nature medicine

LETTERS

https://doi.org/10.1038/s41591-020-0930-4

Check for updates

A combined risk score enhances prediction of type 1 diabetes among susceptible children

Lauric A. Ferrat[1], Kendra Vehik[2], Seth A. Sharp[1], Åke Lernmark[3], Marian J. Rewers[4], Jin-Xiong She[5], Anette-G. Ziegler[6,7,8], Jorma Toppari[9,10], Beena Akolkar[11], Jeffrey P. Krischer[2], Michael N. Weedon[1], Richard A. Oram[1,12,36], William A. Hagopian[13,36] ✉ and TEDDY Study Group*

Type 1 diabetes (T1D)—an autoimmune disease that destroys the pancreatic islets, resulting in insulin deficiency—often begins early in life when islet autoantibody appearance signals high risk[1]. However, clinical diabetes can follow in weeks or only after decades, and is very difficult to predict. Ketoacidosis at onset remains common[2,3] and is most severe in the very young[4,5], in whom it can be life threatening and difficult to treat[6–9]. Autoantibody surveillance programs effectively prevent most ketoacidosis[10–12] but require frequent evaluations whose expense limits public health adoption[13]. Prevention therapies applied before onset, when greater islet mass remains, have rarely been feasible[14] because individuals at greatest risk of impending T1D are difficult to identify. To remedy this, we sought accurate, cost-effective estimation of future T1D risk by developing a combined risk score incorporating both fixed and variable factors (genetic, clinical and immunological) in 7,798 high-risk children followed closely from birth for 9.3 years. Compared with autoantibodies alone, the combined model dramatically improves T1D prediction at ≥2 years of age over horizons up to 8 years of age (area under the receiver operating characteristic curve ≥ 0.9), doubles the estimated efficiency of population-based newborn screening to prevent ketoacidosis, and enables individualized risk estimates for better prevention trial selection.

Type 1 diabetes (T1D) is associated with substantial heritable risk, notably from common human leukocyte antigen (HLA) variants but also from many diverse genetic loci[15]. Environmental factors increase the risk[16]. Recent attempts to predict who will develop T1D and at what age have used islet autoantibodies[17,18], metabolic status[19,20], genetic factors[21–25] and family history[26]. Longitudinal autoantibody measurement has been established as the strongest single predictor of future T1D in first-degree relatives[18] or in general populations either unselected[27] or prescreened for genetic risk[1,18,28]. Combined assessment of both fixed and time-varying risk factors improves both the prediction of T1D progression in first-degree relatives[20,21,23–25] and the accuracy of diabetes diagnoses in adult incident cases[22]. However, no T1D screening or prediction efforts to date have taken full advantage of the complementary information that age, genetic risk, family history and environmental factors offer, when combined with autoantibody status, to estimate future T1D risk in all children. Such combined modeling could markedly improve the prediction of T1D and other childhood diseases throughout early life by allowing risk assessments to reflect each individual's specific age and situation.

The Environmental Determinants of Diabetes in the Young (TEDDY) study screened 425,000 children from the United States, Sweden, Germany and Finland and prospectively studied 8,676 from birth through to 15 years of age[29]. Participants received frequent autoantibody and exposure testing, in addition to physiological and clinical measurements. We used TEDDY data to develop a model predicting T1D during the first 10 years of life. We considered features known to indicate increased T1D risk, including a recently published T1D genetic risk score (GRS2)[30], longitudinal autoantibody measurements and a variety of other medical, demographic and environmental factors[31]. This rich dataset enabled us to develop a combined risk score (CRS), targeting children with high genetic risk, to estimate T1D risk at various landmark ages and over specific time horizons.

Multiple variables are predictive of childhood T1D in univariate analyses of TEDDY data (Extended Data Fig. 1)[32,33]. These include family history in first-degree relatives, the presence of autoantibodies, T1D GRS2 (ref. [30]), the weight *z* score at 1 year of age, sinusitis episodes and country of residence. By 2 years of age, autoantibodies are already highly predictive, with a time-dependent area under the receiver operating characteristic curve (ROC AUC) of 0.75 (95% confidence interval = 0.71–0.78). The GRS2 alone had an AUC of 0.73 (0.70–0.77) despite use in a highly HLA-selected cohort where 94% of the TEDDY cohort had a GRS2 value in the top 20th percentile of a control population. We chose GRS2 because it performed best in TEDDY and other datasets[30] compared with similar GRSs (Extended Data Fig. 2 and Methods). Other T1D-associated variables, such as family history, weight *z* score, sinusitis episodes and country of residence, were less predictive (ROC AUCs of 0.51–0.56).

[1]Institute of Biomedical and Clinical Science, University of Exeter Medical School, Exeter, UK. [2]Health Informatics Institute, Morsani College of Medicine, University of South Florida, Tampa, FL, USA. [3]Department of Clinical Sciences, Lund University and Skåne University Hospital, Malmö, Sweden. [4]Barbara Davis Center for Childhood Diabetes, University of Colorado School of Medicine, Aurora, CO, USA. [5]Center for Biotechnology and Genomic Medicine, Medical College of Georgia, Augusta University, Augusta, GA, USA. [6]Institute of Diabetes Research, Helmholtz Zentrum München, Munich, Germany. [7]Forschergruppe Diabetes, Klinikum Rechts der Isar, Technische Universität München, Munich, Germany. [8]Forschergruppe Diabetes e.V., Munich, Germany. [9]Department of Pediatrics, Turku University Hospital, Turku, Finland. [10]Research Centre for Integrative Physiology and Pharmacology, Institute of Biomedicine, University of Turku, Turku, Finland. [11]National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, USA. [12]Academic Renal Unit, Royal Devon and Exeter NHS Foundation Trust, Exeter, UK. [13]Pacific Northwest Research Institute, Seattle, WA, USA. [36]These authors contributed equally: Richard A. Oram, William A. Hagopian. *A list of members and their affiliations appears at the end of the paper. ✉e-mail: wah@uw.edu

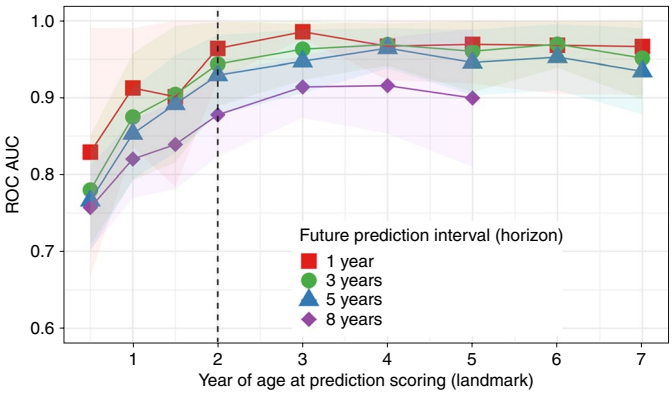


Fig. 1 | Average time-dependent ROC AUCs for the three-variable model by age at prediction scoring. The three model components include autoantibodies, GRS2 and family history. Four different prediction horizons are denoted by different colors. The vertical dotted line corresponds to the landmark age of 2 years featured in Fig. 2a. The shaded region indicates the 95% confidence interval of the mean.

We determined which combination of associated variables from Extended Data Fig. 1 best predicted future T1D at each landmark age using stepwise selection. Overall, a three-variable CRS incorporating autoantibodies, GRS2 and family history performed best in cross-validated time-dependent ROC AUC analysis (Fig. 1) and using the Akaike information criterion. ROC AUCs were all ≥0.92 for landmarks ≥2 years and horizons up to 5 years. Compared with a model using all six associated variables, the three-variable model performed equally well (Fig. 2).

We tested whether additional variables might be eliminated from the three-variable CRS model. Models with GRS2 and family history outperformed GRS2 alone (Extended Data Fig. 3). We asked whether a three-variable CRS was better than autoantibodies alone, the latter being the most established approach for T1D prediction. The three-variable score outperformed autoantibody status alone in univariate Cox regression using the Akaike information criterion, again with higher ROC AUCs upon cross-validation (Extended Data Fig. 4). This effect was greatest at the landmark age of 2 years for all time horizons (Fig. 2a) but also was clear at the landmark age of 4 years for longer horizons (Fig. 2b). Nevertheless, when present, autoantibodies conferred greater hazard ratios for T1D than GRS2 or family history components (Extended Data Fig. 5). The CRS appears to help most for children not yet with autoantibodies, or with only one autoantibody (right side of the ROC curve in Fig. 2 and Extended Data Fig. 6, respectively). In TEDDY, at 2 years of age, 38% of children subsequently developing T1D during follow-up to a median age of 9.3 years will have fewer than two autoantibodies.

The three-variable CRS discrimination in this cohort corresponded to well-separated T1D and non-T1D populations in the plotted distributions of the three-variable CRS. The bimodal CRS distribution among future T1D cases reflects the model's autoantibody term, since many already have ≥1 autoantibodies by the landmark age of 2 years (Fig. 3a) and even more by the age of 4 years (Fig. 3b). Calibration plots for the three-variable model with the same 2- and 4-year landmarks (Fig. 3c,d, respectively) indicate that an increasing CRS generally corresponds to an increasing actual risk of future T1D, with a mild tendency to underestimate the disease risk of children at midrange probabilities.

We generated T1D progression risk estimates for individual children based on the three-variable T1D CRS model, using a 2-year landmark age (Extended Data Fig. 7). At moderately high GRS2 (the 90th percentile of a background population using UK Biobank) and without family history, the risk of T1D in the next 5 years increases by ~14% with one autoantibody and by ~42% with two autoantibodies. Conversely, for a given number of autoantibodies, family history and GRS2 increase the risk fivefold when comparing moderately high GRS2 with no family history versus very high GRS2 with a positive family history (Extended Data Fig. 7 and Supplementary Table 1).

Using TEDDY data, we modeled three population-based screening strategies in incident cases diagnosed by 10 years of age from among all originally screened newborns (Fig. 4). Each was adjusted to achieve a comparable (75%) rate of identification of very high risk ≥4 weeks before onset. The first classic strategy initially selected infants with high GRS2 genetic risk, and followed them closely (defined as quarterly until 3 years of age, then every 6 months until 6 years of age, and then every year thereafter until 8 years of age). The second simple adaptive strategy selected infants with high genetic risk, followed them closely, but then recalculated the T1D CRS at annual landmarks. At each landmark, any child with a T1D probability by 10 years of age of $P < 0.008$ was eliminated from further follow-up. The third advanced adaptive strategy also selected newborn infants at high genetic risk, but then annually recalculated the T1D CRS to reallocate children between a close follow-up group and a reduced follow-up group. Reallocation was based on a calculated T1D probability in the next 2 years of ≥0.006 or <0.006, respectively.

The endpoint of these prediction strategies, via the CRS, is the estimated percentage risk of T1D onset over the stated time horizons. This guides the approach to the family regarding the risk of impending T1D onset in their child, the follow-up schedule for the child in the two adaptive strategies and consideration of prevention therapies. Although related, it is distinct from the more commonly used T1D prediction endpoint of islet autoantibodies.

Consistent with the requirements of the three-variable CRS, each follow-up evaluation updates the status of three autoantibodies (glutamic acid decarboxylase autoantibody (GADA), insulinoma antigen-2 autoantibody (IA2A) and insulin autoantibody (IAA)) and T1D family history. We compared the total number of follow-up evaluations required under each strategy to achieve our goal of 75% advance identification of new-onset cases. Simple and advanced adaptive strategies utilizing the three-variable T1D CRS required 25 and 51% fewer surveillance evaluations, respectively, compared with the standard strategy (Extended Data Figs. 8 and 9).

Effectively, the CRS identifies children requiring frequent evaluation in order to predict impending T1D onset. This includes children with no autoantibodies who nonetheless may be at high risk of early onset. For example, if children are not closely followed from birth to 1 year of age, ten children would not be warned before T1D onset, but this falls to two in ten or zero in ten using the advanced or simple adaptive strategies, respectively. Similarly, if only autoantibody positives are followed quarterly from 1–2 years of age, with others followed yearly, 11 out of 36 children developing T1D during this year would not receive advanced warning to prevent ketoacidosis. This number falls to four in 36 using the advanced adaptive strategy and zero in 36 for the simple adaptive strategy, representing important improvements at these vulnerable ages.

Our results using family history, genotyped risk and autoantibodies highlight that the most accurate disease prediction—particularly of complex disease—will come from integration of multiple risk factors. This has been demonstrated in other settings (for example, Q risk[34]) but to our knowledge has not been demonstrated for a complex childhood disease. It is notable that exposures such as sinusitis, weight and residence country[33], while significant when considered alone, did not appear to add predictive value in a combined model. Using only three variables in our final model lessened the possibility of overfitting, while also minimizing information collection at follow-up evaluations.

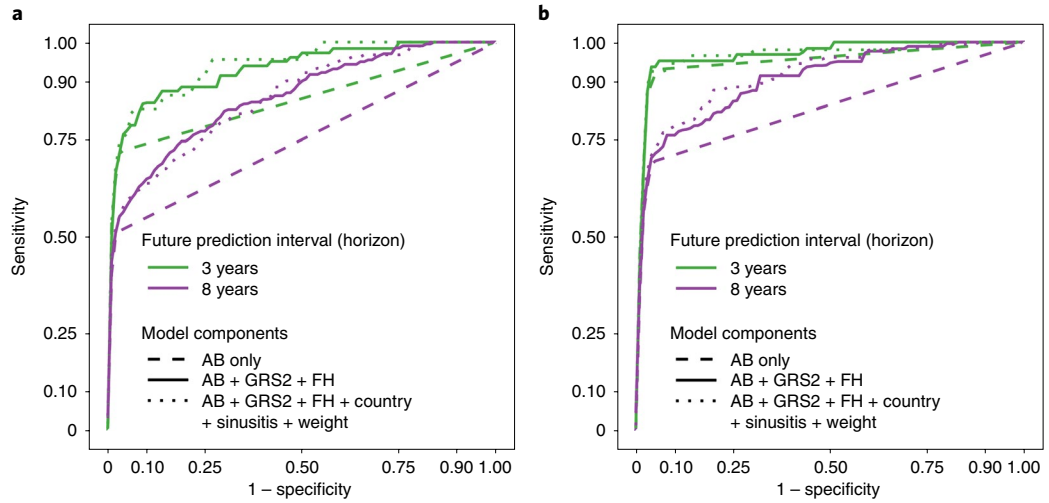


Fig. 2 | ROC curves derived from models incorporating different numbers of variables. a,**b**, Dotted, solid and dashed lines denote the use of all six variables, three variables and autoantibodies only, respectively. The curves in **a** use a landmark age of 2 years, with prediction horizons of 3 or 8 years, as indicated. The curves in **b** use a landmark age of 4 years, also with prediction horizons of 3 or 8 years, as indicated. AB, autoantibodies; FH, family history.

