Distinct Gut Virome Profile of Pregnant Women With Type 1 Diabetes in the ENDIA Study

Ki Wook Kim,1,6 Digby W. Allen,1,4 Thomas Briese,2 Jennifer J. Couper,3 Simon C. Barry,3 Peter G. Colman,6 Andrew M. Cotterill,5 Elizabeth A. Davis,5 Lynne C. Giles,1 Leonard C. Harrison,4 Mark Harris,4 Aveni Haynes,4 Jessica L. Horton,4 Sonia R. Isaacs,5 Komal Jain,4 Walter Ian Lipkin,2 Grant Morahan,5 Claire Morbey,5 Ignatius C. N. Pang,1 Anthony T. Papenfuss,11 Megan A. S. Penno,2 Richard O. Sinnott,16 Georgia Soldatos,16 Rebecca L. Thomson,16 Peter J. Vuillermin,14 John M. Wentworth,8 Marc R. Wilkins,11 William D. Rawlinson,15,b and Maria E. Craig1,16,b; on behalf of the ENDIA Study Group

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

Background. The importance of gut bacteria in human physiology, immune regulation, and disease pathogenesis is well established. In contrast, the composition and dynamics of the gut virome are largely unknown; particularly lacking are studies in pregnancy. We used comprehensive virome capture sequencing to characterize the gut virome of pregnant women with and without type 1 diabetes (T1D), longitudinally followed in the Environmental Determinants of Islet Autoimmunity study.

Methods. In total, 61 pregnant women (35 with T1D and 26 without) from Australia were examined. Nucleic acid was extracted from serial fecal specimens obtained at prenatal visits, and viral genomes were sequenced by virome capture enrichment. The frequency, richness, and abundance of viruses were compared between women with and without T1D.

Results. Two viruses were more prevalent in pregnant women with T1D: picobirnaviruses (odds ratio [OR], 4.2; 95% confidence interval [CI], 1.0–17.1; P = .046) and tobamoviruses (OR, 3.2; 95% CI, 1.1–9.3; P = .037). The abundance of 77 viruses significantly differed between the 2 maternal groups (≥2-fold difference; P < .02), including 8 enterovirus B types present at a higher abundance in women with T1D.

Conclusions. These findings provide novel insight into the composition of the gut virome during pregnancy and demonstrate a distinct profile of viruses in women with T1D.

Keywords. enterovirus; pregnancy; type 1 diabetes; virome capture sequencing.
examined (Supplementary Table 1). ENDIA is a prospective cohort study of children at risk of T1D (have ≥1 first-degree relative with T1D), followed longitudinally from pregnancy to 3 years of life [9]. Women recruited between 2012 and 2016 were included in this analysis. Overall, 59 pregnancies were singleton and 2 were twin pregnancies. Fecal samples were collected during the first (n = 18), second (n = 47), and/or third trimester of pregnancy (n = 59), which were defined as gestational age of 1–14, 15–26, and 27–42 weeks, respectively. All samples were stored at −80°C in aliquots before analysis. For every participant, all available samples were examined. In total, 49 of 61 women had samples for multiple timepoints and 12 of 61 had samples for all 3 trimesters.

The study was reviewed and approved (July 13, 2016) by the study’s lead 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.

**Nucleic Acid Extraction**

Total nucleic acid (NA) was extracted using the MagMAX Total NA Isolation Kit (Thermo Fisher Scientific) on the semi-automated KingFisher FLEX Purification System (Thermo Fisher Scientific), following manufacturer’s guidelines with minor modifications. Thirty percent (w/v) of fecal suspensions were prepared in 1× phosphate-buffered saline and centrifuged for 5 minutes. All spin steps were performed at 13 000 ×g at room temperature. After centrifugation, 175 μL supernatant was transferred to zirconium bead tubes containing 235 μL Lysis/Binding Solution. Bead tubes were shaken at 2400 rpm on the Bioshake iQ (Quantifoil Instruments, Jena, Germany) for 15 minutes, then centrifuged 3 minutes. Into new tubes, 300 μL lysate was transferred and further centrifuged for 6 minutes. Total NA was purified from 200 μL lysate and stored at −80°C.

