

SCIENTIFIC REPORTS



OPEN

Higher abundance of enterovirus A species in the gut of children with islet autoimmunity

Ki Wook Kim^{1,2}, Jessica L. Horton^{1,2}, Chi Nam Ignatius Pang³, Komal Jain⁴, Preston Leung⁵, Sonia R. Isaacs^{1,2}, Rowena A. Bull⁵, Fabio Luciani⁵, Marc R. Wilkins³, Jacki Catteau⁶, W. Ian Lipkin^{4,7}, William D. Rawlinson^{1,2,8}, Thomas Briese^{4,9} & Maria E. Craig^{1,2,6,10}

Enteroviruses (EVs) are prime candidate environmental triggers of islet autoimmunity (IA), with potential as vaccine targets for type 1 diabetes prevention. However, the use of targeted virus detection methods and the selective focus on EVs by most studies increases the risk for substantial investigation bias and an overestimated association between EV and type 1 diabetes. Here we performed comprehensive virome-capture sequencing to examine all known vertebrate-infecting viruses without bias in 182 specimens (faeces and plasma) collected before or at seroconversion from 45 case children with IA and 48 matched controls. From >2.6 billion reads, 28 genera of viruses were detected and 62% of children (58/93) were positive for ≥ 1 vertebrate-infecting virus. We identified 129 viruses as differentially abundant between the gut of cases and controls, including 5 EV-A types significantly more abundant in the cases. Our findings further support EV's hypothesised contribution to IA and corroborate the proposal that viral load may be an important parameter in disease pathogenesis. Furthermore, our data indicate a previously unrecognised association of IA with higher EV-A abundance in the gut of children and provide a catalog of viruses to be interrogated further to determine a causal link between virus infection and type 1 diabetes.

Among the spectrum of environmental factors associated with type 1 diabetes, viruses are postulated to contribute most significantly. Indeed, enterovirus (EV) infection often precedes islet autoimmunity (IA), and rates of EV infection are significantly higher in individuals at type 1 diabetes onset compared to controls^{1–4}. Furthermore, systematic review and meta-analysis of 26 studies with over 4,400 participants confirm a significant epidemiological association between EV and type 1 diabetes⁵.

Mostly transmitted faeco-orally, classical EVs establish primary infection in the gut, providing a potential reservoir for spread into the pancreas^{6,7}. There, EVs can directly infect and damage pancreatic β -cells⁸. This is consistent with the recent detection of EV protein and/or RNA in the pancreata of 6 adults, biopsied within 9 weeks following type 1 diabetes diagnosis⁹. Additionally, persistent EV infection is associated with increased intestinal permeability and inflammation; both pronounced in mice and children with type 1 diabetes¹⁰. Alternatively, there is potential for molecular mimicry between EV peptides and islet antigens¹¹, which could induce a cross-reactive autoimmune response¹². However, the use of targeted virus detection methods and the selective focus on EVs by most studies increases the risk for substantial investigation bias and an overestimated association with type 1 diabetes. Therefore, an unbiased examination of all viruses (the 'virome') is needed to re-evaluate existing associations.

¹School of Women's and Children's Health, University of New South Wales Faculty of Medicine, Sydney, Australia.

²Virology Research Laboratory, Prince of Wales Hospital Randwick, Sydney, Australia. ³School of Biotechnology and Biomedical Sciences, University of New South Wales Faculty of Science, Sydney, Australia. ⁴Center for Infection and Immunity, Mailman School of Public Health, Columbia University, New York, USA. ⁵Systems Medicine, Inflammation and Infection Research Centre, School of Medical Sciences, University of New South Wales Faculty of Medicine, Sydney, Australia. ⁶Institute of Endocrinology and Diabetes, Children's Hospital at Westmead, Sydney, Australia.

⁷Department of Pathology and Neurology, College of Physicians & Surgeons, Columbia University, New York, USA.

⁸Serology and Virology Division, South Eastern Area Laboratory Services Microbiology, Prince of Wales Hospital, Sydney, Australia. ⁹Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, USA. ¹⁰Discipline of Child and Adolescent Health, University of Sydney, Sydney, Australia. Ki Wook Kim and Jessica L. Horton contributed equally. Thomas Briese and Maria E. Craig jointly supervised this work. Correspondence and requests for materials should be addressed to M.E.C. (email: m.craig@unsw.edu.au)

¹¹Discipline of Child and Adolescent Health, University of Sydney, Sydney, Australia. ¹²Discipline of Child and Adolescent Health, University of Sydney, Sydney, Australia. Ki Wook Kim and Jessica L. Horton contributed equally. Thomas Briese and Maria E. Craig jointly supervised this work. Correspondence and requests for materials should be addressed to M.E.C. (email: m.craig@unsw.edu.au)