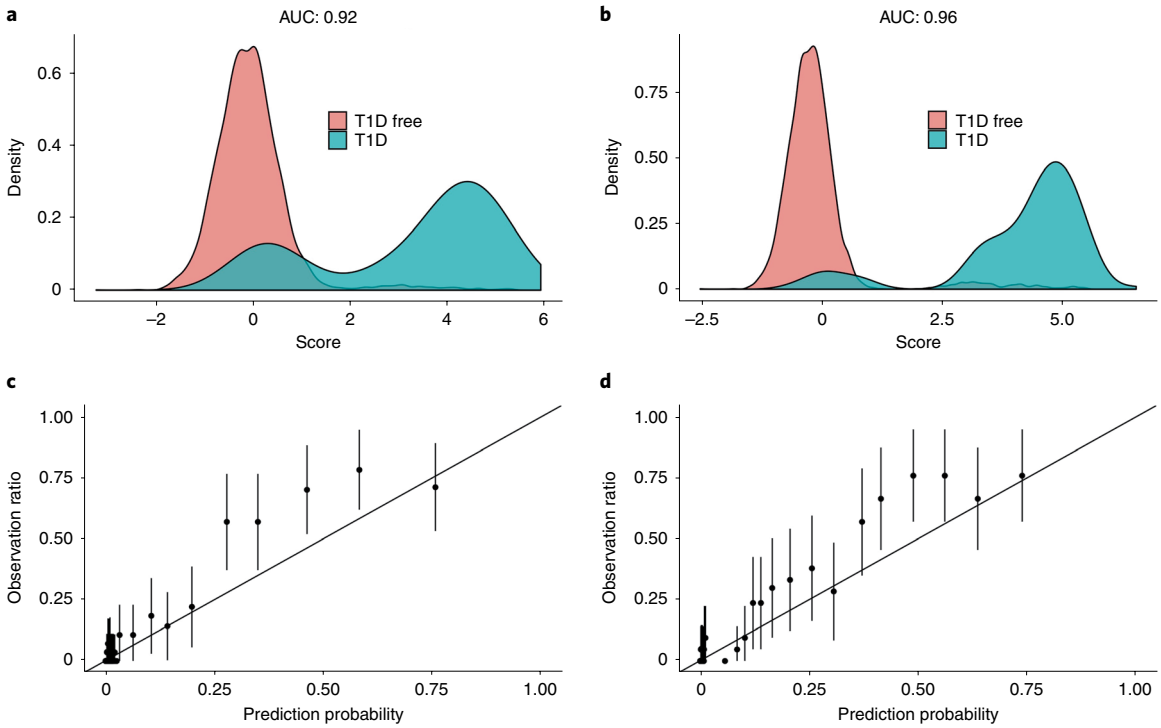


Fig. 3 | Performance of the three-variable model at a 5-year horizon. a,**b**, Score distributions using a 2-year (**a**) or 4-year landmark age (**b**). AUC ROC values are noted. The T1D CRS was generated using the linear predictor of the parametric part of the hazard function of the Cox model. **c**,**d**, Calibration plots using a 2-year (**c**) or 4-year landmark age (**d**). The predictions were grouped into bins corresponding to centiles, then each bin prevalence (the ratio of plots in this bin with observed T1D endpoints to the total number of plots in this bin) was calculated. Each point represents the mean of each bin, and each error bar indicates the 95% confidence interval of that mean. For the 2-year and 4-year landmark ages, a total of $n = 6{,}805$ and $n = 5{,}973$ children with an autoantibody test in the previous 6 months were analyzed, respectively.

An increasing area of interest is whether the prediction of common, complex pediatric diseases can provide practical health benefits at a population level. The identification of babies with rare, treatable diseases, such as phenylketonuria, by postnatal heel prick testing is commonplace in modern healthcare systems, and early treatment is life changing[35]. For T1D, the most life-threatening complication is diabetic ketoacidosis in the very young, which can lead to serious neurological complications and incurs high treatment costs. The detection of islet-specific autoantibodies before onset allows advance warning and close monitoring, which lessens the incidence of diabetic ketoacidosis[10–12]. Successful advance warning in infants requires autoantibody surveillance to start early in life and to occur frequently, since progression from autoantibody positivity to hyperglycemia is most rapid in infants[1,36].

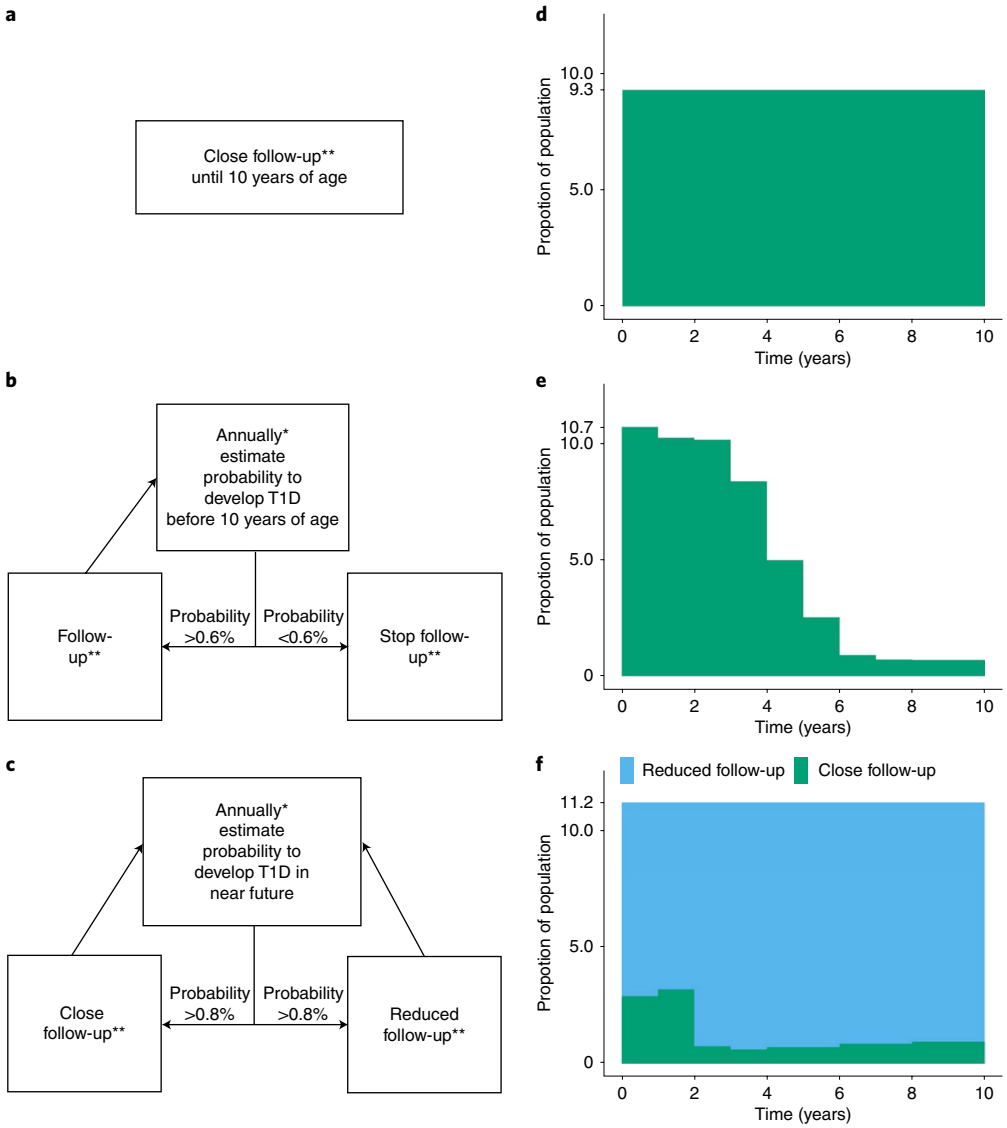


Fig. 4 | Three strategies for population-based newborn screening and surveillance follow-up. a–**c**, Flow charts of the three models, termed classic (**a**), simple adaptive (**b**) and advanced adaptive (**c**). The two adaptive models use the three-variable T1D CRS to dynamically define the follow-up schedule for each individual in the cohort. **d**–**f**, Proportions of children remaining under follow-up in the classic (**c**), simple adaptive (**d**) and advanced adaptive scenarios (**e**), using cohort simulations on TEDDY data ($n = 7{,}798$). Single asterisks indicate that the risk is recalculated annually during the first 4 years life, then every 2 years thereafter. Double asterisks indicate that close follow-up is quarterly to 3 years of age, then biannually to 6 years of age, then annually to 8 years of age, while reduced follow-up is annually to 4 years of age, then every 2 years.

Without a genotyping component, T1D risk prediction either requires surveillance of too many children, or requires selection by family history, which misses most cases. Substituting a polygenic GRS for more commonly used HLA genotypes, and then combining this information with other variables into a CRS for adaptive surveillance, greatly improves the efficiency and therefore may allow reconsideration of public health-based newborn screening for T1D and related autoimmune diseases. In this setting, the ability of the CRS to provide accurate individual risk estimates is an important added benefit, although it must be carefully explained that not all children identified as being at high risk will develop childhood T1D[37].

Greater precision in identifying individuals at high risk of impending T1D may greatly improve the cost and feasibility of early life intervention trials, such as those testing expensive vaccines[14,38], by reducing the number of participants needed to appropriately power early-stage studies[24,30]. It could also lessen potential exposure to immunosuppressive drugs in children less likely to develop T1D. Finally, it opens the possibility of earlier disease mitigation before dysglycemia appears and when more functioning β cells remain.

Our study has several limitations. TEDDY, like many birth cohort studies of T1D, preselected newborns at high HLA risk in order to observe sufficient disease endpoints to achieve study goals. After removal of these HLA effects, the remaining difference in genetic risk is much smaller between TEDDY children who developed T1D and those who did not (Extended Data Fig. 10). Therefore, the CRS yields a lower calculated AUC ROC at landmarks <2 years of age (Fig. 1) than would be expected during general population use. In both Type 1 Diabetes Genetics Consortium and UK Biobank cohorts[30], the GRS2 alone had an AUC ROC of >0.92 for T1D. This implies that a three-variable CRS incorporating GRS2 may have greater ROC AUCs at young ages in an unselected pediatric population than in TEDDY. However, validating the model in children with a wide range of GRS2 risk awaits the availability of

such a dataset. On another note, subtle abnormalities in blood glucose levels by a variety of measures are emerging as an important marker of T1D progression close to diagnosis[19]. These are not typically measured in prediabetes, and were not measured in children lacking multiple autoantibodies in TEDDY, and so cannot be included in our model. Also, the CRS model was less discriminant among children with two or more autoantibodies. Larger studies with more power are needed to study this specific group at very young ages. Finally, the modeled genes and environmental features common to European and US populations may perform differently in other populations with distinct genetic backgrounds or environments. Analyses to date in other cohorts suggest that GRS2 should perform well in all major US ethnicities[39] and that autoantibodies are similarly predictive in this regard[40]. Along these lines, the model validated well when tested in the TEDDY data from each single country using the other three countries to fit the model (Supplementary Table 2). However, to fully address these concerns, external validation in other birth cohorts is an essential next step.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41591-020-0930-4.

Received: 24 October 2019; Accepted: 8 May 2020;
Published online: 07 August 2020

References

1. Ziegler, A. G. et al. Seroconversion to multiple islet autoantibodies and risk of progression to diabetes in children. *J. Am. Med. Assoc.* **309**, 2473–2479 (2013).
2. Dabelea, D. et al. Trends in the prevalence of ketoacidosis at diabetes diagnosis: the SEARCH for Diabetes in Youth Study. *Pediatrics* **133**, e938–e945 (2014).
3. Alonso, G. T. et al. Diabetic ketoacidosis at diagnosis of type 1 diabetes in Colorado children, 2010–2017. *Diabetes Care* **43**, 117–121 (2020).
4. Jefferies, C. et al. 15-year incidence of diabetic ketoacidosis at onset of type 1 diabetes in children from a regional setting (Auckland, New Zealand). *Sci. Rep.* **5**, 10358 (2015).
5. Iovane, B. et al. Diabetic ketoacidosis at the onset of type 1 diabetes in young children: is it time to launch a tailored campaign for DKA prevention in children <5 years? *Acta Biomed.* **89**, 67–71 (2018).
6. Rewers, M. et al. Assessment and monitoring of glycemic control in children and adolescents with diabetes. *Pediatr. Diabetes* **8**, 408–418 (2007).
7. Usher-Smith, J. A., Thompson, M. J., Sharp, S. J. & Walter, F. M. Factors associated with the presence of diabetic ketoacidosis at diagnosis of diabetes in children and young adults: a systematic review. *Br. Med. J.* **343**, d4092 (2011).
8. Desai, D., Mehta, D., Mathias, P., Menon, G. & Schubart, U. K. Health care utilization and burden of diabetic ketoacidosis in the U.S. over the past decade: a nationwide analysis. *Diabetes Care* **41**, 1631–1638 (2018).
9. Glaser, N. et al. Risk factors for cerebral edema in children with diabetic ketoacidosis. *N. Engl. J. Med.* **344**, 264–269 (2001).
10. Barker, J. M. et al. Clinical characteristics of children diagnosed with type 1 diabetes through intensive screening and follow-up. *Diabetes Care* **27**, 1399–1404 (2004).
11. Winkler, C., Schober, E., Ziegler, A. G. & Holl, R. W. Markedly reduced rate of diabetic ketoacidosis at onset of type 1 diabetes in relatives screened for islet autoantibodies. *Pediatr. Diabetes* **13**, 301–306 (2012).
12. Elding Larsson, H. et al. Children followed in the TEDDY study are diagnosed with type 1 diabetes at an early stage of disease. *Pediatr. Diabetes* **15**, 118–126 (2014).
13. Meehan, C., Fout, B., Ashcraft, J., Schatz, D. A. & Haller, M. J. Screening for T1D risk to reduce DKA is not economically viable. *Pediatr. Diabetes* **16**, 565–572 (2015).
14. Bonifacio, E. et al. Effects of high-dose oral insulin on immune responses in children at high risk for type 1 diabetes. *J. Am. Med. Assoc.* **313**, 1541–1549 (2015).
15. Noble, J. A. & Valdes, A. M. Genetics of the HLA region in the prediction of type 1 diabetes. *Curr. Diabetes Rep.* **11**, 533–542 (2011).
16. Rewers, M. & Ludvigsson, J. Environmental risk factors for type 1 diabetes. *Lancet* **387**, 2340–2348 (2016).
17. Beyerlein, A., Wehweck, F., Ziegler, A.-G. & Pflueger, M. Respiratory infections in early life and the development of islet autoimmunity in children at increased type 1 diabetes risk. *JAMA Pediatr.* **167**, 800–807 (2013).
18. Verge, C. et al. Prediction of type I diabetes in first-degree relatives using a combination of insulin, GAD, and ICA512bdc/IA-2 autoantibodies. *Diabetes* **45**, 926–933 (1996).
19. Sosenko, J. M. et al. Glucose and C-peptide changes in the perionset period of type 1 diabetes in the diabetes prevention trial–type 1. *Diabetes Care* **31**, 2188–2192 (2008).
20. Redondo, M. J. et al. A type 1 diabetes genetic risk score predicts progression of islet autoimmunity and development of type 1 diabetes in individuals at risk. *Diabetes Care* **41**, 1887–1894 (2018).
21. Winkler, C. et al. Feature ranking of type 1 diabetes susceptibility genes improves prediction of type 1 diabetes. *Diabetologia* **57**, 2521–2529 (2014).
22. Oram, R. A. et al. A type 1 diabetes genetic risk score can aid discrimination between type 1 and type 2 diabetes in young adults. *Diabetes Care* **39**, 337–344 (2015).
23. Beyerlein, A. et al. Progression from islet autoimmunity to clinical type 1 diabetes is influenced by genetic factors: results from the prospective TEDDY study. *J. Med. Genet.* **56**, 602–605 (2019).
24. Bonifacio, E. et al. Genetic scores to stratify risk of developing multiple islet autoantibodies and type 1 diabetes: a prospective study in children. *PLoS Med.* **15**, e1002548 (2018).
25. Hippich, M. et al. Genetic contribution to the divergence in type 1 diabetes risk between children from the general population and children from affected families. *Diabetes* **68**, 847–857 (2019).
26. Familial risk of type I diabetes in European children. The EURODIAB ACE Study Group & The EURODIAB ACE Substudy 2 Study Group. *Diabetologia* **41**, 1151–1156 (1998).
27. LaGasse, J. M. et al. Successful prospective prediction of type 1 diabetes in schoolchildren through multiple defined autoantibodies: an 8-year follow-up of the Washington State Diabetes Prediction Study. *Diabetes Care* **25**, 505–511 (2002).
28. Ziegler, A. G. et al. Primary prevention of beta-cell autoimmunity and type 1 diabetes—the Global Platform for the Prevention of Autoimmune Diabetes (GPPAD) perspectives. *Mol. Metab.* **5**, 255–262 (2016).
29. Hagopian, W. A. et al. TEDDY—The Environmental Determinants of Diabetes in the Young: an observational clinical trial. *Ann. NY Acad. Sci.* **1079**, 320–326 (2006).
30. Sharp, S. A. et al. Development and standardization of an improved type 1 diabetes genetic risk score for use in newborn screening and incident diagnosis. *Diabetes Care* **42**, 200–207 (2019).
31. Rewers, M. et al. The Environmental Determinants of Diabetes in the Young (TEDDY) study: 2018 update. *Curr. Diabetes Rep.* **18**, 136 (2018).
32. Krischer, J. P. et al. Genetic and environmental interactions modify the risk of diabetes-related autoimmunity by 6 years of age: the TEDDY study. *Diabetes Care* **40**, 1194–1202 (2017).
33. Krischer, J. P. et al. Predicting islet cell autoimmunity and type 1 diabetes: an 8-year TEDDY study progress report. *Diabetes Care* **42**, 1051–1060 (2019).
34. Hippisley-Cox, J., Coupland, C. & Brindle, P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. *Br. Med. J.* **357**, j2099 (2017).
35. National Institutes of Health Consensus Development Panel National Institutes of Health Consensus Development Conference Statement: phenylketonuria: screening and management, October 16–18, 2000. *Pediatrics* **108**, 972–982 (2001).
36. Edge, J. A., Ford-Adams, M. E. & Dunger, D. B. Causes of death in children with insulin dependent diabetes 1990–96. *Arch. Dis. Child.* **81**, 318–323 (1999).
37. Rose, G. Sick individuals and sick populations. *Int. J. Epidemiol.* **14**, 32–38 (1985).
38. Hyöty, H., Leon, F. & Knip, M. Developing a vaccine for type 1 diabetes by targeting coxsackievirus B. *Expert Rev. Vaccines* **17**, 1071–1083 (2018).
39. Redondo, M. J., Oram, R. A. & Steck, A. K. Genetic risk scores for type 1 diabetes prediction and diagnosis. *Curr. Diabetes Rep.* **17**, 129 (2017).
40. Cheng, B.-W. et al. Autoantibodies against islet cell antigens in children with type 1 diabetes mellitus. *Oncotarget* **9**, 16275–16283 (2018).

