Higher frequency of vertebrate-infecting viruses in the gut of infants born to mothers with type 1 diabetes

Ki Wook Kim | Digby W. Allen | Thomas Briese | Jennifer J. Couper
Simon C. Barry | Peter G. Colman | Andrew M. Cotterill | Elizabeth A. Davis
Lyne C. Giles | Leonard C. Harrison | Mark Harris | Aveni Haynes
Jessica L. Horton | Sonia R. Isaacs | Komal Jain | Walter I. Lipkin
Kelly McGorm | Grant Morahan | Claire Morbey | Ignatius C. N. Pang
Anthony T. Papenfuss | Megan A. S. Penno | Richard O. Sinnott
Georgia Soldatos | Rebecca L. Thomson | Peter Vuillermin
John M. Wentworth | Marc R. Wilkins | William D. Rawlinson
Maria E. Craig on behalf of the ENDIA STUDY GROUP

1School of Women's and Children's Health, University of New South Wales, Sydney, New South Wales, Australia
2Center for Infection and Immunity and Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, New York
3Robinson Research Institute and Adelaide Medical School, University of Adelaide, Adelaide, South Australia, Australia
4Department of Diabetes and Endocrinology, The Royal Melbourne Hospital Victoria, Melbourne, Victoria, Australia
5Department of Endocrinology, Queensland Children's Hospital, South Brisbane, Queensland, Australia
6Telethon Kids Institute, The University of Western Australia, Perth, Western Australia, Australia
7School of Public Health, University of Adelaide, Adelaide, South Australia, Australia
8Walter and Eliza Hall Institute and Royal Melbourne Hospital, Melbourne, Victoria, Australia
9Centre for Diabetes Research, Harry Perkins Institute of Medical Research, Perth, Western Australia, Australia
10Hunter Diabetes Centre, Newcastle, New South Wales, Australia
11School of Biotechnology and Biomolecular Science, University of New South Wales, Sydney, New South Wales, Australia
12Department of Computing and Information Systems, University of Melbourne, Melbourne, Victoria, Australia
13Monash Centre for Health Research and Implementation, School of Public Health and Preventive Medicine, Monash University, Melbourne, Victoria, Australia
14School of Medicine, Deakin University, Geelong, Victoria, Australia
15Serology and Virology Division, SEALs Microbiology, Prince of Wales Hospital, Sydney, New South Wales, Australia
16Institute of Endocrinology and Diabetes, The Children's Hospital at Westmead, Sydney, New South Wales, Australia

Correspondence
Prof Maria E. Craig, Institute of Endocrinology and Diabetes, The Children's Hospital at Westmead, Locked Bag 4001, Westmead NSW 2145, Australia.
Email: m.craig@unsw.edu.au

Funding information
Juvenile Diabetes Research Foundation Australia, Grant/Award Number: 4-SRA-130006

Abstract
Background: Microbial exposures in utero and early life shape the infant microbiome, which can profoundly impact on health. Compared to the bacterial microbiome, very little is known about the virome. We set out to characterize longitudinal changes in the gut virome of healthy infants born to mothers with or without type 1 diabetes using comprehensive virome capture sequencing.
Methods: Healthy infants were selected from Environmental Determinants of Islet Autoimmunity (ENDIA), a prospective cohort of Australian children with a first-degree relative with type 1 diabetes, followed from pregnancy. Fecal specimens were collected three-monthly in the first year of life.

Results: Among 25 infants (44% born to mothers with type 1 diabetes) at least one virus was detected in 65% (65/100) of samples and 96% (24/25) of infants during the first year of life. In total, 26 genera of viruses were identified and >150 viruses were differentially abundant between the gut of infants with a mother with type 1 diabetes vs without. Positivity for any virus was associated with maternal type 1 diabetes and older infant age. Enterovirus was associated with older infant age and maternal smoking.

Conclusions: We demonstrate a distinct gut virome profile in infants of mothers with type 1 diabetes, which may influence health outcomes later in life. Higher prevalence and greater number of viruses observed compared to previous studies suggests significant underrepresentation in existing virome datasets, arising most likely from less sensitive techniques used in data acquisition.

