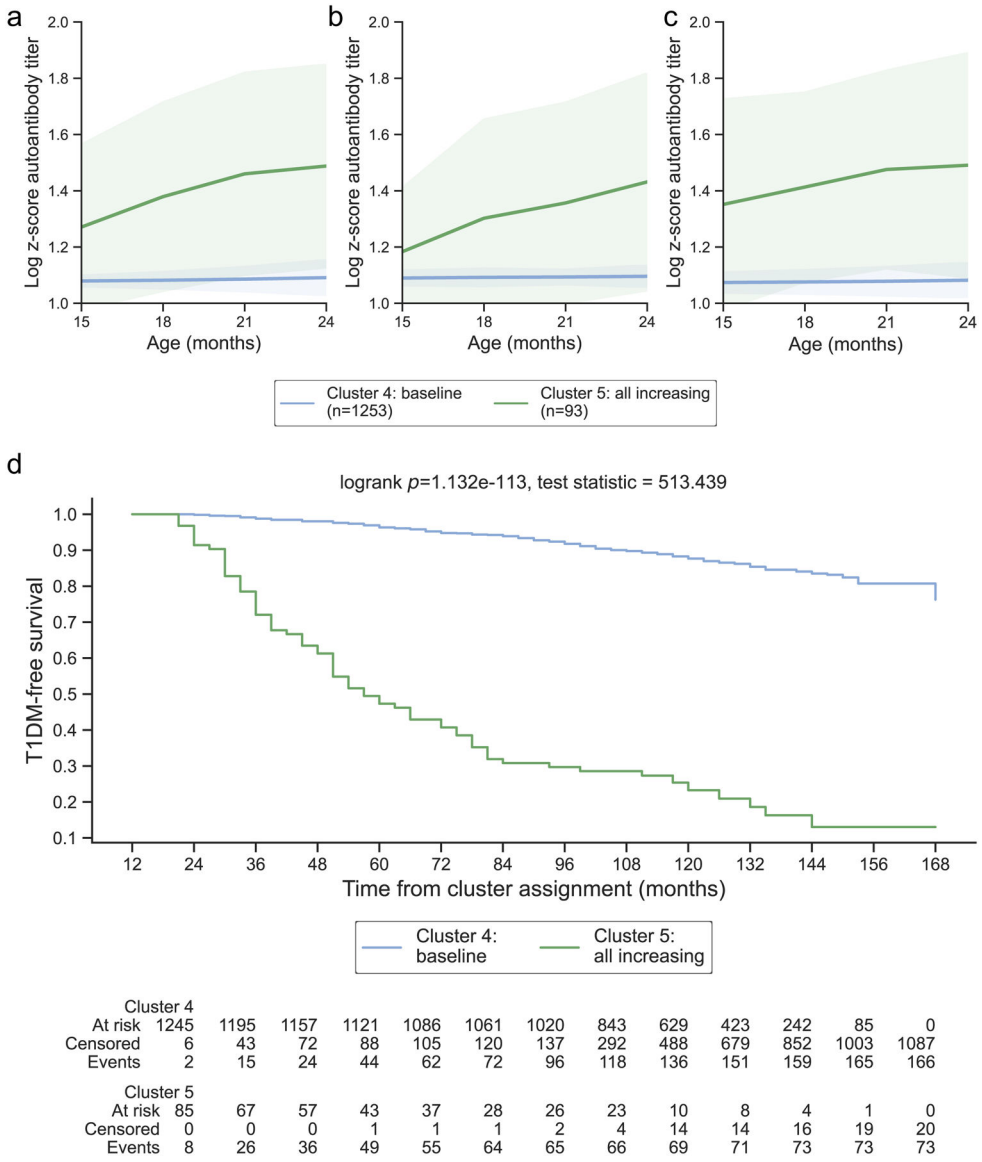
Workflow of methods used to include/exclude participants, preprocess data, develop 1-year cohorts, perform unsupervised clustering, analyze identified clusters, and calculate cluster transitions.



