Charles University, Faculty of Science
Department of Cell Biology
Branch of Study: Immunology

Kateřina Chudá

Role of enterovirus and adenovirus infection in the pathogenesis of celiac disease

Role enterovirových a adenovirových infekcí v rozvoji celiakie

Diploma thesis

Supervisor: prof. MUDr. Ondřej Cinek, Ph.D.
Laboratory of Molecular Genetics
Department of Pediatrics and Institute of Medical Microbiology, Charles University
2nd Faculty of Medicine and Motol University Hospital

Prague 2019

Prohlášení

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze dne 10. 8. 2019

…….

Kateřina Chudá

Poděkování

Na tomto místě bych ráda poděkovala zejména účastníkům studie, sledovaným dětem a také jejich rodičům, za jejich ochotu ke spolupráci. Dále patří poděkování všem, kteří se starali o chod studie, zajišťovali sběr vzorků, vedli dokumentaci a komunikovali s dětmi a jejich rodinami.

Velké díky patří našim kolegům z Norského institutu veřejného zdraví, díky nimž jsem se mohla na předkládaném výzkumném tématu podílet. Zejména děkuji German Tapia, PhD., který naše výsledky statisticky zpracoval a navíc se mi ochotně věnoval a naučil mě s nashromážděnými daty pracovat v programu Stata během naší návštěvy Osla.

Osobně děkuji především svým kolegům z laboratoře Molekulární genetiky 2. lékařské fakulty za to, že mě vzali mezi sebe a věnovali mi svůj čas a péči. Především si vážím laskavého prostředí, které mi naše pracovní skupinka dokázala vytvořit. Děkuji tedy svému školiteli, prof. MUDr. Ondřeji Cinkovi, PhD., a milému kolektivu laboratoře.

Též děkuji svým blízkým, rodině a kamarádům, kteří mě během studia motivovali a podporovali.

Abstract

Celiac disease, a chronic immune-mediated disorder of the small intestine, manifests in a fraction of individuals with genetic predispositions consuming gluten. Environmental factors play an essential role in its triggering. The environmental stimuli may include dietary factors, infections etc. Identification of specific triggers could help in celiac disease prevention.

Our research project focused on common intestinal infections in infancy. We investigated adenoviruses and enteroviruses in stool specimens of children carrying a high-risk HLA genotype for celiac disease. We aimed to determine whether these infections are associated with early markers of celiac autoimmunity, and to identify virus genotypes. To distinguish multiple infections, massive parallel amplicon sequencing was utilized.

During 2001-2007, nearly 50.000 Norwegian newborns were screened within the MIDIA study for the presence of the HLA DR3-DQ2/DR4-DQ8 genotype, which is known to significantly increase the risk of celiac disease. The risk genotype was identified in 912 babies. Up to three years of children´s age, monthly stool specimens were collected and archived. Blood sampling was done every three months up to the age of a year, and then annually. Periodical questionnaires on children´s thrive were collected. During 2014-2016, the cohort was offered participation in a celiac disease study. A total of 220 individuals underwent a celiac disease screening, which led to the diagnosis in 27 cases. Each of patients was matched to two most similar healthy controls from the same cohort. A total of 2.161 stool samples were tested by real-time quantitative PCR for adenovirus and enterovirus and viruses were genotyped. The data were statistically analyzed by mixed effects logistic regression which took into account most of potential confounders.

The development of celiac disease antibodies was preceded by more enterovirus infections than was observed in matched controls. Also, prolonged infections and specimens of high viral load were significantly more frequent before the celiac disease antibody development. Both common enterovirus species, *Enterovirus A* and *Enterovirus B,* were significantly associated with celiac disease. Notably, there was a significant synergistic interaction between the enterovirus effect and the introduction of gluten into infant diet. Adenovirus, the second most frequent enteric virus, was not associated with celiac autoimmunity. The findings may have implications for designing future prevention strategies.

Key words: enterovirus, adenovirus, celiac disease, longitudinal cohort study, gastrointestinal infections, nested case-control study

Abstrakt

Celiakie, chronické imunitně zprostředkované onemocnění tenkého střeva, postihuje část pacientů s genetickými predispozicemi konzumujících lepek. Pro iniciaci nemoci jsou klíčové faktory prostředí. Mohou jimi být například změny ve stravování nebo infekce. Identifikace konkrétních spouštěčů onemocnění by mohla hrát roli v prevenci nemoci.

