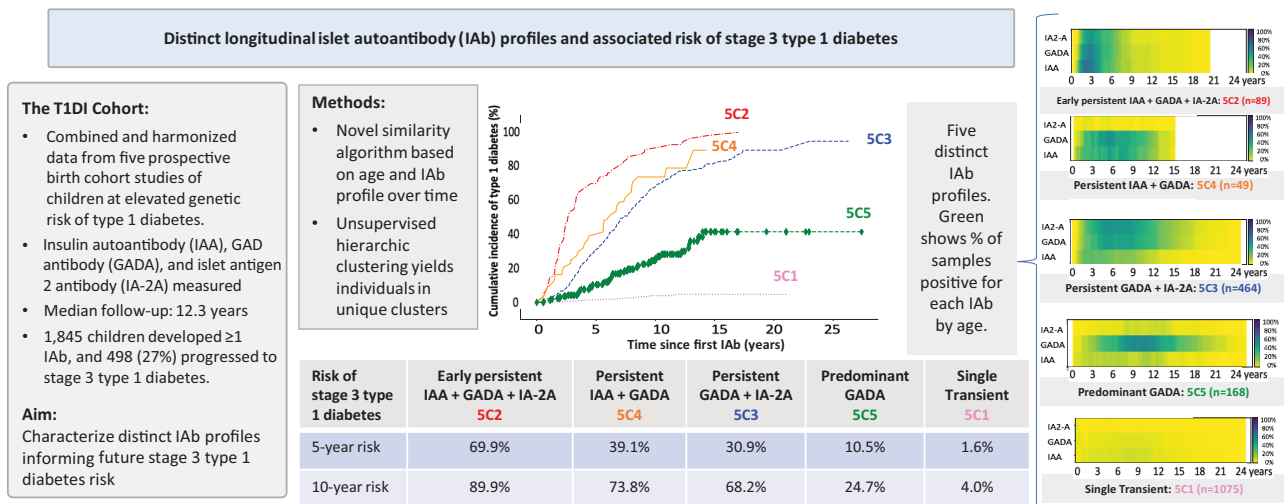


## Data-Driven Phenotyping of Presymptomatic Type 1 Diabetes Using Longitudinal Autoantibody Profiles

Mohamed Ghalwash, Vibha Anand, Kenney Ng, Jessica L. Dunne, Olivia Lou, Markus Lundgren, William A. Hagopian, Marian Rewers, Anette G. Ziegler, and Riitta Veijola, for the T1DI Study Group

Diabetes Care 2024;47(8):1–8 | <https://doi.org/10.2337/dc24-0198>



### ARTICLE HIGHLIGHTS

- Why did we undertake this study?**  
Heterogeneity of islet autoimmunity before diagnosis of type 1 diabetes necessitates further investigation. We established a large combined and harmonized data set of individuals with prospective follow-up and a novel clustering algorithm to define the patterns of islet autoimmunity.
- What is the specific question(s) we wanted to answer?**  
Which are the main longitudinal islet autoantibody profiles and their associations with progression rates from islet autoimmunity to clinical diabetes?
- What did we find?**  
We identified five main clusters of individuals with distinct autoantibody profiles and progression rates to type 1 diabetes.
- What are the implications of our findings?**  
The results strongly support the existence of disease subtypes of type 1 diabetes. These data can be used for planning targeted clinical trials to prevent type 1 diabetes.



# Data-Driven Phenotyping of Presymptomatic Type 1 Diabetes Using Longitudinal Autoantibody Profiles

<https://doi.org/10.2337/dc24-0198>

Mohamed Ghalwash,<sup>1,2</sup> Vibha Anand,<sup>3</sup> Kenney Ng,<sup>1</sup> Jessica L. Dunne,<sup>4</sup> Olivia Lou,<sup>4</sup> Markus Lundgren,<sup>5</sup> William A. Hagopian,<sup>6</sup> Marian Rewers,<sup>7</sup> Anette G. Ziegler,<sup>8</sup> and Riitta Veijola,<sup>9</sup> for the T1DI Study Group

## OBJECTIVE

To characterize distinct islet autoantibody profiles preceding stage 3 type 1 diabetes

## RESEARCH DESIGN AND METHODS

The T1DI (Type 1 Diabetes Intelligence) study combined data from 1,845 genetically susceptible prospectively observed children who were positive for at least one islet autoantibody: insulin autoantibody (IAA), GAD antibody (GADA), or islet antigen 2 antibody (IA-2A). Using a novel similarity algorithm that considers an individual's temporal autoantibody profile, age at autoantibody appearance, and variation in the positivity of autoantibody types, we performed an unsupervised hierarchical clustering analysis. Progression rates to diabetes were analyzed via survival analysis.

## RESULTS

We identified five main clusters of individuals with distinct autoantibody profiles characterized by seroconversion age and sequence of appearance of the three autoantibodies. The highest 5-year risk from first positive autoantibody to type 1 diabetes (69.9%; 95% CI 60.0–79.2) was observed in children who first developed IAA in early life (median age 1.6 years) followed by GADA (1.9 years) and then IA-2A (2.1 years). Their 10-year risk was 89.9% (95% CI 81.9–95.4). A high 5-year risk was also found in children with persistent IAA and GADA (39.1%) and children with persistent GADA and IA-2A (30.9%). A lower 5-year risk (10.5%) was observed in children with a late appearance of persistent GADA (6.1 years). The lowest 5-year diabetes risk (1.6%) was associated with positivity for a single, often reverting, autoantibody.

## CONCLUSIONS

The novel clustering algorithm identified children with distinct islet autoantibody profiles and progression rates to diabetes. These results are useful for prediction, selection of individuals for prevention trials, and studies investigating various pathways to type 1 diabetes.