Cluster 1														
At risk	1344	1289	1224	1179	1130	1089	1056	1014	839	621	419	241	86	0
Censored	4	36	71	98	115	132	147	164	317	514	700	868	1017	1102
Events	1	24	54	72	104	128	146	171	193	214	230	240	246	247
Cluster 2														
At risk	34	28	26	24	24	24	24	24	20	13	8	3	0	0
Censored	0	5	7	9	9	9	9	9	13	20	25	30	33	33
Events	0	1	1	1	1	1	1	1	1	1	1	1	1	1
Cluster 3														
At risk	27	14	13	12	10	10	9	8	7	5	4	2	0	0
Censored	0	0	0	0	0	0	0	1	1	3	3	5	7	7
Events	5	18	19	20	22	22	23	23	24	24	25	25	25	25

Figure 2: Year 1 representative clusters:

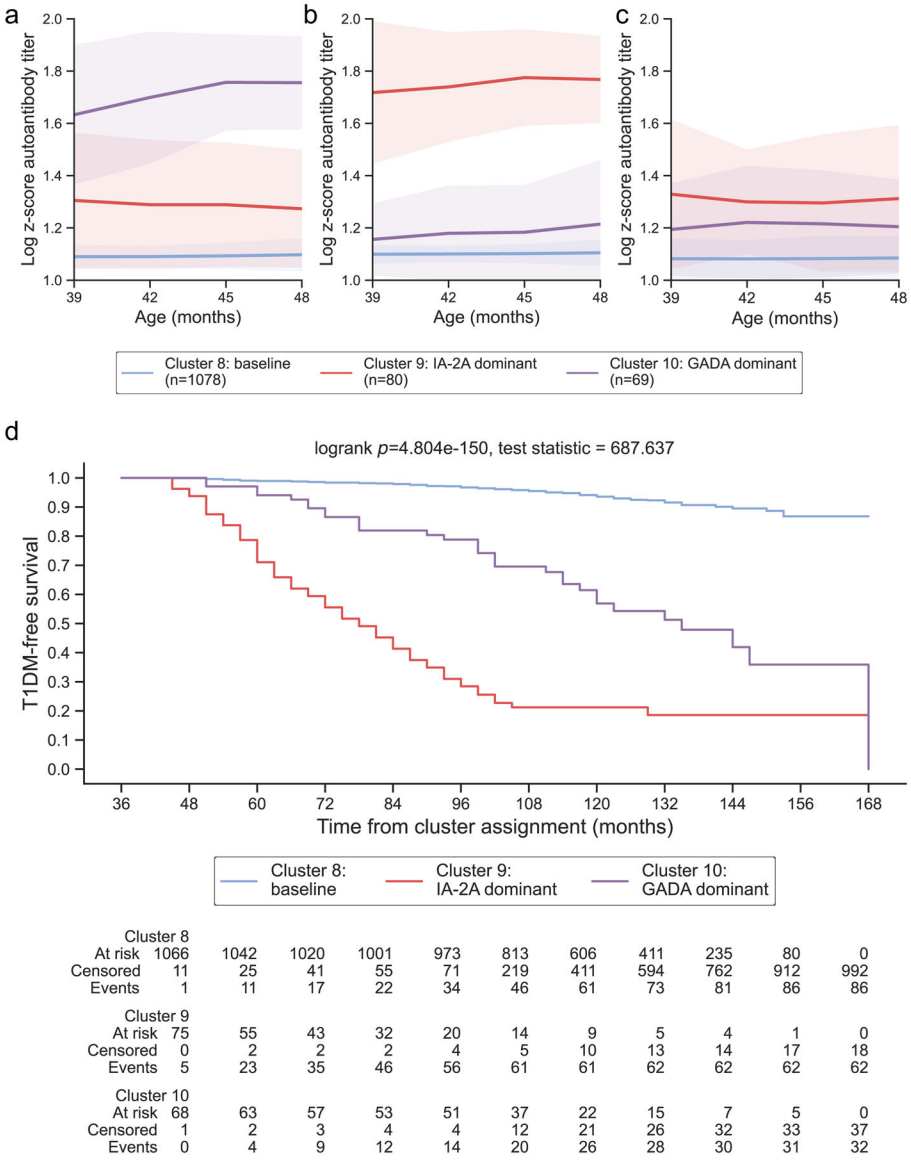
In the first year of life, 3 clusters were identified. Clusters are colored according to the autoantibody cluster center: baseline (blue), all declining (orange), and all increasing (green). Log-scaled z-scores of cluster centers identified through kml3d respective to (a) GADA, (b) IA-2A, and (c) IAA. (d) Survival curves for each identified cluster.