Naše práce se zaměřila na střevní virové infekce běžné v dětství. Vyšetřovali jsme adenoviry a enteroviry ve vzorcích stolice dětí nesoucích pro celiakii vysoce rizikový HLA genotyp. Cílem bylo zjistit, zda jsou tyto infekce asociovány s rozvojem časných markerů celiakální autoimunity, a identifikovat konkrétní genotypy virů. Pro rozlišení vícenásobných infekcí byla použita metoda masivně paralelního sekvenování.

Během let 2001-2007 bylo testováno v rámci norské studie MIDIA téměř 50 000 novorozenců na přítomnost genotypu HLA DR3-DQ2/DR4-DQ8, který významně zvyšuje riziko rozvoje celiakie. Nositeli tohoto genotypu bylo 912 dětí. Každý měsíc do věku tří let byly těmto dětem odebírány vzorky stolice a ukládány v biobance. Krev byla odebírána do jednoho roku věku dítěte každé tři měsíce a poté jednou ročně v kombinaci se sběrem informací o prospívání dětí prostřednictvím dotazníků. Během let 2014-2016 byly děti přizvány k účasti ve studii celiakie. Celkem 220 podstoupilo screening celiakie, což vedlo k stanovení diagnózy ve 27 případech. Ke každému pacientovi byly ze stejné kohorty přiřazeny dvě co nejpodobnější zdravé kontroly. Celkem 2 161 vzorků stolice těchto dětí bylo testováno kvantitativní real-time PCR metodou na přítomnost adenovirů a enterovirů, které byly následně genotypovány. Data byla statisticky zpracována metodou logistické regrese se smíšenými efekty, což umožnilo zohlednit většinu potenciálních zavádějících faktorů.

Rozvoji protilátek typických pro celiakii předcházelo více enterovirových infekcí u případů v porovnání s kontrolami. Také infekce s delším trváním a vzorky o vysoké virové náloži byly signifikantně častější před rozvojem celiakální autoimunity. Oba běžné druhy enteroviru, *Enterovirus A* i *Enterovirus B,* byly signifikantně s celiakií asociovány. Obzvláště pozoruhodná byla signifikantně významná asociace mezi vlivem enterovirové infekce v době zavádění lepku do stravy a rozvojem markerů celiakie. Adenovirus, jakožto druhý nejběžnější střevní virus, nebyl s celiakií asociován. Předložená zjištění mohou být využita při navrhování budoucích preventivních opatření.

Klíčová slova: enterovirus, adenovirus, celiakie, longitudinální kohortová studie, střevní infekce, vnořená studie případů a kontrol

Abbreviations

APC	antigen presenting cell
aOR	adjusted odds ratio
CDA	celiac disease antibodies
CV	coxackievirus
HAdV	human adenovirus
HEV	human enterovirus
CD	celiac disease
CI	confidence interval
CVA	Coxsackie virus A
CVB	Coxsackie virus B
ESPGHAN	The European Society for Pediatric Gastroenterology, Hepatology and Nutrition
GFD	gluten-free diet
HLA	human leukocyte antigen
NGS	new generation sequencing
VP1	enterovirus capsid protein VP1
RT-PCR	reverse transcription polymerase chain reaction
TG2	tissue transglutaminase 2
TGA	tissue transglutaminase antibodies
Th1	T helper 1 cells
WNV	West Nile Virus

Table of contents

1 INTRODUCTION 1

1.1 Aims 1

1.2 Reasons for studying environmental confounders of CD 2

2 CELIAC DISEASE ETIOPATHOLOGY 3

2.1 Genetics 5

2.2 Gluten 7

2.3 Environment 9

3 VIRUSES IN THE PATHOGENESIS OF CELIAC DISEASE 11

3.1 Adenovirus 12
3.1.1 Evidence of adenovirus infections in the pathogenesis of celiac disease 14

3.2 Enterovirus 15
3.2.1 Evidence of enterovirus infections in the pathogenesis of celiac disease 17

3.3 Implications from the evidence of viruses in the context of CD 18

4 MATERIALS AND METHODS 20

4.1 The MIDIA cohort 20

4.2 Implementation of the celiac disease sub-study 22
4.2.1 Celiac disease diagnosis 23

4.3 Processing of stool samples 24
4.3.1 Extraction of nucleic acids 25
4.3.2 Detection of viruses by PCR methods 27
4.3.3 Virus genotyping 28

