

Research Publication Repository

<http://publications.wehi.edu.au/search/SearchPublications>

"This is the peer reviewed version of the following article, which has been published in final form at the link below.

<p>Publication details:</p>	<p>Kim KW, Allen DW, Briese T, Couper JJ, Barry SC, Colman PG, Cotterill AM, Davis EA, Giles LC, Harrison LC, Harris M, Haynes A, Horton JL, Isaacs SR, Jain K, Lipkin WI, McGorm K, Morahan G, Morbey C, Pang ICN, Papenfuss AT, Penno MAS, Sinnott RO, Soldatos G, Thomson RL, Vuillermin P, Wentworth JM, Wilkins MR, Rawlinson WD, Craig ME, Endia Study Group. Higher frequency of vertebrate-infecting viruses in the gut of infants born to mothers with type 1 diabetes. <i>Pediatric Diabetes</i>. 2020 21(2):271-27</p>
<p>Published version is available at:</p>	<p>https://doi.org/10.1111/pedi.12952</p>

Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this manuscript.

This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions. This article may not be enhanced, enriched or otherwise transformed into a derivative work, without express permission from Wiley or by statutory rights under applicable legislation. Copyright notices must not be removed, obscured or modified. The article must be linked to Wiley's version of record on Wiley Online Library and any embedding, framing or otherwise making available the article or pages thereof by third parties from platforms, services and websites other than Wiley Online Library must be prohibited.

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

Short Title: Virome in infants at risk of type 1 diabetes

Authors:

Ki Wook Kim¹†; Digby W. Allen¹†; Thomas Briese²; Jennifer J. Couper³; Simon C. Barry³; Peter G. Colman⁴; Andrew M. Cotterill⁵; Elizabeth A. Davis⁶; Lynne C. Giles⁷; Leonard C. Harrison⁸; Mark Harris⁵; Aveni Haynes⁶; Jessica L. Horton¹; Sonia R. Isaacs¹; Komal Jain²; Walter Ian Lipkin²; Kelly McGorm³; Grant Morahan⁹; Claire Morbey¹⁰; Ignatius C.N. Pang¹¹; Anthony T. Papenfuss⁸; Megan A.S. Penno³; Richard O. Sinnott¹²; Georgia Soldatos¹³; Rebecca L. Thomson³; Peter Vuillermin¹⁴; John M. Wentworth⁸; Marc R. Wilkins¹¹; William D. Rawlinson^{1,15}‡, and Maria E. Craig^{1,16}‡*; on behalf of the ENDIA STUDY GROUP

†Ki Wook Kim and Digby W. Allen are equal first authors.

‡Maria E. Craig and William D. Rawlinson are equal senior authors.

* Maria E. Craig is the corresponding author.

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

²Center for Infection and Immunity and Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, USA

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/pedi.12952

³Robinson Research Institute and Adelaide Medical School, University of Adelaide, Adelaide, Australia

⁴Department of Diabetes and Endocrinology, The Royal Melbourne Hospital Victoria, Melbourne, Australia

⁵Department of Endocrinology, Queensland Children's Hospital, South Brisbane, Australia

⁶Telethon Kids Institute, The University of Western Australia, Perth, Australia

⁷School of Public Health, University of Adelaide, Adelaide, Australia

⁸Walter and Eliza Hall Institute and Royal Melbourne Hospital, Melbourne, Australia

⁹Centre for Diabetes Research, Harry Perkins Institute of 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, 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

Correspondence to:

Prof Maria E. Craig

Institute of Endocrinology and Diabetes, The Children's Hospital at Westmead

Locked Bag 4001, Westmead NSW 2145, Australia

Tel: +61 2 9845 3907; Fax: +61 2 9845 3170; Email: m.craig@unsw.edu.au

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.

Disclosure of interests

All authors declare no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

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.

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 versus 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

Virome; gut; enterovirus; type 1 diabetes; longitudinal

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).

METHODS

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 (Table 1 & 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 (July 13, 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.