Clinical diagnosis of type 1 diabetes is preceded by asymptomatic islet autoimmunity (IA) with the presence of islet autoantibodies (IAbs) such as insulin autoantibody (IAA), antibody against GAD (GADA), and islet antigen 2 antibody (IA-2A). IA predicts development of type 1 diabetes, but heterogeneous islet autoantibody (IAb) profiles and dynamics before diagnosis require further investigation. The rate of progression from the

<sup>1</sup>T.J. Watson Research Center, IBM, Yorktown Heights, NY

<sup>2</sup>Faculty of Science, Ain Shams University, Cairo, Egypt

<sup>3</sup>T.J. Watson Research Center, IBM, Cambridge, MA

<sup>4</sup>JDRF, New York, NY

<sup>5</sup>Department of Clinical Sciences, Lund University/Clinical Research Centre, Skåne University Hospital, Malmö, Sweden

<sup>6</sup>Pacific Northwest Research Institute, Seattle, WA

<sup>7</sup>Department of Pediatrics, Barbara Davis Center for Diabetes, Denver, CO

<sup>8</sup>Institute of Diabetes Research, German Research Center for Environmental Health, Helmholtz Zentrum München, Munich-Neuherberg, Germany

<sup>9</sup>Research Unit of Clinical Medicine, Medical Research Center, Department of Pediatrics, University of Oulu and Oulu University Hospital, Oulu, Finland

Corresponding author: Riitta Veijola, [riitta.veijola@oulu.fi](mailto:riitta.veijola@oulu.fi)

Received 30 January 2024 and accepted 16 May 2024

This article contains supplementary material online at <https://doi.org/10.2337/figshare.25872457>.

© 2024 by the American Diabetes Association. Readers may use this article as long as the work is properly cited, the use is educational and not for profit, and the work is not altered. More information is available at <https://www.diabetesjournals.org/journals/pages/license>.

initiation of IA to type 1 diabetes is highly variable (1–6). Several studies have investigated characteristics of different subgroups of patients who progress differently to type 1 diabetes, which ultimately could be useful for identifying children at risk of rapid or slow progression to the disease. For example, individuals have been categorized by family history of type 1 diabetes, genetic risk, age at development of IA, or type of first appearing IAb. In young children, IAA is the most sensitive and IA-2A is the most specific predictor for development of diabetes, and children who develop multiple IAbs early in life progress faster to diabetes than those who develop multiple IAbs at a later age (7–10). The TEDDY study showed that the presence of IA-2A predicts fast progression to clinical diabetes in young children (11). However, the development of IAbs over time can be heterogeneous; therefore, further investigation is required to define the IAb patterns associated with rapid or slow progression and understand the role of phenomena such as reversion of individual IAbs (12). In our collaboration, the T1DI (Type 1 Diabetes Intelligence) study group, we combined and harmonized data from five prospective studies of type 1 diabetes (13) for such analyses.

In most previous studies, subgroups of IAb profiles have been defined based on clinically predefined hypotheses and not on data-driven analyses. Interestingly, a study from BABYDIAB (14) reported a clustering algorithm for a cohort of 88 children with multiple positive IAbs using their longitudinal autoantibody profiles. However, the clustering algorithm did not account for the age at which the IAb was detected, and it did not differentiate between different IAb types (IAA, IA-2A, GADA), although the specificity of IAbs may influence the rate of progression. In addition, no principled approach to determine the number of clusters was used. In a more recent report, an improved clustering algorithm was presented to take into account the timing of changes in IAb development, and the algorithm was applied to 370 children who developed multiple autoantibody types in the evenly and frequently sampled TEDDY study (15). However, that algorithm may not be applicable to data with irregular measurements, such as the data in the T1DI study or in clinical practice. In the current report, we present a novel clustering algorithm that addresses these limitations.

We applied it to a large cohort of 1,845 IAb-positive children from the T1DI study cohort to identify putative disease subtypes in children with different IAb profiles. This data-driven approach revealed clusters of children with distinct characteristics. In addition, we explored in the youngest children developing IA whether the type of IAb in the first positive sample was associated with progression to type 1 diabetes.

## RESEARCH DESIGN AND METHODS

### Study Populations

We analyzed data from the T1DI study cohort, which combined five prospective longitudinal cohorts: the DIPP (Type 1 Diabetes Prediction and Prevention) study from Finland (4), the DiPiS (Diabetes Prediction in Skåne Study) from Sweden (16), the DAISY (Diabetes Autoimmunity Study in the Young) from the U.S. (17), the DEW-IT (Diabetes Evaluation in Washington) study from the U.S. (18), and the BABYDIAB study from Germany (8). The analyzed study cohort consisted of 1,845 individuals (male 822, female 1,023) who developed at least one IAb and had at least three visits during follow-up. Of these participants, 498 developed type 1 diabetes during follow-up. The median follow-up time was 12.3 years (interquartile range 8.3, 15.4), and the median age at first IAb test was 3.5 months (interquartile range 2.6, 8.9). There were 1,034 individuals from DIPP, 201 from DiPiS, 270 from DAISY, 183 from DEW-IT, and 157 from BABYDIAB. Individuals were categorized into four HLA DR-DQ risk groups based on published odds ratios for type 1 diabetes (A very high risk, B high risk, C slightly elevated risk, and D average to low risk) (13).

### IAb Measurements and Definitions

IAbs were measured from serum or plasma samples by using standard methods, as described in each original study (13). IAb profile here refers to positivity at a single visit. For example, as shown in Supplementary Fig. 1, the IAb profile of participant A at age 3.2 years indicates that GADA was positive but IAA and IA-2A were negative. IAb dynamics refers to the progression of the IAb profile over time (i.e., how the IAb profile changes over time). First appearance of IAb refers to the time of the first measurement where the IAb was observed to be positive. Seroconversion

refers to the positivity of the same IAb in two consecutive visits. Seroconversion age is the age at the first of these two measurements. Persistent IAb refers to the IAb becoming stable positive in two or more consecutive measurements and not reverting to negative. Transient IAb indicates that the IAb was positive once but then reverted to negativity and became stable negative (e.g., pos-neg-neg-neg). Fluctuating IAb indicates that the IAb was switching between positivity and negativity and was unstable (e.g., pos-neg-pos-neg). Reverting IAb indicates that the IAb was found to be stable positive and then reverted to stable negative (e.g., pos-pos-pos-neg-neg-neg).