Cluster 4													
At risk	1245	1195	1157	1121	1086	1061	1020	843	629	423	242	85	0
Censored	6	43	72	88	105	120	137	292	488	679	852	1003	1087
Events	2	15	24	44	62	72	96	118	136	151	159	165	166
Cluster 5													
At risk	85	67	57	43	37	28	26	23	10	8	4	1	0
Censored	0	0	0	1	1	1	2	4	14	14	16	19	20
Events	8	26	36	49	55	64	65	66	69	71	73	73	73

Figure 3: Years 2 – 3 representative clusters:

From ages 2 – 3, 2 clusters were identified and year 2 was selected as representative for this age group. Clusters are colored according to the autoantibody cluster center: baseline (blue) and all increasing (green). Log-scaled z-scores of cluster centers identified through kml3d respective to (a) GADA, (b) IA-2A, and (c) IAA. (d) Survival curves for each identified cluster.



Cluster 8											
At risk	1066	1042	1020	1001	973	813	606	411	235	80	0
Censored	11	25	41	55	71	219	411	594	762	912	992
Events	1	11	17	22	34	46	61	73	81	86	86
Cluster 9											
At risk	75	55	43	32	20	14	9	5	4	1	0
Censored	0	2	2	2	4	5	10	13	14	17	18
Events	5	23	35	46	56	61	61	62	62	62	62
Cluster 10											
At risk	68	63	57	53	51	37	22	15	7	5	0
Censored	1	2	3	4	4	12	21	26	32	33	37
Events	0	4	9	12	14	20	26	28	30	31	32

Figure 4: Years 4 – 8 representative clusters:

From ages 4 – 8, 3 clusters were identified and year 4 was selected as representative for this age group. Clusters are colored according to the autoantibody cluster center: baseline (blue), IA-2A dominant (red), and GADA dominant (purple). Log-scaled z-scores of cluster centers identified through kml3d respective to (a) GADA, (b) IA-2A, and (c) IAA. (d) Survival curves for each identified cluster.