Virome sequencing

Protocols for nucleic acid extraction, cDNA synthesis, sequence-independent amplification and VirCapSeq-VERT have been described previously (15, 16). Briefly, total nucleic acid was extracted using the MagMAX Total Nucleic Acid Isolation Kit (Thermo Fisher Scientific, Waltham, MA) from 30% (w/v) fecal suspensions prepared in 1xPBS. To generate 1µg double-stranded DNA for library synthesis, total nucleic acid was subjected to cDNA synthesis and

Accepted Article

sequence independent pre-amplification using the Transplex Complete Whole Transcriptome Amplification Kit (Sigma-Aldrich, WTA1, St. Louis, MO). Following amplification, purified products were used to prepare Illumina sequencing libraries using the KAPA Hyperplus kit (KAPA Biosystems, Wilmington, MA) with single-index adapters, compatible for Sequence Capture Enrichment (Roche, Basel, Switzerland). VirCapSeq-VERT was performed at Columbia University, New York, USA. Capture was performed according to the Nimblegen SeqCap protocol (Roche, Basel, Switzerland) as outlined previously (14). Post-capture, virus-enriched libraries were purified and amplified before sequencing. To ensure sufficient depth of coverage (~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, CA).

Metagenomic sequence analysis.

Sequence trimming, host sequence filtration, generation of contiguous read assemblies (contigs) and taxonomic classification of reads were performed as previously described (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) (17) to remove the host background. Filtered reads were *de novo* assembled using either SOAPdenovo2 (18), MEGAHIT (19) or MIRA assemblers(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,

Accepted Article
visualization of depth and spread of coverage for individual viruses, both Integrated Genomics Viewer(21) and Geneious (version 9.0.5) (22) were used.

Statistical analysis.

Continuous demographic variables are reported as a mean \pm standard deviation (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) (23). High SES was defined as $>75^{\text{th}}$ percentile and low SES was defined as $<50^{\text{th}}$ percentile (24).

Consistent with previous virome analyses (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 $\sim 0.3\%$ of total reads when using a single index adapters (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), education, sex, breastfeeding, mode of delivery (vaginal or not), gestational age, birth weight, maternal smoking, antibiotics intake, age at sample collection, and siblings as appropriate in each model.

An individual model was developed for each genus of virus, with independent variables that exhibited an association in univariate analyses included in multivariable analyses. The fit of statistical models was compared using the quasi-likelihood under independence model criterion, and the lowest scoring and most parsimonious models are reported. Results are expressed as odds ratio (OR), with 95% confidence intervals (CI). Multivariable analysis was limited to the four most frequent genera of virus detected. Analyses were performed using IBM SPSS Statistics (version 24.0, Chicago, IL) and statistical significance was defined as $p \leq 0.05$.

The richness of virus population was compared between specimens collected at the first and last visit using the Wilcoxon test (paired, nonparametric) in GraphPad Prism version 7 (California, USA). Differential abundance of viruses between infants of mothers with and without type 1 diabetes was examined using the edgeR package (version 3.14.0) (26) in R (version 3.3.0). A matrix of read counts was generated encompassing all samples and detected viruses. Each matrix entry had a count of one added to avoid issues with division by, or log

function of zero (27) before conversion to counts per million. Data were normalized using the Relative Log Expression method with respect to library size (28). Two methods, common (29) and tag-wise (30), were used to estimate the biological coefficient of variation. Samples were divided based on maternal diabetes status, and the “exact” test used to perform hypothesis testing (29). P-values were adjusted to control false discovery rates with the Benjamini-Hochberg multiple testing correction procedure (31). Viruses with an adjusted $P < 0.05$ were identified as displaying statistically significant differential abundance between infants with and without a mother with type 1 diabetes. Heatmaps were generated using `iheatmapr` package in R (32) to visualize viral read abundance across all specimens.

RESULTS

Virome sequencing of 100 fecal specimens from 25 infants generated ~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 six months apart (Figure 2). Viral richness was significantly lower in earliest-in-life specimens compared to the latest timepoint ($P=0.0034$).

Accepted Article

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, $P=0.002$) while the association with maternal type 1 diabetes did not reach statistical significance (OR 2.1, 95% CI 0.9 to 4.7, $P=0.07$). Enterovirus positivity was associated with older age (OR 1.1, 95% CI 1.0 to 1.3, $P=0.032$) and maternal smoking (OR 2.8, 95% CI 1.8 to 4.4, $P<0.0001$), norovirus with low SES (OR 4.0, 95% CI 1.2 to 13.0, $P=0.02$) and anellovirus with greater number of siblings (OR 2.4, 95% CI 1.1 to 5.3, $P=0.03$). Parechovirus was associated with pet ownership (OR 5.5, 95% CI 1.3 to 24.0, $P=0.02$) and older age (OR 1.2, 95% CI 1.0 to 1.3, $P=0.01$). 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, $P=0.003$) after adjustment for infant age. The interaction between infant age and maternal type 1 diabetes was significant ($P=0.005$). 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 versus without type 1 diabetes (Figure S1, Supporting Information). 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 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.