### Time-Aware Clustering Algorithm

This novel algorithm aligns two individuals based on their IAb dynamics. Essentially, it computes and assigns a similarity score based on how closely the sequences of the two IAb profiles match each other. If the IAb profiles in the two individuals progress similarly, a higher similarity score (close to 1) is assigned, indicating that the two individuals have similar IAb dynamics. We extended the clustering algorithm described by Endesfelder (14) and present a novel approach that takes into account 1) the age at which the IAbs were measured, 2) the type of IAbs at that age, and 3) whether the individuals match either positive or negative measurements. Detailed development of the algorithm is described in the Supplementary Materials. In addition, we used the proportion of ambiguously clustered pairs (PAC) method to determine the number of clusters that best describe the data (Supplementary Fig. 3) (19). The novel approach consists of the following steps:

### Profile Similarity Score

Similarity was computed between two IAb profiles by considering which type of IAb was positive and when (what age) it was detected (20). We used individual autoantibody profiles across time to compute the similarity score (Supplementary Fig. 1). A similarity score of 0 indicated that the two individuals were dissimilar with respect to their IAb sequences, whereas a score of 1 indicated that the two sequences of IAb were the same. First, each individual was encoded in a binary matrix where the columns were IAb types, the rows were visits, and the

entries indicated whether each IAb was positive or negative. Because a positive IAb match between two individuals was rarer than a negative IAb match, we assigned a weight to an IAb match based on its prevalence in the data. For example, if the IAb was rarely observed as positive, a higher weight was assigned to emphasize the positive match for that IAb between two profiles, because it was unlikely to happen otherwise. In addition, two IAb profiles that were closer in terms of time (i.e., the age of the two participants) were considered more similar than two profiles that occurred farther apart, given that the profiles were otherwise similar. To address this, the similarity between two profiles was weighted based on the time gap between the profiles and is represented by the time-aware profile distance score.

#### Dynamic Time Warping Alignment

Two IAb profiles were aligned by considering their temporal dynamics using the similarity score between a profile pair of two individuals and the computed time-aware profile distance scores (21).

#### Hierarchical Clustering

Individuals were clustered by using the computed and aligned similarity scores for all pairs of individuals. The hierarchical/agglomerative clustering algorithm was applied.

#### Number of Clusters

The number of stable clusters in the study cohort was determined using the PAC algorithm (19), which uses bootstrapping for stability analysis to identify robust clusters.

#### Statistical Analysis

The characteristics of individuals in each cluster were analyzed separately. Kaplan-Meier survival analysis was used to examine the rate of progression from the first positive autoantibody to diagnosis of stage 3 type 1 diabetes. The log-rank test was used to compute the statistical significance between the progression rates, and a  $P$  value  $<0.05$  was considered significant.

#### Data and Resource Availability

The data supporting these findings are available from the authors on reasonable

request. The data are not publicly available because of privacy regulations.

## RESULTS

When applying the novel data-driven time-aware clustering algorithm to the 1,845 IAb-positive individuals, five large clusters of individuals with different IAb profiles and dynamics were discovered: 5C1, 5C2, 5C3, 5C4, and 5C5 (Fig. 1A–E). The characteristics of children in these clusters, including the 5- and 10-year risks of type 1 diabetes, are presented in Table 1. The cumulative incidences of type 1 diabetes for the five clusters are shown in Fig. 1F. The children in cluster 5C2 ( $n = 89$ ) had a high progression rate to diabetes (5-year risk of 69.9% and 10-year risk of 89.9%). All three IAb were persistently positive in cluster 5C2; however, on average, IAA appeared first and was quickly followed by GADA and IA-2A, and diabetes was diagnosed at a median age of 4.3 years (Fig. 1B and Table 1). The individuals in clusters 5C3 ( $n = 464$ ) and 5C4 ( $n = 49$ ) also had relatively high progression rates to diabetes (5-year risks of 30.9% and 39.1%, respectively), but their median age at diagnosis was older than that in cluster 5C2 (9.2 and 8.2 years, respectively). Clusters 5C3 and 5C4 were characterized by two persistent IAb: 5C3 by persistent IA2A and GADA but less frequent and reverting IAA, and 5C4 by persistent IAA and GADA but rare and transient IA-2A (Fig. 1C and D and Table 1). Most individuals in cluster 5C5 ( $n = 168$ ) (Fig. 1E and Table 1) developed late persistent GADA. During follow-up, IAA was positive in 64.3% of the individuals in 5C5, but it often reverted, and only 7.2% were IAA positive in their last positive sample. IA-2A was positive in 34.5% of the individuals in 5C5, rarely in the first positive sample and infrequently also in the last positive sample. The 5-year risk of type 1 diabetes in cluster 5C5 was 10.5%. Finally, cluster 5C1 ( $n = 1,075$ ) (Fig. 1A and Table 1) consisted of individuals who developed mainly single and transient IAb, most often IAA or GADA, and had an estimated 5-year risk of type 1 diabetes of only 1.6%. The proportions of the four class II HLA genotype groups did not differ between the five main clusters (Table 1).

