Prospective virome analyses in young children at increased genetic risk for type 1 diabetes

Kendra Vehik1,2, Kristian F. Lynch1, Matthew C. Wong3, Xiangjun Tian4, Matthew C. Ross5,6, Richard A. Gibbs3, Nadim J. Ajami2, Joseph F. Petrosino2, Marian Rewers4, Jorma Toppari5,6, Annette G. Ziegler7,8,9, Jin-Xiong She10, Ake Lernmark11, Beena Akolkar12, William A. Hagopian13, Desmond A. Schatz14, Jeffrey P. Krischer1, Heikki Hyöty15,16, Richard E. Lloyd2 and the TEDDY Study Group17

Viruses are implicated in autoimmune destruction of pancreatic islet β cells, which results in insulin deficiency and type 1 diabetes (T1D)1–4. Certain enteroviruses can infect β cells in vitro5, have been detected in the pancreatic islets of patients with T1D and have shown an association with T1D in meta-analyses6. However, establishing consistency in findings across studies has proven difficult. Obstacles to convincingly linking RNA viruses to islet autoimmunity may be attributed to rapid viral mutation rates, the cyclical periodicity of viruses7 and the selection of variants with altered pathogenicity and ability to spread in populations. β cells strongly express cell-surface coxsackie and adenovirus receptor (CXADR) genes, which can facilitate enterovirus infection8. Studies of human pancreata and cultured islets have shown significant variation in enteroviral virulence to β cells between serotypes and within the same serotype9,10. In this large-scale study of known eukaryotic DNA and RNA viruses in stools from children, we evaluated fecally shed viruses in relation to islet autoimmunity and T1D. This study showed that prolonged enterovirus B rather than independent, short-duration enterovirus B infections may be involved in the development of islet autoimmunity, but not T1D, in some young children. Furthermore, we found that fewer early-life human mastadenovirus C infections, as well as CXADR rs6517774, independently correlated with islet autoimmunity.

The Environmental Determinants of Diabetes in the Young (TEDDY) study is the largest prospective observational cohort study of newborns with increased genetic risk for type 1 diabetes (T1D) followed closely in several countries with diverse exposures, including viruses. Two nested–matched case–control studies within TEDDY were designed, with islet autoimmunity and T1D as the respective outcomes. Longitudinal stool samples were examined for virome content before the outcomes were determined. We hypothesized that increased prevalence of enterovirus is associated with increased likelihood of islet autoimmunity and/or T1D development in young children. We further explored the known human fecal virome for other viral associations with islet autoimmunity and T1D.

Metagenomic sequencing was performed on fecal specimens from 383 children with islet autoimmunity and 112 children with T1D (along with nested–matched paired controls) from six TEDDY study sites distributed across the United States, Germany, Sweden and Finland (Supplementary Table 1). Samples were collected approximately monthly, from the age of 3 months until the detection of either condition, totaling 8,654 stools for the islet autoimmunity study and 3,380 stools for the T1D study. To detect known RNA and DNA viruses, total nucleic acid was extracted from the stools, reverse transcribed and subjected to next-generation sequencing (NGS). Additionally, samples of each stool were cultivated on virus-susceptible cells to amplify enteroviruses, then nucleic acids were extracted and subjected to the same analysis using VirMAP18 in parallel. Proportions of stools with viruses were calculated from merged primary and cultured virome outputs containing presence–absence data. Overall, the virome content among the included stools consisted of 621 taxa, representing 96 genera of known eukaryotic virus and 57 genera of bacteriophage, as defined by the International Committee on Taxonomy of Viruses (2017 release). The relative proportions of viruses in the virome included bacteriophages (72%), mammalian viruses (20%) and those associated with the food stream, which were mostly plant viruses (8%) (Fig. 1a).

Of note, 55.8% of the samples were positive for any mammalian virus. Some of the abundant mammalian viruses listed (circovirus and gypivirus) originate in food sources; however, it is unclear whether these can replicate in humans. The most abundant human virus serotypes found in the stools of the children are shown in Supplementary Table 2.

The most prevalent viruses at the species level (>2% of positive stools) were human adenovirus, parechovirus, bocavirus,
enterovirus A (EV-A) and enterovirus B (EV-B) (Fig. 1b,c and Supplementary Table 3). Enterovirus (EV-A or EV-B) was detected in 12.8% of the 8,654 stools, and in at least one stool for 55.4% of the 766 nested-match case–control children in the islet autoimmunity study. An EV-B in at least one stool was observed in 36.6% of the children who developed islet autoimmunity (140/383) and 37.1% of their matched controls (142/383). The number of stools positive for EV-B (for each additional positive sample: odds ratio (OR) = 1.20; 95% confidence interval (CI) = 1.01–1.42; P = 0.04; Fig. 1d) was associated with developing islet autoimmunity. EV-A and EV-B were lower in Finland (Fig. 2a,b; P < 0.001) compared with other sites. Children with consecutive EV-B-positive stools were significantly more likely to have developed islet autoimmunity (OR = 3.05; 95% CI = 1.64–5.69; P = 0.0005; Fig. 1f). This association was similar across all matching strata, including across sites (ORs ≥ 1.7; Fig. 21). Adenovirus (39.6%), parechovirus (25.1%) and enterovirus (19.9%) were detected more frequently in stools from very young children aged 3–6 months (Supplementary Table 4).

Longitudinal analysis of stool samples taken from children before they developed islet autoimmunity revealed variable patterns of EV-B infection and shedding that were associated with islet autoimmunity (Fig. 3; P = 0.005). Sequence data from virus capsid regions enabled the identification of exact serotypes of EV-B in 81.2% of positive samples. A single EV-B infection (that is, a child with only one positive stool) was observed in 16.2% of children who developed islet autoimmunity (n = 62/383) and 20.4% of their matched controls (78/383) (Fig. 3). Next, we examined children with multiple EV-B infections (that is, more than one positive sample) and asked whether these were multiple positive stools for the same strain of one serotype, which would indicate a prolonged shedding period lasting more than 30 d. The definition of ‘same virus strain’ within a serotype was set as an RNA sequence homology of > 98% (that is, the evolution rate of acquired mutations during chronic EV-B infection1). The same EV-B serotype strain in more than one positive sample was observed in 11.8% of children who developed islet autoimmunity (45/383) and 6.5% of their matched controls (25/383). The majority of children who developed islet autoimmunity (n = 62/383) and 20.4% of their matched controls (78/383) (Fig. 3). Next, we examined children with multiple EV-B infections (that is, more than one positive sample) and asked whether these were multiple positive stools for the same strain of one serotype, which would indicate a prolonged shedding period lasting more than 30 d. The definition of ‘same virus strain’ within a serotype was set as an RNA sequence homology of > 98% (that is, the evolution rate of acquired mutations during chronic EV-B infection1). The same EV-B serotype strain in more than one positive sample was observed in 11.8% of children who developed islet autoimmunity (45/383) and 6.5% of their matched controls (25/383). The majority of children who developed islet autoimmunity (77.8%; n = 35/45) and their matched controls (64%; n = 16/25) with prolonged shedding were consecutively positive for a specific serotype. The amount of time shedding the same virus (median (interquartile range (IQR))) was 6.0 (1.5–13.1) months for children who developed islet autoimmunity and 4.1 (1.8–15.9) months for their matched controls. We also identified 3.9% of children who developed islet autoimmunity (15/383) and 1.6% of their matched controls (6/383) who were consecutively positive for EV-B in two or more stools whose virus read homology was just below 98% (homology 95–97%). This was usually due to a lack of sufficient overlapping sequence reads in common regions. These children were probably shedding the same virus, but this was unconfirmed by our definition, thus they were defined as having prolonged shedding in further analyses. The remaining children with multiple, independent, non-consecutive EV-B infections (n = 51) had similar results for islet autoimmunity association as children with a single infection. Children with prolonged shedding of the same EV-B serotype (OR = 2.50; 95% CI = 1.19–5.26) and children with consecutive positive results for different serotypes (OR = 2.18; 95% CI = 0.74–6.46) had similar higher odds of developing islet autoimmunity compared with children negative for EV-B, while children with a single (OR = 0.69; 95% CI = 0.45–1.06) or multiple non-consecutive result(s) (OR = 0.70; 95% CI = 0.30–1.36) did not.