KEYWORDS
enterovirus, gut, longitudinal, type 1 diabetes, virome

1 | INTRODUCTION

Microbial exposures in utero and early life shape the infant microbiome, which can profoundly impact on development, health, and immune maturation.1,2 Although the human microbiome encompasses a diverse population of bacteria, archaea, protists, fungi, viruses, and bacteriophages, current understanding stems mostly from characterization of the gut bacterial population using targeted sequencing of the 16S ribosomal RNA gene.3-7 In contrast to bacteria, viruses lack a universally conserved genetic region that can be targeted for amplification or enrichment. Furthermore, viruses are present at a significantly lower abundance in the gut compared to bacteria or their phages. Thus, it is challenging to characterize the complete population of viruses (“virome”) using conventional metagenomic sequencing.

All published infant virome studies to date have performed metagenomic sequencing without a highly effective method for virus enrichment.8-12 Most used a combination of physical enrichment techniques such as filtration, centrifugation and nuclease treatment that only provide modest increases in viral reads.13 Although these approaches provided sufficient sensitivity to explore the diversity and frequency of the highly abundant bacteriophage population, it is unclear whether this was also the case for non-phage viruses which were seldom detected.

Here we hypothesized that viruses are significantly underrepresented in existing infant virome datasets, and that infants born to a mother with type 1 diabetes have a distinct gut virome profile compared to those from a mother without diabetes. We tested this and further elucidated the composition, richness and the dynamics of viruses in the infant gut through comprehensive virome capture sequencing (VirCapSeq-VERT) of longitudinal fecal samples collected from 25 healthy infants during the first year of life. VirCapSeq-VERT enables effective enrichment of sequences corresponding to all viruses known to infect vertebrates, increasing viral read recovery up to 10,000-fold compared to conventional metagenomic sequencing.14

2 | METHODS

2.1 | Study subjects and sample selection

The case-control study population was nested within Environmental Determinants of Islet Autoimmunity (ENDIA), a longitudinal early life prospective cohort of children with at least one first-degree relative diagnosed with type 1 diabetes followed from pregnancy. We selected 25 infants who had fecal specimens collected at birth and 3, 6, 9, and either 12 or 15 months of age (Tables 1 and 2). In total, 100 longitudinal fecal samples stored at −80 °C were examined by VirCapSeq-VERT. None of the infants have developed type 1 diabetes.

The study was reviewed and approved (13 July 2016) by the Human Research Ethics Committee at the Women’s and Children’s Health Network under the National Mutual Acceptance Scheme (HREC/16/WCHN/66) and at all participating study sites in Australia. All participants provided written informed consent and were free to withdraw from the study at any time. Families were excluded if the mother could not comprehend her participation in the study and therefore was unable to provide informed consent.
**TABLE 1** Characteristics of infants stratified by maternal type 1 diabetes status

<table>
<thead>
<tr>
<th></th>
<th>Infants of mothers with type 1 diabetes (N = 11)</th>
<th>Infants of mothers without type 1 diabetes (N = 14)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Male, n (%)</strong></td>
<td>5 (46)</td>
<td>5 (36)</td>
<td>.7</td>
</tr>
<tr>
<td><strong>Season of birth n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summer</td>
<td>2 (18)</td>
<td>1 (7)</td>
<td>.6</td>
</tr>
<tr>
<td>Autumn</td>
<td>2 (18)</td>
<td>4 (29)</td>
<td>.7</td>
</tr>
<tr>
<td>Winter</td>
<td>3 (28)</td>
<td>4 (29)</td>
<td>.9</td>
</tr>
<tr>
<td>Spring</td>
<td>4 (36)</td>
<td>5 (33)</td>
<td>1.0</td>
</tr>
<tr>
<td><strong>Pet ownership, n (%)</strong></td>
<td>5 (45)</td>
<td>10 (71)</td>
<td>.2</td>
</tr>
<tr>
<td><strong>Siblings, n (%)</strong></td>
<td>4 (36)</td>
<td>8 (57)</td>
<td>.4</td>
</tr>
<tr>
<td>Number of siblings, mean ± SD</td>
<td>0.4 ± 0.5</td>
<td>0.7 ± 0.7</td>
<td>.2</td>
</tr>
<tr>
<td><strong>Obstetric history of infant</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Birth weight (kg), mean ± SD</td>
<td>3.4 ± 0.6</td>
<td>3.5 ± 0.5</td>
<td>.8</td>
</tr>
<tr>
<td>Weight for age z-score, mean ± SD</td>
<td>1.5 ± 2.3</td>
<td>0.2 ± 1.1</td>
<td>.07</td>
</tr>
<tr>
<td>Gestation (wk), mean ± SD</td>
<td>37.2 ± 1.3</td>
<td>39.2 ± 1.8</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Vaginal delivery, n (%)</td>
<td>7 (64)</td>
<td>8 (57)</td>
<td>1.0</td>
</tr>
<tr>
<td><strong>Maternal demographics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age at conception (y), mean ± SD</td>
<td>30.8 ± 3.9</td>
<td>33.5 ± 4.4</td>
<td>.02</td>
</tr>
<tr>
<td>High SES, n (%)</td>
<td>7 (64)</td>
<td>7 (50)</td>
<td>.7</td>
</tr>
<tr>
<td>Pre-pregnancy BMI, mean ± SD</td>
<td>26.2 ± 4.0</td>
<td>25.3 ± 4.8</td>
<td>.3</td>
</tr>
<tr>
<td>Tertiary education, n (%)</td>
<td>10 (91)</td>
<td>12 (86)</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Abbreviations: BMI, body mass index; SES, socioeconomic status.