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 (8-11, 33). 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 (14). Indeed, our application of VirCapSeq-VERT to other pregnancy (15) and childhood (16) 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 (15) and the higher rates of microbial infection observed in individuals with type 1 diabetes versus without (34-36).

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 (33). 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 (37). 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 (38), 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 (39-44). In our recent analysis of the pregnancy gut virome (15), 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 (33). In the present analysis, there was no difference in enterovirus positivity between infants born of a mother with type 1 diabetes versus without. However, the abundance

Accepted Article

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 (16). 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 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 6–12 months of age, which corresponds with the nadir of IgG as maternal antibodies decline.

Consistent with other studies (8-10, 45), 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).

Accepted Article

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 characterisation 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.

REFERENCES

1. Ihekweazu FD, Versalovic J. Development of the Pediatric Gut Microbiome: Impact on Health and Disease. *Am J Med Sci.* 2018; 356:413-23.
2. Lim ES, Wang D, Holtz LR. The Bacterial Microbiome and Virome Milestones of Infant Development. *Trends in microbiology.* 2016; 24:801-10.
3. Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo G, Fierer N, et al. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc Natl Acad Sci U S A.* 2010; 107:11971-5.
4. Palmer C, Bik EM, DiGiulio DB, Relman DA, Brown PO. Development of the Human Infant Intestinal Microbiota. *PLOS Biology.* 2007; 5:e177.
5. Subramanian S, Huq S, Yatsunenko T, Haque R, Mahfuz M, Alam MA, et al. Persistent gut microbiota immaturity in malnourished Bangladeshi children. *Nature.* 2014; 510:417.
6. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al. Human gut microbiome viewed across age and geography. *Nature.* 2012; 486:222.
7. Koenig JE, Spor A, Scalfone N, Fricker AD, Stombaugh J, Knight R, et al. Succession of microbial consortia in the developing infant gut microbiome. *Proceedings of the National Academy of Sciences.* 2011; 108:4578.
8. Reyes A, Blanton LV, Cao S, Zhao G, Manary M, Trehan I, et al. Gut DNA viromes of Malawian twins discordant for severe acute malnutrition. *Proc Natl Acad Sci U S A.* 2015; 112:11941-6.
9. Lim ES, Zhou Y, Zhao G, Bauer IK, Droit L, Ndao IM, et al. Early life dynamics of the human gut virome and bacterial microbiome in infants. *Nature medicine.* 2015; 21:1228-34.

10. McCann A, Ryan FJ, Stockdale SR, Dalmaso M, Blake T, Ryan CA, et al. Viromes of one year old infants reveal the impact of birth mode on microbiome diversity. *PeerJ*. 2018; 6:e4694.
11. Kramna L, Kolarova K, Oikarinen S, Pursiheimo JP, Ilonen J, Simell O, et al. Gut virome sequencing in children with early islet autoimmunity. *Diabetes care*. 2015; 38:930-3.
12. Cinek O, Kramna L, Lin J, Oikarinen S, Kolarova K, Ilonen J, et al. Imbalance of bacteriome profiles within the Finnish Diabetes Prediction and Prevention study: Parallel use of 16S profiling and virome sequencing in stool samples from children with islet autoimmunity and matched controls. *Pediatric diabetes*. 2017; 18:588-98.
13. Hall RJ, Wang J, Todd AK, Bissielo AB, Yen S, Strydom H, et al. Evaluation of rapid and simple techniques for the enrichment of viruses prior to metagenomic virus discovery. *Journal of virological methods*. 2014; 195:194-204.
14. Briese T, Kapoor A, Mishra N, Jain K, Kumar A, Jabado OJ, et al. Virome Capture Sequencing Enables Sensitive Viral Diagnosis and Comprehensive Virome Analysis. *mBio*. 2015; 6:e01491-15.
15. Kim KW, Allen DW, Briese T, Couper JJ, Barry SC, Colman PG, et al. Distinct Gut Virome Profile of Pregnant Women With Type 1 Diabetes in the ENDIA Study. *Open forum infectious diseases*. 2019; 6:ofz025.
16. Kim KW, Horton JL, Pang CNI, Jain K, Leung P, Isaacs SR, et al. Higher abundance of enterovirus A species in the gut of children with islet autoimmunity. *Sci Rep*. 2019; 9:1749.
17. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012; 9:357-9.
18. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience*. 2012; 1:18.
19. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*. 2015; 31:1674-6.
20. Mundry M, Bornberg-Bauer E, Sammeth M, Feulner PGD. Evaluating Characteristics of De Novo Assembly Software on 454 Transcriptome Data: A Simulation Approach. *PloS one*. 2012; 7:e31410.
21. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative Genomics Viewer. *Nat Biotechnol*. 2011; 29:24-6.
22. Kears M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 2012; 28:1647-9.
23. Pink B. Socio-Economic Indexes for Areas (SEIFA) - Technical Paper. In: Australian Bureau of Statistics, (ed). Commonwealth of Australia, Canberra, 2011.
24. Department of Education and Training. Low socio-economic status (SES) - postcode measure. Australian Government, HEIMSHelp, 2015.
25. Kircher M, Sawyer S, Meyer M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res*. 2011; 40:e3-e.
26. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010; 26:139-40.