The children with prolonged shedding of coxsackievirus B (CVB) were more likely to develop islet autoimmunity (OR = 2.49; 95% CI = 1.12–5.54; P = 0.03). However, when individual serotypes were studied, the children positive for CVB4 with no evidence of prolonged shedding were also more likely to develop islet autoimmunity (OR = 2.75; 95% CI = 1.18–6.40; P = 0.02; Fig. 3).

There was no association between age and the timing of the first appearance of EV-B infection, nor was there an association with time of infection before developing islet autoimmunity (Extended Data Fig. 1a–d). There was no overall difference between enterovirus infections and insulin or glumatic acid decarboxylase autoantibodies as the first-appearing autoantibody.

On further discovery of other associated viruses, we found that the number of samples positive for human mastadenovirus F (HAdV-F) up to seroconversion showed evidence of a correlation with the development of islet autoimmunity (OR = 1.33; 95% CI = 1.08–1.54; P = 0.007; false discovery rate (FDR) = 0.08; Fig. 1d). All but 11 of the samples positive for HAdV-F had the serotype HAdV-F41. HAdV-F40 was additionally found in seven samples from children who developed islet autoimmunity and four matched controls. One child was positive for both serotypes. The direction of the association was not consistent across all study sites (Fig. 2d,h).

Next, we examined the association between stool sample infections before the earliest seroconversion (6 months) and islet autoimmunity, based on previous TEDDY publications linking this time period to an increased risk of islet autoimmunity and TID. No association was observed between enterovirus and islet autoimmunity in this timeframe (Fig. 1e and Supplementary Table 4). However, human mastadenovirus C (HAdV-C) was detected in fewer children who developed islet autoimmunity than matched controls (OR = 0.55; 95% CI = 0.38–0.81; P = 0.003; FDR = 0.03) (Fig. 1e and Supplementary Table 4). The direction of this association was consistent across all study sites (Fig. 2c).

The propensity for finding EV-B in stools was strongly associated with the age of the child and the month the sample was collected (Supplementary Table 5). Nevertheless, prolonged shedding for EV-B (OR = 3.70; 95% CI = 1.90–7.22; P = 0.0001; Fig. 4 and Supplementary Table 6) remained strongly associated with islet autoimmunity when controlling for the child’s EV-B propensity and their genetic islet autoimmunity risk. After controlling for EV-B, HAdV-F, human leukocyte antigen (HLA) and specified single nucleotide polymorphisms (SNPs), children with positive

---

Fig. 1 | Stool virome composition up to 36 months of age, and common human viruses related to islet autoimmunity. a, Pie chart indicating the relative proportions of major classes of viruses found in the stool virome (n = 5,725 samples; % positive overall; combined primary and cultured stools) in the first 3 years of life. b,c, Bar graphs indicating the estimated prevalence in the TEDDY cohort (n = 6,890 children at risk for islet autoimmunity and TID who were observed up to 36 months of age) of the 20 most abundant mammalian viruses (b) and enterovirus serotypes found belonging to the two most abundant enterovirus species (EV-A and EV-B) within the same stool dataset (c). The cohort prevalences were estimated by weighting the proportion of stool samples that were positive for the virus, to account for how the children were selected as controls (n = 495) into the two nested case–control studies. d-f, Forest plots showing how common human viruses relate to the odds of developing islet autoimmunity. The effects of increasing numbers of virus-positive samples (d; n = 4,327 matched-pair samples), whether or not children were virus positive from 3–6 months (e; n = 370 matched-pair children) and whether or not children were positive for the common virus in at least two consecutive samples (f; n = 383 matched-pair children) are shown. Circles and bars represent ORs and 95% CIs, respectively, and were calculated using conditional logistic regression models with adjustment for the HLA-DR-DQ genotype. All P values are two sided. OR > 1 indicates a positive correlation between the virus pattern and the development of islet autoimmunity, whereas OR < 1 indicates an inverse correlation. Black circles and CI bars represent non-significant associations. Red circles and CI bars represent statistically significant associations for enterovirus (a priori hypothesis) at P < 0.05 or for other viruses that showed an FDR < 0.05.

1866
HAdV-C in early samples had a lower risk of islet autoimmunity (OR = 0.49; 95% CI = 0.31–0.75; P = 0.001) (Fig. 4 and Supplementary Table 6). The number of positive samples with HAdV-F showed a weak increased risk for islet autoimmunity (P = 0.04, Supplementary Table 6).

We examined the association of EV-B and islet autoimmunity with a panel of SNPs reported to regulate antiviral responses. These SNPs were not associated with islet autoimmunity, although rs2304256 in TYK2 was associated with EV-B presence in stools (OR = 1.36; 95% CI = 1.04–1.78; P = 0.02; Supplementary Table 5).
Letters Nature Medicine

**Letters Nature Medicine**

SNPs rs1539798 (minor allele frequency 0.87; OR 1.47; 95% CI 1.10–1.98; P=0.01; Fig. 4 and Supplementary Table 6), particularly among first-degree relatives (OR 0.80–4.20; P=0.05). Significance was assessed using conditional logistic regression adjusted for the HLA-DR-DQ genotype. All P-values are two sided.

**Fig. 2 | Percentages of children positive for a specific virus between the ages of 3 and 6 months, and percentages of children with consecutive positive samples before islet autoimmunity development, by clinical site.**

Based on the strong association between EV-B and HAdV-C findings and islet autoimmunity, five SNPs typed in the coxsackie and adenovirus receptor gene (CXADR) region (chr21q21.1) were examined. The chromosomal position order was rs6517774, rs1967939, rs1539798, rs2824400 and rs2824404 (ref. 13). The SNPs rs1539798 (minor allele frequency <1%) and rs2824400 (r>0.99 with rs2824404) were excluded. The presence of the minor rs6517774-G allele correlated strongly with a lower propensity for a child to have a stool positive for EV-B (OR=0.66; 95% CI=0.50–0.87; P=0.003; Supplementary Table 5). In contrast, each additional minor rs6517774-G allele increased the odds of islet autoimmunity (OR=1.47; 95% CI=1.10–1.98; P=0.01; Fig. 4 and Supplementary Table 6), particularly among first-degree relatives (OR=3.76; 95% CI=1.62–8.73; interaction: P=0.003). The SNP rs2824404 was associated with islet autoimmunity only in Finland (per additional minor C allele (HLA-adjusted): OR=1.68; 95% CI=1.09–2.57; P=0.02; Finland×SNP interaction: P=0.03). No significant interactions between EV-B and SNPs (either in CXADR or in other gene regions) or HLA were observed.

In the T1D nested-matched case–control study, there was a higher frequency of first-degree relatives with T1D (35.7%) compared with the islet autoimmunity nested-matched case–control study (21.9%) (Supplementary Table 1). In total, 77.7% of children who developed T1D (87/112) and 5.4% of their matched controls (6/112) developed islet autoimmunity. The number of stools positive for EV-B was lower among those with T1D compared with their matched controls (OR=0.73; 95% CI=0.53–0.99; P=0.05; Extended Data Fig. 2a and Supplementary Table 7). This was due to a lower frequency of multiple EV-B infections in those with T1D (Supplementary Table 8). Eight children with T1D had prolonged shedding of the same EV-B serotype. Two children, who were matched controls, also had prolonged shedding of the same EV-B serotype, one of whom developed islet autoimmunity. All eight children with T1D and one matched control who developed islet autoimmunity started their prolonged shedding before seroconversion. Additionally, HAdV-C was less likely to be detected in early stools from children who developed T1D (age ≤6 months) compared with their matched controls (OR=0.45; 95% CI=0.19–1.07; P=0.07; Extended Data Fig. 2b and Supplementary Table 4), as was the case for islet autoimmunity (see above). No association with HAdV-F or other viruses was observed.