**Sequence-Independent Amplification**

Total NA was subjected to complementary deoxyribonucleic acid (cDNA) synthesis and sequence-independent preamplification (SIP) using the Transplex Complete Whole Transcriptome Amplification Kit (WTA1; Sigma-Aldrich, St. Louis, MO), following a published protocol [10]. In brief, 3.5 μL total NA was denatured at 95°C for 5 minutes instead of 70°C before cDNA synthesis to ensure amplification of both DNA and ribonucleic acid (RNA) molecules. Denatured NA was cooled to 18°C, and cDNA was synthesized using the following thermocycling conditions: 18°C 10 minutes, 25°C 10 minutes, 37°C 30 minutes, 42°C 10 minutes, 70°C 20 minutes, and 4°C holding. The entire cDNA library was used as template for SIP using the following cycling conditions repeated 22 times: 94°C 30 seconds and 70°C 5 minutes. After amplification, polymerase chain reaction (PCR) products were visualized on an agarose gel before purification using the ChargeSwitch-Pro PCR CleanUp Kit (Thermo Fisher Scientific).

**Virome Capture Sequencing**

One microgram of double-stranded DNA (dsDNA) was used for library synthesis using the KAPA Hyperplus kit (KAPA Biosystems, Wilmington, MA) with single-index adapters. In brief, dsDNA was enzymatically fragmented to an average of 200 base pairs. Fragments were purified using AmpureXP beads (Beckman Coulter, Brea, CA). Libraries were amplified for 6–9 cycles, quality checked on the LabChip GX Touch 24 Bioanalyzer (PerkinElmer, Waltham, MA), and quantified using the picogreen assay (Thermo Fisher Scientific) on the Victor X2 Fluorescent Microplate Reader (PerkinElmer). Completed libraries were pooled by equal mass for sequence capture. VirCapSeq-VERT was performed according to the Nimblegen SeqCap library protocol (Roche, Basel, Switzerland) as described previously [11]. Postcapture libraries were amplified and amplified before sequencing. To ensure sufficient depth of coverage (approximately 10 million unique sequence reads/sample), uniquely barcoded samples were pooled at a maximum of 20 libraries per pool (20-plex) and sequenced on a lane of HiSeq 4000 (Illumina, San Diego, CA).

**Genome Sequence Analysis**

Genome assembly, contig generation, and taxonomic classification of reads were performed as previously described [11]. Demultiplexed and quality-trimmed sequence reads were aligned against host reference databases from GenBank (National Center for Biotechnology Information) using the Bowtie2 mapping algorithm (version 2.1.0) [12] to remove the host background. Filtered reads were assembled de novo using either SOAPdenovo2 [13], MEGAHIT [14], or MIRA assemblers [15], 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 by BLASTX against the viral GenBank protein database. Viral sequences detected from BLASTX analysis were subjected to another round of BLASTX homology search against the entire GenBank protein database to correct for biased E values and inaccurate taxonomic classifications. For reference-based alignments, to visualize depth and spread of coverage for individual viruses, both Integrated Genomics Viewer [16] and Geneious (version 9.0.5) [17] were used. After taxonomic classification, read counts were corrected to account for sample bleeding due to Illumina index cross-talk, where sequences with single index barcodes are erroneously sorted, resulting in approximately 0.1% of total reads being distributed to the incorrect sample Fastq file. Cutoffs were applied across each pool separately. For each virus, 0.1% of the highest read count in that pool was calculated and subtracted from the
number of reads of that virus in each sample. All resulting read counts below 1 were corrected to zero. This process minimized the risk of false-“positive” identification of viruses in samples. To evaluate virus positivity, a threshold of 100 viral reads matched by Basic Local Alignment Search Tool (BLAST) at the species level, randomly distributed over the target genome, was applied. This threshold was selected for its proximity to the typical limit of detection of targeted quantitative PCR (~100 viral copies/ml), determined based on previous VirCapSeq-VERT experiments using whole blood [11] and feces (unpublished data).

**Statistical Analysis**

The STROBE reporting guidelines for observational cohort studies were followed [18]. Continuous demographic variables are reported as a mean ± standard deviation (SD) for parametric data and median (interquartile range) for skewed data, and categorical data as number (%). Participant characteristics, including demographic variables, lifestyle factors, and comorbidities, are reported according to T1D status and were compared using independent t tests and Fisher’s Exact tests for continuous and categorical data, respectively. The socioeconomic index for areas (index for relative socioeconomic disadvantage) percentile for the postal area in which each patient resided was used as an indicator of socioeconomic status (SES) [19]. High SES was defined as >75th percentile [20]. Virus positivity was determined by a positivity threshold of 100 viral reads matched by BLAST at the species level.