27. McMurdie PJ, Holmes S. Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLOS Computational Biology*. 2014; 10:e1003531.
28. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol*. 2010; 11:R106.
29. Robinson MD, Smyth GK. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics (Oxford, England)*. 2008; 9:321-32.
30. Robinson MD, Smyth GK. Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics*. 2007; 23:2881-7.
31. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior genetics research. *Behav Brain Res*. 2001; 125:279-84.
32. Schep AN, Kummerfeld SK. iheatmapr: Interactive complex heatmaps in R. *Journal of Open Source Software*. 2017.
33. Zhao G, Vatanen T, Droit L, Park A, Kostic AD, Poon TW, et al. Intestinal virome changes precede autoimmunity in type 1 diabetes-susceptible children. *Proceedings of the National Academy of Sciences*. 2017; 114:E6166-E75.
34. Hak E, Rutten GEHM, Gorter KJ, Goudzwaard WL, Muller LMAJ, Hoepelman AIM, et al. Increased Risk of Common Infections in Patients with Type 1 and Type 2 Diabetes Mellitus. *Clin Infect Dis*. 2005; 41:281-8.
35. Carey IM, Critchley JA, DeWilde S, Harris T, Hosking FJ, Cook DG. Risk of Infection in Type 1 and Type 2 Diabetes Compared With the General Population: A Matched Cohort Study. *Diabetes care*. 2018:dc172131.
36. Zhang X, Zhu X, Ji Y, Li H, Hou F, Xiao C, et al. Increased risk of hepatitis B virus infection amongst individuals with diabetes mellitus. *Biosci Rep*. 2019; 39:BSR20181715.
37. Rønningen KS, Wetlesen T, Witsø E, Cinek O, Aldrin M, Grinde B, et al. Predictors of sub-clinical enterovirus infections in infants: a prospective cohort study. *Int J Epidemiol*. 2009; 39:459-68.
38. Qiu F, Liang C-L, Liu H, Zeng Y-Q, Hou S, Huang S, et al. Impacts of cigarette smoking on immune responsiveness: Up and down or upside down? *Oncotarget*. 2016; 8:268-84.
39. Yeung W-CG, Rawlinson WD, Craig ME. Enterovirus infection and type 1 diabetes mellitus: systematic review and meta-analysis of observational molecular studies. *BMJ*. 2011; 342.
40. Richardson SJ, Morgan NG. Enteroviral infections in the pathogenesis of type 1 diabetes: new insights for therapeutic intervention. *Curr Opin Pharm*. 2018; 43:11-9.
41. Kim KW, Ho A, Alshabee-Akil A, Hardikar AA, Kay TW, Rawlinson WD, et al. Coxsackievirus B5 Infection Induces Dysregulation of microRNAs Predicted to Target Known Type 1 Diabetes Risk Genes in Human Pancreatic Islets. *Diabetes*. 2016; 65:996-1003.
42. Honkanen H, Oikarinen S, Nurminen N, Laitinen OH, Huhtala H, Lehtonen J, et al. Detection of enteroviruses in stools precedes islet autoimmunity by several months: possible evidence for slowly operating mechanisms in virus-induced autoimmunity. *Diabetologia*. 2017:1-8.
43. Hyoty H. Viruses in type 1 diabetes. *Pediatric diabetes*. 2016; 17 Suppl 22:56-64.
44. Dunne JL, Richardson SJ, Atkinson MA, Craig ME, Dahl-Jorgensen K, Flodstrom-Tullberg M, et al. Rationale for enteroviral vaccination and antiviral therapies in human type 1 diabetes. *Diabetologia*. 2019.