Lastly, further examination of the CXADR region showed a similar association with T1D as it had for islet autoimmunity (HLA and islet autoimmunity-adjusted OR=1.83; 95% CI=0.80–4.20; P=0.015). There were 27 case–control pairs in the T1D study who had the same islet autoimmunity status during follow-up (23 negative for islet autoimmunity and four positive) and 85 who did not (83 children with T1D and two without were positive for islet autoimmunity while their matched pair was negative).
This large-scale study used NGS to analyze the whole known virome in children with increased genetic risk for T1D, and included analysis of genetically linked SNPs associated with islet autoimmunity. Here, we report that prolonged shedding of the same EV-B serotype in children was strongly associated with islet autoimmunity development, but not the development of T1D. Independent, short-duration EV-B infections in children were not associated with islet autoimmunity or T1D, indicating that the type and duration of an infection may be critical. Additionally, our results showed that having a HAdV-C infection in very early life was associated with a decreased risk of islet autoimmunity and T1D. Independently, children who carried the minor SNP allele rs6517774 in the CXADR gene region were more likely to develop islet autoimmunity.

Previous work examining the stool virome in both small retrospective case–control and larger prospective-based nested case–control studies have been inconclusive. The Finnish Diabetes Prediction and Prevention study has repeatedly reported an association between enterovirus infections and subsequent initiation of islet autoimmunity. This association was recently observed by a large-scale study using NGS to analyze the whole known virome in children with increased genetic risk for T1D, and included analysis of genetically linked SNPs associated with islet autoimmunity. Here, we report that prolonged shedding of the same EV-B serotype in children was strongly associated with islet autoimmunity development, but not the development of T1D. Independent, short-duration EV-B infections in children were not associated with islet autoimmunity or T1D, indicating that the type and duration of an infection may be critical. Additionally, our results showed that having a HAdV-C infection in very early life was associated with a decreased risk of islet autoimmunity and T1D. Independently, children who carried the minor SNP allele rs6517774 in the CXADR gene region were more likely to develop islet autoimmunity.

Previous work examining the stool virome in both small retrospective case–control and larger prospective-based nested case–control studies have been inconclusive. The Finnish Diabetes Prediction and Prevention study has repeatedly reported an association between enterovirus infections and subsequent initiation of islet autoimmunity. This association was recently observed by a large-scale study using NGS to analyze the whole known virome in children with increased genetic risk for T1D, and included analysis of genetically linked SNPs associated with islet autoimmunity. Here, we report that prolonged shedding of the same EV-B serotype in children was strongly associated with islet autoimmunity development, but not the development of T1D. Independent, short-duration EV-B infections in children were not associated with islet autoimmunity or T1D, indicating that the type and duration of an infection may be critical. Additionally, our results showed that having a HAdV-C infection in very early life was associated with a decreased risk of islet autoimmunity and T1D. Independently, children who carried the minor SNP allele rs6517774 in the CXADR gene region were more likely to develop islet autoimmunity.

Previous work examining the stool virome in both small retrospective case–control and larger prospective-based nested case–control studies have been inconclusive. The Finnish Diabetes Prediction and Prevention study has repeatedly reported an association between enterovirus infections and subsequent initiation of islet autoimmunity. This association was recently observed by a large-scale study using NGS to analyze the whole known virome in children with increased genetic risk for T1D, and included analysis of genetically linked SNPs associated with islet autoimmunity. Here, we report that prolonged shedding of the same EV-B serotype in children was strongly associated with islet autoimmunity development, but not the development of T1D. Independent, short-duration EV-B infections in children were not associated with islet autoimmunity or T1D, indicating that the type and duration of an infection may be critical. Additionally, our results showed that having a HAdV-C infection in very early life was associated with a decreased risk of islet autoimmunity and T1D. Independently, children who carried the minor SNP allele rs6517774 in the CXADR gene region were more likely to develop islet autoimmunity.

Previous work examining the stool virome in both small retrospective case–control and larger prospective-based nested case–control studies have been inconclusive. The Finnish Diabetes Prediction and Prevention study has repeatedly reported an association between enterovirus infections and subsequent initiation of islet autoimmunity. This association was recently observed by a large-scale study using NGS to analyze the whole known virome in children with increased genetic risk for T1D, and included analysis of genetically linked SNPs associated with islet autoimmunity. Here, we report that prolonged shedding of the same EV-B serotype in children was strongly associated with islet autoimmunity development, but not the development of T1D. Independent, short-duration EV-B infections in children were not associated with islet autoimmunity or T1D, indicating that the type and duration of an infection may be critical. Additionally, our results showed that having a HAdV-C infection in very early life was associated with a decreased risk of islet autoimmunity and T1D. Independently, children who carried the minor SNP allele rs6517774 in the CXADR gene region were more likely to develop islet autoimmunity.
in dendritic cells\(^1\). Another mechanism could be that EV-B is prone to 5’ terminal genomic deletions, leading to prolonged non-lytic infections\(^2\). Alternatively, prolonged shedding could be caused by weaker immune protection in children with disease. Prolonged enterovirus shedding, as previously recognized\(^26-28\), might be a bio-
sample (all previously reported to be associated with islet autoimmunity in the TeDDY cohort\(^40\), and confirmed as the main receptor to infect cells, and this tight junction-type mol-
cule is strongly expressed by \(\beta\) cells, particularly in the membranes of insulin granules\(^32\). CXADR is also used by adenoviruses. There is enhanced expression of the CXADR gene in the pancreas of patients with both islet autoimmunity and T1D\(^32\). This receptor is important for virus internalization into cells\(^31\), and might induce inflammation and tissue damage\(^51\). As supportive analysis, we found that children carrying the minor rs6517774-G allele within the CXADR region had an increased risk for islet autoimmunity, but a lower number of EV-B infections. Children who had both minor alleles for rs6517774 and rs2824404 (primarily in Finland) were more likely to develop islet autoimmunity in the absence of EV-B, suggesting greater penetrance of these SNPs in low-threshold exposure to EV-B. Caution is advised when interpreting these SNP findings, as they need to be examined independently.

We report a potentially interesting finding that adenovirus is associated with islet autoimmunity. Both coxsackievirus and adenovirus use CXADR to infect cells. Nevertheless, we found a protective association with HAdV-C. The mechanism of action for this association is unknown. However, early HAdV-C infections before the age of 6 months were associated with a low risk of islet autoimmunity. No other detected virus was associated with a low risk of islet autoimmunity, which suggests that some HAdV-C factors may be involved. Since HAdV-C and CVB use the same receptor, which has been shown to be associated with increased islet autoimmunity risk, we speculate that competition in receptor binding might be a

| Controls | Cases |
|----------|-------|
| n (%)    | n (%) |
| P-value  |       |

| HLA-DR-DQ genotype versus other |
|---------------------------------|
| DR3-DQ2/DR4-DQ8 | 146 (38.1) | 202 (52.7) | <0.0001 |
| DR4-DQ8/DR4-DQ8 | 70 (18.3)  | 62 (16.2)  | 0.04    |
| DR4-DQ8/DR8-DQ4 | 62 (16.2)  | 59 (15.4)  | 0.25    |

| SNPs (minor allele frequency) associated with islet autoimmunity |
|---------------------------------------------------------------|
| rs2476601 in PTPN22 | 10.8 | 18.0 | 0.0001 |
| rs 1004446 in INS | 36.8 | 32.2 | 0.02  |
| rs 2292239 in ERBB3 | 30.8 | 37.9 | 0.001 |
| rs3184504 in SH2B3 | 46.1 | 52.5 | 0.008 |
| rs6517774 in CXADR | 36.3 | 42.3 | 0.01  |

| Viruses associated with islet autoimmunity |
|-------------------------------------------|
| EV-B prolonged shedding (yes versus no)   |
| HAdV-C in stool samples aged 3–6 months (yes versus no) | 119 (32.2) | 84 (22.7) | 0.001 |
| One sample positive for HAdV-F versus none | 77 (20.1)  | 74 (19.3) | 0.43  |
| Two samples positive for HAdV-F versus none | 37 (9.7)  | 33 (8.6)  | 0.06  |
| Three or more samples positive for HAdV-F versus none | 33 (8.6)  | 44 (11.5) | 0.006 |