Characteristic	Case (n = 45)	Control (n = 48)	P value
Male, n (%)	25 (56)	27 (56)	0.95
Risk HLA Class II genotype†, n (%) ^a	17 (38)	13 (27)	0.27
DR3/DR4 ^b	8 (18)	3 (6)	0.19
DR3/DR3 ^b	0	1 (2)	0.50
DR4/DR4 ^b	2 (4)	3 (6)	0.65
Mean age at time of sampling (SD)	5.7 (3.7)	5.4 (3.5)	0.64
Relative with type 1 diabetes, n (%)^c			
Mother	17 (38)	25 (52)	
Father	14 (31)	15 (31)	
Sibling/s	15 (33)	9 (19)	
Progression to type 1 diabetes	3	0	
Samples, n	91	91	
Stool	32	32	
Plasma	59	59	

Table 1. Summary of study participant characteristics. * Any of the following Human Leukocyte Antigen (HLA) genotypes: (1) HLA-DQB1*02/DQB1*0302; (2) HLA-DQB1*0302/x (x not DQB1*02, DQB1*0301 or DQB1*0602); (3) HLA-DQA1*05-DQB1*02/y (y not DQA1*0201-DQB1*02, DQB1*0301, DQB1*0602 or DQB1*0603); (4) HLA-DQA1*03-DQB1*02/y (y not DQA1*0201-DQB1*02, DQB1*0301, DQB1*0602 or DQB1*0603). ^aHLA Class II information not available for 3 cases and 4 controls. ^bPrecise HLA-DRB1 and -DQB1 typing available for 48/93 participants (26 cases, 22 controls). ^cOne case and one control child had both a parent and sibling with type 1 diabetes.

Previously, the Finnish Diabetes Prediction and Prevention (DIPP) and The Environmental Determinants of Diabetes in the Young (TEDDY) studies examined the gut and plasma virome of children with type 1 diabetes, respectively^{13,14}. Additionally, a recent study examined the gut virome of 11 infants with type 1 diabetes¹⁵. However, data on vertebrate-infecting viruses from these studies remain difficult to interpret, partly due to the limited sensitivity of conventional virome sequencing caused by the overwhelming abundance of host and bacterial nucleic acid in most clinical specimens¹⁶.

Here, we used Virome-Capture-Sequencing for Vertebrate-infecting viruses (VirCapSeq-VERT) as an enhanced and unbiased method to characterise the virome of children with IA. With sensitivity on par or greater than quantitative real-time PCR (qPCR), VirCapSeq-VERT utilises ~2 million capture probes covering entire genomes of every known vertebrate-infecting virus for enrichment prior to high-throughput sequencing. This produces up to 10,000-fold increase in viral reads compared to conventional methods¹⁶.

Methods

Subjects and samples. This case-control study was nested within the Australian Viruses In the Genetically at Risk (VIGR) prospective birth cohort of children with ≥ 1 first-degree relative with type 1 diabetes⁴. We examined the virome of 45 cases with IA, defined as seropositivity for ≥ 1 islet autoantibodies (GAD65, IA-2 and IAA) above the cut-off (0.6 units/mL, 0.8 units/mL, 53 nU/mL, respectively) in two consecutive visits. Every case sample was paired with an age and gender matched persistently IA-negative control sample (Table 1). HLA class II information was available for 86/93 children, of whom 30 (17 cases, 13 controls) were categorised as having a type 1 diabetes risk genotype.

Overall, 64 faeces and 118 plasma collected and stored at -80°C between 2006–2015 were analysed. Of these, 32 faecal samples from 20 cases collected at seroconversion and/or 15 ± 6 (mean \pm s.d.) months prior, and 59 plasma samples from 41 cases collected at seroconversion and/or within 13 ± 4 (mean \pm s.d.) months prior were examined; of available samples this concerned always the one(s) collected closest to seroconversion. Of 93 participants, 25 were included in both the gut and blood virome analysis (16 cases, 9 controls). There were 26 visits at which both sample types were collected (Supplementary Table 1). The project was approved by the Sydney Children's Hospital Network Human Research Ethics Committee (HREC#12SCHN225) and all experimental methods were performed in accordance with the relevant guidelines and regulations. For all child participants in the study, informed consent was provided in writing by a parent or guardian on the child's behalf, consenting for research participation, collection of biological specimens and use of clinical data in this publication.

VirCapSeq-VERT. Performed as previously described¹⁶ with following modifications. Total NA was extracted from 235 μL of plasma using the MagMAX Isolation Kit on KingFisher (ThermoFisher). For faeces, 30% (w/v) suspension in 1 x PBS was centrifuged (5 min, 13,000 x g) and 175 μL of supernatant was used. Purified NA was amplified using the Whole Transcriptome Amplification Kit 2 (Sigma-Aldrich)¹⁷. Products were purified using the ChargeSwitch-Pro Kit and quantified by PicoGreen (ThermoFisher). Capture-compatible libraries were prepared with KAPA Hyperplus (Roche) and pooled (20 and 24 libraries/pool for faeces and plasma, respectively).