### 2.2 Virome sequencing

Protocols for nucleic acid extraction, cDNA synthesis, sequence-independent amplification and VirCapSeq-VERT have been described previously.\textsuperscript{15,16} Briefly, total nucleic acid was extracted using the MagMAX Total Nucleic Acid Isolation Kit (Thermo Fisher Scientific, Waltham, Massachusetts) from 30% (w/v) fecal suspensions prepared in 1x PBS. To generate 1 μg double-stranded DNA for library synthesis, total nucleic acid was subjected to cDNA synthesis and sequence-independent pre-amplification using the Transplex Complete Whole Transcriptome Amplification Kit (Sigma-Aldrich, WTA1, St. Louis, Missouri). Following amplification, purified products were used to prepare Illumina sequencing libraries using the KAPA Hyperplus kit (KAPA Biosystems, Wilmington, Massachusetts) with single-index adapters, compatible for Sequence Capture Enrichment (Roche, Basel, Switzerland). VirCapSeq-VERT was performed at Columbia University, New York. Capture was performed according to the Nimblegen SeqCap protocol (Roche, Basel, Switzerland) as outlined previously.\textsuperscript{14} Post-capture, virus-enriched libraries were purified and amplified before sequencing. To ensure sufficient depth of coverage (about 10 million raw sequence reads/sample), uniquely barcoded samples were pooled at a maximum of 20 libraries per pool for sequence capture and each pool was sequenced single-end on an individual lane of HiSeq2500 (Illumina, San Diego, California).

### 2.3 Metagenomic sequence analysis

Sequence trimming, host sequence filtration, generation of contiguous read assemblies (contigs) and taxonomic classification of reads were performed as previously described.\textsuperscript{15,16} In brief, de-multiplexed and quality-trimmed sequence reads were aligned against host reference databases from GenBank (NCBI) using the Bowtie2 mapping algorithm (version 2.1.0)\textsuperscript{17} to remove the host background. Filtered reads were de novo assembled using either SOAPdenovo2,\textsuperscript{18} MEGAHIT\textsuperscript{39} or MIRA assemblers,\textsuperscript{20} then contigs and unique singletons were subjected to homology search at the nucleotide level using megablast. Sequences that exhibited poor or no homology at the nucleotide level were screened further using BLASTX against the viral GenBank protein database. For reference-based alignments, visualization of depth and spread of coverage for individual viruses, both Integrated Genomics Viewer\textsuperscript{21} and Geneious (version 9.0.5)\textsuperscript{22} were used.

