



Gut Virome Sequencing in Children With Early Islet Autoimmunity

DOI: 10.2337/dc14-2490

Lenka Kramná,¹ Kateřina Kolářová,¹
Sami Oikarinen,² Juha-Pekka Pursiheimo,³
Jorma Ilonen,⁴ Olli Simell,⁵
Mikael Knip,^{6,7,8,9} Riitta Veijola,¹⁰
Heikki Hyöty,^{2,11} and Ondrej Cinek¹

OBJECTIVE

This study used next-generation sequencing (NGS) technologies to characterize the gut virome at the onset of islet autoimmunity.

RESEARCH DESIGN AND METHODS

We conducted a case-control study nested within the Finnish Diabetes Prediction and Prevention cohort. The stool virome in 19 case children, who turned islet autoantibody positive before the age of 2 years and later developed clinical type 1 diabetes, and 19 tightly matched control subjects was analyzed using NGS performed from stool samples collected 3, 6, and 9 months before the onset of islet autoimmunity. Human virus findings were verified using real-time PCR.

RESULTS

One or more human viruses were present in 10.4% and bacteriophages in 54% of the samples. The virome composition showed no association with islet autoimmunity. NGS was less sensitive and specific than real-time PCR.

CONCLUSIONS

The present data suggest no dramatic changes in the gut virome shortly before the emergence of islet autoimmunity and emphasize the need of verification of mass sequencing results when viral exposure is assessed in association studies.

The advent of massive parallel sequencing, also called next-generation sequencing (NGS), has revolutionized many fields of science, including molecular virology. However, the application of NGS in virology is challenging because there is no sequence signature common to all viruses. Certain virus infections have been associated with the initiation of the process that damages β -cells and leads to type 1 diabetes, but the overall role of viruses has remained unresolved (1). Here, we used NGS to characterize the whole gut virome in children shortly before the appearance of islet autoimmunity.

RESEARCH DESIGN AND METHODS

Subjects and Stool Samples

Our case-control study was nested within the Finnish Diabetes Prediction and Prevention (DIPP) birth cohort that recruits families with a newborn infant carrying HLA-conferred susceptibility to type 1 diabetes (2). The current case group comprised 19 children (11 girls) with early appearance of islet autoimmunity. We assumed that subjects with early autoimmunity might have encountered a powerful environmental stimulus within the time covered by frequent stool sample collection. All case subjects developed two or more islet autoantibodies before the age of

¹Department of Pediatrics, University Hospital Motol, and 2nd Faculty of Medicine, Charles University in Prague, Prague, Czech Republic

²Department of Virology, School of Medicine, University of Tampere, Tampere, Finland

³Department of Medical Biochemistry and Genetics, University of Turku, Turku, Finland

⁴Department of Immunogenetics, University of Turku, Turku, Finland

⁵Department of Pediatrics, Turku University Central Hospital, Turku, Finland

⁶Children's Hospital, University of Helsinki, and Helsinki University Central Hospital, Helsinki, Finland

⁷Diabetes and Obesity Research Program, University of Helsinki, Helsinki, Finland

⁸Department of Pediatrics, Tampere University Hospital, Tampere, Finland

⁹Folkhälsan Research Center, Helsinki, Finland

¹⁰Department of Pediatrics, Oulu University Hospital, and University of Oulu, Oulu, Finland

¹¹Fimlab Laboratories, Pirkanmaa Hospital District, Tampere, Finland

Corresponding author: Ondrej Cinek, ondrej.cinek@lfmotol.cuni.cz.

Received 21 October 2014 and accepted 6 January 2015.

This article contains Supplementary Data online at <http://care.diabetesjournals.org/lookup/suppl/doi:10.2337/dc14-2490/-/DC1>.

© 2015 by the American Diabetes Association. Readers may use this article as long as the work is properly cited, the use is educational and not for profit, and the work is not altered.