45. Kapusinszky B, Minor P, Delwart E. Nearly Constant Shedding of Diverse Enteric Viruses by Two Healthy Infants. *J Clin Microbiol.* 2012; 50:3427.
46. Gerner P, Oettinger R, Gerner W, Falbrede J, Wirth S. Mother-to-infant transmission of TT virus: prevalence, extent and mechanism of vertical transmission. *The Pediatric Infectious Disease Journal.* 2000; 19:1074-7.
47. Komatsu H, Inui A, Sogo T, Kuroda K, Tanaka T, Fujisawa T. TTV infection in children born to mothers infected with TTV but not with HBV, HCV, or HIV. *Journal of Medical Virology.* 2004; 74:499-506.
48. De Vlamincq I, Khush KK, Strehl C, Kohli B, Luikart H, Neff NF, et al. Temporal response of the human virome to immunosuppression and antiviral therapy. *Cell.* 2013; 155:1178-87.
49. Blatter JA, Sweet SC, Conrad C, Danziger-Isakov LA, Faro A, Goldfarb SB, et al. Anellovirus loads are associated with outcomes in pediatric lung transplantation. *Pediatr Transplant.* 2018; 22.
50. Young JC, Chehoud C, Bittinger K, Bailey A, Diamond JM, Cantu E, et al. Viral metagenomics reveal blooms of anelloviruses in the respiratory tract of lung transplant recipients. *Am J Transplant.* 2015; 15:200-9.
51. Zhang T, Breitbart M, Lee WH, Run J-Q, Wei CL, Soh SWL, et al. RNA Viral Community in Human Feces: Prevalence of Plant Pathogenic Viruses. *PLOS Biology.* 2005; 4:e3.
52. Moreno-Gallego JL, Chou SP, Di Rienzi SC, Goodrich JK, Spector TD, Bell JT, et al. Virome Diversity Correlates with Intestinal Microbiome Diversity in Adult Monozygotic Twins. *Cell host & microbe.* 2019; 25:261-72.e5.
53. Couper JJ, Haller MJ, Greenbaum CJ, Ziegler A-G, Wherrett DK, Knip M, et al. ISPAD Clinical Practice Consensus Guidelines 2018: Stages of type 1 diabetes in children and adolescents. *Pediatr Diabetes.* 2018; 19:20-7.
54. Hippich M, Beyerlein A, Hagopian WA, Krischer JP, Vehik K, Knoop J, et al. Genetic Contribution to the Divergence in Type 1 Diabetes Risk Between Children From the General Population and Children From Affected Families. *Diabetes.* 2019.

TABLES

Table 1. Characteristics of infants stratified by maternal type 1 diabetes status

	Infants of mothers with type 1 diabetes (N=11)	Infants of mothers without type 1 diabetes (N=14)	P-value
Male, <i>n</i> (%)	5 (46)	5 (36)	0.7
Season of birth <i>n</i> (%)			
Summer	2 (18)	1 (7)	0.6
Autumn	2 (18)	4 (29)	0.7

	Winter	3 (28)	4 (29)	1.0
	Spring	4 (36)	5 (33)	1.0
Pet ownership, <i>n</i> (%)		5 (45)	10 (71)	0.2
Siblings, <i>n</i> (%)		4 (36)	8 (57)	0.4
Number of siblings, mean \pm SD		0.4 \pm 0.5	0.7 \pm 0.7	0.2
Obstetric history of infant				
Birth weight (kg), mean \pm SD		3.4 \pm 0.6	3.5 \pm 0.5	0.8
Weight for age z-score, mean \pm SD		1.5 \pm 2.3	0.2 \pm 1.1	0.07
Gestation (weeks), mean \pm SD		37.2 \pm 1.3	39.2 \pm 1.8	<0.0001
Vaginal delivery, <i>n</i> (%)		7 (64)	8 (57)	1.0
Maternal demographics				
Age at conception (years), mean \pm SD		30.8 \pm 3.9	33.5 \pm 4.4	0.02
High SES, <i>n</i> (%)		7 (64)	7 (50)	0.7
Pre-pregnancy BMI, mean \pm SD		26.2 \pm 4.0	25.3 \pm 4.8	0.3
Tertiary education, <i>n</i> (%)		10 (91)	12 (86)	1.0