### 2.4 Statistical analysis

Continuous demographic variables are reported as a mean ± SD (SD) for symmetrically distributed variables and median [IQR] for skewed variables. Categorical variables are summarized as number (%). Participant characteristics, including demographic variables, lifestyle factors, and comorbidities were stratified based on maternal type 1 diabetes and compared using independent t tests and Fisher’s exact tests for continuous and categorical variables, respectively. The socioeconomic index for areas (SEIFA) percentile for the postal area in which each participant resided was used as an indicator of socioeconomic status (SES).\textsuperscript{23} High SES was defined as >75th percentile and low SES was defined as <50th percentile.\textsuperscript{24}

Consistent with previous virome analyses,\textsuperscript{15,16} the virus positivity threshold was set at ≥100 viral reads matched by BLAST at the species level after a conservative 1% correction for potential sequence bleed-through. The estimated rate of bleed-through on the Illumina platform is ~0.3% of total reads when using a single index adapters.\textsuperscript{25} Associations between virus positivity and explanatory variables were examined using univariate and multivariable generalized estimating equations (GEE). Logistic regression models for the binary outcome of virus positivity at genus level were fit, with GEE used to account for the correlation among serial observations from the same infant. The major explanatory variable considered was maternal type 1 diabetes. Other explanatory variables investigated were maternal age at conception, parity, pet ownership, SES, pre-pregnancy body mass index (BMI),
Virome sequencing of 100 fecal specimens from 25 infants generated approximately 1.8 billion raw reads. This translated to 14.8 ± 7.8 M reads per sample following host and primer sequence removal. In total, 26 genera of viruses were detected and 65% (65/100) of samples were positive for nucleic acid corresponding to at least one virus. This equated to 96% (24/25) of infants being virus-positive in the first year of life. Noroviruses (28% of all samples positive), enteroviruses (26%), parechoviruses (14%), anelloviruses (11%), and bocaparvoviruses (7%) were the most frequently detected, and multiple viruses were detected in 31% of samples (Figure 1). In 60% (15/25) of infants, viruses of the same genus were detected at multiple timepoints across consecutive visits or 6 months apart (Figure 2). Viral richness was significantly lower in earliest-in-life specimens compared to the latest timepoint \((P = .0034)\).

Of the 25 infants examined, 11 were born to mothers with type 1 diabetes; characteristics stratified by maternal diabetes status are shown in Table 1. Mothers with type 1 diabetes were younger at conception and had shorter gestation. Total virus positivity was higher in infants of mothers with type 1 diabetes compared to those without (75% vs 59%). In univariate analysis, positivity for any virus was associated with older infant age (OR 1.2, 95% CI 1.1 to 1.3, \(P = .002\)) while the association with maternal type 1 diabetes did not reach statistical significance (OR 2.1, 95% CI 0.9 to 4.7, \(P = .07\)). Enterovirus positivity was associated with older age (OR 1.1, 95% CI 1.0 to 1.3, \(P = .032\)) and maternal smoking (OR 2.8, 95% CI 1.8 to 4.4, \(P < .0001\)), norovirus with low SES (OR 4.0, 95% CI 1.2 to 13.0, \(P = .02\)) and anellovirus with greater number of siblings (OR 2.4, 95% CI 1.1 to 5.3, \(P = .03\)). Parvovirus was associated with pet ownership (OR 5.5, 95% CI 1.3 to 24.0, \(P = .02\)) and maternal smoking (OR 2.8, 95% CI 1.0 to 1.3, \(P = .01\)). In multivariable GEE analysis, positivity for any virus was associated with maternal type 1 diabetes (OR 9.2, 95% CI 2.2 to 39.0, \(P = .003\)) after adjustment for infant age. The interaction between infant age and maternal type 1 diabetes was significant \((P = .005)\). No other examined variables including breastfeeding, mode of delivery and antibiotics intake were significantly associated with virus positivity.

Differential abundance analysis identified 165 viruses with a ≥ 2-fold difference in viral read abundance between infants of mothers with vs without type 1 diabetes (Figure S1). However, some of these differences were based on viral reads detected from a single sample. When only viruses detected in ≥3 samples were included in the analysis, 17 differentially abundant viruses