Reads were de-multiplexed, filtered (primer and host sequences) and *de novo* assembled. Contigs and unique singletons were homology searched at both the nucleotide and protein level¹⁶. Correcting for index cross-talk,

0.1% of the highest read count for each virus was subtracted across all samples within each pool. Two positivity thresholds applied: (1) 100 viral reads matched at the species level, distributed randomly across the target sequence; (2) 50 viral reads per 100,000 raw reads¹³. Former threshold was empirically established based on its proximity to the detection limit of qPCR in whole blood¹⁶ and faeces (Supplementary Fig. 1).

Statistical analyses. Participant characteristics and virus frequencies compared using Chi-squared and Fisher's exact tests, respectively. Differential abundance was examined by edgeR¹⁸, using a matrix of read counts encompassing all samples and detected viruses. Before conversion to counts per million, each matrix entry had 1 added to avoid issues with division by, or log function of 0. Data normalised using the Relative Log Expression method with respect to library size¹⁹. Common and tag-wise methods were used to estimate the biological coefficient of variation. Samples were divided into case and control groups, and the "exact" test was used to perform hypothesis testing²⁰. *P* values were adjusted to control false discovery rates²¹.

EV and Multiplex qPCR. First strand cDNA was synthesised using SuperScript III (Invitrogen) with random hexamers and 8 μ L total NA. qPCR performed with EV-specific primers F: 5'-TCC TCC GGC CCC TGA ATG CG-3' and R: 5'-ATT GTC ACC ATA AGC AGC CA-3', and probe 5'-[6FAM]TGT GTC GTA ACG GGT AAC TCT G[BHQ1]-3' on the LightCycler 480II (Roche), using 4 μ L cDNA, 12.5 μ L 2X Sensimix II Probe (Bioline), 0.25 μ M probe and 0.5 μ M of each primer. Cycle conditions: 95°C 5 min, 40 cycles 95°C 5 s; 55°C 45 s; 68°C 45 s. Norovirus, astrovirus, rotavirus and sapovirus were assayed with the Allplex™ GI-Virus Assay kit (GI9701Y; Seegene) on the CFX96™ (Bio-Rad).

Infection history. Acute infections, illnesses or symptoms indicative of potential viral infection experienced by children between study visits were collected with the corresponding dates by parents and provided to study personnel at every visit (usually every 6 months), starting from age 6 months. The association between virus positivity determined by VirCapSeq-VERT and past infection was examined using Chi-squared analysis on infection history recorded within 3 months preceding sample collection.

Results

Virus positivity determined by VirCapSeq-VERT. We characterised the virome of 93 children (45 cases with IA and 48 matched controls; Table 1) in the VIGR cohort [8]. Setting 100 reads as the positivity threshold, 75% of faeces (48/64) and 38% of plasma (45/118) were positive for ≥ 1 vertebrate-infecting virus (Fig. 1A,B), and 62% of children (58/93) were virus-positive. Of them, 43% reported a history of infection within 3 months prior to sample collection compared with 11% of virus-negative children ($P = 0.002$). There was no difference in virus positivity between cases and controls in the gut (66% vs 75%; $P = 0.59$) or plasma (41% vs 36%; $P = 0.70$), even when samples collected before and at seroconversion were analysed separately (Supplementary Table 2). We found no difference in virus positivity between children with type 1 diabetes risk HLA class II genotypes vs without (60% vs 68%, OR 0.71, 95% CI 0.30–1.78, $P = 0.49$).

In total, 28 viral genera were detected (Fig. 1). Anellovirus, EV and picobirnavirus were the most frequent viruses in faeces (Fig. 1A). In plasma, anellovirus, EV and erythroparvovirus were most frequent (Fig. 1B). Cardioviruses and noroviruses were detected exclusively in controls. There were minor differences in the frequency of specific viruses between cases and controls, but none reached statistical significance. In 19 children that had faeces and plasma collected on the same visit, EV was the only virus detected in both samples from the same timepoint (Supplementary Table 1); although only in 1/26 visits.

The prevalence of EV in faeces was identical between cases and controls. However, at the species level, EV-A and EV-B types were more frequent in cases (Fig. 1C). In some samples, multiple EV types were detected (Supplementary Table 3). In one case faecal specimen, reads matched to five different EVs (Rhinovirus C, ECHOvirus E30, Coxsackievirus B3; CVB3, CVA5 and CVA6) with 25–100% reference genome coverage and high sequence identity ($>95\%$), which may indicate co-infection and/or recombinant genomes^{22–24}. Presence of EV was also shown by qPCR in 5/5 EV-positive case faeces and 3/5 control samples (Supplementary Fig. 1). One negative qPCR result (KWK-291) was due to the presence of rhinovirus (Supplementary Table 3), not targeted by the qPCR. Norovirus, sapovirus and astrovirus positivity in faeces was likewise confirmed by multiplex qPCR (Supplementary Fig. 1).