2 years, and all eventually progressed to clinical type 1 diabetes. One nondiabetic autoantibody-negative control child was selected for each case, matched for date and place of birth, for the HLA-conferred risk level of type 1 diabetes and sex. The project was approved by the Pirkanmaa Hospital District Ethical Committee, Tampere, Finland (approval no. 97193M), and parents or guardians expressed their consent in writing.

Where available, three pairs of stool samples were analyzed for each case-control matching pair. Samples were taken 3, 6, and 9 months before the date of the first islet autoantibody positivity in the case patient. A total of 96 samples was collected through the years 1998 to 2007 and stored at -80°C until analyzed.

Sequencing and Analysis Pipeline

The laboratory protocol is detailed in the Supplementary Data. Briefly, the virus fraction of the stool was mechanically enriched by a series of filtering and centrifugation steps. The contaminating ribosomal fraction was reduced by partial digestion of unprotected nucleic acids by a nuclease cocktail. A sequence-independent single-primer amplification of all nucleic acids followed, as adapted from previous studies (3,4). Each sample was processed in duplicate (i.e., with two different tag sequences). The preamplified libraries underwent size selection and purification and were further processed using the Nextera XT DNA Sample Preparation Kit (Illumina, San Diego, CA) in batches of 24 samples for each multiplexed sequencing run. The libraries were sequenced on an MiSeq instrument (Illumina) using the 2×250 -cycles sequencing kit.

The resulting reads were trimmed and filtered using the Galaxy suite (5). Three lines of downstream analyses were used, each yielding a specific type of information:

- A. The overall virus content was characterized at the genus level by querying de novo contigs and singleton sequences against a database of full-length virus sequences deposited in Genbank as of February 2014.
- B. Genotypes were identified for the most prevalent human viruses by mapping the reads to collections of VP1 sequences.

- C. Overall taxonomic identification was performed in a random selection of 100,000 raw NGS reads from every sample, using the MEGAN5 program (6).

Specific Quantitative PCR

The human virus genera indicated by NGS were systematically retested using virus-specific primers and probes in one-tube real-time RT-PCR assays (or DNA real-time PCR) as described (7–10). The testing was performed from the nucleic acid extracts, which served as the starting material for the NGS, and from the final libraries before pooling and sequencing.

The agreement between the quantitative PCR as a gold standard and the NGS was suboptimal. The current NGS protocol mostly failed to detect samples whose quantity was low in virus-specific PCR (PCR threshold cycle >30). The NGS on the MiSeq platform was also prone to producing sequencing artifacts caused by a signal bleed-through originating in samples with very high virus quantities. The observations led us to establishing a positivity threshold at 50 reads per 100,000 (i.e., 0.05% of the read count).

Statistical Analysis

The association of virus positivity with islet autoimmunity was tested at the genus level by applying conditional logistic regression models. Analyses were performed in Stata 12.1 software (StataCorp LP, College Station, TX).

RESULTS

Human Viruses

With the positivity threshold set at 50 of 100,000 NGS reads, the mass sequencing identified 10 samples (10.4%) positive for at least one human virus: parechovirus (five samples [5.2%], all genotype 1), bocavirus (three samples [3.1%], two of genotype 2 and one genotype 1), anellovirus (two samples [2.1%], Torque teno mini virus 5 and an unclassified Torque teno mini virus), enterovirus (one sample [1.0%], echovirus 25), and sapovirus (one sample [1.0%], genogroup GII). All instances of human virus positivity were limited to one follow-up sample, indicating relatively short excretion of these viruses to the stools. Although not reaching our arbitrary positivity threshold, we noted also the presence of sequences belonging to *Norovirus*, *Rotavirus*,

Papillomavirus, and *Rhinovirus*, each in a single sample only. The agreement of NGS to virus-specific PCR was incomplete (Fig. 1). Notably, PCR was more sensitive and specific in samples with low virus loads.