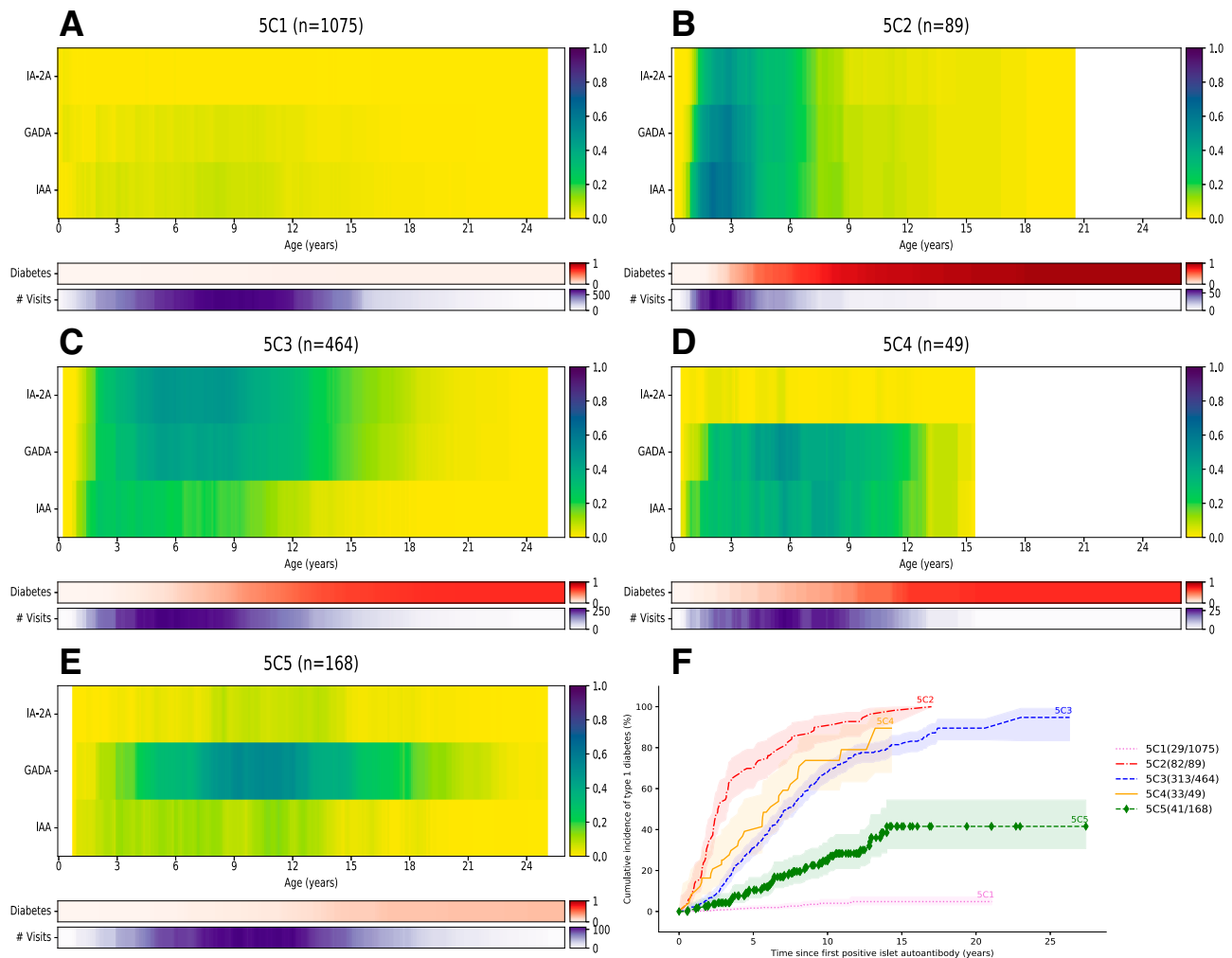
The PAC analysis also revealed 18 smaller but still stable clusters of individuals with typical progression patterns

of IA. These 18 clusters were subclusters for the five main clusters (Fig. 2A). The cumulative type 1 diabetes incidence in the children in the 18 subclusters is shown in Fig. 2B. Subclusters 18C1, 18C2, and 18C4 originated from 5C1 and were characterized with positivity for a single, often reverting, IAb (IA-2A, IAA, or GADA, respectively) and low 5- and 10-year diabetes risks (Supplementary Fig. 4A–C and Supplementary Table 2). Eight of the remaining 15 subclusters (18C7, 18C10, 18C12, 18C13, 18C14, 18C16, 18C17, and 18C18) had at least 10 participants per cluster, and their IAb dynamics and clinical characteristics are also shown in Supplementary Fig. 4D–K and Supplementary Table 2).

We also validated the hypothesis that the long-term risk of type 1 diabetes is different based on the first appearing IAb in early life. We observed that children who seroconverted before age 2 years and developed two or more IAb ( $n = 254$ ) represented two groups based on IAA status in the first positive sample. In group A ( $n = 194$ ), IAA was positive with or without other IAb in the first positive sample, whereas in group B ( $n = 60$ ), IAb other than IAA were present in the first positive sample. Progression rates to diabetes in these two groups were significantly different (log-rank test  $P = 0.0002$ ) (Fig. 3). Children who were IAA positive in their first positive sample had a 10-year risk of 71.4% (95% CI 64.5–77.9), 15-year risk of 82.7% (95% CI 76.1–88.3), and 20-year risk of 92.1% (95% CI 81.3–97.9), whereas children without IAA in their first positive sample had a 10-year risk of 47.6% (95% CI 35.2–61.8), 15-year risk of 53.9% (95% CI 40.2–68.7), and 20-year risk of 61.6% (95% CI 44.0–79.3).

## CONCLUSIONS

Identification of IAb temporal patterns that are associated with progression rates to type 1 diabetes is crucial to improve prediction, understand various disease pathways, and design targeted prevention trials. For example, young children with IAb profiles and dynamics associated with a high 5-year progression rate need early interventions specific for their disease subtype. In addition, information on distinct IAb profiles and expected risk of progression to clinical diabetes will be valuable when providing counseling on diabetes risk to individuals with IAb



**Figure 1**—Distinct longitudinal IAb profiles and associated risk of stage 3 type 1 diabetes. *A–E*: Longitudinal patterns of the three islet autoantibodies (IAA, GADA, and IA-2A, shown on x-axis of each panel) in the five clusters of children with distinct dynamics of IA. Age (years) is shown on x-axis. Green indicates fraction of positivity for each antibody across all measurements at each age (dark green indicates mostly positive samples, light green indicates that only a small proportion of samples were positive, and yellow shows that samples measured at corresponding age were negative). Middle section in each panel (Diabetes) shows in red the cumulative proportion of children progressing to stage 3 type 1 diabetes. For example, light red color in panel *B* (cluster 5C2) indicates that children start to progress to diabetes from age  $\sim 2$  years, and dark red shows that most of them progress to diabetes during follow-up. Similarly, bottom section in each panel (#Visits) shows in purple the number of measurements collected at each age. *F*: Cumulative incidence of stage 3 type 1 diabetes (% with 95% CI) for the five clusters of children with distinct IAb patterns discovered by the novel clustering algorithm. Number of individuals progressing to type 1 diabetes and total number of participants in each cluster are reported for each curve.

positivity. Here we present a novel clustering algorithm that matches similar children based on their IAb dynamics and accounts for the various types of IAbs and the age when an IAb first appeared. Moreover, we show that this algorithm can be applied to individuals with different follow-up protocols. We applied the algorithm to a cohort of 1,845 IAb-positive children and discovered five main groups of children with different IAb profiles and dynamics, which were strongly associated with different progression rates to stage 3 type 1 diabetes. Two of the discovered clusters had similar patterns of IAb development and diabetes risk reported in earlier studies, but three of the clusters

represented less well-described novel patterns of IA.