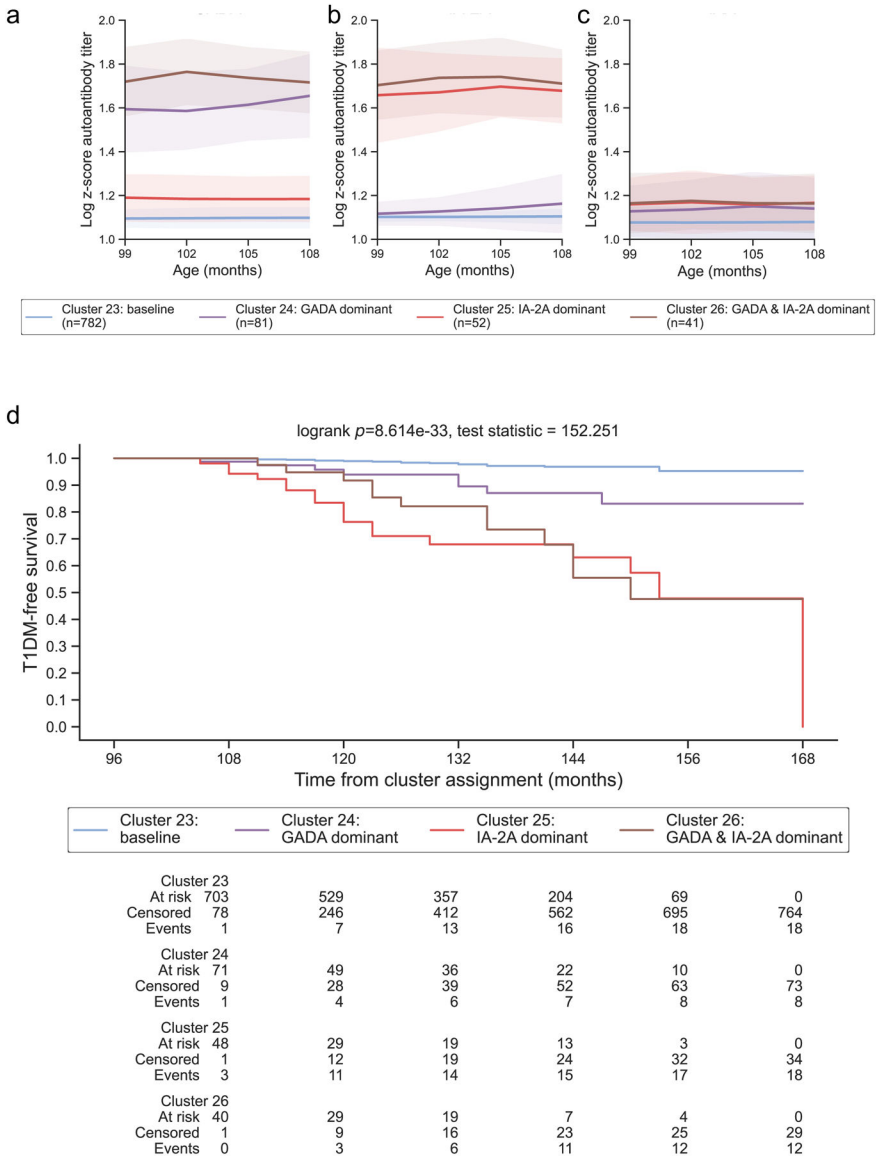


Figure 5: Years 9 – 12 representative clusters:

From ages 9 – 12, 4 clusters were identified and year 9 was selected as representative for this age group. Clusters are colored according to the autoantibody cluster center: baseline (blue), IA-2A dominant (red), GADA dominant (purple), and IA-2A and GADA dominant (brown). Log-scaled z-scores of cluster centers identified through kml3d respective to (a) GADA, (b) IA-2A, and (c) IAA. (d) Survival curves for each identified cluster.