The differential abundance of viruses between mothers with and without T1D was examined using the edgeR package (version 3.14.0) [21] in R (version 3.3.0). A matrix of read counts, corrected for index cross-talk, was generated encompassing all samples and detected viruses. Each matrix entry had a count of 1 added to avoid issues with division by or log function of zero [22] before conversion to counts per million. Data were normalized using the Relative Log Expression method with respect to library size [23]. Two methods, common and tag-wise [24], 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 the biological coefficient of variation. Samples were divided into case and control groups, and the “exact” test was used to perform the biological coefficient of variation. Data were normalized using the Relative Log Expression method with respect to library size [23]. Two methods, common and tag-wise [24], 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 the biological coefficient of variation. Samples were divided into case and control groups, and the “exact” test was used to perform the biological coefficient of variation.

**RESULTS**

The pregnancy gut virome of women with and without T1D was characterized using VirCapSeq-VERT on 124 fecal specimens collected from 61 mothers (n = 35 with T1D, n = 26 without T1D) in the ENDIA study (Supplementary Table 1). The mean (±SD) age of mothers at conception was 32 ± 4 years and BMI was 27 ± 6 kg/m². Compared to women without T1D, women with T1D gave birth at a significantly younger age, after a shorter gestational length, and had fewer children (Supplementary Table 1).

High-throughput sequencing generated ~2 billion raw reads, which reduced to 1.6 billion reads after filtration of host and primer sequences. This equated to 12.7 ± 4.2 million filtered reads per sample. In total, 29 genera of eukaryotic viruses were detected, and 63% of samples (78 of 124) tested positive for at least 1 virus (Figure 1). Members of the Picobirnavirus, Parechovirus, and Enterovirus (EV) genera were among the most frequent vertebrate-infecting viruses sequenced. Although nonvertebrate-infecting viruses were excluded from...
VirCapSeq-Vert enrichment, tobamoviruses were frequently detected, suggesting that plant viruses are highly abundant in the gut during pregnancy and prevalent in human feces [30]. Although not reaching statistical significance, there was a trend to higher virus positivity in mothers with T1D versus without (64% vs 50%; P = .14). Rarefaction analysis revealed no difference in the richness of vertebrate-infecting viruses between women with and without T1D (Supplementary Figure 1), suggesting that all participants were exposed to a comparable community of viruses, independent of their T1D status.

Examination of longitudinal samples from 49 of 61 participants (n = 28 with T1D, n = 21 without T1D) identified alphapapillomaviruses, circoviruses, parechoviruses, and picobirnaviruses in multiple trimesters of pregnancy within individuals (Figure 1).
This may indicate persistent or recurring infection by closely related strains. There was no difference in the proportion of virus-positive samples across the 3 trimesters ($P = .95$). Chicken anaemia virus, genus Gyrovirus, and plant virus, pepper mild mottle virus, genus Tobamovirus were also detected across multiple trimesters, but these most likely originated from dietary intake [30]. Two viruses were more prevalent in women with T1D: picobirnaviruses (33% vs 9%; OR = 4.2; 95% CI, 1.0–17.1; $P = .046$) and tobamoviruses (22% vs 9%; OR = 3.2; 95% CI, 1.1–9.3; $P = .037$). In multivariable GEE models, the higher odds of having picobirnaviruses and tobamoviruses in women with T1D remained significant after adjustment for maternal age. In addition, there was a trend towards higher rates of gyroviruses, chloroviruses, and carlaviruses in women with T1D, but differences did not reach statistical significance. The frequency of EV did not differ between the 2 maternal groups; however, significant differences in EV types were observed. Coxsackievirus A2 (CVA2), CVB4, CVB5, Rhinovirus B, and ECHOviruses were detected exclusively in women with T1D, whereas CVA6, CVA10, CVA14, and EV71 were present only in mothers without T1D (Supplementary Table 2).

Differential abundance analysis identified 77 virus types with ≥2-fold significant difference ($P < .02$) between pregnant women with T1D versus those without, with a false discovery rate <5% (Figure 3 and Supplementary Data). Among the top 15 differentially abundant viruses were 3 EV-B types (CVB4, CVB3, and ECHOvirus E18), all present at higher abundance in women with T1D (Table 1). In contrast, 4 EV-A types (CVA10, CVA16, CVA5, and CVA14) were more abundant during pregnancy in women without T1D.