Earlier efforts to identify IAb temporal patterns had some limitations and were applied in much smaller cohorts than our T1D cohort. We developed a novel clustering algorithm that addresses several limitations of the earlier algorithms and applied it to a large cohort of individuals with one or more positive IAb (IAA, GADA, and/or IA-2A). The T1D cohort is the largest cohort of IAb-positive children currently available and represents individuals from five prospective studies from Europe and the U.S. This also imposed some challenges in using the combined data, such as how to cluster children with different

sampling frequencies (0.6–2.0 years in the five original cohorts). Our novel time-aware clustering algorithm addresses the variation in sampling interval and matches children based on types and combinations of IAbs over time and by the ages at which the IAbs were observed. In addition, the novel algorithm accounts for the overall variation in positivity for the various autoantibody types in the cohort. The prevalence of IAA in the entire T1D cohort was 5%, and that of GADA was 8%, showing that IAA was less often observed than GADA (Supplementary Table 1). Having two measurements with positive IAA should therefore be weighted more heavily than two measurements with

**Table 1—Characteristics of children positive for IABs in five main clusters**

	5C1	5C2	5C3	5C4	5C5
<i>N</i> of individuals	1,075	89	464	49	168
Male sex	492	32	199	22	77
<i>N</i> of individuals with type 1 diabetes	29	82	313	33	41
Age at diagnosis, years	8.3 (5.7, 11.1)	4.3 (3.0, 7.5)	9.2 (6.3, 12.2)	8.2 (4.4, 11.2)	12.5 (9.8, 14.5)
HLA class II group, %					
A	17	34	30	43	22
B	43	45	51	33	48
C	18	11	7	12	12
D	21	10	13	12	17
Follow-up time, years	13.4 (10.0, 15.5)	4.2 (3.0, 6.8)	10.3 (7.0, 13.7)	8.5 (4.5, 11.5)	14.5 (11.3, 17.8)
First sample positive*					
IAA	4.3 (600)	1.6 (89)	3.1 (374)	3.0 (48)	6.0 (108)
GADA	4.6 (481)	1.9 (89)	3.9 (425)	3.1 (48)	6.1 (159)
IA-2A	3.7 (142)	2.1 (89)	4.3 (464)	4.7 (9)	8.1 (58)
Last sample positive*					
IAA	6.3 (600)	4.2 (89)	8.4 (374)	8.5 (48)	10.3 (108)
GADA	6.0 (481)	4.2 (89)	10.1 (425)	8.4 (48)	14.2 (159)
IA-2A	4.4 (142)	4.2 (89)	10.3 (464)	5.5 (9)	11.1 (58)
Age at seroconversion, year <sup>†</sup>	5.0 (2.0, 8.1)	1.5 (1.0, 2.6)	3.5 (1.8, 6.0)	3.1 (1.5, 5.3)	6.1 (3.1, 9.1)
<i>N</i> of seroconverted individual <sup>†</sup>	250	89	464	49	162
IAb profile <sup>‡</sup>					
IAA only	49.9/50.0	49.4/0.0	19.2/0.0	55.1/2.0	22.0/0.6
GADA only	37.2/38.7	10.1/0.0	29.5/0.2	24.5/4.1	57.7/79.2
IA-2A only	8.5/9.2	1.1/0.0	8.8/19.4	0.0/0.0	0.6/8.9
IAA + GADA, negative IA-2A	1.2/0.7	28.1/0.0	15.7/0.0	20.4/93.9	16.7/5.4
IAA + IA-2A, negative GADA	0.8/0.4	6.7/1.1	6.2/15.1	0.0/0.0	0.0/1.2
GADA + IA-2A, negative IAA	2.1/0.9	3.4/0.0	10.1/46.6	0.0/0.0	2.4/4.8
IAA + GADA + IA-2A	0.3/0.1	1.1/98.9	10.3/18.8	0.0/0.0	0.6/0.0
Risk of type 1 diabetes, %					
5 year	1.6 (1.0–2.7)	69.9 (60.0–79.2)	30.9 (26.8–35.5)	39.1 (26.6–54.8)	10.5 (6.6–16.6)
10 year	4.0 (2.7–5.9)	89.9 (81.9–95.4)	68.2 (63.3–72.9)	73.8 (59.2–86.4)	24.7 (18.0–33.4)