### TABLE 2 Characteristics of infants stratified by sample time point

<table>
<thead>
<tr>
<th>Maternal T1D (Y/N)</th>
<th>Sample 1</th>
<th>Sample 2</th>
<th>Sample 3</th>
<th>Sample 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maternal diabetes</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Age (mo), mean ± SD</td>
<td>3.1 ± 3.1</td>
<td>1.8 ± 2.1</td>
<td>5.2 ± 1.3</td>
<td>4.9 ± 2.3</td>
</tr>
<tr>
<td>Breastfeeding, n (%)</td>
<td>10 (91)</td>
<td>14 (100)</td>
<td>8 (73)</td>
<td>14 (100)</td>
</tr>
<tr>
<td>Season, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summer</td>
<td>3 (27)</td>
<td>5 (36)</td>
<td>1 (8)</td>
<td>4 (29)</td>
</tr>
<tr>
<td>Autumn</td>
<td>2 (18)</td>
<td>1 (7)</td>
<td>4 (37)</td>
<td>6 (43)</td>
</tr>
<tr>
<td>Winter</td>
<td>4 (37)</td>
<td>3 (21)</td>
<td>2 (18)</td>
<td>2 (14)</td>
</tr>
<tr>
<td>Spring</td>
<td>2 (18)</td>
<td>5 (36)</td>
<td>4 (37)</td>
<td>2 (14)</td>
</tr>
<tr>
<td>Maternal smoking, n (%)</td>
<td>0 (0)</td>
<td>1 (7)</td>
<td>1 (9)</td>
<td>1 (7)</td>
</tr>
<tr>
<td>Pets, n (%)</td>
<td>4 (37)</td>
<td>10 (71)</td>
<td>5 (45)</td>
<td>10 (71)</td>
</tr>
</tbody>
</table>
remained. Among the 15 most differentially abundant viruses, human bocavirus, and rotavirus A were more abundant in the infants of mothers with type 1 diabetes (Table 3). Conversely, human parechoviruses, coxsackievirus A6, Rhinovirus C, and torque teno viruses were significantly less abundant. Some noroviruses were more abundant in infants of mothers with type 1 diabetes and others less abundant.

4 | DISCUSSION

We detected a diverse range of viruses in the infant gut during the first year of life, many of which were more prevalent than reported in previous virome studies. Although various factors can influence virus positivity during infancy, our observations likely reflect the enhanced sensitivity of VirCapSeq-VERT for detecting vertebrate-infecting viruses over conventional metagenomic sequencing. Indeed, our application of VirCapSeq-VERT to other pregnancy and childhood specimens also indicated higher virus positivity compared to previous virome studies.

Total virus positivity was positively associated with maternal type 1 diabetes and older infant age, with a significant interaction between these two variables. We speculate that infants of mothers with type 1 diabetes are more likely to harbor viruses in the gut compared to infants whose mothers do not have type 1 diabetes. In contrast, total virus positivity was not associated with the number or the presence of siblings. This suggests that a mother with type 1 diabetes may be a significant source of viruses for infants in the first year of life. The absence of matching timepoint maternal samples precluded investigation of whether mothers with type 1 diabetes are in general more likely to harbor viruses compared to those without diabetes. However, this is certainly possible given the trend to higher virus positivity in pregnant women with type 1 diabetes compared to pregnant women without diabetes and the higher rates of microbial infection observed in individuals with type 1 diabetes vs without.

Enteroviruses were detected in 26% of samples and in 76% (19/25) of infants during the first year of life. This represents more than double the prevalence of enteroviruses found in the gut virome analysis of 22 infants in the DIABIMMUNE study. In univariate analysis, enterovirus presence was associated with older infant age. This is consistent with the progressive increase of enterovirus prevalence in fecal samples in the first year of life. Although our sample size was small, enterovirus positivity was also associated with maternal smoking. As smoking impairs responsiveness to viral infections and cause immune dysfunction, the effects of maternal smoking on virus susceptibility in the offspring should be explored and validated in larger cohorts.

A large body of evidence supports the role of enteroviruses as prime environmental triggers of islet autoimmunity and type 1 diabetes. In our recent analysis of the pregnancy gut virome, the frequency of enteroviruses did not differ significantly between pregnant women with and without type 1 diabetes. Similarly, examination of the gut virome changes preceding the development of islet autoimmunity and type 1 diabetes in 11 case children by Zhao et al found no significant differences in enterovirus abundance or prevalence between cases and controls. In the present analysis, there was no difference in enterovirus positivity between infants born of a mother with type 1 diabetes vs without. However, the abundance of some enteroviruses differed significantly, including a higher abundance of
coxsackievirus A6 in infants of mothers without type 1 diabetes. Interestingly, we previously reported higher abundance enterovirus A species in the gut of children with islet autoimmunity. Therefore, infants positive for enterovirus A in this study will be followed closely for their progression to islet autoimmunity.