Previously, 50 viral reads per 100,000 raw reads (50P100K) was used to indicate virus positivity in the DIPP virome study¹³. Using this threshold, faecal virus positivity reduced from 75% to 36% (23/64) (Fig. 2). Despite ~50% reduction, this still resulted in a >3 -fold higher positivity than reported in the DIPP study (10%; 10/96), supporting the notion of an enhanced sensitivity of VirCapSeq-VERT. Frequencies of most viruses remained unchanged, except for a noticeable decrease in the number of EV-positive control faeces, rendering EV 4-fold higher in the gut of cases.

We found no association between EV in faeces and the first seropositivity for IAA (OR 0.66, 95% CI 0.13–3.58, $P = 1.00$) nor GADA (OR 1.41, 95% CI 0.33–6.27, $P = 1.00$).

Differential abundance of gut viruses. In total, 129 viruses had ≥ 2 -fold difference ($P < 0.02$) in abundance between the gut of cases and controls, with false discovery rate $<5\%$ (Fig. 3). One EV-A (CVA2) and two EV-Bs (ECHOvirus E30 and CVB3) were among the 10 most differentially abundant viruses (Supplementary Data), and every differentially abundant EV-A (CVA2, 5, 6, 8 and 14; Fig. 3 in red) was significantly more abundant in cases than in controls ($P < 0.00001$), even when samples collected before or at seroconversion were analysed separately (Supplementary Fig. 2).

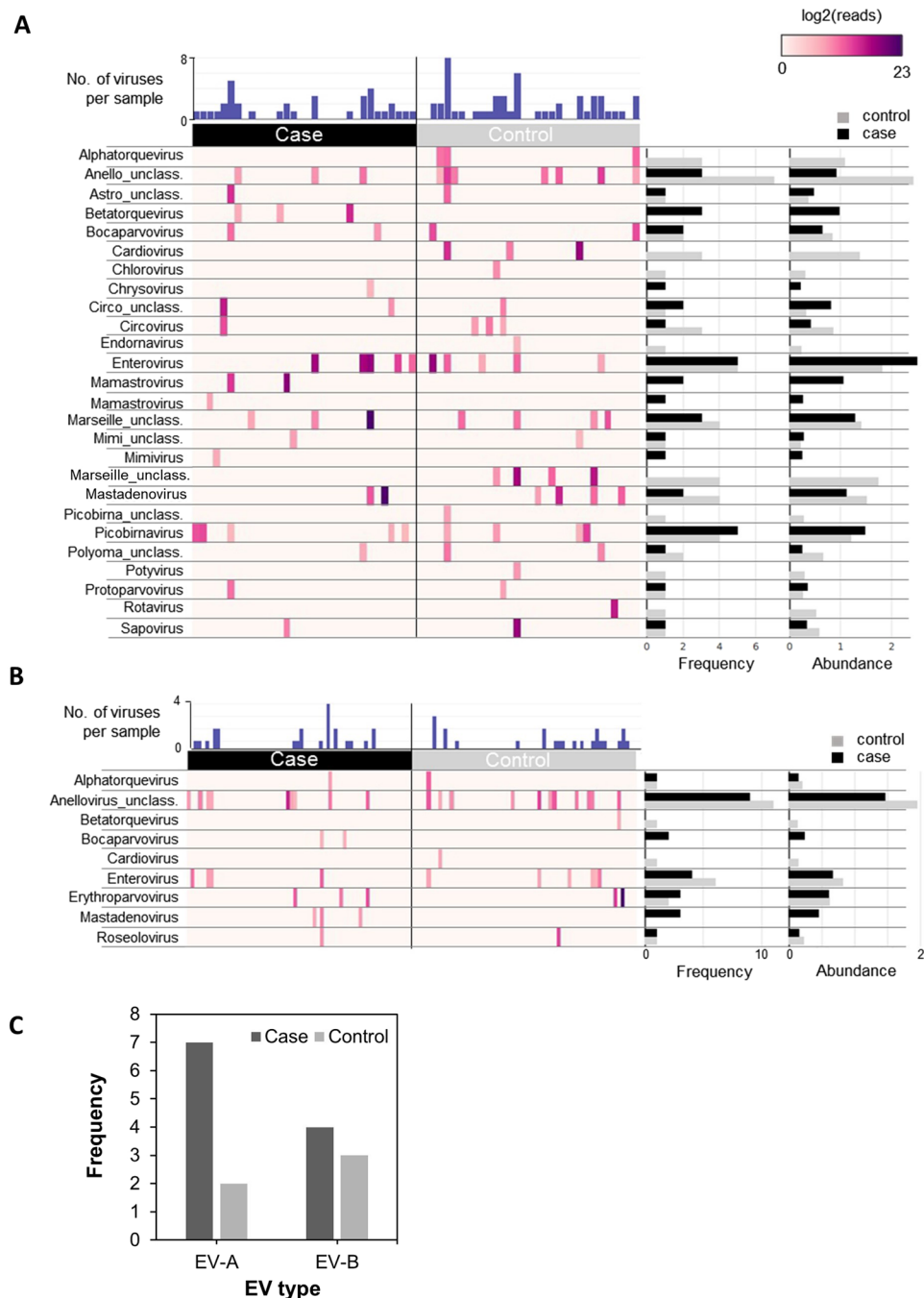


Figure 1. Viruses detected by VirCapSeq-VERT. (A) Heatmap of viral reads (log₂ scale) detected in 32 case and control faeces (n = 64). Only viruses with ≥ 100 reads matched by BLAST at the species level were included and represented at the genus level. Number of viruses detected per specimen, frequency of each virus within the case/control group, and the mean log read counts (Abundance) are summarised by bar charts. (B) Heatmap of viral reads detected in 59 case and control plasma specimens (n = 118). (C) Frequency of EV-A and EV-B types sequenced from faeces of case and control children.

