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.