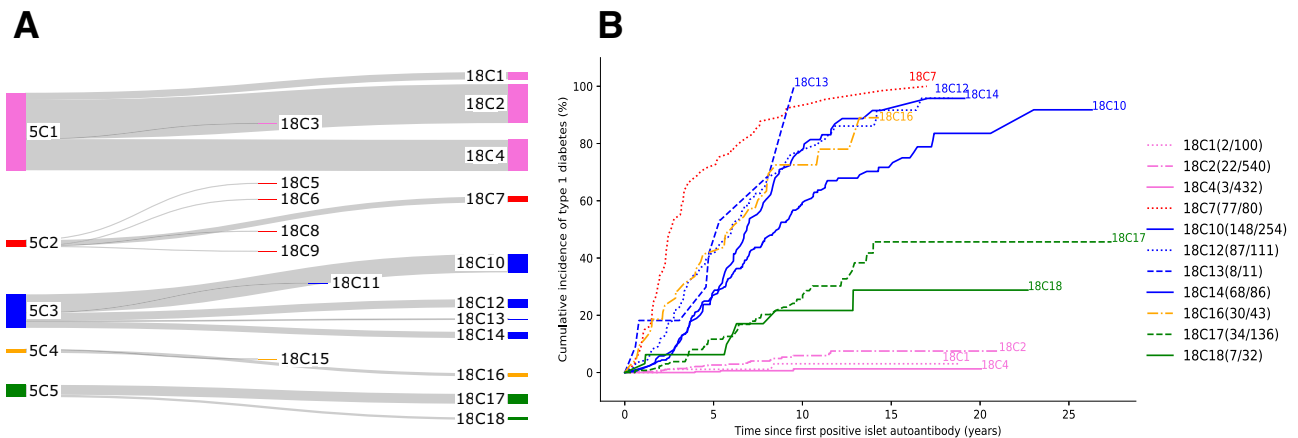
Data are given as median (IQR) or % (95% CI) unless otherwise indicated. \*Median age (*n* of individuals positive for each IAb), years. <sup>†</sup>Seroconversion was defined as first of two consecutive visits with positivity for same type of IAb. <sup>‡</sup>IAb profile included seven mutually exclusive possibilities in first/last positive sample (%).

positive GADA. We addressed this issue by weighting the positivity match differently for different IAb types. Furthermore, we used a principled way to determine the number of stable clusters in the cohort.

We discovered five distinct clusters, each with a typical IAb pattern and timing. In addition, the novel algorithm distinguished a total of 18 smaller subclusters that were stable and represented subgroups of children from the five main clusters. Cluster 5C2 and subcluster 18C7 included children with early initiation of IA, with persistent IAA, GADA, and IA-2A, and diagnosis of type 1 diabetes at age ~4 years. This group represents a specific and aggressive subtype of type 1 diabetes that has also been recognized in earlier studies (1,10,15,22). Interestingly, the 464 children in cluster 5C3 were also characterized by

persistent GADA, IA-2A, and IAA at an early age, but they frequently lost IAA during follow-up and were diagnosed with diabetes later, at age ~9 years. This pattern in cluster 5C3 has been less described in the literature, with inconsistent conclusions. The pattern of being positive for the three IABs and then losing IAA was also described by Endesfelder et al. (14), who reported a lower 10-year risk of 23% in children losing IAA compared with 76% in those who maintained IAA positivity. In contrast, the TEDDY study reported that in multipositive children, IAA reversion had little effect on the risk of type 1 diabetes (12). Cluster 5C4 included children with persistent positivity for IAA and GADA from age 3 years, but only few of them were ever positive for IA-2A. This is also a novel group, with a high 10-year risk of type 1 diabetes of 73.8%,

and demonstrates that although the presence of IA-2A has been shown to be an important predictor of diabetes in young children (11,22), it is not necessary for disease development. The third relatively novel cluster included children in cluster 5C5 (and subcluster 18C17) who typically developed persistent GADA at age ~6 years, less frequently had positivity for IAA or IA-2A, and progressed to type 1 diabetes rather late, at age ~12 years. Single positivity for GADA is commonly observed in adults diagnosed with type 1 diabetes but is less described in children. The pattern in cluster 5C5 is in line with observations from the TEDDY study, reporting a significantly higher 5-year risk of diabetes in children with single but stable positivity for IAA than in those with single and stable GADA positivity (15). The largest cluster,



**Figure 2**—Evolution of 18 subclusters from the five main clusters of children with distinct patterns of IA. *A*: Evolution from the five main clusters with distinct color codes to 18 subclusters. *B*: Cumulative incidence of stage 3 type 1 diabetes (%) for 11 subclusters with distinct IAB patterns discovered by the novel clustering algorithm. Red indicates subcluster 18C7, which represents a majority of individuals from cluster 5C2 and has the highest risk of progression to type 1 diabetes. Blue represents subclusters 18C10, 18C12, 18C13, and 18C14, which are associated with high risk of progression. Similarly, yellow represents subcluster 18C16 and is also linked to high risk of progression. Green depicts subclusters 18C17 and 18C18, with intermediate risk of progression. Purple indicates three subclusters with positivity for single autoantibody and associated low risk of progression. Data for seven subclusters that included <10 children are not included. Number of individuals progressing to type 1 diabetes and total number of individuals in each subcluster are reported for each curve.

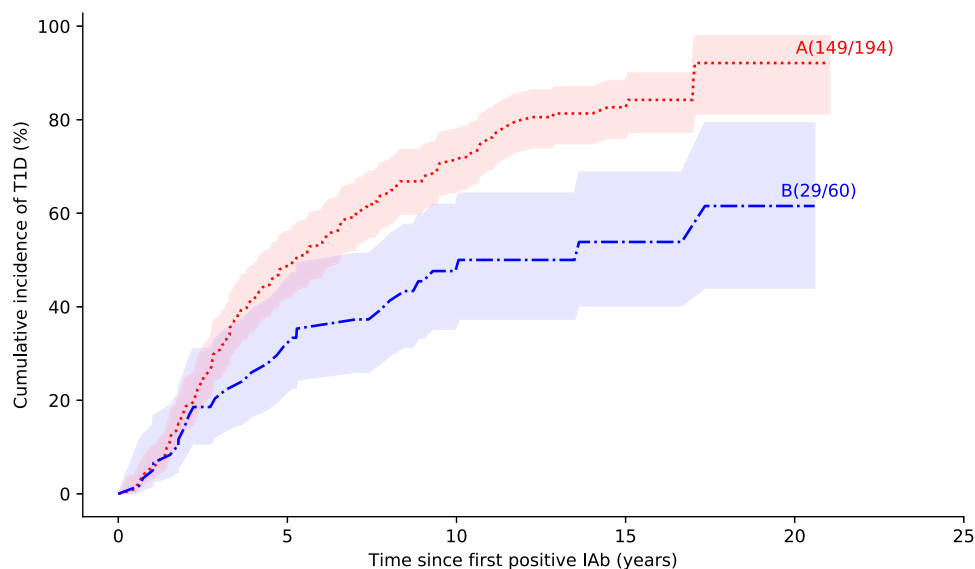
5C1, represented individuals with mainly single and transient IAbs associated with a low risk of progression to diabetes. This group is well known from earlier studies but can only be identified by observing IAbs in consecutive samples after first detection.