Discussion

This is the first examination all vertebrate-infecting viruses in a cohort at risk of type 1 diabetes using VirCapSeq-VERT. Analysis of faecal specimens from prior to or at seroconversion revealed significantly higher abundance of EV-A species in the gut of cases, strengthening the evidence that EVs contribute to the development of IA and that the risk for type 1 diabetes may be closely linked to viral load²⁵. The lack of overlap between the gut and plasma virome is in keeping with the short duration of viremia following EV infection vs prolonged gut excretion²⁶, corroborating the model that gut is a major regulator of early inflammation in type 1 diabetes²⁷, and a reservoir from which EVs can spread into the pancreas^{28,29}. Sufficient viral load may be important for EVs to establish persistent enteric infection, and thereby induce intestinal changes that contribute to the initiation

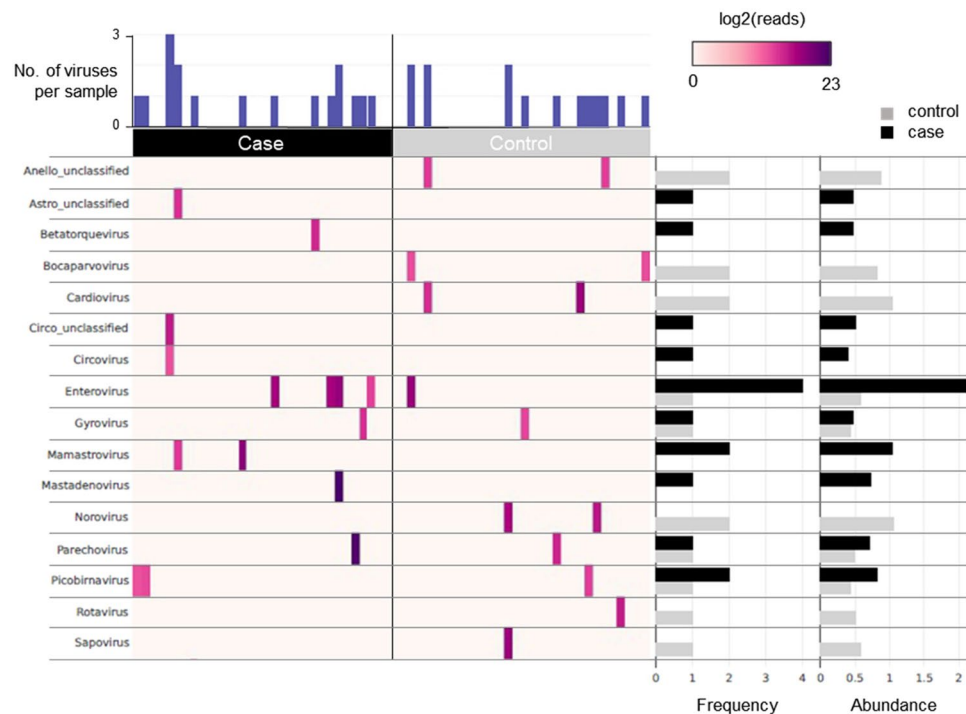


Figure 2. Viruses detected by VirCapSeq-VERT in faecal specimens based on the 50 reads per 100,000 raw counts (50P100K) threshold. Heatmap of virus reads (log₂ scale) detected by VirCapSeq-VERT in 32 case and 32 control faeces (n = 64). Viruses represented at the genus level. Number of viruses detected per specimen, frequency of each virus within the case or control group, and the mean log read counts (Abundance) are summarised by bar charts.