DISCUSSION

We demonstrated that eukaryotic viruses are prevalent in the gut of women during pregnancy, and that women with T1D are more likely to harbor picobirnaviruses and tobamoviruses compared with women without T1D. Furthermore, we found significant differences in viral abundance between women with and without T1D, including 8 EV B types that were all present at a higher abundance in women with T1D. These results demonstrate a distinct profile of viruses in women with T1D in pregnancy.

The pathogenicity of picobirnaviruses in humans remains to be definitively established. A weak association with gastrointestinal disorders in animals has been found, whereas in humans they are only considered as possible opportunistic pathogens [31].

**Figure 2.** Longitudinal changes in the gut virome during pregnancy. Presence-absence heatmap of viruses detected over multiple trimesters of pregnancy (T1, T2, and T3) in women with type 1 diabetes (n = 28 individuals) and without (n = 21 individuals).

**Figure 3.** Viruses differentially abundant between the gut of women with and without type 1 diabetes during pregnancy. Volcano plot of viruses with ≥2-fold difference (marked by vertical dotted lines) in abundance between pregnant women with and without T1D. Only differences with false discovery rate below 5% ($P < .05$) as determined by edgeR are represented. Species A (EV-A) and B enteroviruses (EV-B) are marked in red and blue, respectively. All other viruses represented in gray.
Most recently, picobirnaviruses were detected at high levels in patients with human immunodeficiency virus [32] and graft-versus-host disease [33], leading to the proposal that they may serve as a biomarker of immunosuppression. Thus, it is plausible that a higher prevalence of picobirnaviruses in women with T1D could be reflective of impaired antiviral defence. Tobamoviruses are not known to be pathogenic to humans and are commonly thought to be introduced to the gut through diet [30, 34]. Therefore, their higher prevalence in women with T1D during pregnancy may reflect differences in diet or consumption of contaminated drinking water [35]. Alternatively, there may be other factors involved such as gut permeability and intestinal inflammation, which are both increased in individuals with T1D [36] and may prevent effective clearance of dietary viruses. Our frequent detection of tobamoviruses in feces is consistent with other virome studies, including a recent study of 5 mother-infant pairs [30, 37].

A recent study examining the intestinal virome changes that precede the development of autoimmunity in T1D-susceptible children detected circoviruses at a greater abundance and prevalence in controls, suggesting that infection with this virus may offer protection from the development of T1D [48]. Consistent with this hypothesis, circoviruses were exclusively detected in women without T1D in our investigation (Figure 1). However, our sample size was too small to detect a statistically significant difference, and the case participants examined in our study were all women who had a long-standing T1D.

We examined potential confounding factors, in addition to T1D, that may influence the risk of virus infection during pregnancy. In univariate analysis, younger maternal age was associated with picobirnaviruses, older maternal age and no tertiary education was associated with EVs, and low SES was associated with gyroviruses. In multivariable GEE models, all of the aforementioned relationships remained significant except between maternal age and picobirnaviruses. Glycemic control may also influence susceptibility to infection; however, the majority of women with T1D in our study achieved glycemic targets for pregnancy.

To the best of our knowledge, this is the first study to examine the longitudinal gut virome across all 3 trimesters of pregnancy, providing novel baseline data for future gut virome investigations. Another major strength of this study is the application of virome capture sequencing [11, 49], which is the most sensitive and comprehensive sequence-based virome characterization tool currently available for vertebrate-infecting viruses. This method specifically targets all known viruses capable of infecting humans

### Table 1. Top 15 Differentially Abundant Species of Viruses between the Gut of Pregnant Women With Type 1 Diabetes Versus Without