Norovirus was present in 28% of samples and associated with low SES in our cohort. As both SES and norovirus infection are commonly associated with a reduced level of hygiene, infants in low SES families may be exposed more frequently to noroviruses compared to infants from high SES households. In contrast, no other viruses were

### FIGURE 2
Longitudinal dynamics of the infant gut virome in the first year of life. Presence-absence heatmap of viruses detected in 25 infants (11 from mothers with type 1 diabetes) over four consecutive study visits (V1-4). Viruses are represented at the genus level.

### TABLE 3
Top 15 differentially abundant viruses (detected in ≥3 samples) between the gut of infants from mothers with vs without type 1 diabetes.

<table>
<thead>
<tr>
<th>Virus</th>
<th>log2FD</th>
<th>Abs(log2FD)</th>
<th>P</th>
<th>FDR</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Higher in infants of mothers with T1D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human bocavirus</td>
<td>16.8</td>
<td>16.8</td>
<td>4.0E-35</td>
<td>8.3E-34</td>
<td>1</td>
</tr>
<tr>
<td>Norovirus GII/Hu/JP/2011</td>
<td>5.4</td>
<td>5.4</td>
<td>3.6E-21</td>
<td>2.5E-20</td>
<td>8</td>
</tr>
<tr>
<td>Norovirus GII.4</td>
<td>4.7</td>
<td>4.7</td>
<td>2.1E-08</td>
<td>5.0E-08</td>
<td>9</td>
</tr>
<tr>
<td>Norovirus Hu/GII.4</td>
<td>4.5</td>
<td>4.5</td>
<td>1.5E-10</td>
<td>5.4E-10</td>
<td>10</td>
</tr>
<tr>
<td>Human rotavirus A</td>
<td>4.5</td>
<td>4.5</td>
<td>5.4E-08</td>
<td>1.0E-07</td>
<td>11</td>
</tr>
<tr>
<td>Lower in infants of mothers with T1D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human parechovirus 4</td>
<td>−14.4</td>
<td>14.4</td>
<td>3.1E-23</td>
<td>3.2E-22</td>
<td>2</td>
</tr>
<tr>
<td>Human parechovirus</td>
<td>−10.3</td>
<td>10.3</td>
<td>1.2E-12</td>
<td>4.9E-12</td>
<td>3</td>
</tr>
<tr>
<td>Coxsackievirus A6</td>
<td>−9.0</td>
<td>9.0</td>
<td>1.9E-08</td>
<td>4.9E-08</td>
<td>4</td>
</tr>
<tr>
<td>Rhinovirus C</td>
<td>−7.6</td>
<td>7.6</td>
<td>2.0E-10</td>
<td>6.0E-10</td>
<td>5</td>
</tr>
<tr>
<td>Torque teno mini virus ALA22</td>
<td>−6.8</td>
<td>6.8</td>
<td>1.6E-16</td>
<td>8.3E-16</td>
<td>6</td>
</tr>
<tr>
<td>Human parechovirus 1</td>
<td>−6.4</td>
<td>6.4</td>
<td>4.2E-06</td>
<td>6.7E-06</td>
<td>7</td>
</tr>
<tr>
<td>Norovirus Hu/GII.4/PA363</td>
<td>−4.2</td>
<td>4.2</td>
<td>2.8E-07</td>
<td>4.9E-07</td>
<td>12</td>
</tr>
<tr>
<td>Torque teno virus</td>
<td>−4.0</td>
<td>4.0</td>
<td>4.9E-08</td>
<td>1.0E-07</td>
<td>13</td>
</tr>
<tr>
<td>Norovirus Hu/GII.4/NSW684S</td>
<td>−3.9</td>
<td>3.9</td>
<td>1.8E-04</td>
<td>2.5E-04</td>
<td>14</td>
</tr>
<tr>
<td>Norovirus Hu/GII.4/GII4-HK01</td>
<td>−3.3</td>
<td>3.3</td>
<td>8.5E-04</td>
<td>1.1E-03</td>
<td>15</td>
</tr>
</tbody>
</table>

Abbreviations: Abs, absolute; FD, fold-difference; FDR, false-discovery rate; T1D, type 1 diabetes.
associated with SES. Parechovirus was associated with pet ownership but a zoonotic link between parechoviruses in animals and humans has yet to be demonstrated. Positivity of parechoviruses peaked between 6 and 12 months of age, which corresponds with the nadir of IgG as maternal antibodies decline.