No association with islet autoimmunity was observed for any of the observed human viruses, regardless of whether NGS or specific (RT-)PCR results were taken as predictors (all nominal P values > 0.15).

Other Sequence Content

The reads belonged mostly to bacteria (14%) and eukaryotes (10%); specifically, 2% belonged to human sequences and 0.3% to *Viridiplantae*. Viruses made up to 2% of all identified sequences, with major predominance of bacteriophages identifiable in 52 of 96 samples (54.2%). No significant association with islet autoimmunity was found in an unplanned exploratory association analysis of the eight most frequently observed bacteriophage groups. On average, 74% of the reads could not be matched to any known sequence deposited in public databases.

CONCLUSIONS

Technical Aspects of NGS

By a systematic evaluation of NGS results against virus-specific PCR as a gold standard, we demonstrated that the present state of the NGS protocols does not allow complete characterization of all viruses present in the sample, leading to false-negative findings when virus titers are low (amplification threshold cycle of 30 and later; Fig. 1). Thus, despite the clear advantage of NGS as an exploratory tool, the limitations in its sensitivity should be taken into account when virome association studies are conducted.

Furthermore, several samples were negative in virus-specific PCR but still were assigned several reads matching that virus in NGS (open circles along the vertical axis in Fig. 1). Because cross-contamination between samples during the earlier steps of library preparation was excluded by PCR tests from sequencing libraries, we hypothesize that the reason is a signal bleed-through in the NGS sequencer: all false-positive samples shared one of their sequencing indices with at least one sample highly positive for the respective virus (data not shown). It has indeed been reported

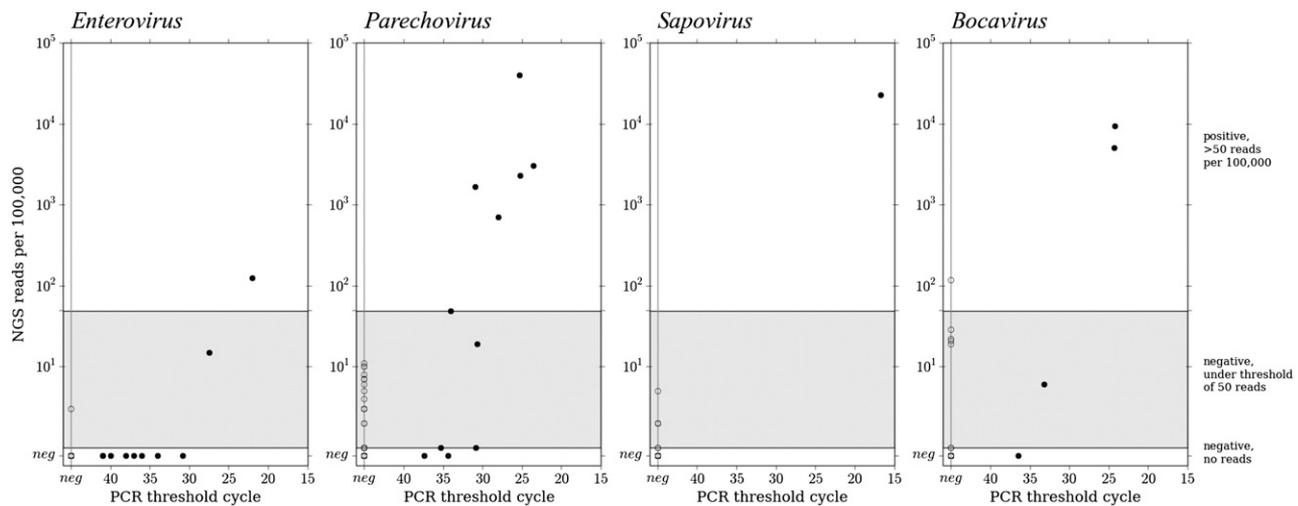


Figure 1—Results from NGS compared with specific real-time PCR. The horizontal axis of the graphs shows threshold cycles of the real-time PCR in a reversed scale: increasing quantities of the virus correspond to lower threshold cycles. The vertical axis plots the number of reads assigned to the virus genus in the NGS analysis, normalized per 100,000 reads. The area below our threshold of NGS positivity (50 reads per 100,000) is shaded. Values below the gray zone are entirely negative (no reads found). The gray zone marks 1–49 normalized reads (deemed as negative in our analysis), and samples higher than 50 reads were deemed positive. Individual samples are plotted as ● if positive by real-time PCR, and ○ if negative by real-time PCR.