Fig. 4 | Multivariable conditional logistic regression of EV-B, HAdV-C and HAdV-F on islet autoimmunity status. A total of 378 children who developed islet autoimmunity, along with their matched controls, were included. This model includes the effects on islet autoimmunity status of HLA-DR-DQ genotypes, SNPs in PTPN22, INS, ERBB3 or SH2B3 (all previously reported to be associated with islet autoimmunity in the TEDDY cohort\(^40\), and confirmed as the main receptor to infect cells. Nevertheless, we found a protec-
tion of low-grade enteroviral infection in the islets of patients with T1D\(^36,31\). Thus, functionally 'defective' enterovirus might focally attack and persist within \(\beta\) cells with low levels of viral replication\(^6\). Viral tropism depends largely on the expression of viral receptors by the cell. Interestingly, among enteroviruses, only CVB uses CXADR as the main receptor to infect cells, and this tight junction-type mol-
ecule is strongly expressed by \(\beta\) cells, particularly in the membranes
mechanism for protection by HAdV-C. Conversely, HAdV-C infections were frequent and they are known to persist for long periods; for example, in tonsillar tissues, where they may cause long-lasting activation in the innate immune system of the respiratory tract. This could protect against other viruses that replicate in the same tissues, including islet autoimmunity-associated enteroviruses. In fact, this viral interference has been suggested between adenoviruses and rhinoviruses, and interferon-λ signaling in the gut activated by astrovirus can protect against norovirus infection in the mouse. However, in the present study, we did not find significant interactions between HAdV-C and EV-B; thus, further studies are needed to determine whether viral interference could explain our adenovirus finding.

Although our sample size was large, there were limitations to the power of our study, especially in the evaluation of virus serotypes. This may be attributed to the constant geographic movement and evolution of enteroviral serotypes, which shift between variants with altered pathogenicity. Notably, the association between prolonged shedding of EV-B and islet autoimmunity risk was observed among stools with higher compliance in children up to the age of 36 months. We were unable to adequately measure virus abundance due to the timing of infection onset being unknown relative to when the stool was collected, as well as the abundance measures being inappropriate for the sequencing methodology. We deliberately cultivated on cells to increase the enterovirus hits by amplifying the virus signal, but all viruses were considered. Both primary and cultured virome data were combined before analysis, but they also showed the same association between EV-B and islet autoimmunity when analyzed separately. Stool collection began at age 3–4 months, and compliance declined from the age of 2 years onwards. Therefore, the number of children with consecutive positive/shedding may have been underestimated.

Overall, the observed virome composition aligned with the existing knowledge about viral epidemiology in children. The most frequently detected human viruses (human mastadenovirus, parachoervirus, bocavirus, mammastrovirus, norovirus and enterovirus) were common in previous studies using similar technologies. It will be important to identify different immunological response profiles to viral infection. While some EV-B serotypes were variably associated with islet autoimmunity, future studies will require more targeted and sensitive approaches to detect prolonged shedding, such as AmpliSeq. Both epidemiological and bench investigations are needed to identify the serotypes that are most prone to triggering islet autoimmunity or protecting against it, and to discern the host and viral mechanisms that lead to prolonged infections and the lack of virus clearing in the gut. Although coxsackievirus vaccines are already in development, more experimental work is needed to elucidate the mechanisms by which long-term or persistent virus infections promote immunological dysregulation leading to islet autoimmunity.

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/s41591-019-0667-0.

Received: 18 June 2019; Accepted: 25 October 2019; Published online: 2 December 2019

References

1. Van der Werf, N., Kroese, F. G. M., Rozing, J. & Hillebrandt, J. L. Viral infections as potential triggers of type 1 diabetes. Diabetes Metab. Res. Rev. 23, 169–183 (2007).

2. Stene, L. C. et al. Enterovirus infection and progression from islet autoimmunity to type 1 diabetes: the Diabetes and Autoimmunity Study in the Young (DAISY). Diabetes 59, 3174–3180 (2010).

3. Laitinen, O. H. et al. Coxsackievirus B1 is associated with induction of β-cell autoimmunity that portends type 1 diabetes. Diabetes 63, 446–455 (2014).

4. Yeung, W. C., Rawlinson, W. D. & Craig, M. Enterovirus infection and type 1 diabetes mellitus: systematic review and meta-analysis of observational molecular studies. Br. Med. J. 342, d35 (2011).

5. Anagandula, M. et al. Infection of human islets of Langerhans with two strains of coxsackie B virus serotype 1: assessment of virus replication, degree of cell death and induction of genes involved in the innate immune pathway. J. Med. Virol. 86, 1402–1411 (2014).

6. Krogh, L. et al. Detection of a low-grade enteroviral infection in the islets of Langerhans of living patients newly diagnosed with type 1 diabetes. Diabetes 64, 1682–1687 (2015).

7. Abedi, G. R. et al. Enterovirus and human parechovirus surveillance—United States, 2009–2013. Morb. Mortal. Weekly Rep. 64, 949–954 (2015–2016).

8. Heathcote, E. et al. Unexpected subcellular distribution of a specific isofrom of the coxsackie and adenovirus receptor, CAR-SIV, in human pancreatic beta cells. Diabetologia 61, 2344–2355 (2018).

9. Roivainen, M. et al. Mechanisms of coxsackievirus-induced damage to human pancreatic beta-cells. J. Clin. Endocrinol. Metab. 85, 432–440 (2000).

10. Roivainen, M. et al. Functional impairment and killing of human beta cells by enteroviruses: the capacity is shared by a wide range of serotypes, but the extent is a characteristic of individual virus strains. Diabetologia 45, 693–702 (2002).

11. Ajami, N. J., Wong, M. C., Ross, M. C., Lloyd, R. E. & Petrosino, J. F. Maximal viral information recovery from sequence data using VirMAP. Nat. Commun. 9, 3205 (2018).

12. Laasri, M. et al. Evolution of echovirus 11 in a chronically infected immunodeficient patient. PLoS Pathog. 14, e1006943 (2018).

13. Schneider, V. A. et al. Evaluation of GRC3h8 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. Genome Res. 27, 849–864 (2017).

14. Tapias, G. et al. Human enterovirus RNA in monthly fecal samples and islet autoimmunity in Norwegian children with high genetic risk for type 1 diabetes: the MIDIA study. Diabetes Care 34, 151–155 (2011).

15. Honkanen, H. et al. Detection of enteroviruses in stools precedes islet autoimmunity by several months: possible evidence for slowly operating mechanisms in virus-induced autoimmunity. Diabetologia 60, 424–431 (2017).

16. Simonen-Tikka, M. L. et al. Human enterovirus infections in children at increased risk for type 1 diabetes: the Babyladit study. Diabetologia 54, 2995–3002 (2011).

17. Oikarinen, S. et al. Enterovirus RNA in blood is linked to the development of type 1 diabetes. Diabetes 64, e276–e279 (2015).

18. Salminen, K. et al. Enterovirus infections are associated with the induction of beta-cell autoimmunity in a prospective birth cohort study. J. Med. Virol. 91, 98–93 (2008).