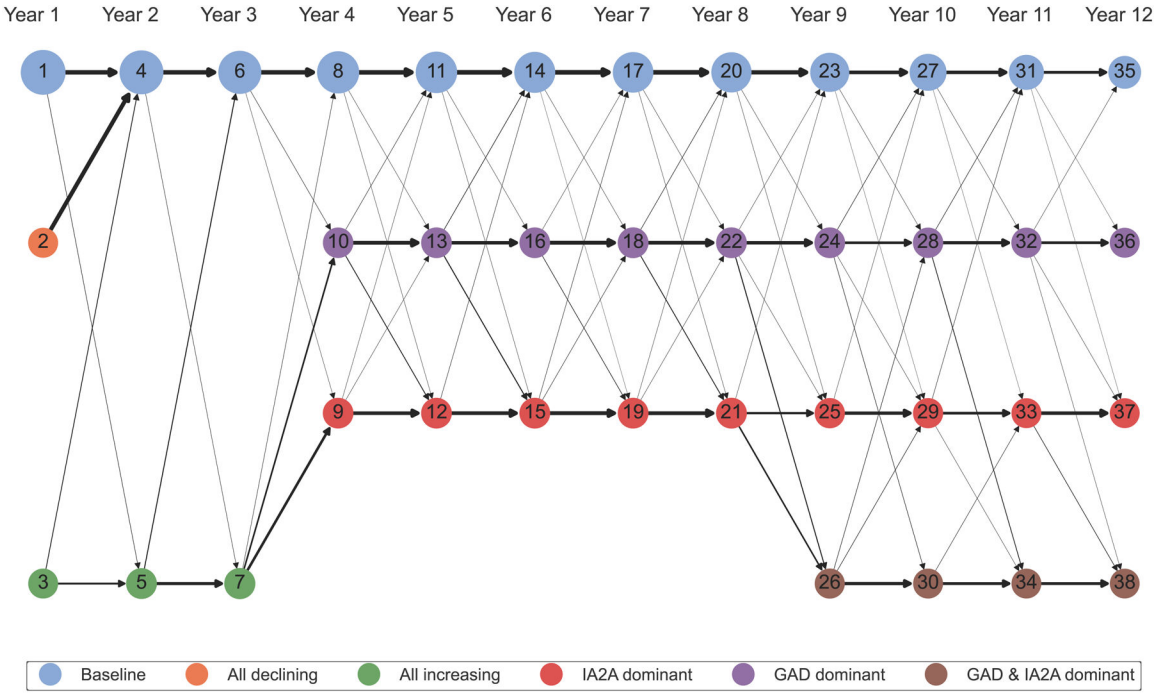


Figure 6: Transitions in cluster membership across years 1 – 12:

Network diagram of transitions in cluster membership across Years 1 – 12. The nodes represent cluster membership in each year and are colored according to the autoantibody cluster center: baseline (blue), all declining (orange), all increasing (green), IA-2A dominant (red), GADA dominant (purple), or IA-2A and GADA dominant (brown). The numbers in the nodes represent cluster number. The size of the node correlates to the scaled number of individuals assigned to that cluster. The black arrows represent the scaled number of participants that transitioned from a given cluster to a different cluster in the subsequent year.

Table 1:

Type 1 Diabetes Risk and Titer Percentiles for Each Islet Autoantibody Cluster:

Clusters are labelled according to the autoantibody cluster center: baseline (blue, a), all declining (orange, b), all increasing (green, c), IA-2A dominant (red, d), GADA dominant (purple, e), or IA-2A and GADA dominant (brown, f). Risk for type 1 diabetes derived from time-to-event analysis for each autoantibody cluster in each year. The 1-year and 5-year incidences are presented as percentages with the 95% confidence interval (CI). “N/A” indicates years that did not have 5-year islet autoantibody data available. GADA, IA-2A, and IAA titers for the cluster centers were converted to z-scores and percentiles are reported for the 3-month intervals used in the clustering for each year.