that the Illumina platform may misidentify a certain proportion of reads, assigning them to a wrong sample due to signal bleed-through between samples (11). This may lead to false-positive findings in NGS studies that use sample multiplexing.

Viruses Detected in Stool Samples From the Study Participants, and the Lack of Association With Islet Autoimmunity

The stool virome did not significantly differ between children who developed islet autoimmunity and closely matched control subjects. The most frequently detected human viruses belonged to the *Picornaviridae* family: *Human parechovirus* (12,13) and *Human enterovirus*, which is the most frequently studied viral candidate trigger of islet autoimmunity and type 1 diabetes (1). Nonhuman viruses merit further analyses, because especially the phage population was abundant and notably more stable over time than human viruses.

Limitations of the Study and Previous Works on the Human Virome in Type 1 Diabetes Etiology

The limitations of our study lie in the size of the data set (sample count and sample intervals). We lack power to detect subtle changes in the frequencies of specific virus genotypes, and the sample frequency is too sparse to reliably detect all short infections. Moreover, a

significant role may be played by viruses replicating outside the gut or by new yet unrecognized viruses hiding among the unidentified reads.

To our best knowledge, this is the first study of the fecal virome in relation to diabetes. The Environmental Determinants of Diabetes in the Young (TEDDY) study used NGS for detection of the viruses in plasma from 14 children who rapidly progressed to type 1 diabetes and from 14 control children (14). The type of molecular evidence for virus positivity was not reported, and confirmatory tests were not done, making it difficult to compare their results from plasma with ours from stool.

In conclusion, for association studies, NGS can effectively provide a comprehensive picture of prevailing human viruses and bacteriophages in a given virome and yield important data on yet unidentified organisms. However, we presently believe that NGS should be complemented with virus-specific PCR as the gold standard of molecular detection of viruses, at least until more effective NGS protocols are available.

Acknowledgments. Prof. Eric Delwart and Dr. Linlin Li, of the Blood Research Institute, San Francisco, CA, and Dr. Eric F. Donaldson, of the Department of Epidemiology, University of North Carolina, Chapel Hill, NC, are sincerely thanked for their kind explanation of several steps in their published protocols and for

providing the sequences of their tagged primers. Nikola Ptáková, Department of Biology and Medical Genetics, University Hospital Motol, and Tanja Kuusela, Department of Virology, University of Tampere, are gratefully acknowledged for expert technical assistance.

Funding. Support was received by the Ministry of Health of the Czech Republic (IGA MZ 11465-5), the Academy of Finland (to H.H.), Project for the Conceptual Development 00064203 and EU CZ.2.16/3.1.00/24022OPPK (to University Hospital Motol, Prague). Further support was received from the JDRF, the Academy of Finland (grants 129448, 255770, and 132362 to H.H.), Centre of Excellence in Molecular Systems Immunology and Physiology Research 2012-2017 and 250114 (to M.K. and O.S.), Sigrid Jusélius Foundation, the Competitive Research Funding of University Hospitals in Oulu, Tampere, and Turku, and the European Union's Seventh Framework Programme (the PEVNET Study Group grant agreement No. 261441). The recruitment of study subjects and collection of stool samples in the DIPP study was funded by grants from JDRF and the Academy of Finland.

Duality of Interest. H.H. and M.K. are minor shareholders (<5%) in Vactech Ltd., a small biotech company developing picornavirus vaccines. No other potential conflicts of interest relevant to this article were reported.

