nature portfolio

Corresponding author(s): Dong D. Wang

Last updated by author(s): Aug 5, 2026

Reporting Summary

Nature Portfolio 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 Portfolio policies, see our Editorial Policies 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	Metagenomic sequencing data were processed by Bowtie2 v2.2.3, Casava v1.8.2 (Illumina), cutadapt v1.9dev2, Trim Galore v0.2.8 (Babraham Bioinformatics), PRINSEQ v0.20.3, MetaPhlAn2 v2.6.0, HUMAnN2 v0.9.4. Genetic data were processed using KING (v2.3.0) and PLINK (v1.90b7). All the data analysis and plotting were performed using R v4.2.3 (Main package: vegan v2.6-4, Maaslin2 v1.12.0, ggplot2 v3.4.2, lmerTest v3.1-3, traj v2.0.1, survival v3.5-5) The analysis-specific programs are publicly available through https://github.com/biobakery/teddy_genetics

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 and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio 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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

TEDDY microbiome whole-genome sequencing and SNP data that support the findings of this study are available in the NCBI database of Genotypes and Phenotypes (dbGaP) under the primary accession code phs001442.v4.p3, in accordance with the dbGaP controlled-access authorization process. Data from The Environmental Determinants of Diabetes in the Young (https://doi.org/10.58020/y3jk-x087) reported here are available upon request through the NIDDK Central Repository (NIDDK-CR) website, Resources for Research (R4R), https://repository.niddk.nih.gov/. Processed taxonomic profiles and metadata from the DIABIMMUNE three country cohort are available through the DIABIMMUNE website (https://diabimmune.broadinstitute.org/diabimmune).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	For the TEDDY cohort, sex was recorded at enrolment based on parental report for all participants. Gender was not assessed in this study. The study included 403 (45.4%) female and 484 (54.6%) male participants. Sex was included as a covariate in all analyses to account for potential sex-related influences on the outcomes. In addition, sex-stratified analyses of the primary findings yielded results consistent with the main analyses. For the DIABIMMUNE three-country cohort, which was used as a secondary replication cohort specifically for the microbiome maturation pattern analysis, sex was recorded as reported in the original study and was not used as a stratification variable.
Reporting on race, ethnicity, or other socially relevant groupings	The study population are 887 mostly white, non-hispanic children from six different clinical centers in the U.S. (Colorado, Georgia/Florida, and Washington) and Europe (Finland, Germany, and Sweden). For the DIABIMMUNE cohort, race and ethnicity were not systematically collected or used in the replication analysis.
Population characteristics	The study included children from the TEDDY and DIABIMMUNE cohorts. TEDDY participants were infants and young children with increased genetic risk for type 1 diabetes based on HLA genotype, including children from the general population and those with a first-degree relative with type 1 diabetes. DIABIMMUNE participants were followed from birth to three years of age and carried HLA alleles associated with increased risk of autoimmunity.
Recruitment	Families with a newborn in participating clinical centers with HLA-conferred genetic predisposition for T1D or first-degree relative(s) with T1D were invited to join the TEDDY study. The DIABIMMUNE three-country cohort recruited newborn infants from Finland, Estonia, and Russia with increased HLA-conferred genetic susceptibility. The recruitment procedures for the published datasets were described in the corresponding papers.
Ethics oversight	The study was approved by local US Institutional Review Boards and European Ethics Committee Boards in Colorado's Colorado Multiple Institutional Review Board, Georgia's Medical College of Georgia Human Assurance Committee (2004–2010), Georgia Health Sciences University Human Assurance Committee (2011–2012), Georgia Regents University Institutional Review Board (2013–2015), Augusta University Institutional Review Board (2015–present), Florida's University of Florida Health Center Institutional Review Board, Washington state's Washington State Institutional Review Board (2004–2012) and Western Institutional Review Board (2013–present), Finland's Ethics Committee of the Hospital District of Southwest Finland, Germany's Bayerischen Landesärztekammer (Bavarian Medical Association) Ethics Committee, Sweden's Regional Ethics Board in Lund, Section 2 (2004–2012) and Lund University Committee for Continuing Ethical Review (2013–present). The study is monitored by an External Advisory Board formed by the National Institutes of Health. The DIABIMMUNE study was conducted according to the guidelines laid down in the Declaration of Helsinki, and all procedures involving human subjects were approved by the local ethical committees of the participating hospitals. The parents and/or study subjects gave their written informed voluntary consent prior to sample collection.

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

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	All stool samples with metagenomic data (N = 12,151 stool samples) from subjects in TEDDY islet autoimmunity and type 1 diabetes case-control cohorts are being analyzed for maximal power. This includes all participants with islet autoimmunity or type 1 diabetes and their

	matched controls in TEDDY study as of May 31, 2012 (N = 887 participants; host genetic data available for 877 participants). For the DIABIMMUNE three-country cohort, 785 stool metagenomic samples from 212 infants and young children were included for the microbiome maturation pattern analysis.
Data exclusions	The exclusion criteria were pre-specified. At the sample level, metagenomes failing quality control were excluded before analysis. At the feature level, microbial features were retained only if they were present in at least 10% of samples, with minimum relative abundance thresholds of 0.0002 for species and pathways and 0.00002 for level-4 EC categories; these filters were specified in advance to avoid spurious associations arising from sparsely detected features. For genetic analyses, sample-level and SNP-level quality control was performed using KING --autoQC and PLINK, and individuals identified as genetic outliers by principal component analysis were removed. For trajectory clustering analyses, individuals with insufficient longitudinal sampling (<4 metagenomic samples) were excluded because reliable developmental patterns could not be inferred.
Replication	Replication of the microbiome maturation pattern was performed in an independent birth cohort (DIABIMMUNE). Other analyses were conducted in the primary TEDDY cohort only.
Randomization	Controls were matched individually to cases as described in detail in the manuscript text. Cases were sampled until diagnosis of T1D and matched control samples were included up until the corresponding day of life.
Blinding	No blinding used, TEDDY is an observational follow-up study

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 and archaeology
☒	☐ Animals and other organisms
☐	☒ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

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

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	https://repository.niddk.nih.gov/studies/teddy/
Data collection	The TEDDY study involves six clinical centers across four countries: the United States (Colorado, Georgia/Florida, and Washington), Finland, Sweden, and Germany. Participant recruitment took place between September 2004 and February 2010, enrolling 8,676 infants aged 3–4 months who were identified through newborn genetic screening for HLA genotypes associated with increased risk for type 1 diabetes (T1D). Blood samples were collected every three months until participants reached the age of four, transitioning to biannual collections thereafter. Stool samples were collected monthly from three months to 48 months of age, then every three months until 10 years of age, and biannually thereafter. Stool sample collection ceased for all participants in August 2018
Outcomes	Appearance of one or more islet cell autoantibodies: GADA, IAA, or IA-2A confirmed at two consecutive visits. And development of Type 1 Diabetes

Plants

Seed stocks	Not applicable
Novel plant genotypes	Not applicable
Authentication	Not applicable