<table>
<thead>
<tr>
<th>Virus</th>
<th>LogFD</th>
<th>PValue</th>
<th>FDR</th>
<th>Rank (Magnitude of FD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Higher in Women with T1D</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human coxsackievirus B4</td>
<td>11.8</td>
<td>1.70E-27</td>
<td>1.50E-26</td>
<td>1</td>
</tr>
<tr>
<td>Human coxsackievirus B3</td>
<td>11.2</td>
<td>1.40E-26</td>
<td>1.10E-25</td>
<td>3</td>
</tr>
<tr>
<td>Bathycoccus sp RCC1105 virus BpV2</td>
<td>7.1</td>
<td>2.90E-21</td>
<td>1.50E-20</td>
<td>7</td>
</tr>
<tr>
<td>Human adenovirus A</td>
<td>7</td>
<td>3.20E-21</td>
<td>1.60E-20</td>
<td>8</td>
</tr>
<tr>
<td>ECHOVirus E18</td>
<td>6</td>
<td>1.60E-19</td>
<td>5.90E-19</td>
<td>15</td>
</tr>
<tr>
<td><strong>Lower in Women With T1D</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coxsackievirus A10</td>
<td>−11.5</td>
<td>1.70E-38</td>
<td>1.80E-36</td>
<td>2</td>
</tr>
<tr>
<td>Brandmavirus UC1</td>
<td>−9.3</td>
<td>2.90E-35</td>
<td>1.50E-33</td>
<td>4</td>
</tr>
<tr>
<td>Phaeocystis globosa virus</td>
<td>−7.7</td>
<td>1.30E-31</td>
<td>3.50E-30</td>
<td>5</td>
</tr>
<tr>
<td>Porcine picobirnavirus</td>
<td>−7.4</td>
<td>3.40E-22</td>
<td>1.90E-21</td>
<td>6</td>
</tr>
<tr>
<td>Tomato mosaic virus</td>
<td>−7</td>
<td>2.30E-30</td>
<td>4.80E-29</td>
<td>9</td>
</tr>
<tr>
<td>Ostreococcus lucimarinus virus OIV4</td>
<td>−6.9</td>
<td>1.70E-29</td>
<td>2.20E-28</td>
<td>10</td>
</tr>
<tr>
<td>Coxsackievirus A16</td>
<td>−6.9</td>
<td>1.70E-29</td>
<td>2.20E-28</td>
<td>11</td>
</tr>
<tr>
<td>Paramecium bursaria Chlorella virus</td>
<td>−6.8</td>
<td>7.70E-30</td>
<td>1.40E-28</td>
<td>12</td>
</tr>
<tr>
<td>Coxsackievirus A5</td>
<td>−6.4</td>
<td>4.30E-28</td>
<td>4.80E-27</td>
<td>13</td>
</tr>
<tr>
<td>Coxsackievirus A14</td>
<td>−6.4</td>
<td>4.60E-28</td>
<td>4.80E-27</td>
<td>14</td>
</tr>
</tbody>
</table>

Abbreviations: FD, fold difference; FDR, false discovery rate; T1D, type 1 diabetes.
and other vertebrates, significantly reducing sequences produced from host and bacterial background, allowing up to a 10,000-fold increase in the number of viral reads recovered compared with conventional virome sequencing methods. Furthermore, our method enabled the examination of both RNA and DNA viruses simultaneously [11]. Despite these strengths, the interpretation of our virome data is limited by the fact that sequencing cannot differentiate between the presence of viral genomes in the gut versus actively replicating viruses. In addition, the absence of a nonpregnant control group precluded the analysis of the effect of pregnancy on virus infection in this study.

Given the results of our recent systematic review and meta-analysis of 2992 women and children that demonstrated a significant association between maternal virus infection in pregnancy and T1D in the offspring [8], future studies could be aimed at examining the impact of maternal virus infections on the development of islet autoimmunity and T1D in the offspring. For this purpose, the offspring of women examined in this study are being followed longitudinally for these 2 outcomes as part of the ENDIA study, a prospective cohort study following at risk children. The characterization of the gut virome in these mother-infant pairs will allow identification of potential vertical transmission of viruses (currently underway). The impact of diet on the gut virome will also be examined. The virome of other potential sources of vertical transmission should also be investigated such as the oral, skin, breastmilk, and the vaginal virome, which has been recently shown to be of clinical importance for its potential contribution to preterm birth [50].

CONCLUSIONS

In conclusion, our findings provide novel insight into the diversity and dynamics of the gut virome during pregnancy and identify T1D and maternal age as key factors influencing virus infection in pregnancy. We show a novel potential association between T1D and picobirnaviruses and demonstrate a distinct profile of viruses during pregnancy in women with T1D, providing novel targets for prevention studies.

Supplementary Data

Supplementary materials are available at Open Forum Infectious Diseases online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Acknowledgments

We thank Nishit P. Bhuva (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. Financial support. This work was funded by the National Health and Medical Research Council (NHMRC) Practitioner fellowships (APP1045777 [to M. E. C.] and APP1044694 [to W. D. R.]), Juvenile Diabetes Research Foundation Australia (JDRF)/NHMRC Centre of Research Excellence for the Protection of Pancreatic Beta Cells (APP1078106), and the National Institutes of Health (Center for Research in Diagnostics and Discovery) (U19 AI109761 [to T. B. and W. E. L.]). The ENDIA Study 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.

Potential conflicts of interest. All authors: No reported conflicts of interest.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest.

References