TEDDY Study Group

Colorado Clinical Center

Marian Rewers[4], Aaron Barbour[4], Kimberly Bautista[4], Judith Baxter[4], Daniel Felipe-Morales[4], Kimberly Driscoll[4], Brigitte I. Frohnert[4], Marisa Stahl[4], Patricia Gesualdo[4], Michelle Hoffman[4], Rachel Karban[4], Edwin Liu[4], Jill Norris[4], Stesha Peacock[4], Hanan Shorrosh[4], Andrea Steck[4], Megan Stern[4], Erica Villegas[4] and Kathleen Waugh[4]

Finland Clinical Center

Jorma Toppari[9,10], Olli G. Simell[9,10], Annika Adamsson[9,10], Suvi Ahonen[14,15,16], Mari Åkerlund[14,15,16], Leena Hakola[14], Anne Hekkala[17,18], Henna Holappa[17,18], Heikki Hyöty[14,15], Anni Ikonen[17,18], Jorma Ilonen[9,19], Sinikka Jäminki[14,15], Sanna Jokipuu[9,10], Leena Karlsson[9,10], Jukka Kero[9,10], Miia Kähönen[17,18], Mikael Knip[14,15], Minna-Liisa Koivikko[17,18], Merja Koskinen[14,15], Mirva Koreasalo[14,15,16], Kalle Kurppa[14,15], Jarita Kytölä[14,15], Tiina Latva-aho[17,18], Katri Lindfors[14], Maria Lönnrot[14,15], Elina Mäntymäki[15], Markus Mattila[14], Maija Miettinen[16], Katja Multasuo[17,18], Teija Mykkänen[17,18], Tiina Niininen[14,15], Sari Niinistö[15,16], Mia Nyblom[14,15], Sami Oikarinen[14,15], Paula Ollikainen[17,18], Zhian Othmani[9,10], Sirpa Pohjola[17,18], Petra Rajala[9,10], Jenna Rautanen[15,16], Anne Riikonen[14,15,16], Eija Riski[9,10], Miia Pekkola[14,15], Minna Romo[9,10], Satu Ruohonen[9,10], Satu Simell[9,10], Maija Sjöberg[9,10], Aino Stenius[17,18], Päivi Tossavainen[17,18], Mari Vähä-Mäkilä[9,10], Sini Vainionpää[9,10], Eeva Varjonen[9,10], Riitta Veijola[17,18], Irene Viinikangas[17,18] and Suvi M. Virtanen[14,15,16,18]

Georgia/Florida Clinical Center

Jin-Xiong She[5], Desmond Schatz[20], Diane Hopkins[5], Leigh Steed[5], Jennifer Bryant[5], Katherine Silvis[5], Michael Haller[20], Melissa Gardiner[5], Richard McIndoe[5], Ashok Sharma[5], Stephen W. Anderson[21], Laura Jacobsen[20], John Marks[20] and P. D. Towe[20]

Germany Clinical Center

Anette-G. Ziegler[6,7,8], Ezio Bonifacio[22], Cigdem Gezginci[6,7,8], Anja Heublein[6,7,8], Eva Hohoff[23], Sandra Hummel[6,7,8], Annette Knopff[6,7,8], Charlotte Koch[6,7,8], Sibylle Koletzko[24], Claudia Ramminger[6,7,8], Roswith Roth[25], Jennifer Schmidt[6,7,8], Marlon Scholz[6,7,8], Joanna Stock[6,7,8], Katharina Warncke[6,7,8], Lorena Wendel[6,7,8] and Christiane Winkler[6,7,8]

Sweden Clinical Center

Åke Lernmark[3], Daniel Agardh[3], Carin Andrén Aronsson[3], Maria Ask[3], Rasmus Bennet[3], Corrado Cilio[3], Susanne Dahlberg[3], Helene Engqvist[3], Emelie Ericson-Hallström[3], Annika Björne Fors[3], Lina Fransson[3], Thomas Gard[3], Monika Hansen[3], Hanna Jisser[3], Fredrik Johansen[3], Berglind Jonsdottir[3], Helena Elding Larsson[3], Marielle Lindström[3], Markus Lundgren[3], Marlena Maziarz[3], Maria Månsson-Martinez[3], Jessica Melin[3], Zeliha Mestan[3], Caroline Nilsson[3], Karin Ottosson[3], Kobra Rahmati[3], Anita Ramelius[3], Falastin Salami[3], Anette Sjöberg[3], Birgitta Sjöberg[3], Carina Törn[3] and Åsa Wimar[3]

Washington Clinical Center

William A. Hagopian[13], Michael Killian[13], Claire Cowen Crouch[13], Jennifer Skidmore[13], Masumeh Chavoshi[13], Arlene Meyer[13], Jocelyn Meyer[13], Denise Mulenga[13], Nole Powell[13], Jared Radtke[13], Matei Romancik[13], Shreya Roy[13], Davey Schmitt[13] and Sarah Zink[13]

Pennsylvania Satellite Center

Dorothy Becker[26], Margaret Franciscus[26], MaryEllen Dalmagro-Elias Smith[26], Ashi Daftary[26], Mary Beth Klein[26] and Chrystal Yates[26]

Data Coordinating Center

Jeffrey P. Krischer[2], Sarah Austin-Gonzalez[2], Maryouri Avendano[2], Sandra Baethke[2], Brant Burkhardt[2], Martha Butterworth[2], Joanna Clasen[2], David Cuthbertson[2], Christopher Eberhard[2], Steven Fiske[2], Jennifer Garmeson[2], Veena Gowda[2], Kathleen Heyman[2], Belinda Hsiao[2], Christina Karges[2], Francisco Perez Laras[2], Qian Li[2], Shu Liu[2], Xiang Liu[2], Kristian Lynch[2], Colleen Maguire[2], Jamie Malloy[2], Cristina McCarthy[2], Hemang Parikh[2], Cassandra Remedios[2], Chris Shaffer[2], Laura Smith[2], Susan Smith[2], Noah Sulman[2], Roy Tamura[2], Dena Tewey[2], Michael Toth[2], Ulla Uusitalo[2], Kendra Vehik[2], Ponni Vijayakandipan[2], Keith Wood[2] and Jimin Yang[2]

Project scientist

Beena Akolkar[11]

Autoantibody Reference Laboratories

Liping Yu[4], Dongmei Miao[4], Polly Bingley[27], Alistair Williams[27], Kyla Chandler[27], Ilana Kelland[27], Yassin Ben Khoud[27], Huma Zahid[27] and Matthew Randell[27]

HLA Reference Laboratory

William A. Hagopian[13], Masumeh Chavoshi[13], Jared Radtke[13] and Sarah Zink[13]

Repository

Sandra Ke[28] and Niveen Mulholland[28]

SNP Laboratory

Stephen S. Rich[29], Wei-Min Chen[29], Suna Onengut-Gumuscu[29], Emily Farber[29], Rebecca Roche Pickin[29], Jonathan Davis[29], Jordan Davis[29], Dan Gallo[29], Jessica Bonnie[29] and Paul Campolieto[29]

Metagenomics and Microbiome Laboratory

Joseph F. Petrosino[30], Nadim J. Ajami[30], Richard E. Lloyd[30,31], Matthew C. Ross[30], Jacqueline L. O'Brien[30], Diane S. Hutchinson[30], Daniel P. Smith[30], Matthew C. Wong[30], Xianjun Tian[30], Tulin Ayvaz[30], Auriole Tamegnon[30], Nguyen Truong[30], Hannah Moreno[30], Lauren Riley[30], Eduardo Moreno[30], Tonya Bauch[30], Lenka Kusic[30], Ginger Metcalf[32], Donna Muzny[32], Harshavardhan Doddapaneni[32] and Richard Gibbs[32]

Other contributors

Kasia Bourcier[33], Thomas Briese[34], Suzanne Bennett Johnson[35] and Eric Triplett[20]

Committees

Ancillary Studies

Anette G. Ziegler[6,7,8], Marian Rewers[4], Jorma Toppari[9,10], Jin-Xiong She[5], Åke Lernmark[3], William A. Hagopian[13], Jeffrey P. Krischer[2], Roy Tamura[2] and Beena Akolkar[11]

Diet

Jill Norris[4], Suvi M. Virtanen[14,15,16,18], Brigitte I. Frohnert[4], Patricia Gesualdo[4], Mirva Koreasalo[14,15,16], Maija Miettinen[16], Sari Niinistö[15,16], Anne Riikonen[14,15,16], Katherine Silvis[5], Eva Hohoff[6,7,8], Sandra Hummel[6,7,8], Christiane Winkler[6,7,8], Carin Andrén Aronsson[3], Jennifer Skidmore[13], MaryEllen Dalmagro-Elias Smith[26], Martha Butterworth[2], Qian Li[2], Xiang Liu[2], Roy Tamura[2], Ulla Uusitalo[2] and Jimin Yang[2]

Genetics

William A. Hagopian[13], Stephen S. Rich[29], Jill Norris[4], Andrea Steck[4], Jorma Ilonen[9,10], Jin-Xiong She[5], Anette G. Ziegler[6,7,8], Åke Lernmark[3], Carina Törn[3], Qian Li[2], Xiang Liu[2], Hemang Parikh[2], Beena Akolkar[11], Wei-Min Chen[29] and Suna Onengut-Gumuscu[29]

Human Subjects/Publicity/Publications

Marian Rewers[4], Jorma Toppari[9,10], Jin-Xiong She[5], Desmond Schatz[20], Anette G. Ziegler[6,7,8], Åke Lernmark[3], William A. Hagopian[13], Jeffrey P. Krischer[2], Kendra Vehik[2] and Beena Akolkar[11]

Immune Markers

Åke Lernmark[3], Corrado Cilio[3], Ezio Bonifacio[22], Marian Rewers[4], Mikael Knip[14,15], Desmond Schatz[20], William A. Hagopian[13], Jeffrey P. Krischer[2], Brant Burkhardt[2], Kristian Lynch[2], Kendra Vehik[2], Beena Akolkar[11], Liping Yu[4], Polly Bingley[27] and Kasia Bourcier[33]

Infectious Agents

Heikki Hyöty[14,15], Eric Triplett[35], Richard Lloyd[32], Marian Rewers[4], Patricia Gesualdo[4], Kathleen Waugh[4], Maria Lönnrot[14,15], Åke Lernmark[3], Daniel Agardh[3], Corrado Cilio[3], Helena Elding Larsson[3], William A. Hagopian[13], Michael Killian[13], Brant Burkhardt[2], Kristian Lynch[2], Kendra Vehik[2], Beena Akolkar[11] and Thomas Briese[34]

Laboratory Implementation

Kathleen Waugh[4], William A. Hagopian[13], Desmond Schatz[20], Michael Killian[13] and Beena Akolkar[11]

Psychosocial

Suzanne Bennett Johnson[35], Roswith Roth[25], Judith Baxter[4], Kimberly Driscoll[4], Desmond Schatz[20], Joanna Stock[6,7,8], Åke Lernmark[3], Steven Fiske[2], Xiang Liu[2], Kristian Lynch[2], Laura Smith[2] and Kendra Vehik[2]

Quality Assurance

Marian Rewers[4], Judith Baxter[4], Åke Lernmark[3], Jeffrey P. Krischer[2], William A. Hagopian[13] and Beena Akolkar[11]

Steering

Marian Rewers[4], Åke Lernmark[6,7,8], William A. Hagopian[13], Jorma Toppari[9,10], Jin-Xiong She[5], Anette-G. Ziegler[6,7,8], Jeffrey P. Krischer[2] and Beena Akolkar[11]