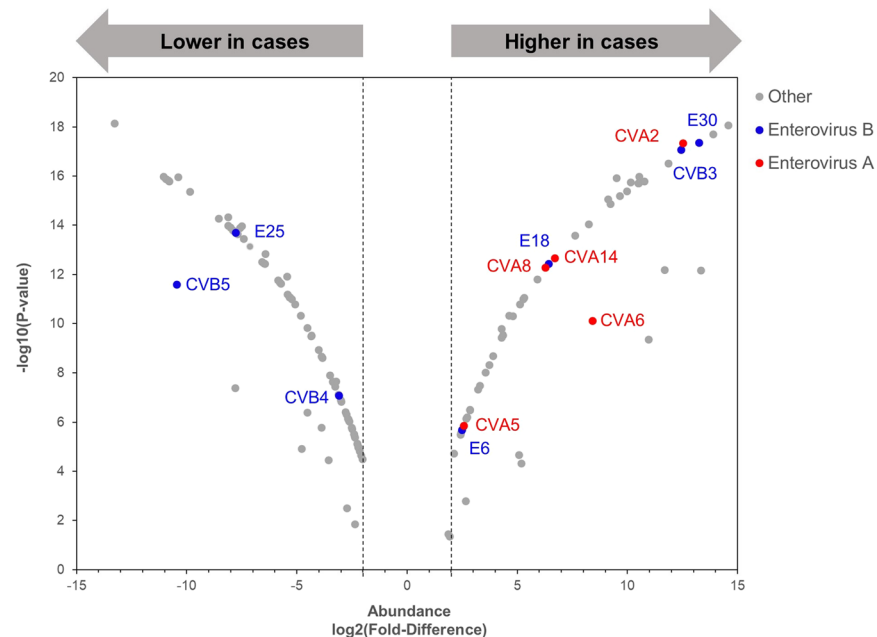


Figure 3. Higher abundance of enterovirus A in the gut of children with islet autoimmunity. Volcano plot of viruses differentially abundant between case and control in the gut. Only viruses with ≥ 2 -fold difference (marked by dotted vertical line) and false discovery rate $< 5\%$ ($q < 0.05$) as determined by edgeR are represented. Enterovirus A types represented in red: Coxsackievirus A2 (CVA2), CVA5, CVA6, CVA8 and CVA14. Enterovirus B types represented in blue: ECHOvirus E6 (E6), E18, E25, E30, CVB3, CVB4 and CVB5. All other viruses represented in grey.

of IA and the progression to type 1 diabetes. This is consistent with murine models showing the requirement of preceding intestinal inflammation for the development of β -cell autoimmunity³⁰. Higher EV load may translate to higher replication and transmission to the pancreas, which can impose greater cellular stresses on pancreatic islets, contributing to inaccurate protein translation and production of aberrant insulin polypeptides that render β -cells immunogenic³¹.

The predominance of EV-A in VIGR children differed from previous associations of EV-B with type 1 diabetes^{4,5}. However, it is consistent with the 10-year follow-up of 129 cases and 282 controls, which identified CVA2, A4 and A16 as the most common EV types in children with type 1 diabetes³². Furthermore, it fits the high prevalence of EV-A in natural circulation among infants³³.

In this study, cardioviruses and noroviruses were detected exclusively in control samples. The biological significance of this is unclear, especially for cardioviruses which have unknown pathogenicity in humans³⁴. Neither viruses have been associated with IA or type 1 diabetes. Indeed, RNA testing of human cardioviruses in 27 cases with IA vs 53 matched controls in the Norwegian longitudinal birth cohort study found no association between Cardiovirus in faeces and IA³⁵.

Compared to case-control studies that used targeted detection methods on a selection of candidate viruses, our sample size is relatively small. However, it is comparable to other type 1 diabetes virome studies, and is the first to examine both the gut and plasma virome in children with IA. Low sample numbers potentially masked statistically significant differences in virus frequency between cases and controls. To ensure sufficient statistical power, future validation studies should examine a greater sample size and include multiple longitudinal time-points preceding IA if possible. Nevertheless, we found a significant association between IA and higher gut EV-A abundance, and identified >100 additional viruses differentially abundant between the gut of cases and controls.

Another potential limitation is that participants were not matched for their HLA genotype. However, there is an increasing proportion of individuals with non-risk HLA genotypes developing type 1 diabetes, particularly in the Australian population^{36–38}, and although a range of HLA genotypes are associated with increased susceptibility and persistence of some viruses, there is no evidence supporting their influence on the outcome of EV infections. Examination of over 7,000 longitudinal faecal specimens collected from 419 children with HLA-DR4/DR3 genotype and 373 without found no association between EV RNA detection and HLA risk³⁹. Therefore, we consider it unlikely that our EV-A detection is primarily linked to HLA genotype.

In conclusion, studies into environmental triggers of type 1 diabetes have to date maintained a strong focus on identifying statistically significant differences in the frequencies of select viruses, especially CVBs⁵. Our findings highlight the importance of additionally examining the differential abundance of viruses. We demonstrate an association between IA and higher EV-A abundance in the gut of children, and shortlist additional viruses as potential contributors in the development of IA. These results should be validated in other cohorts with greater sample size and more longitudinal timepoints preceding IA, and likely candidates be targeted by mechanistic studies to determine if a causal link with type 1 diabetes exists, as it is paramount to define the full breadth diabetogenic viruses for future vaccine development.