19. Cinek, O. et al. Enterovirus RNA in longitudinal blood samples and risk of islet autoimmunity in children with a high genetic risk of type 1 diabetes: the MIDIA study. Diabetologia 57, 1270–1272 (2014).

20. Lee, H. S. et al. Next-generation sequencing for viruses in children with rapid-onset type 1 diabetes. Diabetes 56, 1705–1711 (2013).

21. Sioofy-Khojine, A. B. et al. Coxsackievirus-B1 infections are associated with the initiation of insulin-driven autoimmunity that progresses to type 1 diabetes. Diabetes 61, 1193–1202 (2012).

22. Oikarinen, S. et al. Virus antibody survey in different European populations indicates risk association between coxsackievirus-B1 and type 1 diabetes. Diabetes 63, 655–662 (2014).

23. Bessaud, M., Joffret, M. L., Blondel, B. & Depeyroux, E. Exchanges of genomic domains between poliovirus and other cocirculating species C enteroviruses reveal a high degree of plasticity. Sci. Rep. 6, 38831 (2016).

24. Hamalainen, S. et al. Coxsackievirus-B1 reveals strain-specific differences in plasmacytoid dendritic cell-mediated immunogenicity. J. Med. Virol. 86, 1411–1420 (2014).

25. Leveque, N. et al. Functional consequences of RNA 5′-terminal deletions on coxsackievirus B3 RNA replication and ribonucleoprotein complex formation. J. Virol. 91, e00423-17 (2017).

26. Chung, P. W., Huang, Y. C., Chang, L. Y., Lin, T. Y. & Ning, H. C. Duration of enterovirus shedding in stool. J. Microbiol. Immunol. Infect. 49, 167–170 (2005).

27. Alexander, J. P. Jr., Gary, H. E. Jr. & Pallansch, M. A. Duration of poliovirus excretion and its implications for acute flaccid paralysis surveillance: a review of the literature. J. Infect. Dis. 175, S176–S182 (1997).

28. Melnick, J. L. & Rennick, V. Infectivity titers of enterovirus as found in human stools. J. Med. Virol. 5, 205–220 (1980).

29. Ylipaasto, P. et al. Enterovirus infection in human pancreatic islet cells, islet tropism in vivo and receptor involvement in cultured islet beta cells. Diabetologia 47, 225–239 (2004).

30. Richardson, S. J., Leete, P., Bone, A. J., Fouillis, A. K. & Morgan, N. G. Expression of the enteroviral capsid protein VP1 in the islet cells of patients with type 1 diabetes is associated with induction of protein kinase R and downregulation of MeCP1. Diabetologia 56, 185–193 (2013).
31. Busse, N. et al. Detection and localization of viral infection in the pancreas of patients with type 1 diabetes using short fluorescently-labelled oligonucleotide probes. *Oncotarget* 8, 12620–12636 (2017).

32. Hodiak, M. et al. Coxsackie–adenovirus receptor expression is enhanced in pancreas from patients with type 1 diabetes. *BMJ Open Diabetes Res. Care* 4, e000219 (2016).

33. Shafren, D. R. et al. Coxsackieviruses B1, B3, and B5 use decay accelerating factor as a receptor for cell attachment. *J. Virol.* 69, 3873–3877 (1995).

34. Itou, M. et al. Expression of coxsackievirus and adenovirus receptor in hearts of rats with experimental autoimmune myocarditis. *Circ. Res.* 86, 275–280 (2000).

35. Garnett, C. T. et al. Latent species C adenoviruses in human tonsil tissues. *J. Virol.* 83, 2417–2428 (2009).

36. Wang, Z. et al. Broad spectrum respiratory pathogen analysis of throat swabs from military recruits reveals interference between rhinoviruses and adenoviruses. *Microb. Ecol.* 59, 623–634 (2010).

37. Ingle, H. et al. Viral complementation of immunodeficiency confers protection against enteric pathogens via interferon-λ. *Nat. Microbiol.* 4, 1120–1128 (2019).

38. Messacar, K., Abzug, M. J. & Dominguez, S. R. 2014 outbreak of enterovirus D68 in North America. *J. Med. Virol.* 88, 739–745 (2016).

39. Greninger, A. L. et al. A novel outbreak enterovirus D68 strain associated with acute flaccid myelitis cases in the USA (2012–14): a retrospective cohort study. *Lancet Infect. Dis.* 15, 671–682 (2015).

40. 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).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019
Methods

Six clinical research centers—three in the United States (Colorado, Georgia/Florida and Washington) and three in Europe (Finland, Germany and Sweden)—participated in population-based human leukocyte antigen (HLA) screening of newborns between 2008 and 2010. Children who were at-risk HLA genotypes (as described) were enrolled (n = 8,676) and prospectively followed from 3 months of age to 15 years. Study visits included a blood draw every 3 months until 4 years, and every 3 or 6 months thereafter, depending on islet autoimmunity positivity. HLA (DR-DQ isoform) genotypes were confirmed by reverse blot hybridization at the Center for Health Research Laboratory at Roche Molecular Systems53. SNPs were genotyped by the Center for Public Health Genomics at the University of Virginia using the Illumina IllumiChip, which is a custom array for genotyping SNPs selected from regions of the human genome firmly associated with autoimmune diseases54. The final selection of SNPs, including 196,524 SNPs in 186 regions for 12 autoimmune diseases, was decided by the IllumiChip Consortium. Stool samples were collected monthly from the age of 3–48 months, and then quarterly throughout the age of 10 years, as previously described55-57.

Persistent confirmed autoimmunity was defined by the presence of confirmed islet autoimmunity (glutamic acid decarboxylase, insulinoma-associated 2 or insulin) in each of the two TEDDY laboratories on two or more consecutive visits45. T1D diagnosis was defined according to American Diabetes Association criteria46. The detailed study design and methods have been published previously58-60. The protocol was approved by each local institutional review board. A two-step consent process was used in which the first consent was given by the legal guardian(s) and the second assent by the TEDDY child completed at an appropriate age as determined by the local institutional review board/ethics board.

Nested-matched case–control population

A nested-matched case–control population was selected through risk-set sampling using metadata and islet autoimmunity sample results from 31 May 2012, as previously detailed61. In a separate nested-matched case–control population, for each child diagnosed with T1D, a control was selected based on the same family history of T1D (general population or first-degree relative) at the time of diagnosis47. Due to the lack of matching stool sample availability, there was a 9% reduction in the number of pairs for the 1:1 viral metagenomic study for the islet autoimmunity nested-matched case–control population, which left 383 pairs (n = 4,327 age-matched stool samples in each group). There was also a 2% reduction in the number of pairs for the 1:1 T1D nested-matched case–control population, resulting in 112 pairs (n = 1,690 age-matched stool samples in each group). All participants were at least 6 months of age at the time of risk-set sampling. Of the 383 children who developed islet autoimmunity, 93 were included with incident, persistent, confirmed islet autoimmunity who were matched to a control selected on the same family history of T1D (general population or first-degree relative) at the time of diagnosis47. Due to the lack of matching stool sample availability, there was a 2% reduction in the number of pairs for the 1:1 viral metagenomic study for the islet autoimmunity nested-matched case–control population, which left 383 pairs (n = 4,327 age-matched stool samples in each group). There was also a 2% reduction in the number of pairs for the 1:1 T1D nested-matched case–control population, resulting in 112 pairs (n = 1,690 age-matched stool samples in each group).