Study Coordinators

Judith Baxter[4], Michael Killian[13], Kimberly Bautista[4], Patricia Gesualdo[4], Michelle Hoffman[4], Rachel Karban[4], Jill Norris[4], Kathleen Waugh[4], Annika Adamsson[9,10], Miia Kähönen[17,18], Tiina Niininen[14,15], Aino Stenius[17,18], Eeva Varjonen[9,10], Diane Hopkins[5], Leigh Steed[5], Jennifer Bryant[5], Melissa Gardiner[5], John Marks[20], Claudia Ramminger[6,7,8], Joanna Stock[6,7,8], Christiane Winkler[6,7,8], Carin Andrén Aronsson[3], Berglind Jonsdottir[3], Jessica Melin[3], Michael Killian[13], Claire Cowen Crouch[13], Denise Mulenga[13], Cristina McCarthy[2], Laura Smith[2], Susan Smith[2], Roy Tamura[2] and Suzanne Bennett Johnson[35]

Celiac Disease

Daniel Agardh[3], Edwin Liu[4], Sibylle Koletzko[24], Kalle Kurppa[14,15], Marisa Stahl[4], Michelle Hoffman[4], William A. Hagopian[13], Kalle Kurppa[14,15], Katri Lindfors[3], Satu Simell[9,10], Leigh Steed[5], Carin Andrén Aronsson[3], Michael Killian[13] and Roy Tamura[2]

Clinical Implementation

Michael Haller[20], Helena Elding Larsson[3], Brigitte I. Frohnert[4], Patricia Gesualdo[4], Michelle Hoffman[4], Andrea Steck[4], Jorma Toppari[9,10], Miia Kähönen[17,18], Riitta Veijola[17,18], Leigh Steed[5], Laura Jacobsen[20], John Marks[20], Joanna Stock[6,7,8], Katharina Warncke[6,7,8], Markus Lundgren[3], Åsa Wimar[3], William A. Hagopian[13], Claire Cowen Crouch[13], Xiang Liu[2], Roy Tamura[2] and Kendra Vehik[2]

[14]Tampere University, Tampere, Finland. [15]Tampere University Hospital, Tampere, Finland. [16]National Institute for Health and Welfare, Helsinki, Finland. [17]University of Oulu, Oulu, Finland. [18]Oulu University Hospital, Oulu, Finland. [19]University of Kuopio, Kuopio, Finland. [20]University of Florida, Gainesville, FL, USA. [21]Pediatric Endocrine Associates, Atlanta, GA, USA. [22]Center for Regenerative Therapies, TU Dresden, Dresden, Germany. [23]Department of Nutritional Epidemiology, University of Bonn, Bonn, Germany. [24]Department of Gastroenterology, Dr. von Hauner Children's Hospital, Ludwig Maximillians University, Munich, Germany. [25]Karl-Franzens-Universität Graz, Graz, Germany. [26]Children's Hospital of Pittsburgh, Pittsburgh, PA, USA. [27]Bristol Medical School, University of Bristol, Bristol, UK. [28]NIDDK Biosample Repository at Fisher BioServices, Rockville, MD, USA. [29]Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA. [30]Alkek Center for Metagenomics and Microbiome Research, Baylor College of Medicine, Houston, TX, USA. [31]Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, TX, USA. [32]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA. [33]National Institute of Allergy and Infectious Diseases, Bethesda, MD, USA. [34]Columbia University, New York, NY, USA. [35]Florida State University, Tallahassee, FL, USA.

Methods

TEDDY participants. TEDDY is a prospective cohort study designed to identify environmental causes of T1D[41]. From 2004–2010, 424,788 newborns were screened at six US and European centers for high-risk HLA genotypes. TEDDY then enrolled 8,676 eligible infants with the intent to follow them until 15 years of age. The three major eligible HLA DR–DQ haplotypes are DR3–DQA1*0501–DQB1*0201, DR4–DQA1*0301–DQB1*0302 and DR8–DQA1*0401–DQB1*0402. These are referred to by their DR haplotypes and form the four TEDDY-eligible haplogenotypes DR3/4, DR4/4, DR4/8 and DR3/3. The frequencies of all eligible HLA haplogenotypes for each center have been published[42]. Historical data from the TEDDY centers suggest that ~50% of childhood T1D cases carry one of these four included haplogenotypes.

Follow-up. TEDDY children were followed prospectively from 3–4 months of age, with visits every 3 months until 4 years of age. Each evaluation tested the three islet antibodies (GADA, IA2A and IAA), changes in family history, as well as other measurements specified by the TEDDY protocol. After 4 years of age, children with any islet autoantibodies remained on quarterly visits, while antibody-negative children were evaluated every 6 months. Of 8,676 TEDDY enrollees, 7,798 were analyzed herein on the basis of full autoantibody testing, single-nucleotide polymorphism (SNP) genotyping on the ImmunoChip array, and carrying one of the four major TEDDY-eligible HLA haplogenotypes. At the time of analysis, the median follow-up was 9.3 years (range: 1–168 months; interquartile range: 54–132 months), covering 65,331 person-years of observation. Children were followed prospectively until 15 years of age or until T1D onset, as defined using the American Diabetes Association's criteria for diagnosis[41]. In this dataset, 305 children developed T1D. Local Institutional Review Board approval, parental informed consent, and child assent where relevant, were obtained for all participants without exception. The study was also monitored by an External Evaluation Committee of the US National Institutes of Health. More details can be found in the Life Sciences Reporting Summary.

TEDDY study measurements. The TEDDY study measures a wide range of background information and environmental exposures on the cohort. Background information includes self-reported race and ethnicity, geographic residence country and TEDDY clinical center. TEDDY registers family history of T1D in the mother, father or sibling. Medical history includes pregnancy factors such as infections and cesarean section, and childhood factors such as medications and illnesses. Parental questionnaires captured incidences of the child's febrile illnesses, respiratory infections (common cold, sinus infection, ear infection, bronchitis and pneumonia) and gastrointestinal infections[43]. Serum collected at each clinic visit was analyzed for the presence of autoantibodies to GADA, IA2A and IAA in two separate core laboratories using harmonized radiobinding assays incorporating extensive quality control[44]. Only persistent autoantibodies positive for the same antigen, as confirmed by both core laboratories in two consecutive samples, were considered in the current analyses.

Generation of the T1D GRS. TEDDY cohort children were genotyped on an Illumina Infinium ImmunoChip SNP array[45]. Before imputation, we performed SNP variant quality control filtering on SNP genotype missingness (<1%), Hardy–Weinberg equilibrium ($P < 1 \times 10^{-6}$) and minor allele frequency (<1%). For variants in HLA (chromosome 6: 27–35 megabase pairs), due to the HLA-based cohort selection[42], we omitted Hardy–Weinberg equilibrium-based filtering in order to retain key variants. Sample quality control checks for sex discordance, individual genotype missingness (<1%) and principal component analysis resulted in the exclusion of 85 subjects. After quality control, 167,350 SNPs and 7,798 individuals were available for analysis. The ImmunoChip data were then imputed to the 1000 Genomes reference panel, yielding 37.1 million SNPs at an imputation quality of >0.8. Independent of this TEDDY dataset, we have described three T1D GRSs: our original 30-SNP T1D GRS[22]; the 40-SNP combined TEDDY GRS[21]; and a recently published more comprehensive 67-SNP T1D GRS2 (ref. [30]). The latter was newly generated in TEDDY for this analysis. Most of the GRS2 SNPs were genotyped directly, but 21 were imputed with $r^2 \geq 0.75$ and four were imputed with $r^2 = 0.358$–0.544 (Supplementary Table 3). These SNP genotypes were used to generate continuous numerical risk score values for GRS2.

Variable selection. A broad list of features were evaluated for potential use in a CRS. We evaluated whether incorporating features that change in an individual over time (for example, the development of autoantibodies), along with fixed characteristics such as GRS, could improve the prediction of future T1D risk. Evaluations used time-dependent ROC AUC and *P* values (two-sided Wald test) computed at a landmark age of 2 years and a horizon of 8 years on $n = 6{,}805$ individuals. All were required to be available in a typical clinical setting, such as initial genetic screening for a panel of SNP markers followed by a standard blood sample and medical history during each follow-up evaluation. To reduce the chance of false discovery and overfitting, we also required each variable to be previously established as associated with T1D in published TEDDY analyses and in the background literature (Extended Data Fig. 1). The numbers of diabetic and nondiabetic children in each of these variable categories are shown in Supplementary Table 3.

Simplification of risk factors. We combined information on T1D in a sibling, father or mother to create a single 'any first-degree relative with T1D' variable denoted family history. Likewise, we combined the GADA, IA2A and IAA variables to create a single variable representing the number of persistent islet autoantibodies. We then compared the model performance using each summary variable versus that using all of the corresponding fully detailed variables. The summary variables, family history and autoantibodies were each equally as informative as their individual components (data not shown). T1D GRS2 (ref. [30]) outperformed previous GRSs used in TEDDY[21] and elsewhere[22] at all landmarks and for horizon time prediction, with an average univariate AUC of 0.73 versus 0.63 (Extended Data Fig. 2) using $n = 7{,}798$ individuals. Therefore, only GRS2 was considered in our modeling, which left us with six variables to consider (family history, autoantibodies, GRS2, weight *z* score, sinusitis episodes and residence country).

CRS model construction. We used an approach where CRS generation occurred at fixed time points at and after birth, using all of the information available up to that time. Participants were assumed negative for islet autoantibodies at birth based on extrapolation from published TEDDY incidence data[46]. The score was revised at each later time point as information became updated. This approach has been termed landmarking[47,48] and takes advantage of the TEDDY study design, where risk factors are measured repeatedly in an individual at different time points during childhood. Only patients without T1D at the landmark age of interest are included in analyses. The visit was assigned to occur at the formal visit age if it complied with the protocol-approved visit window. Selected landmark ages were at different visit times: birth and then 1, 1.5, 2, 3, 4, 5, 6 and 7 years of age, representing nine different models. Another important feature of survival analyses is the future prediction time interval after the landmark, termed the horizon time. The numbers of children at each landmark age were: 7,798 (birth), 7,563 (1 year), 7,123 (1.5 years), 6,805 (2 years), 6,316 (3 years), 5,973 (4 years), 5,706 (5 years), 5,517 (6 years) and 5,323 (7 years). For example, a landmark at 2 years and horizon time of 5 years means that we aim to predict whether a child will develop T1D by age 7 years using a CRS generated on a nondiabetic child at 2 years of age. Horizon times used in this study were 1, 3, 5 and 8 years.

The CRSs were generated using a Cox regression model. Our goal was to maximize the predictive accuracy while minimizing the number of variables required. We initially selected variables that were independently significant using the Wald test[49]. At each landmark, we sought to find the best combination of variables to predict T1D, by performing bidirectional stepwise selection with the Bayesian information criterion[50].

Model evaluation. Since TEDDY is a prospective cohort study where participants progress to T1D over time and are subject to censoring, time dependency must be incorporated into the predictive assessment of the CRS. We used time-dependent analysis of ROC AUC to evaluate model performance at the various landmark ages and horizon times. We used threefold cross-validation (repeated ten times) to assess model precision and to reduce overestimation of model performance. To compare models, we used the R timeROC package developed by Blanche et al.[51]. Overall, we selected a set of variables that gave optimal prediction at the various landmarks and horizon times according to the best average AUC derived by cross-validation.

Screening simulation. We compared a strategy of selecting high-risk children from birth and following them all, irrespective of their changing probability of T1D (classic strategy), versus two strategies that allowed us to either stop (simple adaptive) or modulate (advanced adaptive) later follow-up visits in those individuals with a lower probability of T1D. Our goal was to test whether we could detect in advance the same number of childhood T1D cases (75%) with fewer follow-up visits using one of the latter strategies. The three strategies are detailed in Fig. 4.

We used UK Biobank to estimate, for initial newborn screening, T1D GRS2 cut-off values, which achieved various targets for the proportion of future cases included in the initially followed cohort. This is described in the first table in the published description of the GRS2 (ref. [30]). These specific target proportions were matched to the specific sensitivities of the follow-up performance of each overall strategy, to achieve a net 75% pre-onset case detection.

For follow-up, our baseline schedule comprised quarterly evaluations through 3 years of age, then every 6 months until 6 years of age, and then annually thereafter. This strategy was chosen because TEDDY included samples at each of these ages, and for the TEDDY cohort, using this schedule missed very few children. In TEDDY, all high-risk children remain in follow-up. This schedule misses very few children, with a median of 51 d (interquartile range: 84–18.5 d; range: 1–384 d) from the last visit to T1D presentation.

The classic strategy used this baseline follow-up schedule. The simple adaptive strategy also used the baseline follow-up schedule in all children remaining in follow-up. The advanced adaptive strategy used the baseline follow-up strategy for

those children in the close follow-up subgroup, but annual follow-up for those in the reduced follow-up subgroup. Each follow-up strategy was simulated separately over the TEDDY dataset. We compensated for right censoring in TEDDY data by using an inverse weighting estimator.

The optimum cut-offs for initial genetic inclusion and for retention or reassignment in the surveillance group were chosen using a grid optimization. We selected the cut-off among the values from 0.001–0.010 with steps of 0.001 to determine the optimum value, defined as that minimizing the number of follow-up evaluations while ensuring that 75% of the population cases had an adequate follow-up. Optimization was performed independently for each design strategy. For the simple adaptive strategy, it led to a CRS landmark cut-off of T1D probability of ≥0.6% (up to 10 years of age) for continued follow-up, which required 10.7% of the screened newborn population to be included in the initially followed cohort. For the advanced adaptive strategy, it led to a CRS landmark cut-off of T1D probability (within 2 years) of ≥0.8% for assignment to the frequently followed subgroup, which required 11.2% of screened newborns to be included in the initially followed cohort. Summary statistics of each strategy are shown in Extended Data Fig. 8.

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

Data availability

Clinical metadata and GRS genotyping data analyzed for this study are available in the NIDDK Central Repository at https://www.niddkrepository.org/studies/teddy, in accordance with the NIDDK's controlled-access authorization process.

Code availability

The R code used in these analyses is available in the NIDDK Central Repository at https://www.niddkrepository.org/studies/teddy, in accordance with the NIDDK's controlled-access authorization process.

References

41. TEDDY Study Group The Environmental Determinants of Diabetes in the Young (TEDDY) study. *Ann. NY Acad. Sci.* **1150**, 1–13 (2008).
42. Hagopian, W. A. et al. The Environmental Determinants of Diabetes in the Young (TEDDY): genetic criteria and international diabetes risk screening of 421,000 infants. *Pediatr. Diabetes* **12**, 733–743 (2011).
43. Lönnrot, M. et al. A method for reporting and classifying acute infectious diseases in a prospective study of young children: TEDDY. *BMC Pediatr.* **15**, 24 (2015).
44. Bonifacio, E. et al. 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**, 3360–3367 (2010).
45. Cortes, A. & Brown, M. A. Promise and pitfalls of the immunochip. *Arthritis Res. Ther.* **13**, 101 (2011).
46. Krischer, J. P. et al. The 6 year incidence of diabetes-associated autoantibodies in genetically at-risk children: the TEDDY study. *Diabetologia* **58**, 980–987 (2015).
47. Dafni, U. Landmark analysis at the 25-year landmark point. *Circ. Cardiovasc. Qual. Outcomes* **4**, 363–371 (2011).
48. Anderson, J. R., Cain, K. C. & Gelber, R. D. Analysis of survival by tumor response. *J. Clin. Oncol.* **1**, 710–719 (1983).
49. Klein, J. P. & Moeschberger, M. L. *Survival Analysis. Techniques for Censored and Truncated Data* (Springer, 2003).
50. Venables, W. N. & Ripley, B. *Modern Applied Statistics with S* (Springer, 2002).
51. Blanche, P., Dartigues, J.-F. & Jacqmin-Gadda, H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat. Med.* **32**, 5381–5397 (2013).