Data Availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

References

1. Lonnrot, M. *et al.* Enterovirus infection as a risk factor for beta-cell autoimmunity in a prospectively observed birth cohort: the Finnish Diabetes Prediction and Prevention Study. *Diabetes* **49**, 1314–1318 (2000).
2. Oikarinen, S. *et al.* Enterovirus RNA in Blood Is Linked to the Development of Type 1 Diabetes. *Diabetes* **60**, 276 (2010).
3. Lin, H.-C. *et al.* Enterovirus infection is associated with an increased risk of childhood type 1 diabetes in Taiwan: a nationwide population-based cohort study. *Diabetologia* **58**, 79–86, <https://doi.org/10.1007/s00125-014-3400-z> (2015).
4. Yeung, W. C. *et al.* Children with islet autoimmunity and enterovirus infection demonstrate a distinct cytokine profile. *Diabetes* **61**, 1500–1508, <https://doi.org/10.2337/db11-0264> (2012).
5. Yeung, W.-C. G., Rawlinson, W. D. & Craig, M. E. Enterovirus infection and type 1 diabetes mellitus: systematic review and meta-analysis of observational molecular studies. *BMJ* **342**, <https://doi.org/10.1136/bmj.d35> (2011).
6. Vaarala, O. Is the origin of type 1 diabetes in the gut? *Immunology and cell biology* **90**, 271–276, <https://doi.org/10.1038/icb.2011.115> (2012).
7. Oikarinen, M. *et al.* Type 1 diabetes is associated with enterovirus infection in gut mucosa. *Diabetes* **61**, 687–691, <https://doi.org/10.2337/db11-1157> (2012).
8. Hyoty, H. Viruses in type 1 diabetes. *Pediatric diabetes* **17**(Suppl 22), 56–64, <https://doi.org/10.1111/pedi.12370> (2016).
9. Krogvold, 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, <https://doi.org/10.2337/db14-1370> (2015).
10. Li, X. & Atkinson, M. A. The role for gut permeability in the pathogenesis of type 1 diabetes—a solid or leaky concept? *Pediatric diabetes* **16**, 485–492, <https://doi.org/10.1111/pedi.12305> (2015).
11. Atkinson, M. A. *et al.* Cellular immunity to a determinant common to glutamate decarboxylase and coxsackie virus in insulin-dependent diabetes. *The Journal of clinical investigation* **94**, 2125–2129, <https://doi.org/10.1172/jci117567> (1994).
12. Coppieters, K. T., Wiberg, A. & Von Herrath, M. G. Viral infections and molecular mimicry in type 1 diabetes. *APMIS: acta pathologica, microbiologica, et immunologica Scandinavica* **120**, 941–949, <https://doi.org/10.1111/apm.12011> (2012).
13. Kramna, L. *et al.* Gut virome sequencing in children with early islet autoimmunity. *Diabetes care* **38**, 930–933, <https://doi.org/10.2337/dc14-2490> (2015).
14. Lee, H. S. *et al.* Next-generation sequencing for viruses in children with rapid-onset type 1 diabetes. *Diabetologia* **56**, 1705–1711, <https://doi.org/10.1007/s00125-013-2924-y> (2013).
15. Zhao, G. *et al.* Intestinal virome changes precede autoimmunity in type I diabetes-susceptible children. *Proc Natl Acad Sci USA* **114**, E6166–e6175, <https://doi.org/10.1073/pnas.1706359114> (2017).
16. Briese, T. *et al.* Virome Capture Sequencing Enables Sensitive Viral Diagnosis and Comprehensive Virome Analysis. *mBio* **6**, e01491–01415, <https://doi.org/10.1128/mBio.01491-15> (2015).