Consistent with other studies, anelloviruses were frequently detected in infancy. In our analysis, anellovirus was associated with greater number of siblings, supportive of the notion that a key source of its exposure is through contact with other infants. This is consistent with the lack of evidence for the vertical transmission of anelloviruses.46,47 Although their pathogenicity in humans remains uncertain, the expansion of anelloviruses in the gut and blood correlates with diminished immune status in immunocompromised patients.48-50

Interpretation of these results should take into consideration the following limitations. Firstly, as with all sequence-based virus detection, positivity for viral nucleic acid is a marker of, not proof of, infection. Viruses or viral nucleic acid may pass through the gut without causing a productive infection, as it is the case with plant viruses and other diet derived viruses.51 Second, we specifically focused on vertebrate-infecting viruses, excluding from our analysis other virus, bacteriophage and bacterial population coinhabiting the infant gut. This was deliberate to maximize the detection sensitivity for potentially human-infectious viruses. Although it would be desirable to obtain a wholistic view of the infant gut microbiome, recent evidence suggests that viruses have no influence on the bacterial microbiome or bacteriophage frequency.52 Third, all infants had a first degree relative with type 1 diabetes, who have a higher than baseline risk for development of islet autoimmunity and type 1 diabetes.53,54 Therefore, some of the associations identified in our univariate analysis may not reflect the general population. None of the infants have developed type 1 diabetes to date. Finally, our sample size precludes exploration of multiple variables that may act as confounders or effect modifiers in multivariable analyses. However, even if we select a more conservative p value of 0.01, the association with maternal diabetes and older age remains significant.

Strengths of our study include the first application of the highly sensitive VirCapSeq-VERT in a healthy infant cohort, the unbiased detection of all viruses simultaneously and the longitudinal nature of our cohort, which will enable validation of our findings using larger numbers through infancy and childhood. Furthermore, the availability of corresponding maternal pregnancy samples allows future investigation of whether any viruses detected in the present study were potentially vertically transmitted.

In conclusion, we provide a comprehensive and unbiased characterization of the gut virus population in infants during the first year of life and evaluate differences based on maternal type 1 diabetes. We identify older infant age, SES, sibling, pets, maternal smoking and maternal type 1 diabetes as factors influencing the gut virome in infants. Moreover, the higher prevalence and number of viruses observed compared to previous studies suggests an underrepresentation in previously reported virome datasets.

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 Basic Local Alignment Search Tool (BLAST) result summaries. The authors acknowledge funding from the National Health and Medical Research Council (NHMRC) Practitioner fellowships (APP1045777 M.E.C. and APP1044694 W.D.R.), Juvenile Diabetes Research Foundation Australia (JDRF)/NHMRC Centre of Research Excellence for the Protection of Pancreatic Beta Cells (APP1078106), JDRF International (3-SRA-2017-417-A-N) and the National Institutes of Health (Center for Research in Diagnostics and Discovery) (U19 AI10976 T.B. and W.I.L.). The ENDIA Study (ANZCTR # ACTRN12613000794707) is supported by JDRF Australia, the recipient of the Australian Research Council Special Research Initiative in Type 1 Juvenile Diabetes, The Leona M. and Harry B. Helmsley Charitable Trust, JDRF International, and the NHMRC.

CONFLICT OF INTEREST
The authors declare no potential conflict of interest.

AUTHOR CONTRIBUTIONS
M.E.C., W.D.R., T.B., W.I.L., and K.W.K. designed the study. K.W.K., S.R.I., and J.L.H. performed the experiments. I.C.N.P. performed the differential abundance analysis, K.J. and T.B. performed the de novo assembly of Illumina sequence reads and BLAST analysis. K.W.K., D.W.A., and M.E.C. performed the univariate and multivariable GEE analyses. All authors contributed to the interpretation of results. K.W.K., D.W.A., and M.E.C. wrote the manuscript, and all authors edited the manuscript.

ORCID
Jennifer J. Couper https://orcid.org/0000-0003-4448-8629
Aveni Haynes https://orcid.org/0000-0001-9954-5016
Rebecca L. Thomson https://orcid.org/0000-0002-7807-4144
Maria E. Craig https://orcid.org/0000-0001-6004-576X

REFERENCES


SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.