Acknowledgements

The TEDDY study is included in ClinicalTrials.gov under the identifier NCT00279318. The TEDDY Study Group is funded by U01 DK63829, U01 DK63861, U01 DK63821, U01 DK63865, U01 DK63863, U01 DK63836, U01 DK63790, UC4 DK63829, UC4 DK63861, UC4 DK63821, UC4 DK63865, UC4 DK63863, UC4 DK63836, UC4 DK95300, UC4 DK100238, UC4 DK106955, UC4 DK112243, UC4 DK117483 and UC4 DK100238, and by contract number HHSN267200700014C from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases (NIAID), Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), National Institute of Environmental Health Sciences (NIEHS), Centers for Disease Control and Prevention (CDC) and JDRF. This work is supported in part by the National Institutes of Health/National Center for Advancing Translational Sciences Clinical and Translational Science Awards (UL1 TR000064 (University of Florida) and UL1 TR001082 (University of Colorado)) and Diabetes Research Center (5P30 DK017047; University of Washington). R.A.O. is supported by a Diabetes UK Harry Keen Fellowship (16/0005529). S.A.S. is supported by a Diabetes UK PhD studentship (17/0005757). M.N.W. is supported by the Wellcome Trust Institutional Strategic Support Fund (WT097835MF). R.A.O., L.A.F., W.A.H. and K.V. are supported by a JDRF strategic research agreement (3-SRA-2019-827-S-B). The authors thank the following former staff members at TEDDY Study Group centers: Michael Abbondondolo, Lori Ballard, David Hadley, Hye-Seung Lee, Wendy McLeod, Steven Meulemans, Henry Erlich, Steven J. Mack and Anna Lisa Fear.

Author contributions

L.A.F., W.A.H., R.A.O., K.V. and S.A.S. designed the study, contributed to analysis and wrote the manuscript. Å.L., M.J.R., J.-X.S., A.-G.Z., J.T., B.A., J.P.K. and M.N.W. contributed to analysis and reviewed the manuscript. All authors contributed to discussions and editing of the manuscript.

Competing interests

R.A.O. holds a UK Medical Research Council Institutional Confidence in Concept grant to develop a ten-SNP biochip T1D genetic test in collaboration with Randox. A.-G.Z. is a co-applicant on patent application WO/2019/002364 Al covering the use of a GRS to identify and treat individuals with high T1D genetic risk. Neither of these genetic risk tests is identical to the more extensive GRS2 used in this paper. The other authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41591-020-0930-4.

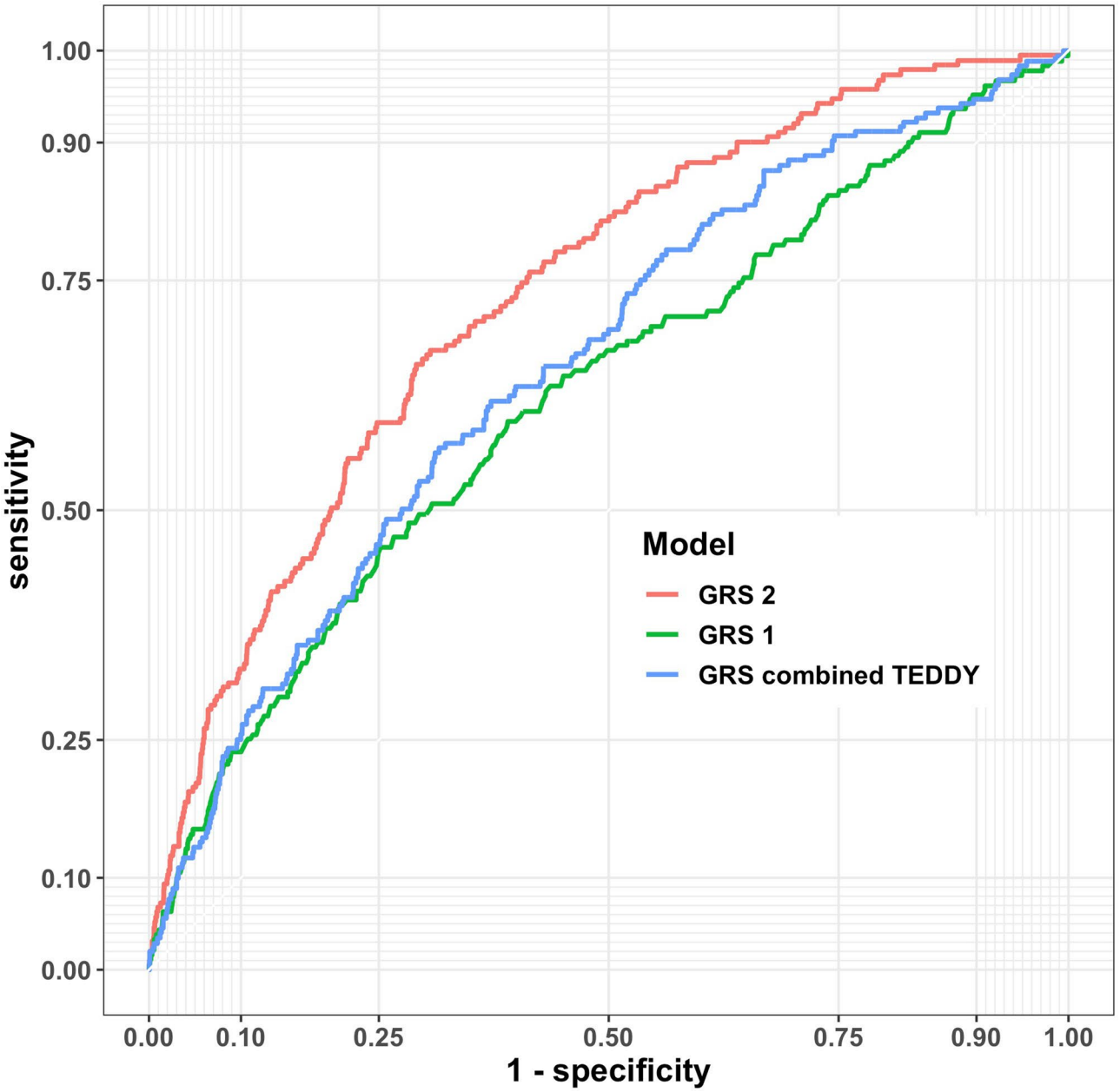
Supplementary information is available for this paper at https://doi.org/10.1038/s41591-020-0930-4.

Correspondence and requests for materials should be addressed to W.A.H.

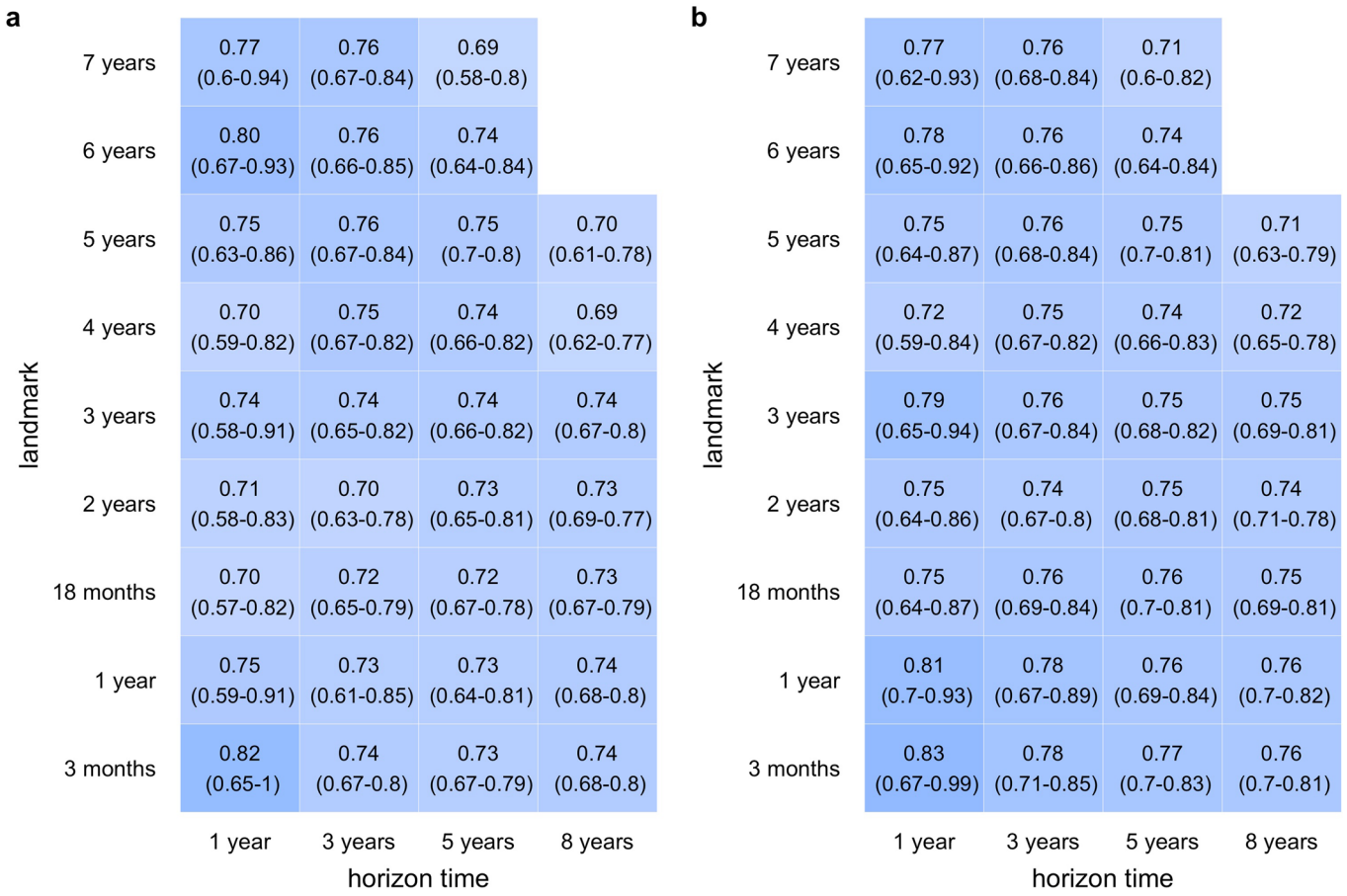
Peer review informatio Saheli Sadanand was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Variable name	Significant effect in TEDDY (ref)	Effect outside of TEDDY (ref)	ROC AUC (95%CI) to predict T1D	p-value for Cox model
Sex	1	2,3	0.50 (0.47-0.54)	0.88
Probiotic	1, 4	5	0.51 (0.49-0.53)	0.66
Common cold days	6	7, 8	0.50 (0.47-0.56)	0.098
Influenza episodes	9	7, 8	0.51 (0.49-0.52)	0.69
Respiratory episodes	9	7, 8	0.51 (0.47-0.55)	0.32
Fever episodes	9	7	0.51 (0.47-0.55)	0.80
Laryngitis tracheitis episodes	9	7, 8	0.51 (0.48-0.53)	0.45
Acute sinusitis episodes	9	7, 8	0.51 (0.49-0.53)	**0.03**
Caesarean section	10	11-15	0.52 (0.48-0.55)	0.27
Gestational respiratory infections	10	16, 17	0.52 (0.49-0.56)	0.99
Mother T1D	6	3, 18	0.51 (0.49-0.53)	**0.03**
Father T1D	6	3, 18	0.53 (0.51-0.56)	**1.2e-05**
Siblings T1D	6	3, 18	0.54 (0.52-0.56)	**1.7e-08**
Z-score weight 1y	6, 19	20, 21	0.54 (0.51-0.59)	**0.006**
Country	6, 22	2, 23, 24	0.55 (0.51-0.59)	**0.03**
Any FH T1D	1	3, 18	0.56 (0.53-0.59)	**1e-08**
IA2A	25	26-28	0.59 (0.56-0.62)	**7.9e-57**
GAD	25	26-28	0.65 (0.62-0.68)	**1.8e-81**
GRS2	-	29	0.73 (0.7-0.77)	**2.8e-36**
IAA	1	28, 30	0.73 (0.69-0.76)	**3.7e-119**
AB number	1	26-28	0.75 (0.71-0.78)	**5.5e-136**

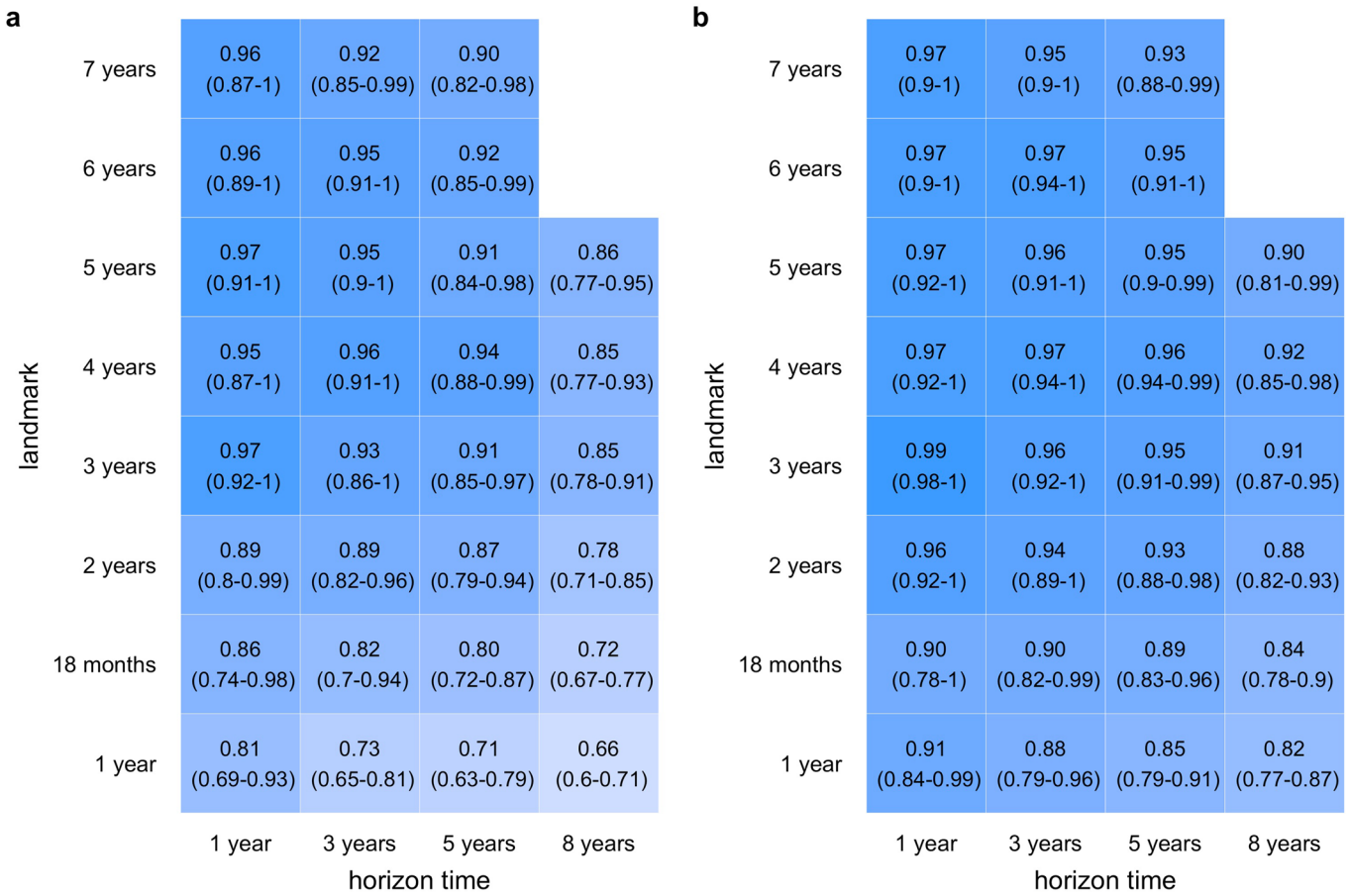
Extended Data Fig. 1 | Variables previously shown or susceptible to be associated with T1D auto-immunity evaluated in univariate analysis. Time ROC AUC and p-value (two side Wald test) are computed at landmark age 2 years and horizon of 8 years (n = 6,805). Abbreviations: Type 1 diabetes (T1D), Family history (FH), Islet Autoantibodies (AB), insulinoma Antigen-2 Autoantibody (IA2A), Glutamic Acid Decarboxylase Autoantibody (GADA), Insulin AutoAntibody (IAA), Genetic Risk score (GRS2). The referent sex is female. A concise list of references for this table is provided in the Supplementary Information file associated with this paper.