SD = standard deviation; SES = socioeconomic status; BMI = body mass index

Table 2. Characteristics of infants stratified by sample time point

Maternal T1D (Y/N)	Sample 1		Sample 2		Sample 3		Sample 4	
	Y	N	Y	N	Y	N	Y	N
Age (months), mean \pm SD	3.1 \pm 3.1	1.8 \pm 2.1	5.2 \pm 1.3	4.9 \pm 2.3	8.9 \pm 1.5	8.1 \pm 2.5	12.9 \pm 1.6	11.4 \pm 2.4
Breastfeeding, <i>n</i> (%)	10 (91)	14 (100)	8 (73)	14 (100)	7 (64)	10 (71)	6 (55)	8 (57)
Season, <i>n</i> (%)								
Summer	3 (27)	5 (36)	1 (8)	4 (29)	2 (18)	2 (14)	3 (27)	1 (7)
Autumn	2 (18)	1 (7)	4 (37)	6 (43)	3 (27)	4 (29)	1 (8)	2 (14)
Winter	4 (37)	3 (21)	2 (18)	2 (14)	4 (37)	5 (36)	3 (27)	4 (29)
Spring	2 (18)	5 (36)	4 (37)	2 (14)	2 (18)	3 (21)	4 (38)	7 (50)
Maternal smoking, <i>n</i> (%)	0 (0)	1 (7)	1 (9)	1 (7)	1 (9)	1 (7)	0 (0)	1 (7)
Pets, <i>n</i> (%)	4 (37)	10 (71)	5 (45)	10 (71)	5 (45)	9 (64)	5 (45)	8 (57)

SD = standard deviation

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

Virus	log ₂ FD	Abs(log ₂ FD)	<i>P</i>	FDR	Rank
Higher in infants of mothers with T1D					
Human bocavirus	16.8	16.8	4.0E-35	8.3E-34	1
Norovirus GII/Hu/JP/2011	5.4	5.4	3.6E-21	2.5E-20	8
Norovirus GII.4	4.7	4.7	2.1E-08	5.0E-08	9
Norovirus Hu/GII.4	4.5	4.5	1.5E-10	5.4E-10	10
Human rotavirus A	4.5	4.5	5.4E-08	1.0E-07	11
Lower in infants of mothers with T1D					
Human parechovirus 4	-14.4	14.4	3.1E-23	3.2E-22	2
Human parechovirus	-10.3	10.3	1.2E-12	4.9E-12	3
Coxsackievirus A6	-9.0	9.0	1.9E-08	4.9E-08	4
Rhinovirus C	-7.6	7.6	2.0E-10	6.0E-10	5
Torque teno mini virus ALA22	-6.8	6.8	1.6E-16	8.3E-16	6

Human parechovirus 1	-6.4	6.4	4.2E-06	6.7E-06	7
Norovirus Hu/GII.4/PA363	-4.2	4.2	2.8E-07	4.9E-07	12
Torque teno virus	-4.0	4.0	4.9E-08	1.0E-07	13
Norovirus Hu/GII.4/NSW684S	-3.9	3.9	1.8E-04	2.5E-04	14
Norovirus Hu/GII.4/GII4-HK01	-3.3	3.3	8.5E-04	1.1E-03	15

FD = fold-difference; Abs = absolute; FDR = false-discovery rate; T1D = type 1 diabetes

FIGURE LEGENDS

Figure 1. Viruses detected by VirCapSeq-VERT. Heatmap of viral read counts (log2 scale) measured using VirCapSeq-VERT in 100 fecal samples collected from infants of mothers with (n=44) and without (n=56) type 1 diabetes. Viruses are represented at the genus level and the number of viruses detected within each sample is summarized by a bar chart above.

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.