It is tempting to hypothesize that the initiatory factors of early and aggressively progressing disease subtypes are different from those triggering a different IAb pattern later in childhood. Therefore, for example,

the children in clusters 5C2 and 5C5 should be studied in greater detail to identify possible specific initiators of the disease, such as whether they were environmental or genetic or a combination of both. The effective preventive interventions may be different for children in these groups.

It has been hypothesized that there are different subtypes of type 1 diabetes based on the order of IAb appearance (23,24). IAA as the first appearing IAb has been associated with fast progression to

type 1 diabetes, whereas GADA as the first appearing IAb has been associated with a more moderate progression rate. This was also apparent in the clusters discovered by our novel algorithm. In our validation analysis of individuals with early seroconversion before age 2 years, positivity for IAA alone or in combination with other IAbs in the first positive sample was associated with a much higher progression rate to diabetes compared with IAA negativity in the first positive sample.



**Figure 3**—Cumulative incidence of type 1 diabetes (T1D; % with 95% CI) for children who seroconverted before age 2 years and developed positivity for two or more islet autoantibodies. Red (group A;  $n = 194$ ) represents children with IAA positivity in first positive sample. Blue (group B;  $n = 60$ ) represents children who were negative for IAA in first positive sample. Progression rates were significantly different between groups A and B (log-rank test  $P = 0.0002$ ). For group A, 5-year diabetes risk was 48.5% (95% CI 41.6–55.8), and 10-year risk was 71.3% (95% CI 64.5–77.8). For group B, 5-year diabetes risk was 31.5% (95% CI 21.1–45.3), and 10-year risk was 47.6% (95% CI 35.2–61.8).



The strengths of this study include a large study cohort of individuals with IAb positivity, a long median follow-up time of 12.3 years, and measurements of IAA, GADA, and IA-2A from more than 260,000 individual samples. The T1DI cohort also includes a variety of class II HLA genotypes (13), and therefore, the T1DI study population resembles the general population more than the cohorts included in previous clustering analyses. There are also some limitations in our study. We did not include ZnT8A data, because the measurements were not systematically performed in the individual study cohorts. ZnT8A measurements may further improve risk stratification in the future. Current knowledge suggests that ZnT8A is not often seen at the initiation of IA but appears later during follow-up. If ZnT8A had been available for our analyses, it is likely that we would have seen even more heterogeneity in IAb patterns compared with our current results. No external validation cohort was available to repeat our results, and this will require further work in the future, before potential implementation of the algorithm for risk stratification. The clustering algorithm could potentially be developed to include additional factors, such as IAb levels and comprehensive genetic risk profiles of the participants for similarity matching. However, our results clearly demonstrate that IA in children, although heterogeneous, can still be used to group by disease subtype. It is important to learn more about disease subtypes in the future. Large data sets and data-driven analyses, as exemplified here, are excellent tools to clarify distinct pathogenetic pathways of type 1 diabetes and inform future research and practice.

In conclusion, we developed a clustering algorithm to define temporal patterns of IA, applied it in a large cohort of IAb-positive children, and discovered distinct patterns of IA that were strongly associated with varying progression rates to type 1 diabetes. These findings have implications for type 1 diabetes prediction and future prevention trials. Furthermore, in ongoing screening programs of IA, these findings may guide the interpretation of individual test results.

**Acknowledgments.** The authors thank the participants of the DAISY, DiPiS, DIPP, DEW-IT,

and BABYDIAB studies and the dedicated study personnel at the study sites.

**Funding.** This work was supported by funding from JDRF (IBM: 1-RSC-2017-368-I-X, 1-IND-2019-717-I-X; DAISY: 1-SRA-2019-722-I-X, 1-RSC-2017-517-I-X, 5-ECR-2017-388-A-N; DiPiS: 1-SRA-2019-720-I-X, 1-RSC-2017-526-I-X; DIPP: 1-RSC-2018-555-I-X, 1-SRA-2019-721-I-X; DEW-IT: 1-SRA-2019-719-I-X, 1-RSC-2017-516-I-X) as well as from the National Institutes of Health (DAISY: DK032493, DK032083, DK104351; DiPiS: DK26190). DIPP was funded by JDRF (1-SRA-2016-342-M-R, 1-SRA-2019-732-M-B), the European Union (BMH4-CT98-3314), the Novo Nordisk Foundation, the Academy of Finland (decision no. 292538) and the Centre of Excellence in Molecular Systems Immunology and Physiology Research 2012-2017 (decision no. 250114), special research funds for University Hospitals in Finland, the Diabetes Research Foundation (Finland), and the Sigrid Juselius Foundation (Finland). BABYDIAB was supported by funds from the German Federal Ministry of Education and Research to the German Center for Diabetes Research. DiPiS was funded by the Swedish Research Council (14064), the Swedish Childhood Diabetes Foundation, the Swedish Diabetes Association, the Nordisk Insulin Fund, SUS funds, the Lions Club International (district 101-S), the Royal Physiographic Society, the Skåne County Council Foundation for Research and Development, LUDC-IRC/EXODIAB funding from the Swedish Foundation for Strategic Research (decision no. IRC15-0067), and the Swedish Research Council (decision no. 2009-1039). DEW-IT was funded by the Centers for Disease Control and Prevention (URG/CCU017247), with additional support from the University of Washington Diabetes Research Center (P30 DK017047), the Hussman Foundation, and the Washington State Life Science Discovery Fund.

**Duality of Interest.** M.G., V.A., and K.N. are employees of IBM. J.L.D. performed this work as an employee of JDRF and is now an employee of Sanofi. O.L. is an employee of JDRF. No other potential conflicts of interest relevant to this article were reported.