Clustering		Type 1 Diabetes Risk		GADA Titer Percentiles				IA-2A Titer Percentiles				IAA Titer Percentiles			
Year	Cluster	1-year Risk (CI)	5-year Risk (CI)	3-month	6-month	9-month	12-month	3-month	6-month	9-month	12-month	3-month	6-month	9-month	12-month
1	Cluster 1 [a]	1.80% (1.21, 2.68)	10.07% (8.53, 11.86)	46.04	45.40	45.76	46.68	46.79	46.70	46.84	47.12	46.26	44.93	45.06	45.91
	Cluster 2 [b]	3.12% (0.449, 20.18)	3.12% (0.449, 20.18)	99.99	97.66	70.75	50.53	50.09	50.52	48.07	47.52	57.87	47.46	45.98	46.77
	Cluster 3 [c]	56.25% (40.19, 73.54)	68.75% (52.66, 83.62)	45.68	50.03	68.71	86.70	47.80	47.63	50.58	71.70	48	76.42	99.78	100
2	Cluster 4 [a]	1.21% (0.73, 2.01)	6.10% (4.87, 7.62)	45.80	46.10	46.56	47.19	47.03	47.32	47.47	47.75	45.17	45.38	45.67	46.11
	Cluster 5 [c]	27.96% (19.98, 38.26)	69.19% (59.65, 78.29)	69.72	82.21	89.59	91.62	58.60	73.56	79.82	87.22	79.25	85.56	90.75	91.81
3	Cluster 6 [a]	0.52% (0.229, 1.16)	4.80% (3.690, 6.239)	46.57	46.91	47.67	48.17	47.72	47.80	48.12	48.56	46.22	46.27	46.21	46.82
	Cluster 7 [c]	20.87% (14.51, 29.5)	67.73% (59.06, 76.14)	90.90	92.77	92.96	91.74	83.97	92.52	95.05	97.45	79.35	81.86	81.15	82.84
4	Cluster 8 [a]	1.03% (0.570, 1.859)	4.50% (3.39, 5.96)	47.02	47.07	47.42	47.93	48.14	48.27	48.44	48.82	46.14	46.06	46.18	46.45
	Cluster 9 [d]	28.93% (20.27, 40.27)	78.78% (68.95, 87.19)	73.88	71.92	71.90	70.01	99.42	99.59	99.78	99.75	76.70	73.22	72.73	74.72
	Cluster 10 [e]	5.93% (2.27, 15.03)	30.45% (20.82, 43.14)	98.08	99.23	99.70	99.69	55.13	58.07	58.59	62.55	59.97	63.38	62.72	61.31

Clustering		Type 1 Diabetes Risk		GADA Titer Percentiles				IA-2A Titer Percentiles				IAA Titer Percentiles			
Year	Cluster	1-year Risk (CI)	5-year Risk (CI)	3-month	6-month	9-month	12-month	3-month	6-month	9-month	12-month	3-month	6-month	9-month	12-month
5	Cluster 11 [a]	0.60% (0.27, 1.329)	4.34% (3.16, 5.93)	47.12	47.31	47.72	48.27	48.38	48.46	48.50	48.60	45.76	45.84	45.90	45.94
	Cluster 12 [d]	25.49% (17.59, 36.09)	74.06% (63.67, 83.44)	75.16	75.75	73.24	71.83	99.22	99.57	99.68	99.74	68.58	67.61	65.90	67.39
	Cluster 13 [e]	6.11% (2.59, 14.06)	28.49% (19.41, 40.58)	98.77	99.25	99.38	99.42	55.20	59.38	61.88	63.19	58.92	60.10	58.56	57.68
6	Cluster 14 [a]	0.63% (0.279, 1.39)	5.41% (3.980, 7.340)	47.66	47.75	47.92	48.17	48.75	48.82	48.96	48.99	45.79	45.74	45.81	46.29
	Cluster 15 [d]	20.88% (13.86, 30.75)	63.40% (52.09, 74.66)	79.87	79.76	77.94	78.04	99.29	99.68	99.67	99.71	61.49	61.40	61.03	60.08
	Cluster 16 [e]	4.86% (1.849, 12.44)	25.06% (16.09, 37.76)	98.72	99.06	99.32	99.32	54.15	55.09	55.96	59	54.60	54.76	55.84	56.22
7	Cluster 17 [a]	0.66% (0.3, 1.46)	5.05% (3.52, 7.21)	47.58	47.65	47.81	47.94	48.77	48.77	48.91	49.07	45.74	45.74	45.77	45.92
	Cluster 18 [e]	4.38% (1.67, 11.26)	31.44% (20.23, 46.77)	98.11	98.81	99.09	98.97	54.83	56.11	58.03	60.33	54.41	56.74	55.40	56.91
	Cluster 19 [d]	24.42% (16.66, 34.94)	62.73% (50.33, 75.14)	80.75	80.33	80.44	81.50	99.57	99.57	99.68	99.71	57.70	59.28	59.37	61.28
8	Cluster 20 [a]	0.70% (0.31, 1.55)	N/A	47.75	47.77	48	48.30	48.54	48.58	48.68	48.77	45.76	45.70	45.96	45.92
	Cluster 21 [d]	24.08% (16.56, 34.23)	N/A	80.31	80.16	80.05	79.28	98.96	99.30	99.39	99.31	60.04	59.90	60.97	62.60
	Cluster 22 [e]	4.81% (1.83, 12.32)	N/A	98.81	99.24	99.05	99.14	56.32	58.97	61.60	63.91	52.72	52.14	52.66	53.59
9	Cluster 23 [a]	1.05% (0.5, 2.19)	N/A	47.64	47.81	48	48.05	48.52	48.49	48.62	48.79	45.52	45.49	45.60	45.77
	Cluster 24 [e]	6.09% (2.3, 15.61)	N/A	96.97	96.68	97.59	98.55	50.28	51.52	53.36	55.99	51.62	52.61	54.46	53.18
	Cluster 25 [d]	23.71% (13.81, 38.9)	N/A	59.46	58.75	58.61	58.67	98.61	98.84	99.20	98.95	55.61	56.79	55.50	56.44