Metagenomic whole-genome shotgun (WGS) sequencing and analysis

The nested-matched case–control populations for both the islet autoimmunity and T1D designs included 8,654 and 3,380 matched total stool samples that passed quality filtering and were included in subsequent blinded testing, as previously published45. Filtreates of stool samples were aliquoted, snap-frozen and stored at −80°C, then passed through a 1.2-μm filter) were directly extracted for total nucleic acids, then incubated on mixtures of four cell lines (Hela cells, Vero cells, RD cells expressing coxsackievirus and adenosine receptor, and HEK-293 cells; 25% each, plated at 40% overall confluency) in Dulbecco’s modified Eagle medium containing 2% calf serum for 6 d, to amplify the viruses present at very low levels. Cell lines were chosen for the breadth of virus replicative efficiency of type A, B and C enteroviruses, as well as other common viruses. Infected cultures were not passaged. Cells and supernatants were collected for total nucleic acid extraction and analysis.

Total nucleic acids were extracted and sequenced independently in parallel from both primary stool and cultured stool samples (cells and supernatants) using a MagMAX Viral RNA Isolation Kit (catalog number AM1939; Thermo Fisher Scientific) without employing DNase to prevent DNA extraction. Extracted viral RNA was reverse transcribed using SuperScript II Reverse Transcriptase (catalog number 18064410; Thermo Fisher Scientific) and random hexamers. After short molecule and random hexamer removal using a ChargeSwitch kit (catalog number C512000; Thermo Fisher Scientific), molecules were amplified and tagged with a 12-base-pair (bp) barcode tag containing a V8A2 semi-random primer (BC12-V8A2 construct using AccuPrime Taq DNA Polymerase and cleaned with the ChargeSwitch kit). Tags were attached via a barcoded, semi-random primer construct43, resulting in dual-barcoded (same barcode on both sides) molecules used for Illumina. Double barcoded amplicons derived from each sample were pooled at 30 per sequencing lane, and a single WGS library was generated on the pool without shearing. Separate negative controls were introduced during the extraction, amplification and library preparation steps. Positive controls (a mix of four laboratory strain viruses, poliovirus, simian rotavirus, adenosine virus and mouse hepatitis virus) were included blindly in certain wells of all of the sample plates. We performed a single WGS library preparation per sequencing lane of pooled, pre-barcoded samples, minimizing carryover as each lane only had a single index. Additionally, because the samples all carried secondary internal barcodes, they were not subject to the carryover or cross-breed sometimes observed from run to run with library indices on the same platform. The raw sequence data were processed using custom pipelines employing a Bioanalyzer to ensure the appropriate range for the platform (~200–1,000 bp), then the library was loaded in an Illumina HiSeq 2000 (Illumina) and sequenced using 2 × 100 bp chemistry at the Human Genome Sequencing Center at Baylor College of Medicine. Stool samples were received blinded, without metadata, from the NIDDK Repository in defined ‘runs’ that were generally sequenced on the same sequencing flowcell in different lanes. Reads were demultiplexed into a sample bin using the barcode prefixing read-1 and read-2, allowing zero mismatches. Demultiplexed reads were further processed by trimming off barcodes, semi-random primer sequences and Illumina adapters. This process utilized a custom demultiplexer and the BBduk algorithm included in BBDMap62.

Virome analysis

The resulting trimmed primary and cultured virome datasets were analyzed using a pipeline created at the Alkek Center for Metagenomics and Microbiome Research at Baylor College of Medicine, employing a clustering algorithm (ViruMAP) that reconstitutes putative viral genomes using a mapping assembly strategy that leverages both nucleotide and translated nucleotide
alignment information. ViMAP uses a custom-formatted version of gbk/GenBank (virus) and gbpfh/GenBank (Phage); however, to filter out false positives, all other GenBank organismal divisions were used as a master database. Viral taxonomies were acquired using a scoring system that incorporates and translated nucleotide alignment results in a bits-per-base fashion and optimizes for the highest-resolution taxonomic rank, generating a taxonomy ID aggregate bit score output. ViMAP testing and evaluation with in-house and public metagenomic viral datasets indicated that the 300 aggregate bit-score threshold provides superior accuracy and sensitivity. ViMAP has undergone extensive validation utilizing mock viral communities and publically available reference datasets that also report PCR analysis.

Comparative analysis of differences in the determination of virus positivity was performed at four ViMAP aggregate bit scores (100, 400, 700 and 1,000), which returned 1,487,1,280, 1,078 and 971 enterovirus-positive stools, respectively. From this analysis, an aggregate bit-score of 400 was chosen as the most conservative for accuracy, which eliminated false positives, reduced the risk of low-level cross-sample contamination, and increased the accuracy of enterovirus serotype determinations, yet maintained sensitivity for output for statistical analysis. The distribution of enterovirus reads in enterovirus-positive stools at the 400 aggregate bit-score level ranged from 12–27,396,504 reads, with a median of 13,117 and a mean of 15,233 reads. ViMAP analysis showed that EV-A or EV-B was present in 12.8% of stools, which compared well with PCR analysis of stools in a stool bit-score level ranged from 12–27,396,504 reads, with a median of 13,117 and a distribution of enterovirus reads in enterovirus-positive stools at the 400 aggregate bit-score level were converted to ‘yes’ or ‘no’ scores (1 versus 0). Virus read counts could range from 1 to 400 aggregate bit-scores, which were used to calculate the odds ratio of EV-B results (Extended Data Fig. 4).

The two samples detected by PCR were also detected by NGS (both primary and cultured samples). The prevalence of viruses in stools of children up to 36 months of age was evaluated using conditional logistic regression models adjusting for HLA genotypes. The magnitudes of the associations were assessed by ORs with 95% CIs. It was an a priori hypothesis to examine the role of enteroviruses, given the evidence from previous studies. For the remaining common viruses (n = 12), and with consideration of multiple comparisons, an FDR <10% was the criteria for further evaluation.

Significant viral associations were next examined, adjusting for genetic factors by including the associated SNP risk factors in the conditional logistic model. SNPs in MD221 (IFIH1), PTPN2, TYK2 and BACH2, as well as SNP previously associated with either iatrogenic autoimmunity or T1D/1.1, were considered. The influence of potential non-genetic confounders at the sampling and subject level was controlled for by including in the model a propensity score for whether a child would have a stool sample positive for the virus. The propensity score was created to adjust for risk factors associated with the probability of EV-B, and to better estimate the average effect of EV-B on the risk of iatrogenic autoimmunity without making strong assumptions on how iatrogenic autoimmunity is related to the risk factors associated with EV-B. The virus propensity score was calculated from a logistic random-effects model that regressed stool sample virus positivity on the age of the child, the year and month of stool sample collection, matching factors, case status, and common enterovirus reads. The score for each stool sample was estimated using the fixed effects in the final model, assuming that all children were controls, and then calculating the marginal prediction. The scores were then averaged across samples of the child to create subject-level propensity scores.

Specifically, the regression adjustment and calculation of the propensity score from the case–control data were performed by fitting a model for \( P(E) = 1[D,D] \) and predicting the exposure for each stool sample as though it were a control. The discordance between cases and matched controls, both in the number of stool samples positive for each virus and the length of the prolonged shedding period (duration) was not significantly associated with case status in the children with consecutive positive EV-B results (Extended Data Fig. 4).

For analysis, the virus-positive reads above the ViMAP 400 aggregate bit-score level were converted to ‘yes’ or ‘no’ scores (1 versus 0). Virus read counts could not be attributed to infection since the stool samples are unknown; thus, we could only reliably determine whether or not the virus was present in the particular stool. Reduction of the information content and dimensionality of the data was carried out to quantify the association between the presence of viral taxa and the development of iatrogenic autoimmunity. The methods employed, starting from sample collection and ending with sequencing, preclude a quantitative analysis of the viral taxa found mainly due to the use of the untargeted approaches in the data-generation workflow. Briefly, the extraction process aims to produce clean nucleic acids indiscriminately from any source present. Although the PCR amplification process uses a semi-random primer, it does not provide enough specificity to distinguish viral from non-viral genomes. Moreover, the sequencing step involves a complex purification process, and the child’s samples were used for the SNP rs6517774 in CXADR but having a minor allele of the SNP rs23042556 in TYK2 all showed a higher propensity to have an EV-B-positive stool sample. The average was taken across stool samples of the child for use as an
overall propensity score. Supplementary Table 9 shows how the EV-B groups were distributed across the overall propensity score ranges. As expected, there were more children with consecutive positive EV-B results at higher propensity score ranges. This propensity was included in the conditional logistic model to adjust for non-genetic T1D risk factors that may confound the association.