17. Conceicao-Neto, N. *et al.* Modular approach to customise sample preparation procedures for viral metagenomics: a reproducible protocol for virome analysis. *Sci Rep* **5**, 16532, <https://doi.org/10.1038/srep16532> (2015).
18. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140, <https://doi.org/10.1093/bioinformatics/btp616> (2009).
19. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome biology* **11**, R106, <https://doi.org/10.1186/gb-2010-11-10-r106> (2010).
20. Robinson, M. D. & Smyth, G. K. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics* **9**, <https://doi.org/10.1093/biostatistics/kxm030> (2008).
21. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc B* **57** (1995).
22. Simmonds, P. & Welch, J. Frequency and Dynamics of Recombination within Different Species of Human Enteroviruses. *Journal of virology* **80**, 483 (2006).
23. Lukashev, A. N. *et al.* Recombination in Circulating Enteroviruses. *Journal of virology* **77**, 10423–10431, <https://doi.org/10.1128/JVI.77.19.10423-10431.2003> (2003).
24. Isaacs, S. R. *et al.* Amplification and next generation sequencing of near full-length human enteroviruses for identification and characterisation from clinical samples. *Sci Rep* **8**, 11889, <https://doi.org/10.1038/s41598-018-30322-y> (2018).
25. Kanno, T. *et al.* Group B Coxsackievirus Diabetogenic Phenotype Correlates with Replication Efficiency. *Journal of virology* **80**, 5637–5643, <https://doi.org/10.1128/jvi.02361-05> (2006).
26. Chung, P. W., Huang, Y. C., Chang, L. Y., Lin, T. Y. & Ning, H. C. Duration of enterovirus shedding in stool. *Journal of microbiology, immunology, and infection = Wei mian yu gan ran za zhi* **34**, 167–170 (2001).
27. Vaarala, O. The gut as a regulator of early inflammation in type 1 diabetes. *Current opinion in endocrinology, diabetes, and obesity* **18**, 241–247, <https://doi.org/10.1097/MED.0b013e3283488218> (2011).
28. Oikarinen, M. *et al.* Type 1 Diabetes Is Associated With Enterovirus Infection in Gut Mucosa. *Diabetes* **61**, 687–691, <https://doi.org/10.2337/db11-1157> (2012).
29. Tauriainen, S., Oikarinen, S., Oikarinen, M. & Hyöty, H. Enteroviruses in the pathogenesis of type 1 diabetes. *Seminars in Immunopathology* **33**, 45–55, <https://doi.org/10.1007/s00281-010-0207-y> (2011).
30. Graham, S. *et al.* Enteropathy precedes type 1 diabetes in the BB rat. *Gut* **53**, 1437–1444, <https://doi.org/10.1136/gut.2004.042481> (2004).
31. Kracht, M. J. *et al.* Autoimmunity against a defective ribosomal insulin gene product in type 1 diabetes. *Nature medicine* **23**, 501–507, <https://doi.org/10.1038/nm.4289> (2017).
32. 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*, 1–8, <https://doi.org/10.1007/s00125-016-4177-z> (2017).
33. Witso, E. *et al.* High Prevalence of Human Enterovirus A Infections in Natural Circulation of Human Enteroviruses. *Journal of clinical microbiology* **44**, 4095–4100, <https://doi.org/10.1128/jcm.00653-06> (2006).
34. Himeda, T. & Ohara, Y. Saffold virus, a novel human Cardiovirus with unknown pathogenicity. *Journal of virology* **86**, 1292–1296, <https://doi.org/10.1128/jvi.06087-11> (2012).
35. Tapia, G. *et al.* Saffold Virus, a Human Cardiovirus, and Risk of Persistent Islet Autoantibodies in the Longitudinal Birth Cohort Study MIDIA. *PLoS one* **10**, e0136849, <https://doi.org/10.1371/journal.pone.0136849> (2015).
36. Hermann, R. *et al.* Temporal changes in the frequencies of HLA genotypes in patients with Type 1 diabetes—indication of an increased environmental pressure? *Diabetologia* **46**, 420–425, <https://doi.org/10.1007/s00125-003-1045-4> (2003).
37. Fourlanos, S. *et al.* The Rising Incidence of Type 1 Diabetes Is Accounted for by Cases With Lower-Risk Human Leukocyte Antigen Genotypes. *Diabetes care* **31**, 1546–1549, <https://doi.org/10.2337/dc08-0239> (2008).
38. Gillespie, K. M. *et al.* The rising incidence of childhood type 1 diabetes and reduced contribution of high-risk HLA haplotypes. *Lancet (London, England)* **364**, 1699–1700, [https://doi.org/10.1016/s0140-6736\(04\)17357-1](https://doi.org/10.1016/s0140-6736(04)17357-1) (2004).
39. Witso, E. *et al.* Genetic Determinants of Enterovirus Infections: Polymorphisms in Type 1 Diabetes and Innate Immune Genes in the MIDIA Study. *Viral immunology* **28**, 556–563, <https://doi.org/10.1089/vim.2015.0067> (2015).

Acknowledgements

We thank Nishit P. Bhuvu (Columbia University, NY) for performing the virome capture hybridization and submitting samples for sequencing; Cheng Guo (Columbia University, NY) for assisting in the processing of metagenomic data and generation of BLAST result summaries; Xiaoyu J. Che (Columbia University, NY) for assisting with the rarefaction analysis; Kelly Spooner (Prince of Wales Hospital, Sydney) for assisting with VIGR sample and data collection. This study was supported by the Diabetes Australia Research Grant (K.W.K., W.D.R. and M.E.C.), Rebecca L. Cooper Medical Research Grant (K.W.K., W.D.R. and M.E.C.), Australian National Health and Medical Research Council (NHMRC) Practitioner Fellowship (M.E.C. and W.D.R.), the US National Institutes of Health (Center for Research in Diagnostics and Discovery; T.B., K.J. and W.I.L.).

Author Contributions

M.E.C., W.D.R., T.B. W.I.L. and K.W.K. designed the study. K.W.K., J.L.H. and S.R.I. performed the experiments. C.N.P. and M.R.W. performed the differential abundance analysis, K.J. and T.B. performed the *de novo* assembly of Illumina sequence reads and BLAST analysis. P.L., R.A.B. and F.L. provided expert opinion and assisted in the metagenomic analysis. All authors contributed to the interpretation of results. K.W.K., M.E.C. and J.L.H. wrote the manuscript, and all authors edited the manuscript. M.E.C. initiated and developed the VIGR cohort, and contributed to sample and data collection. J.C. coordinated the VIGR study recruitment and sample collection. M.E.C. is the guarantor of this work and, as such, had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-38368-8>.

Competing Interests: The authors declare no competing interests.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019