Author Contribution. L.K. prepared the libraries, performed sequencing, analyzed data, and wrote the manuscript. K.K. helped develop the analysis pipeline and analyzed data. S.O. and H.H. conceived the study, selected the subjects and their samples, participated in the laboratory and data analysis and interpretation, and revised the manuscript. J.-P.P. performed sequencing during the protocol development. J.I., O.S., M.K., and R.V. participated in recruiting study subjects and in writing the manuscript. O.C. supervised the molecular analysis protocol, wrote parts of the data analysis pipeline, analyzed data, and helped in writing the

manuscript. O.C. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Prior Presentation. Parts of this study were presented in abstract form at the 40th Annual Conference of the International Society for Pediatric and Adolescent Diabetes, 3–6 September 2014, Toronto, Canada; the 50th Annual Meeting of the European Association for the Study of Diabetes, 15–19 September 2014, Vienna, Austria; and the 17th Annual Meeting of the European Society for Clinical Virology, 28 September–1 October 2014, Prague, the Czech Republic; and were published as abstracts (Abstract O29. *Pediatric Diabetes* 2014;15(Suppl.)29; and Abstract OP29.169. *Diabetologia* 2014;57 (Suppl. 1)S78).

References

1. Tauriainen S, Oikarinen S, Oikarinen M, Hyöty H. Enteroviruses in the pathogenesis of type 1 diabetes. *Semin Immunopathol* 2011;33: 45–55
2. Ilonen J, Hammas A, Laine AP, et al. Patterns of β -cell autoantibody appearance and genetic associations during the first years of life. *Diabetes* 2013;62:3636–3640
3. Donaldson EF, Haskew AN, Gates JE, Huynh J, Moore CJ, Frieman MB. Metagenomic analysis of the viromes of three North American bat species: viral diversity among different bat species that share a common habitat. *J Virol* 2010;84: 13004–13018
4. Victoria JG, Kapoor A, Li L, et al. Metagenomic analyses of viruses in stool samples from children with acute flaccid paralysis. *J Virol* 2009;83:4642–4651
5. Goecks J, Nekrutenko A, Taylor J; Galaxy Team. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* 2010;11:R86
6. Huson DH, Mitra S, Ruscheweyh HJ, Weber N, Schuster SC. Integrative analysis of environmental sequences using MEGAN4. *Genome Res* 2011;21:1552–1560
7. Honkanen H, Oikarinen S, Pakkanen O, et al. Human enterovirus 71 strains in the background population and in hospital patients in Finland. *J Clin Virol* 2013;56:348–353
8. Corless CE, Guiver M, Borrow R, et al. Development and evaluation of a ‘real-time’ RT-PCR for the detection of enterovirus and parechovirus RNA in CSF and throat swab samples. *J Med Virol* 2002;67:555–562
9. Kantola K, Sadeghi M, Antikainen J, et al. Real-time quantitative PCR detection of four human bocaviruses. *J Clin Microbiol* 2010;48: 4044–4050
10. Pang XL, Preiksaitis JK, Lee BE. Enhanced enteric virus detection in sporadic gastroenteritis using a multi-target real-time PCR panel: a one-year study. *J Med Virol* 2014;86:1594–1601
11. Kircher M, Sawyer S, Meyer M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res* 2012;40:e3
12. Kolehmainen P, Oikarinen S, Koskiniemi M, et al. Human parechoviruses are frequently detected in stool of healthy Finnish children. *J Clin Virol* 2012;54:156–161
13. Tapia G, Cinek O, Witsø E, et al. Longitudinal observation of parechovirus in stool samples from Norwegian infants. *J Med Virol* 2008;80: 1835–1842
14. Lee HS, Briese T, Winkler C, et al.; TEDDY study group. Next-generation sequencing for viruses in children with rapid-onset type 1 diabetes. *Diabetologia* 2013;56:1705–1711