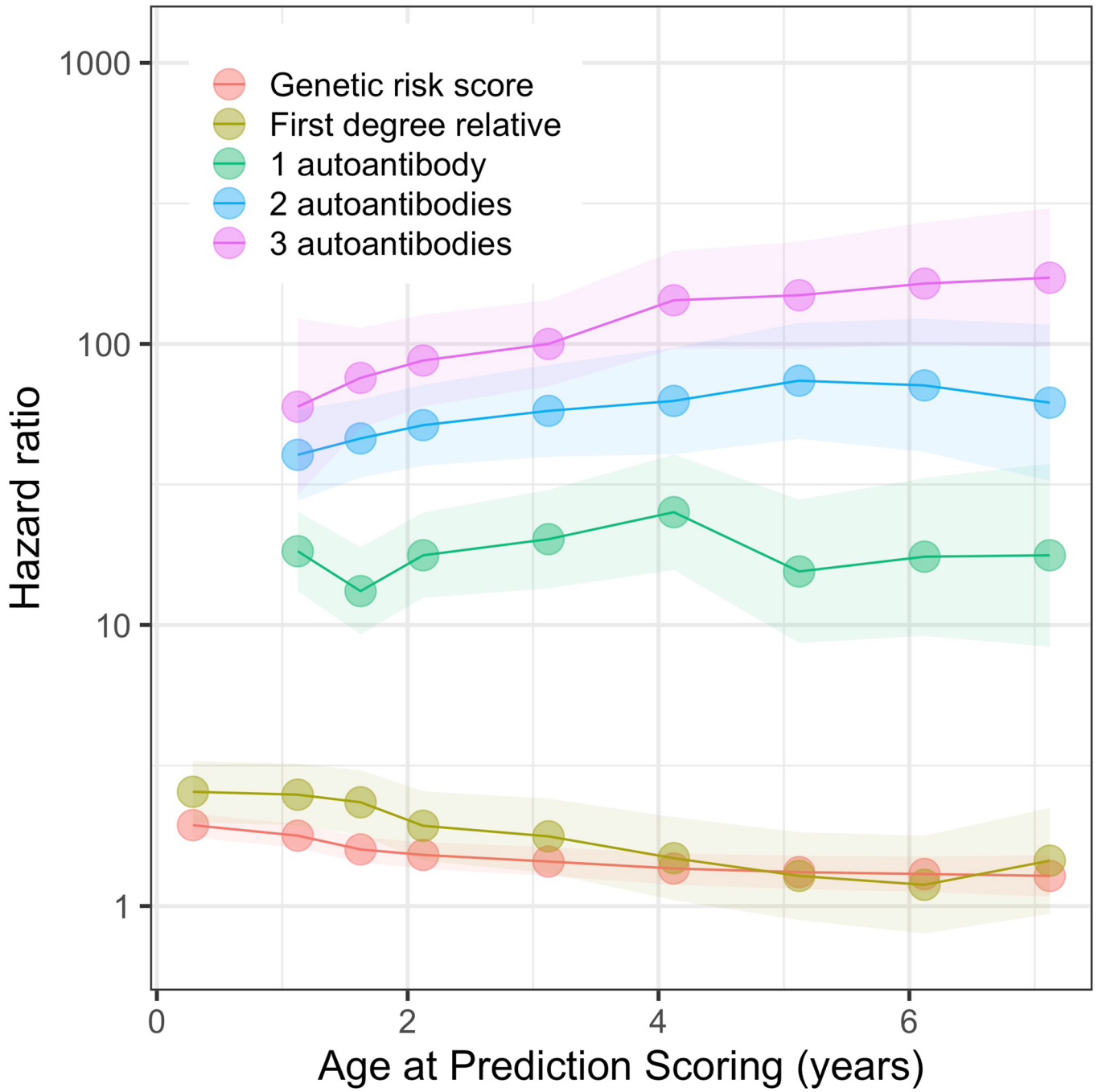
Extended Data Fig. 2 | Time dependent ROC curves comparing the performance of various genetic risk scores in the TEDDY cohort. Shown are curves for GRS1, GRS2 and the combined TEDDY GRS to predict T1D from a landmark age of birth and a horizon interval of 8 years (n= 7,798).



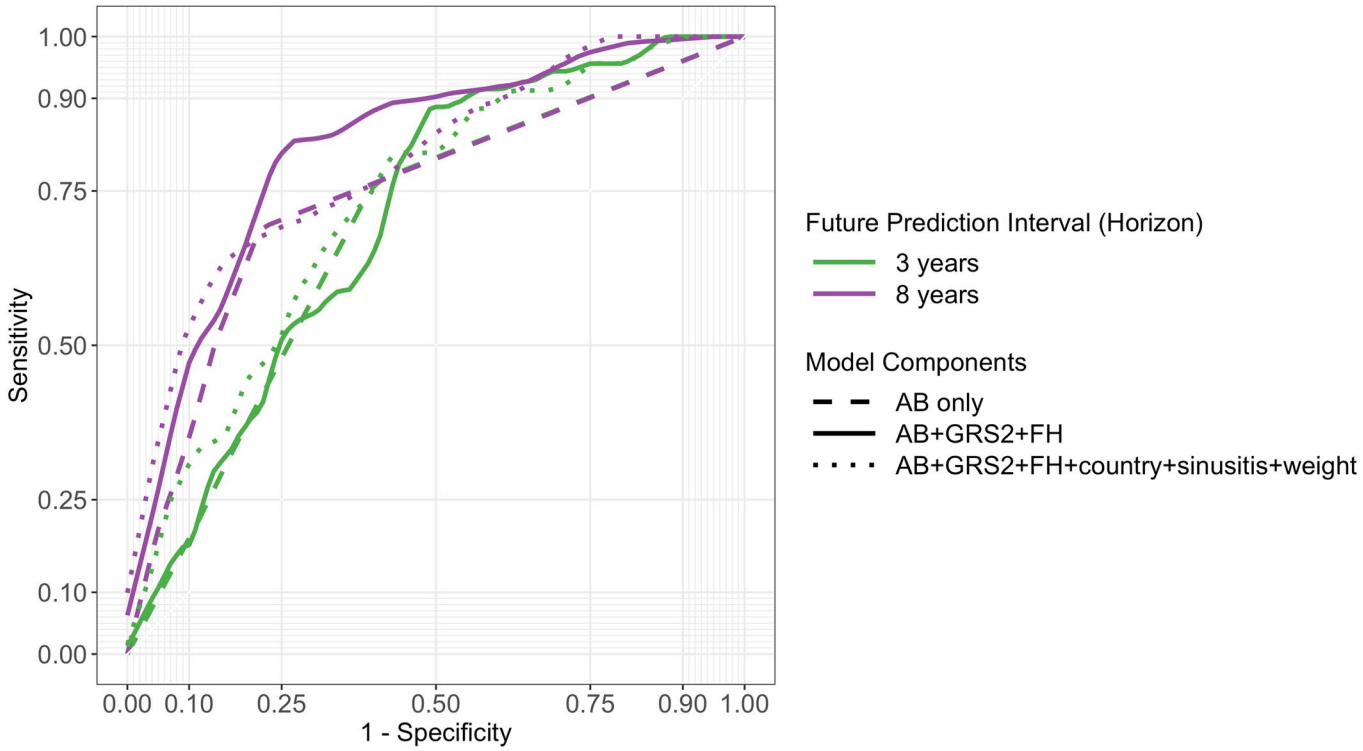
Extended Data Fig. 3 | Family history adds predictive power to the T1D GRS2. T1D GRS2 alone (**a**) is compared to T1D GRS2 + FH (**b**) at nine different landmark scoring ages and over four different horizon times. Although 95% confidence intervals always overlapped, among 34 total combinations, T1D GRS2 + FH gave a larger AUC ROC in 24 of these combinations. Results were similar in 9 combinations, and in only one instance was T1D GRS2 better. T1D GRS2 + FH superiority was greatest at landmarks ≤3 years of age. The number of children at each landmark age were 7798 (birth), 7563 (1 year), 7123 (1.5 years), 6805 (2 years), 6316 (3 years), 5973 (4 years), 5706 (5 years), 5517 (6 years) and 5323 (7 years).



Extended Data Fig. 4 | T1D GRS2 and family history add predictive power to AB. AB alone (**a**) is compared to the three-variable model of AB, GRS2 and FH. (**b**) at eight different landmark scoring ages and over four different horizon times. Although 95% confidence intervals overlapped, among 30 total combinations, the three-variable model yielded larger AUC ROC in 28 of these combinations and similar results in the remaining 2 combinations. The differences were often substantial, especially at landmarks ≤4 years of age. The number of children at each landmark age were 7798 (birth), 7563 (1 year), 7123 (1.5 years), 6805 (2 years), 6316 (3 years), 5973 (4 years), 5706 (5 years), 5517 (6 years) and 5323 (7 years).



Extended Data Fig. 5 | Hazard ratio for each variable at different ages at prediction scoring landmarks. Each point represents the hazard ratio at a landmark age (x abscises), the shaded region its respective 95% confidence interval. The number of children at each landmark age were 7798 (birth), 7563 (1 year), 7123 (1.5 years), 6805 (2 years), 6316 (3 years), 5973 (4 years), 5706 (5 years), 5517 (6 years) and 5323 (7 years).



Extended Data Fig. 6 | Time dependent ROC of different models now considering only children positive for at least one AB (n = 252). The landmark age is 2 years. At the 3 year time horizon the CRS (AB+GRS2+FH) performs similarly to AB only, but at the 8 year horizon the CRS is more predictive.

Autoantibody status	Genetic risk	Family history	1 year horizon (%)	3 year horizon (%)	5 year horizon (%)
0	++	No	0.1 [0.0 - 0.1]	0.2 [0.1 - 0.2]	0.3 [0.2 - 0.4]
0	++	Yes	0.2 [0.1 - 0.2]	0.3 [0.2 - 0.5]	0.5 [0.3 - 0.8]
0	+++	No	0.1 [0.1 - 0.2]	0.2 [0.2 - 0.3]	0.4 [0.3 - 0.5]
0	+++	Yes	0.2 [0.1 - 0.3]	0.5 [0.3 - 0.7]	0.8 [0.5 - 1.0]
0	++++	No	0.3 [0.2 - 0.4]	0.6 [0.5 - 0.8]	1.1 [0.8 - 1.3]
0	++++	Yes	0.6 [0.4 - 0.8]	1.2 [0.8 - 1.7]	2.0 [1.4 - 2.7]
1	++	No	1.4 [0.7 - 2.1]	3.0 [1.6 - 4.4]	4.8 [2.6 - 7.0]
1	++	Yes	2.7 [1.2 - 4.2]	5.6 [2.7 - 8.5]	9.1 [4.6 - 13.1]
1	+++	No	2.0 [1.1 - 2.9]	4.2 [2.5 - 5.9]	6.8 [4.1 - 9.4]
1	+++	Yes	3.8 [1.9 - 5.7]	8.0 [4.3 - 11.5]	12.7 [7.1 - 17.9]
1	++++	No	5.3 [3.1 - 7.3]	10.9 [7.1 - 14.4]	17.1 [11.8 - 22.2]
1	++++	Yes	9.9 [5.5 - 14]	19.9 [12.2 - 26.9]	30.4 [19.6 - 39.7]
2	++	No	4.0 [1.9 - 6.1]	8.4 [4.4 - 12.2]	13.3 [7.3 - 19.0]
2	++	Yes	7.6 [3.6 - 11.5]	15.5 [8.0 - 22.4]	24.1 [13.1 - 33.7]
2	+++	No	5.7 [3.1 - 8.3]	11.8 [7.0 - 16.3]	18.5 [11.5 - 25.0]
2	+++	Yes	10.7 [5.7 - 15.5]	21.4 [12.6 - 29.3]	32.6 [20.2 - 43.0]
2	++++	No	14.5 [9.3 - 19.5]	28.4 [20.2 - 35.7]	42.1 [31.8 - 50.8]
2	++++	Yes	26.1 [16.6 - 34.5]	47.4 [34.2 - 58]	65.0 [50.5 - 75.3]
3	++	No	6.8 [3.2 - 10.1]	13.8 [7.3 - 19.9]	21.6 [11.8 - 30.3]
3	++	Yes	12.6 [5.7 - 18.9]	24.9 [12.6 - 35.4]	37.4 [20.0 - 51.0]
3	+++	No	9.5 [5.1 - 13.7]	19.1 [11.4 - 26.2]	29.3 [18.2 - 39.0]
3	+++	Yes	17.5 [9.0 - 25.2]	33.6 [19.4 - 45.2]	48.8 [30.0 - 62.5]
3	++++	No	23.4 [15.0 - 31.0]	43.3 [31.3 - 53.2]	60.4 [46.4 - 70.8]
3	++++	Yes	40.2 [25.2 - 52.1]	66.4 [48.5 - 78.1]	83.2 [66.7 - 91.6]

Extended Data Fig. 7 | Individual estimated future T1D risk probability percentages (and 95% confidence intervals) for 24 different scenarios combining a GRS risk level and FH background with different AB status calcluated at age 2 years. "++" represents a T1D genetic risk score at 80th percentile of the general (UK) population. "+++" represents a T1D genetic risk score at 90th percentile of the general (UK) population. "++++" represents a T1D genetic risk score at 99th percentile of the general (UK) population.