Clustering		Type 1 Diabetes Risk		GADA Titer Percentiles				IA-2A Titer Percentiles				IAA Titer Percentiles			
Year	Cluster	1-year Risk (CI)	5-year Risk (CI)	3-month	6-month	9-month	12-month	3-month	6-month	9-month	12-month	3-month	6-month	9-month	12-month
	Cluster 26 [f]	8.27% (2.73, 23.61)	N/A	99.44	99.74	99.58	99.40	99.28	99.58	99.61	99.35	56.18	57.60	56.23	56.11
10	Cluster 27 [a]	1.39% (0.66, 2.91)	N/A	47.90	48.01	48.11	48.43	48.45	48.47	48.60	48.72	45.56	45.41	45.52	45.82
	Cluster 28 [e]	3.74% (0.940, 14.24)	N/A	97.28	97.94	98.39	98.14	51.36	52.05	53.04	56.22	55.03	55.95	53.97	54.82
	Cluster 29 [d]	21.65% (11.79, 37.8)	N/A	61.44	62.45	63.14	62.27	99.12	99.32	99.03	98.98	55.09	55.52	56.26	59.57
	Cluster 30 [f]	13.5% (5.28, 32.14)	N/A	99.71	99.56	99.58	99.42	99.68	99.55	99.54	99.56	55.47	53.65	53.64	56.65
11	Cluster 31 [a]	1.25% (0.52, 2.98)	N/A	47.98	48.12	48.30	48.31	48.75	48.73	48.73	48.85	45.27	45.40	45.56	46.13
	Cluster 32 [e]	4.18% (1.06, 15.72)	N/A	97.15	97.74	97.96	98.14	49.80	49.65	50.16	52.44	50.40	49.74	49.50	49.60
	Cluster 33 [d]	14.82% (5.81, 34.96)	N/A	60.80	61.63	61.70	61.91	99.37	99.49	99.68	99.50	53.53	53.94	52.55	52.02
	Cluster 34 [f]	25.00% (10.93, 51.08)	N/A	99.24	99.38	99.28	99.03	98.59	99.12	99.41	99.16	62.82	64	61.84	62.95
12	Cluster 35 [a]	0.00% (0.0, 0.0)	N/A	47.73	47.62	47.89	48.12	48.31	48.31	48.30	48.33	44.94	44.86	44.78	44.83
	Cluster 36 [e]	3.85% (0.549, 24.31)	N/A	97.57	98.68	97.50	97.76	48.77	50.37	49.58	49.67	46.78	46.16	47	46.35
	Cluster 37 [d]	30.93% (14.78, 57.50)	N/A	60.55	60.10	59.90	58.18	99.09	98.90	98.40	98.51	52.22	51.84	51.90	50.98
	Cluster 38 [f]	23.61% (9.48, 51.73)	N/A	98.97	98.16	98.50	98.16	99.11	98.78	98.60	98.45	57.63	57.05	56.91	57.77