**Author Contributions.** M.G. conceptualized and developed the time-aware clustering algorithm, was responsible for all data analyses, and drafted and revised the manuscript. V.A. contributed to the interpretation of data, revised the manuscript, and is the guarantor of data integration, as a representative of IBM. K.N. contributed to the interpretation of data and revised the methods section of the manuscript. J.L.D., O.L., M.L., W.A.H., M.R., and A.G.Z. contributed to the interpretation of data and revision of the manuscript. M.L., W.A.H., M.R., A.G.Z., and R.V., as principal investigators of the individual studies, verify the underlying data from original study sites. R.V. contributed to the analysis plan and interpretation of data and drafted and revised the manuscript. All authors contributed to the discussion and/or revision of the final draft of the manuscript. V.A. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

**Handling Editors.** The journal editors responsible for overseeing the review of the manuscript were Steven E. Kahn and Thomas Danne.

## References

- Ziegler AG, Rewers M, Simell O, et al. Seroconversion to multiple islet autoantibodies and risk of progression to diabetes in children. *JAMA* 2013;309:2473–2479
- Gorus FK, Balti EV, Messaoui A, et al.; Belgian Diabetes Registry. Twenty-year progression rate to clinical onset according to autoantibody profile, age, and HLA-DQ genotype in a registry-based group of children and adults with a first-degree relative with type 1 diabetes. *Diabetes Care* 2017;40:1065–1072
- Insel RA, Dunne JL, Atkinson MA, et al. Staging presymptomatic type 1 diabetes: a scientific statement of JDRF, the Endocrine Society, and the American Diabetes Association. *Diabetes Care* 2015;38:1964–1974
- Bauer W, Veijola R, Lempainen J, et al. Age at seroconversion, HLA genotype, and specificity of autoantibodies in progression of islet autoimmunity in childhood. *J Clin Endocrinol Metab* 2019;104:4521–4530
- Bingley PJ, Wherrett DK, Shultz A, Rafkin LE, Atkinson MA, Greenbaum CJ. Type 1 Diabetes TrialNet: a multifaceted approach to bringing disease-modifying therapy to clinical use in type 1 diabetes. *Diabetes Care* 2018;41:653–661
- Beyerlein A, Bonifacio E, Vehik K, et al.; TEDDY Study Group. Progression from islet autoimmunity to clinical type 1 diabetes is influenced by genetic factors: results from the prospective TEDDY study. *J Med Genet* 2019;56:602–605
- Kukko M, Kimpimäki T, Korhonen S, et al. Dynamics of diabetes-associated autoantibodies in young children with human leukocyte antigen-conferred risk of type 1 diabetes recruited from the general population. *J Clin Endocrinol Metab* 2005;90:2712–2717
- Giannopoulou EZ, Winkler C, Chmiel R, et al. Islet autoantibody phenotypes and incidence in children at increased risk for type 1 diabetes. *Diabetologia* 2015;58:2317–2323
- Hummel M, Bonifacio E, Schmid S, Walter M, Knopff A, Ziegler AG. Brief communication: early appearance of islet autoantibodies predicts childhood type 1 diabetes in offspring of diabetic parents. *Ann Intern Med* 2004;140:882–886
- Parikka V, Näntö-Salonen K, Saarinen M, et al. Early seroconversion and rapidly increasing autoantibody concentrations predict prepubertal manifestation of type 1 diabetes in children at genetic risk. *Diabetologia* 2012;55:1926–1936
- Jacobsen LM, Larsson HE, Tamura RN, et al.; TEDDY Study Group. Predicting progression to type 1 diabetes from ages 3 to 6 in islet autoantibody positive TEDDY children. *Pediatr Diabetes* 2019;20:263–270
- Vehik K, Lynch KF, Schatz DA, et al.; TEDDY Study Group. Reversion of  $\beta$ -cell autoimmunity changes risk of type 1 diabetes: TEDDY study. *Diabetes Care* 2016;39:1535–1542
- Anand V, Li Y, Liu B, et al.; T1DI Study Group. Islet autoimmunity and HLA markers of pre-symptomatic and clinical type 1 diabetes: joint analyses of prospective cohort studies in Finland, Germany, Sweden, and the U.S. *Diabetes Care* 2021;44:2269–2276
- Endesfelder D, Hagen M, Winkler C, et al. 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* 2016;59:2172–2180
15. Endesfelder D, Zu Castell W, Bonifacio E, et al.; TEDDY Study Group. Time-resolved autoantibody profiling facilitates stratification of preclinical type 1 diabetes in children. *Diabetes* 2019;68:119–130
16. Elding Larsson H. A Swedish approach to the prevention of type 1 diabetes. *Pediatr Diabetes* 2016;17(Suppl. 22):73–77
17. Rewers M, Bugawan TL, Norris JM, et al. Newborn screening for HLA markers associated with IDDM: Diabetes Autoimmunity Study in the Young (DAISY). *Diabetologia* 1996;39:807–812
18. Wion E, Brantley M, Stevens J, et al. Population-wide infant screening for HLA-based type 1 diabetes risk via dried blood spots from the public health infrastructure. *Ann N Y Acad Sci* 2003;1005:400–403
19. Şenbabaoğlu Y, Michailidis G, Li JZ. Critical limitations of consensus clustering in class discovery. *Sci Rep* 2014;4:6207
20. Köhler M, Beyerlein A, Vehik K, et al.; TEDDY Study Group. Joint modeling of longitudinal autoantibody patterns and progression to type 1 diabetes: results from the TEDDY study. *Acta Diabetol* 2017;54:1009–1017
21. Sakoe H, Chiba S. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans Acoust Speech Signal Process* 1978; 26:43–49
22. Vehik K, Bonifacio E, Lernmark Å, et al.; TEDDY Study Group. Hierarchical order of distinct autoantibody spreading and progression to type 1 diabetes in the TEDDY study. *Diabetes Care* 2020;43:2066–2073
23. Krischer JP, Lynch KF, Schatz DA, et al.; TEDDY Study Group. The 6 year incidence of diabetes-associated autoantibodies in genetically at-risk children: the TEDDY study. *Diabetologia* 2015;58:980–987
24. Ilonen J, Hämäläinen A, Laine AP, et al. Patterns of  $\beta$ -cell autoantibody appearance and genetic associations during the first years of life. *Diabetes* 2013;62:3636–3640