Viruses showing an association with islet autoimmunity (or T1D, if appropriate) were further characterized by testing for interactions with HLA and SNPs. All P-values were two sided, and \( P < 0.05 \) was considered significant unless otherwise stated. SAS 9.4 (SAS Institute) was used for the statistical analysis, and GraphPad Prism 8.0 was used for the figures.

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

**Data availability**
TEDDY virome sequencing data that support the findings of this study have been deposited in the NCBI database of Genotypes and Phenotypes (dbGaP) with the primary accession code phs001442, in accordance with the dbGaP controlled-access authorization process. Clinical metadata and virome results data analyzed for the current study will be made available in the NIDDK Central Repository at https://www.niddkreppository.org/studies/teddy.

**Code availability**
VirMAP was used to generate the virome data and has been deposited in GitHub (https://github.com/cmnr/virmap). All of the software code and dependencies are listed on the GitHub site. SAS 9.4 (SAS Institute) was used for the statistical analysis, and GraphPad Prism 8.0 was used to create the figures.

**References**
41. 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. *Diabetes*. 62, 733–743 (2011).
42. Parkes, M., Cortes, A., van Heel, D. A. & Brown, M. A. Genetic insights into common pathways and complex relationships among immune-mediated diseases. *Nat. Rev. Genet.* 14, 661–673 (2013).
43. Stewart, C. J. et al. Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature*. 562, 583–588 (2018).
44. Vatanen, T. et al. The human gut microbiome in early-onset type 1 diabetes from the TEDDY study. *Nature*. 562, 589–594 (2018).
45. Bonifacio, E. et al. Harmonization of glutamic acid decarboxylase and islet antigen-2 autoantibody assays for the National Institute of Diabetes and Digestive and Kidney Diseases consortia. *J. Clin. Endocrinol. Metab.* 95, 3360–3367 (2010).
46. American Diabetes Association. Diagnosis and classification of diabetes mellitus. *Diabetes Care* 37, S81–S90 (2014).
47. Lee, H. S. et al. Biomarker discovery study design for type 1 diabetes in the Environmental Determinants of Diabetes in the Young (TEDDY) study. *Diabetes Metab. Res. Rev.* 30, 424–434 (2014).
48. TEDDY Study Group. The Environmental Determinants of Diabetes in the Young (TEDDY) study. *Ann. NY Acad. Sci.* 1150, 1–13.
49. TEDDY Study Group. The Environmental Determinants of Diabetes in the Young (TEDDY) study: study design. *Pediatr. Diabetes* 8, 286–298 (2007).
50. Vehik, K. et al. Methods, quality control and specimen management in an international multicentre investigation of type 1 diabetes: TEDDY. *Diabetes Metab. Res. Rev.* 29, 557–567 (2013).
51. Clem, A. L., Sims, J., Telang, S., Eaton, J. W. & Chesney, J. Virus detection and identification using random multiplex (RT)-PCR with 3′-locked random primers. *Virol. J.* 4, 65 (2007).
52. Bushnell, B. *BBMap Short Read Aligner* (Univ. California, 2016).
53. Soofooy-Khojine, A. B. et al. Molecular epidemiology of enteroviruses in young children at increased risk of type 1 diabetes. *PLoS ONE* 13, e0201959 (2018).
54. Krischer, J. P. et al. The influence of type 1 diabetes genetic susceptibility regions, age, sex, and family history on the progression from multiple autoantibodies to type 1 diabetes: a TEDDY study report. *Diabetes* 66, 3122–3129 (2017).
55. *Handbook of Modern Statistical Methods* (Taylor & Francis Group, 2018).
56. Usuitalo, U. et al. Association of early exposure of probiotics and islet autoimmunity in the TEDDY study. *JAMA Pediatr.* 170, 20–28 (2016).

**Acknowledgements**
The TEDDY Study is funded by U01 DK63829, U01 DK63861, U01 DK63821, U01 DK63865, U01 DK63863, U01 DK63790, UC4 DK63829, UC4 DK63861, UC4 DK63821, UC4 DK63865, UC4 DK63863, UC4 DK63863, UC4 DK95300, UC4 DK100238, UC4 DK106955, UC4 DK112243, UC4 DK117483 and contract number HHSN26720070014C from the NIDDK, National Institute of Allergy and Infectious Diseases, National Institute of Child Health and Human Development, National Institute of Environmental Health Sciences, Centers for Disease Control and Prevention and JDRF. This work was supported in part by the NIH/NCATS Clinical and Translational Science Awards to the University of Florida (UL1 TR000064) and the University of Colorado (UL1 TR001082).

**Author contributions**
K.V., K.F.L., M.R., J.T., A.G.Z., J.-X.S., A.L., A.W.H., D.A.S., J.-X.S., A.L., B.A., W.A.H., D.A.S., J.P.K., H.H. and R.E.L. designed the study. M.R., W.A.H., J.T., A.G.Z., J.-X.S., A.L., A.W.H., D.A.S., J.P.K., H.H. and R.E.L. performed the data analysis, data interpretation and figure generation. K.V., K.F.L., H.H. and R.E.L. wrote the paper. All authors contributed to critical revisions and approved the final manuscript.

**Competing interests**
H.H. is a shareholder and chairman of the board of Vactech, and a member of the Scientific Advisory Board of Provention Bio, which develops vaccines against picornaviruses and CVA. The authors have no other relevant affiliations, nor financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

**Additional information**
Extended data is available for this paper at https://doi.org/10.1038/s41591-019-0667-0.
Supplementary information is available for this paper at https://doi.org/10.1038/s41591-019-0667-0.
Correspondence and requests for materials should be addressed to K.V.
Peer review information Jennifer Sargent was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.
Reprints and permissions information is available at www.nature.com/reprints.
Extended Data Fig. 1 | Percentage of stool samples at age of first appearance of Enterovirus B. Panel a shows sample positivity and Panel b sample consecutive positivity. Panels c and d show months prior to autoantibody seroconversion of Enterovirus B for sample positivity (c) and sample consecutive positivity (d) by autoantibody case status (n=383 matched pair children). Blue line represents control samples and red line represents case samples. The timing of the first appearance of an Enterovirus B infection from enrollment (3 months of age) or months prior to islet autoimmunity showed no obvious trend by age of child.
Extended Data Fig. 2 | Common human viruses related to type 1 diabetes (T1D). The three forest plots (a–c) show how common human viruses relate to the odds of children being diagnosed with T1D. The results were shown as odds ratios (OR, circle) and 95% confidence intervals (CI, bars) and were calculated using conditional logistic regression models with adjustment for HLA-DR-DQ genotype. OR>1 indicated a positive correlation between virus pattern and diagnosis with T1D, OR<1 indicated an inverse correlation. Plot (a) examined if an increase in the number of samples positive for virus was correlated with T1D (n=112 matched pair children). Plot (b) examined if children positive for the virus between 3 and 6 months of age were related to T1D (n=103 matched pair children). Plot (c) examined if children positive for the common virus in at least two consecutive samples (yes versus no) were related to T1D (n=112 matched pair children). Black circles and CI bars represent non-significant associations. Red circles and CI bars represent significant association with T1D. The number of positive stool samples for Enterovirus B was lower among T1D cases compared to matched controls. Human mastadenvirus C, similar to islet autoimmunity cases, was less likely to be detected in early stool samples (3–6 months of age) compared to the matched control for T1D cases. All p-values were two-sided.
Extended Data Fig. 3 | Heatmaps of contig alignments of successive stools (n=6 children). Heatmaps showing percent homology of alignments of enterovirus contigs isolated from successive stools from the same child. Stool collection date (successive days in the study) are shown, the serotype for the enterovirus aligned, all are aligned to an enterovirus genome map with scale of nucleotides at the bottom. Heatmap color is assigned on ~7 nt/pixel, heatmap color scale of percent homology is shown at the top.
Extended Data Fig. 4 | Children consecutive positive for Enterovirus B with prolonged shedding of same serotype. Categorical months of shedding by number of children for islet autoimmunity cases (n=45) and controls (n=25). Red bars denote cases and blue bars denote controls. Length of prolonged shedding period (duration) was not associated with case status in the children with consecutive positive Enterovirus B. Conditional logistic regression was used to evaluate significance; test was two-sided.
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.