Cohort design	% of all screened newborns followed	% of cases among all screened newborns in followed cohort	% of cases not detected pre-onset in followed cohort	Net % of cases in screened newborns detected pre-onset	Follow-up evaluation per newborn child screened	Follow-up evaluation per T1D case detected	Efficiency gain vs simple approach
Simple	9.3%	75.0%	0.0%	75.0%	1.67	743	-
Simple adaptative	10.7%	79.8%	4.8%	75.0%	1.26	560	24.5%
Advanced adaptative	11.2%	81.6%	6.6%	75.0%	0.83	367	50.6%

Extended Data Fig. 8 | Comparison of newborn screening strategies aiming to predict ≥75% of the children who will develop T1D before age 10. In the "Classic" design, the 9.3% of screened newborn population containing 75% of the T1D cases, are all followed for 10 years. In the "Simple Adaptive" design, 10.7% of the screened newborns containing 79.8% of the T1D cases, are followed for variable lengths determined by CRS-based risk, and 4.8% of T1D cases miss AB detection before onset, leaving 75% detected in advance. In the "Advanced Adaptive" design, 11.2% of the screened newborns containing 81.6 % of T1D cases are followed closely or less closely determined by CRS-based risk, 6.6% of cases miss AB detection before onset, again leaving 75% detected. Numbers are computed by using the performance of each strategy on TEDDY data. Tests per child are computed using TEDDY data and simulation to take into account right censoring in TEDDY data.

A)

Landmark /horizon	T1D children caught	T1D children missed	% of T1D caught per period	% of people in the cohort	cumulative caught	cumulative missed	cumulative evaluations
0/1	10	0	100%	100%	10	0	7798
1/1	36	0	100%	100%	46	0	38050
2/1	31	0	100%	100%	77	0	65270
3/1	20	0	100%	100%	97	0	77902
4/1	34	0	100%	100%	131	0	89848
5/1	24	0	100%	100%	155	0	101260
6/1	20	0	100%	100%	175	0	106777
7/1	27	0	100%	100%	202	0	112100
8/2	46	0	100%	100%	248	0	117234

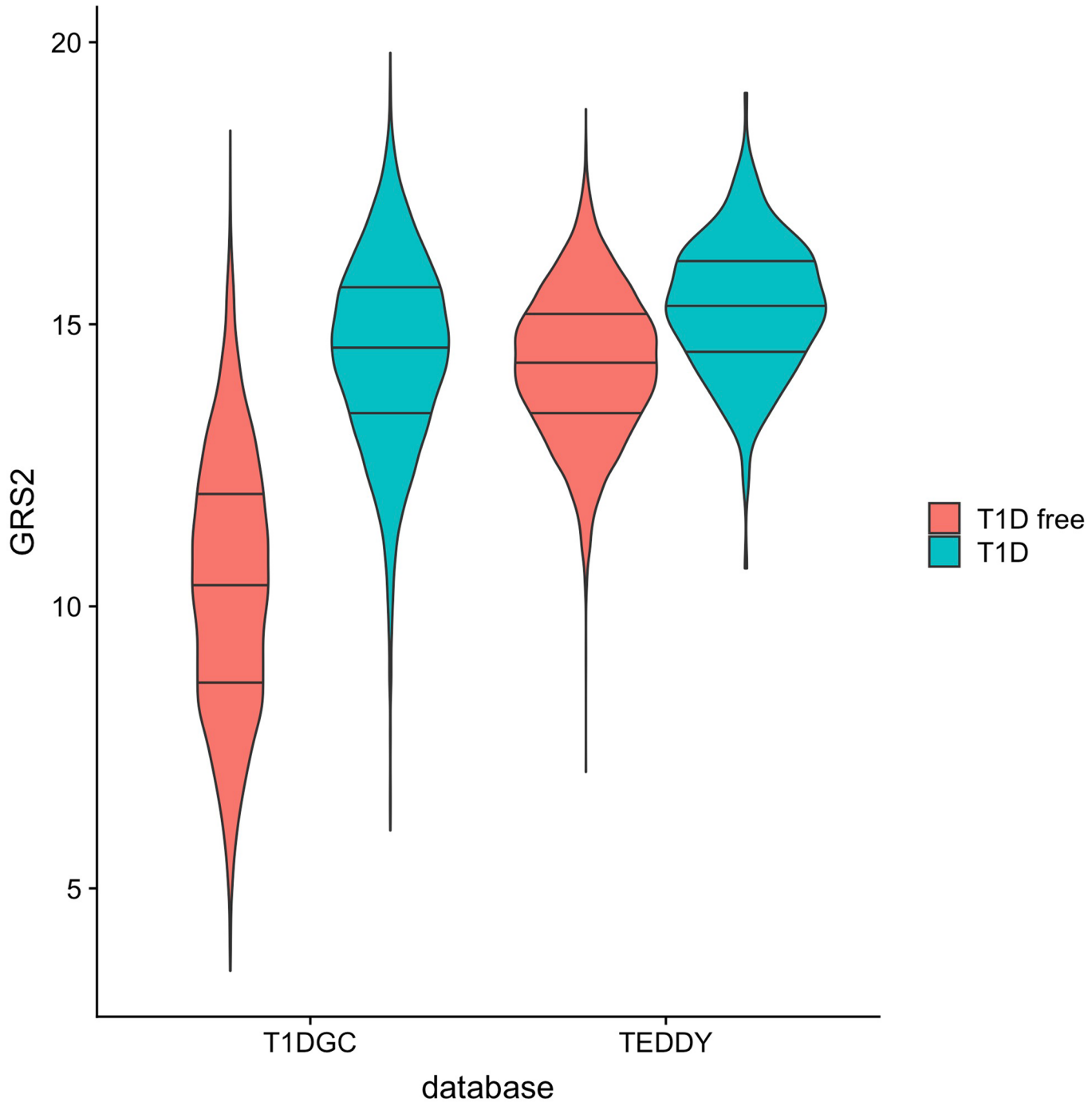
B)

Landmark /horizon	T1D children caught	T1D children missed	% of T1D caught per period	% of people in the cohort	cumulative caught	cumulative missed	cumulative evaluations
0/1	10	0	100%	100	10	0	7798
1/1	36	0	100%	95.7%	46	0	36742
2/1	31	0	100%	94.9%	77	0	62630
3/1	20	1	95.2%	78.2%	97	1	72552
4/1	34	6	85%	46.6%	131	7	78140
5/1	24	2	92.3%	23.6%	155	9	80838
6/1	20	3	87.0%	8.5%	175	12	81305
7/1	22	2	91.7%	6.7%	197	14	81640
8/2	36	1	97.3%	6.5%	233	15	81937

c)

Landmark /horizon	T1D children caught	T1D children missed	% of T1D caught per period	% of people in the cohort	cumulative caught	cumulative missed	cumulative evaluations
0/1	8	2	80.0%	25.8%	8	2	7798
1/1	32	4	88.9%	28.4%	40	6	21811
2/1	28	3	90.3%	6.7%	68	9	29993
3/1	18	2	90.0%	5.5%	86	11	36656
4/2	56	2	96.6%	6.4%	142	13	43772
6/2	44	3	93.6%	7.7%	186	16	49714
8/2	42	4	91.3%	8.5%	228	20	50150

Extended Data Fig. 9 | Visit number calculation for each design. Table A. Visit number calculations for the "Classic" design. Infants initially selected for high. GRS2 genetic risk were all followed quarterly until age 3, and every 6 months until age 6, then annually thereafter. This simulation was made on the TEDDY dataset. **Table B**. Visit number calculations for the "Simple Adaptive" design. Infants selected for high genetic risk were initially followed as in the Classic strategy, but the T1D CRS was recalculated at annual landmarks, at which time any child whose T1D probability by age 10 had decreased to <0.8% was eliminated from further follow-up. Of new cases, 94% had high risk detected before onset. This simulation was made on the TEDDY dataset. **Table C**. Visit number calculations for the "Advanced Adaptive" design. Infants selected for high genetic risk were initially followed as in the Classic strategy, but at birth and annually thereafter, a T1D CRS calculation was used to reallocate children among the quarterly or annual surveillance groups based on T1D probability in 2 years of ≥0.6% or <0.6%, respectively. Of new cases, 92% had high risk detected before onset. Simulation made on the TEDDY dataset.



Extended Data Fig. 10 | GRS2 violin plot in the Type 1 Diabetes Genetics Consortium (T1DGC) and TEDDY datasets. T1DGC is more representative of the general background population. The genetic pre- selection in TEDDY based on the major T1D risk locus HLA-DR-DQ, renders the T1D GRS2 higher in TEDDY, even in T1D free subjects. Further, the separation between T1D and non-T1D subjects in TEDDY is much less. There are 7,798 observations in TEDDY including 305 with T1D. There are 15729 observations in T1DGC including 6483 with T1D. The lines in the violin plots respectively indicate the 25th, 50th and 75th percentiles, while the lowest and the highest point of each violin plot indicates the minimum and the maximum, respectively, for each group of individuals.

nature research

Corresponding author(s): William Hagopian

Last updated by author(s): 24/03/2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection.
Data analysis	analysis Gcta v1.91.1beta - Principal component generation and analysis KING v2.1- Relatedness analysis Eagle v2.4 - Genotype phasing Impute2 v2.3.2 - Genotype imputation Plink v1.9 - Genotype quality control R v3.6 was used for the statistical analysis

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data Availability. Clinical metadata including islet antibody status and diabetes diagnoses, and GRS genotyping data, analyzed for the current study are available in the NIDDK Central Repository at https://www.niddkrepository.org/studies/teddy.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Out of 8,676 TEDDY enrollees, 7,883 were analyzed herein on the basis of full AB testing, SNP genotyping on the ImmunoChip array, and carrying one of the four major TEDDY eligible HLA haplogenotypes. At the time of analysis, median follow-up was 9.3 years (range 1-168 months, interquartile range [IQR] 54 to 132 months) covering 65,331 person-years of observation. 305 children developed T1D. No sample size calculation was performed for this particular study but the Data set TEDDY was designed to follow up enough children to ensure a good sample size for a variety of studies.
Data exclusions	The exclusion criteria were pre-estabilised; Children with poor quality genotyping data based on missingness or mismatched sex. After quality control 7,798 individuals were available for analysis.
Replication	Observational cohort. No replication.
Randomization	Not relevant for the aims of this observational study.
Blinding	TEDDY is an observational follow-up study, thus no overall blinding was used.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☐	☒ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	Children samples were obtained from six geographical locations (Finland, Germany, Sweden in Europe; and Washington State, Colorado and Georgia in the United States). These children were selected to be at high HLA genetic risk for developing T1D.
Recruitment	Children were recruited based on specific type 1 diabetes risk human leukocyte antigen (HLA) genotypes and/or family history of T1D risk. Recruitment began in September of 2004 and was completed in February 2010. Six clinical centers took part. Three were in the USA (Colorado, Washington State and Florida/Georgia) and 3 in Europe (Germany, Sweden, and Finland). N=424,788 newborns were randomly screened at birth in hospitals in all centers, of which 418,367 were general population infants and 6,421 were first-degree relatives (FDR) of a family member with type 1 diabetes. N=20,152 general population and 1,437 FDR were HLA eligible. The latter represent about half of the subjects within the originally screened cohort who would be expected to develop type 1 diabetes, but are all among the future diabetes patients with the greatest HLA risk. This HLA bias is considered in the main text. A total of 7,709 general population children (38%) and 967 FDR children (67%) had parents who consented to enrollment in the follow-up surveillance study. There was a bias towards FDR participation, since these families may be more motivated towards diabetes research. Ethnicity differed between sites, with more African-background participants in Georgia, more Hispanic participants in Colorado and more Asian participants in Seattle. At all these sites, participation in follow-up was greater among non-Hispanic Whites. European sites were not allowed to collect race or ethicity data. TEDDY placed significant study burden on participants, and there may be an unmeasurable bias that people likely to complete the study had a greater interest in a healthy lifestyle, or that they may be of a higher socioeconomic status, than those choosing not to participate.

Ethics oversight	The samples and clinical information at all clinical sites were in all cases obtained under IRB or local ethics board approval, in all cases using informed consent, and also with the initial and ongoing approval of a study-specific National Institute of Diabetes and Digestive and Kidney Diseases External Evaluation Committee.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	NCT00279318
Study protocol	Full protocol can be accessed at https://teddy.epi.usf.edu/documents/TEDDY_Protocol.pdf.
Data collection	Six clinical research centers - three in the U.S. (Colorado, Georgia/Florida, Washington State), and three in Europe (Finland, Germany, and Sweden) participated in a population-based HLA screening of newborns between 2004 and 2010. Families with children with high risk HLA genotypes were invited to enroll in follow-up, and n=8,676 did this. They were then prospectively followed from three months of age until either developing type 1 diabetes (T1D) or until an intended age of 15 years old, with study visits that included a blood draw every 3 months until 4 years of age, and every 3 or 6 months thereafter for islet autoantibody positive or negative subjects, respectively. Stool samples were collected monthly from ages 3-48 months and then quarterly until age 10 years.
Outcomes	T1D diagnosis was defined according to American Diabetes Association criteria.

nature medicine

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41591-020-0930-4

In the format provided by the authors and unedited.

A combined risk score enhances prediction of type 1 diabetes among susceptible children

Lauric A. Ferrat[1], Kendra Vehik[2], Seth A. Sharp[1], Åke Lernmark[3], Marian J. Rewers[4], Jin-Xiong She[5], Anette-G. Ziegler[6,7,8], Jorma Toppari[9,10], Beena Akolkar[11], Jeffrey P. Krischer[2], Michael N. Weedon[1], Richard A. Oram[1,12,36], William A. Hagopian[13,36] ✉ and TEDDY Study Group*

[1]Institute of Biomedical and Clinical Science, University of Exeter Medical School, Exeter, UK. [2]Health Informatics Institute, Morsani College of Medicine, University of South Florida, Tampa, FL, USA. [3]Department of Clinical Sciences, Lund University and Skåne University Hospital, Malmö, Sweden. [4]Barbara Davis Center for Childhood Diabetes, University of Colorado School of Medicine, Aurora, CO, USA. [5]Center for Biotechnology and Genomic Medicine, Medical College of Georgia, Augusta University, Augusta, GA, USA. [6]Institute of Diabetes Research, Helmholtz Zentrum München, Munich, Germany. [7]Forschergruppe Diabetes, Klinikum Rechts der Isar, Technische Universität München, Munich, Germany. [8]Forschergruppe Diabetes e.V., Munich, Germany. [9]Department of Pediatrics, Turku University Hospital, Turku, Finland. [10]Research Centre for Integrative Physiology and Pharmacology, Institute of Biomedicine, University of Turku, Turku, Finland. [11]National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, USA. [12]Academic Renal Unit, Royal Devon and Exeter NHS Foundation Trust, Exeter, UK. [13]Pacific Northwest Research Institute, Seattle, WA, USA. [36]These authors contributed equally: Richard A. Oram, William A. Hagopian. *A list of members and their affiliations appears at the end of the paper. ✉e-mail: wah@uw.edu

References used in the Table of Extended Data 1

1. Krischer, J. P. *et al.* The 6 year incidence of diabetes-associated autoantibodies in genetically at-risk children: the TEDDY study. *Diabetologia* **58**, 980–987 (2015).
2. Karvonen, M. Incidence and trends of childhood Type 1 diabetes worldwide 1990–1999. *Diabet. Med.* **23**, 857–866 (2006).
3. Harjutsalo, V., Podar, T. & Tuomilehto, J. Cumulative Incidence of Type 1 Diabetes in 10,168 Siblings of Finnish Young-Onset Type 1 Diabetic Patients. *Diabetes* **54**, 563–569 (2005).
4. Uusitalo, U. *et al.* Association of Early Exposure of Probiotics and Islet Autoimmunity in the TEDDY Study. *JAMA Pediatr.* **170**, 20 (2016).
5. Gülden, E., Wong, F. S. & Wen, L. The gut microbiota and Type 1 Diabetes. *Clin. Immunol.* **159**, 143–153 (2015).
6. Krischer, J. P. *et al.* Genetic and Environmental Interactions Modify the Risk of Diabetes-Related Autoimmunity by 6 Years of Age: The TEDDY Study. *Diabetes Care* **40**, 1194–1202 (2017).
7. Beyerlein, A., Wehweck, F., Ziegler, A.-G. & Pflueger, M. Respiratory Infections in Early Life and the Development of Islet Autoimmunity in Children at Increased Type 1 Diabetes Risk. *JAMA Pediatr.* **167**, 800 (2013).
8. Beyerlein, A., Donnachie, E., Jergens, S. & Ziegler, A.-G. Infections in Early Life and Development of Type 1 Diabetes. *JAMA* **315**, 1899 (2016).
9. Lönnrot, M. *et al.* Respiratory infections are temporally associated with initiation of type 1 diabetes autoimmunity: the TEDDY study. *Diabetologia* **60**, 1931–1940 (2017).
10. Lynch, K. F. *et al.* Gestational respiratory infections interacting with offspring HLA and CTLA-4 modifies incident β-cell autoantibodies. *J. Autoimmun.* **86**, 93–103 (2018).
11. Bonifacio, E., Warncke, K., Winkler, C., Wallner, M. & Ziegler, A.-G. Cesarean Section and Interferon-Induced Helicase Gene Polymorphisms Combine to Increase Childhood Type 1 Diabetes Risk. *Diabetes* **60**, 3300–3306 (2011).
12. Puff, R. *et al.* Compromised immune response in infants at risk for type 1 diabetes born by Caesarean Section. *Clin. Immunol.* **160**, 282–285 (2015).
13. Cardwell, C. R. *et al.* Caesarean section is associated with an increased risk of childhood-onset type 1 diabetes mellitus: a meta-analysis of observational studies. *Diabetologia* **51**, 726–735 (2008).
14. Stene, L. C., Magnus, P., Lie, R. T., Søvik, O. & Joner, G. No Association between Preeclampsia or Cesarean Section and Incidence of Type 1 Diabetes among Children: A Large, Population-Based Cohort Study. *Pediatr. Res.* **54**, 487–490 (2003).
15. Sevelsted, A., Stokholm, J., Bonnelykke, K. & Bisgaard, H. Cesarean Section and Chronic Immune Disorders. *Pediatrics* **135**, e92–e98 (2015).
16. Dahlquist, G. G., Ivarsson, S., Lindberg, B. & Forsgren, M. Maternal Enteroviral Infection During Pregnancy as a Risk Factor for ChildHood IDDM: A Population-Based Case-Control Study. *Diabetes* **44**, 408–413 (1995).
17. Viskari, H. R. *et al.* Maternal First-Trimester Enterovirus Infection and Future Risk of Type 1 Diabetes in the Exposed Fetus. *Diabetes* **51**, 2568–2571 (2002).
18. EURODIAB ACE Group 1998. Familial Risk of Type I diabetes in European Children. *Diabetologia* **41**, 1151–1156 (1998).

19. Steck, A. K. *et al.* 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**, 808–813 (2015).
20. Knerr, I. *et al.* The 'accelerator hypothesis': relationship between weight, height, body mass index and age at diagnosis in a large cohort of 9,248 German and Austrian children with type 1 diabetes mellitus. *Diabetologia* **48**, 2501–2504 (2005).
21. Verbeeten, K. C., Elks, C. E., Daneman, D. & Ong, K. K. Association between childhood obesity and subsequent Type 1 diabetes: a systematic review and meta-analysis. *Diabet. Med.* **28**, 10–18 (2011).
22. Yang, J. *et al.* Prevalence of obesity was related to HLA-DQ in 2–4-year-old children at genetic risk for type 1 diabetes. *Int. J. Obes.* **38**, 1491–1496 (2014).
23. Knip, M. *et al.* Environmental triggers and determinants of type 1 diabetes. *Diabetes* **54**, S125–S136 (2005).
24. Diaz-Valencia, P. A., Bougnères, P. & Valleron, A.-J. Global epidemiology of type 1 diabetes in young adults and adults: a systematic review. *BMC Public Health* **15**, 255 (2015).
25. Bonifacio, E. *et al.* Genetic scores to stratify risk of developing multiple islet autoantibodies and type 1 diabetes: A prospective study in children. *PLOS Med.* **15**, e1002548 (2018).
26. Verge, C. F. *et al.* Prediction of type I diabetes in first-degree relatives using a combination of insulin, GAD, and ICA512bdc/IA-2 autoantibodies. *Diabetes* **45**, 926–933 (1996).
27. Kupila, A. *et al.* Feasibility of genetic and immunological prediction of Type I diabetes in a population-based birth cohort. *Diabetologia* **44**, 290–297 (2001).
28. Ziegler, A. G. *et al.* Seroconversion to Multiple Islet Autoantibodies and Risk of Progression to Diabetes in Children. *JAMA* **309**, 2473 (2013).
29. Sharp, S. A. *et al.* Development and Standardization of an Improved Type 1 Diabetes Genetic Risk Score for Use in Newborn Screening and Incident Diagnosis. *Diabetes Care* **42**, 200–207 (2019).
30. Ziegler, A. G., Hummel, M., Schenker, M. & Bonifacio, E. Autoantibody Appearance and Risk for Development of Childhood Diabetes in Offspring of Parents With Type 1 Diabetes. *Diabetes* **48**, (1999).

Auto-antibody status	Genetic risk	Family history	1 year horizon		3 year horizon		5 year horizon	
			Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
0	++	No	100	5.8	100	5.8	100	5.8
0	++	Yes	100	33.8	98.8	33.8	98.4	33.9
0	+++	No	100	17.6	100	17.7	100	17.7
0	+++	Yes	97.8	57.3	95.3	57.4	93.6	57.7
0	++++	No	97.8	76.9	91.8	77.3	87.9	77.5
0	++++	Yes	87.0	93.9	85.1	94.2	78.6	94.6
1	++	No	80.3	96.5	80.5	96.9	75.5	97.3
1	++	Yes	80.3	97.0	78.1	97.3	73.1	97.7
1	+++	No	80.3	96.7	80.5	97.1	75.5	97.5
1	+++	Yes	76.0	97.4	73.6	97.8	69.3	98.2
1	++++	No	73.7	97.8	71.2	98.2	66.9	98.6
1	++++	Yes	58.6	98.5	58.9	98.8	58.7	99.2
2	++	No	76.0	97.5	73.6	97.8	69.3	98.2
2	++	Yes	67.2	98.2	66.8	98.6	63.9	99.0
2	+++	No	73.7	97.9	71.2	98.3	66.9	98.6
2	+++	Yes	54.1	98.6	51.9	98.9	54.1	99.3
2	++++	No	41.2	99.0	34.8	99.2	40.4	99.6
2	++++	Yes	13.3	99.6	9.0	99.7	11.8	99.8
3	++	No	67.2	98.1	66.8	98.4	63.9	98.8
3	++	Yes	49.8	98.9	42.7	99.1	46.4	99.4
3	+++	No	58.6	98.5	58.9	98.8	58.7	99.2
3	+++	Yes	36.8	99.2	31.4	99.3	36.5	99.6
3	++++	No	21.9	99.5	18.0	99.6	20.2	99.7
3	++++	Yes	4.5	100	2.3	99.9	2.3	100

Supplementary Table 1. Sensitivity and specificity given future T1D risk probabilities for 24 different scenarios combining GRS and FH risk levels with different AB status for 2-year-old children. "++" represents a genetic risk score at 80th percentile of the general (UK) population. "+++" represents a genetic risk score at 90th percentile of the general (UK) population. "++++" represents a genetic risk score at 99th percentile of the general (UK) population.

The cut-offs used for sensitivity and specificity for analyses were the CRS given for the value of AB, GRS2 and FH described in the first 3 columns. The CRS is given by the linear part of the hazard function of the Cox proportional hazard model, which has the form $\lambda(t)e^{X_i B}$, where X_i represent the values of the number of AB, GRS2 and FH of an individual i and B are the parameters of the model estimated when fitting the model. The CRS of the individual i is equal to $X_i B$.

Horizon time	*Original (cross validation)*	*Fitted on others; tested on Finland*	*Fitted on others; tested on Germany*	*Fitted on others; tested on Sweden*	*Fitted on others; tested on USA*
1 year	*0.96*	*0.95*	*0.91*	*0.94*	*0.94*
3 years	*0.94*	*0.94*	*0.96*	*0.91*	*0.94*
5 years	*0.93*	*0.92*	*0.96*	*0.91*	*0.93*
8 years	*0.87*	*0.84*	*0.95*	*0.87*	*0.86*

Supplementary Table 2. Summary of AUC ROC results for 2-year landmark, fitted on three countries to predict on the fourth.

variable		non T1D (7493)	T1D (305)
Country	USA	3143	103
	Finland	1612	89
	Germany	507	28
	Sweden	2231	85
First degree relative with T1D	no	6691	221
	yes	802	84
Mother T1D	no	7214	283
	yes	279	22
Father T1D	no	7131	260
	yes	362	45
Siblings T1D	no	7381	280
	yes	112	25
HLA genotype	other	340	17
	DR4/DR3	2835	166
	DR4/DR4	1464	55
	DR4/DR8	1277	42
	DR3/DR3	1577	25
Sex	Female	3684	148
	Male	3809	157
Caesarean section	no	5542	223
	yes	1951	82

Supplementary Table 3 Numbers of T1D and non-T1D children in the cohort by model variable.

Chromosome	SNP	Locus	MAF	Minor allele	Major allele	r^2
1	rs2476601	PTPN22	0.1083	A	G	Genotyped
1	rs3024505	PTPN22	0.1509	A	G	Genotyped
2	rs3087243	CTLA4	0.3982	A	G	Genotyped
2	rs2111485	IFIH1	0.4118	A	G	Genotyped
4	rs17388568	ADAD1	0.3064	A	G	Genotyped
6	rs9259118	HLA-A*0301	0.1268	T	C	0.943852
6	rs1281934	Intergenic DRB1-DQA1	0.0039	G	A	0.544
6	rs12153924	HLA-A*0201	0.3001	A	G	0.979
6	rs9500974	HLA-A*0205	0.006	T	G	Genotyped
6	rs72848653	HLA-A*24	0.0686	T	C	0.915
6	rs144530872	HLA-A*29	0.016	A	G	0.959
6	rs371250843	HLA-B*18	0.0463	T	TG	0.996
6	rs540653847	HLA-B*3906	0.0073	GC	G	0.808
6	rs2524277	HLA-B*44	0.0104	A	G	Genotyped
6	rs16899379	HLA-B*45	0.0025	A	G	0.933
6	rs149663102	HLA-B*57	0.010	T	TG	0.994
6	rs12189871	HLA-C*06	0.033	T	C	Genotyped
6	rs17211699	HLA-DQ2.2	0	T	C	-
6	rs9273369	HLA-DQ2.5	0.4008	C	T	Genotyped
6	rs12527228	HLA-DQ4.2	0.0814	T	C	0.981
6	rs10947332	HLA-DQ5.1	0.1088	A	G	Genotyped
6	rs1794265	HLA-DQ5.3	0	A	C	-
6	rs117806464	HLA-DQ6.1	0	A	G	-
6	rs17843689	HLA-DQ6.2	0	C	T	-
6	rs62406889	HLA-DQ6.3	0	T	G	-
6	rs16822632	HLA-DQ6.9	0	A	G	-
6	rs1281935	HLA-DQ7.3	0	T	G	-
6	rs9469200	HLA-DQ7.5	0	C	T	-
6	rs7454108	HLA-DQ8.1	0.4878	G	A	Genotyped
6	rs28746898	HLA-DQ9.2	0	G	A	-
6	rs9405117	HLA-DQ9.3	0.0073	A	C	0.358
6	rs9271346	XL9 Regulatory	0.0011	C	T	0.892
6	rs2567287	HLA-DPB1*1501	0.0086	A	G	0.998
6	rs9378176	HLA-DPB1*0501	0.0154	G	A	0.999
6	rs75658393	Intergenic BTNL2-DRA1	0.0318	C	T	0.992
6	rs9269173	Intergenic DRA1-DRB1	0.0528	A	T	0.977
6	rs116522341	BTNL2	0.061	G	C	0.791

6	rs6934289	HLA-DPB1*0402	0.1117	C	T	0.998
6	rs17214657	HLA-DPB1*0101	0.1346	C	T	Genotyped
6	rs72928038	BACH2	0.1573	A	G	Genotyped
6	rs559242105	DPB1*0301	0.1583	CTA	C	0.944
6	rs1738074	TAGAP	0.4308	T	C	Genotyped
6	rs9388489	CENPW	0.4641	G	A	Genotyped
7	rs4948088	COBL	0.0457	A	C	Genotyped
9	rs6476839	GLIS3	0.4213	A	T	Genotyped
10	rs41295121	IL2RA	0.0092	T	C	Genotyped
10	rs61839660	IL2RA	0.0745	T	C	Genotyped
10	rs60888743	RNLS	0.25	G	A	0.998
11	rs3842753	INS	0.254	T	G	0.968
12	rs11170466	ITGB7	0.0635	T	C	0.466
12	rs4759229	ERBB3	0.3286	A	G	Genotyped
12	rs653178	ATXN2	0.4486	C	T	Genotyped
12	rs10492166	CLEC1	0.4889	A	G	Genotyped
13	rs9585056	IRF7	0.2507	C	T	Genotyped
14	rs56994090	MEG3	0.4421	C	T	Genotyped
15	rs2289702	CTSH	0.1026	T	C	0.988
15	rs72727394	RASGRP1	0.2064	T	C	Genotyped
16	rs9924471	CCDC101	0.145	A	G	Genotyped
16	rs12708716	CLEC16A	0.3349	G	A	Genotyped
18	rs1893217	PTPN2	0.1612	G	A	Genotyped
18	rs1615504	CD226	0.4821	T	C	Genotyped
19	rs144309607	TYK2	0.007	T	C	0.514
19	rs425105	PRKD2	0.1585	C	T	Genotyped
20	rs2281808	SIRPG	0.3306	T	C	0.996
21	rs9981624	UBASH3A	0.3277	G	C	Genotyped
22	rs5763779	HORMAD2	0.3403	A	G	Genotyped
22	rs229541	C1QTNF6	0.4207	A	G	Genotyped

Supplementary Table 4. The 67 SNPs used in the T1D GRS2 with their Minor Allele Frequency (MAF) in this cohort, and when not genotyped directly, the imputation score r^2. Note that a total of 8 SNPs from the published GRS2 mark HLA-DQ haplotypes are not present in this cohort, and their MAFs are therefore each shown as 0."