- 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   | VirMAP v1.0 (paper version) was used to generate the virome data and has been deposited in GitHub, https://github.com/cmmr/virmap. diamond version 0.9.22 BBMap version 37.58 blastn: 2.7.1+ lbzip2 version 2.5 MEGAHIT v1.1.3 khmer 2.0+706.g1745464 (normalize-by-median.py) pigg 2.3.3 vsearch v2.9.1_linux_x86_64 zstd v1.3.1 perl v5.24.0 All of the code and dependencies are listed on the GitHub site. |
| SAS 9.4 (SAS Institute, Cary, NC, USA) was used for the statistical analysis and GraphPad Prism 8.0 (San Diego, CA, USA) was used to for the figures. |

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.
TEDDY virome sequencing data that support the findings of this study have been deposited in the NCBI database of Genotypes and Phenotypes (dbGaP) with the primary accession code phs001442.v1.p1, in accordance with the dbGaP controlled-access authorization process. Clinical metadata and virome results data analyzed for the current study will be made 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 | Longitudinal stool samples from 3 months of age from 383 matched islet autoimmunity (IA) case-control pairs and 112 matched type 1 diabetes (T1D) case-control pairs were analyzed by metagenomic sequencing (IA, n=8,654; T1D, n=3,380) and processed using VirMap. The design includes two matched nested case-control studies. All persistent confirmed IA positive children as of May 31, 2012 were included as case in the IA nested case control study. In a 1:1 matched-nested case-control study, a control was matched to each case based on the case's event time. The control was required to be event-free within +/−45 days of the matching case’s event-time. The matching criteria were clinical site, sex and family history of T1D. A similar 1:1 matched-nested case-control was designed for T1D where all children who developed T1D as of May 31, 2012 were included in the nested case-control study with a matching control that was chosen on the case’s event-time (within +/−45 days of case event-time) matched on clinical site, sex and family history of T1D. Power calculations showed both matched-nested case-control studies had 80% power at a significance level of 5% to detect an OR>=2.28 for the IA and an OR>=4.58 for the T1D matched-nested case-control. |
|---|
| Data exclusions | For the nested case-control analysis, some samples were removed so that exactly the same number of samples was included between case and control pairs. This prevented skewing data due to the generally increased number of samples from cases and missing matched pair control data. This lack of matching stool sampling availability resulted in a 9% reduction in the number of pairs for the 1:1 viral metagenomic study for the islet autoimmunity nested-matched case-control, which left 383 pairs (n = 4,327 age matched stool samples in each group) and a 2% reduction in the number of pairs for the 1:1 T1D nested-matched case-control resulting in 112 pairs (n=1,690 age matched stool samples in each group). All outcome analyses were conducted on matched-pair sample data. |
| Replication | Observational cohort. No replication. |
| Randomization | Controls were matched individually to cases as described in the manuscript Methods text. Cases were sampled based on specific case-control design (i.e., until either development of islet autoimmunity or diagnosis of T1D) and matched control samples were included up until the corresponding age of event. |
| Blinding | TEDDY is an observational follow-up study, thus no overall blinding was used. However, the selection of samples sent to the Alkek Center for Metagenomics and Microbiome Research at the Baylor College of Medicine, Houston, TX, were determined by the Data Coordinating Center (USF Health Informatics Institute, Tampa, FL) without the laboratory knowing the case-control status. |

**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 |

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)
- Hela - ATCC
- Vero – Mary Estes, Baylor College of Medicine
- RD-CAR – Nora Chapman, University of Nebraska
- HEK-293 – ATCC
- Hela - ATCC
- Vero – Mary Estes, Baylor College of Medicine
- RD-CAR – Nora Chapman, University of Nebraska
- HEK-293 – ATCC

Authentication
- VERO cells and RD-CAR were not authenticated. HEK-293 and HeLa are ATCC authenticated.

Mycoplasma contamination
- All cell lines are certified mycoplasma-free by PCR assay.

Commonly misidentified lines
- No commonly misidentified cells were used. Finally, no experiments were performed with cells, they were only used to amplify virus.

Human research participants

Policy information about studies involving human research participants

Population characteristics
- Children were a mean±(stddev) of 2.1(1.2) years of age in the IA case-control and 2.7(1.4) years of age in the T1D case-control.
- Children samples were obtained from six geographical locations (Finland, Germany, Sweden in Europe and Washington, Colorado and Georgia in the United States). These children are at high HLA genetic risk for developing IA or T1D, with half of the cases for IA or T1D and the other half controls (Supplementary Table 1).

Recruitment
- Children were recruited based on specific type 1 diabetes risk human leukocyte antigen (HLA) genotypes and family history of T1D risk. Recruitment began in September of 2004 and was completed in February 2010. Six clinical centers in the USA (Colorado, Washington, and Florida/Georgia) and Europe (Germany, Sweden, and Finland) randomly HLA-screened 424,788 children at birth in hospitals in the four countries. A total of 418,367 general population infants were screen, of which 20,152 (4.8%) were eligible, and 1,437 of the 6,421 screened infants (22.4%) with a first-degree relative with type 1 diabetes were eligible. There were 8,676 children enrolled as participants in the study.

Ethics oversight
- The samples and clinical information were obtained under conditions of informed consent and with the approval of the National Institute of Diabetes and Digestive and Kidney Diseases External Executive Committee and participating clinical center Institutional Review Boards.

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), and three in Europe (Finland, Germany, and Sweden) participated in a population-based HLA screening of newborns between 2004 and 2010. Children with high risk HLA genotypes were enrolled (n=8,676) and prospectively followed from three months of age to 15 years with study visits that include a blood draw every three months until four years and every three or six months thereafter depending on islet autoimmunity positivity. Stool samples were collected monthly from ages 3 to 48 months and then quarterly until the age of 10 years. A nested-matched case-control was conducted through risk-set sampling using metadata and islet autoimmunity sample results as of 31 May 2012. In a separate nested-matched case-control, each child diagnosed with T1D had a control selected based on their event time from birth. Metadata were collected using validated questionnaires that have been either published or
extensively scrutinized by experts. TEDDY provides many tools, such as ‘The TEDDY book’, to the parents to assist in real-time
collection of all events in their child’s life to ensure bias and error are minimized.

Outcomes

Persistent confirmed autoimmunity was defined by the presence of a confirmed islet autoimmunity (glutamic acid decarboxylase
(GADA), insulinoma-associated 2 (IA-2A) or insulin (IAA)) at each of the two TEDDY laboratories on two or more consecutive
visits. T1D diagnosis was defined according to American Diabetes Association criteria. Discordance between cases and matched
controls, both in the number of stool samples positive for each common virus (positive >2% of samples) and in the presence of
consecutive samples positive for the virus (allowing for one missed monthly visit), were evaluated using conditional logistic
regression models adjusting for HLA genotypes. The magnitude of the associations were assessed by odds ratios with 95%
confidence intervals.