4.4 Statistical analysis 36

4.5 Materials 38
4.5.1 Chemicals 38
4.5.2 Laboratory equipments 38

6 RESULTS 40

6.1 Enterovirus 40

6.2 Adenovirus 47

7 DISCUSSION 52

7.1 Enterovirus and adenovirus in the pathogenesis of celiac disease 52
7.1.1 Strengths of the study 53
7.1.2 Limitations of the study 54

7.2 Interpretation 55

7.3 Future prospects 57

8 CONCLUSION 59

9 REFERENCES 60

10 APPENDIX 66

10.1 Paper I 66
10.1.1 Supplementary materials of paper I 75

10.2 Paper II 91
10.2.1 Supplementary materials of paper II 101

1 INTRODUCTION

Celiac disease (CD) is an incurable chronic inflammatory disorder of the small intestine. The driver of the pathogenic inflammation is gluten, which is present in a common diet. Currently, life-long gluten-free diet is the only solution. The disease develops most often in children, although can occur also in adulthood. Causes of the development of CD are multifactorial. Both genetic and environmental components are required for the disease phenotype. CD develops almost exclusively in subjects with HLA-DQ2 and/or HLA-DQ8 molecules, who eat gluten-containing cereals and have likely gone through some environmental stimuli. These environmental triggers could be infections, changes in a diet or other stressors. Suggested candidates are mainly those agents which come into contact with the intestinal mucosa, the place of CD pathogenesis. If factors that impact CD risk is uncovered and modified, the prevention of CD might be introduced (Tye-Din et al. 2018). The presented thesis is focused on gastrointestinal adenovirus and enterovirus infections in early childhood.

1.1 Aims

We aimed to test whether the development of celiac disease is preceded by an increased number of enterovirus or adenovirus intestinal infections. Additionally, we aimed to test the effects of symptomatic infections, infections with high viral loads or prolonged virus shedding.

We aimed:

- to develop a reliable high-throughput extraction of nucleic acids and avoid the cross-contamination in a 96-well format,
- to prepare a protocol for virus genotyping using massive parallel amplicon sequencing,
- to test whether adenovirus or enterovirus infections may be associated with the later celiac disease,
- to characterize the nature of infections,

- to determine their serotypes and identify those associated with the CD development.

1.2 Reasons for studying environmental confounders of CD

The incidence of CD is increasing over the last decades (Lohi et al. 2007). The current global prevalence is estimated as 1 % of the whole human population, which means approximately 77.2 million of people worldwide (Singh et al. 2018, https://worldometers.info 3.8.2019).

Its prevalence differs among countries. For instance, CD affects almost 3 % of the Finnish population (Singh et al. 2018). Generally, Scandinavian countries are more affected, whereas some world areas are less. This remarkable disproportion is not possible to explain just by genetic predispositions or by amount of consumed gluten. Why do some genetically predisposed individuals develop CD, whereas the others do not, is not clear. The increasing prevalence is obviously induced by environmental conditions (Abadie et al. 2011).

Although there have been numerous attempts to develop curative therapies, currently the only solution is a lifelong gluten-free diet (GFD), which is difficult to follow and is expensive. Nevertheless, adherence to the diet is enough for the disappearance of symptoms in the vast majority of cases (Alhassan et al. 2019). In addition, CD diagnosis is associated with high morbidity, which could imply socioeconomic disadvantages and relates also to lower work productivity (Crocker et al. 2018). CD also increase the risk of other diseases (Heikkilä et al. 2015). Thus, the environmental determinants need to be elucidated.

2 CELIAC DISEASE ETIOPATHOLOGY

Like many autoimmune diseases, the CD phenotype results from a complex interplay of several factors, genetic and environmental. The known causative factor is gluten, whose consumption leads to a villous atrophy of the small intestine tissue (Sollid et Jabri 2013). The scheme illustrating the disease etiology (***Figure 1***) and immunopathology (***Figure 2***) is shown.

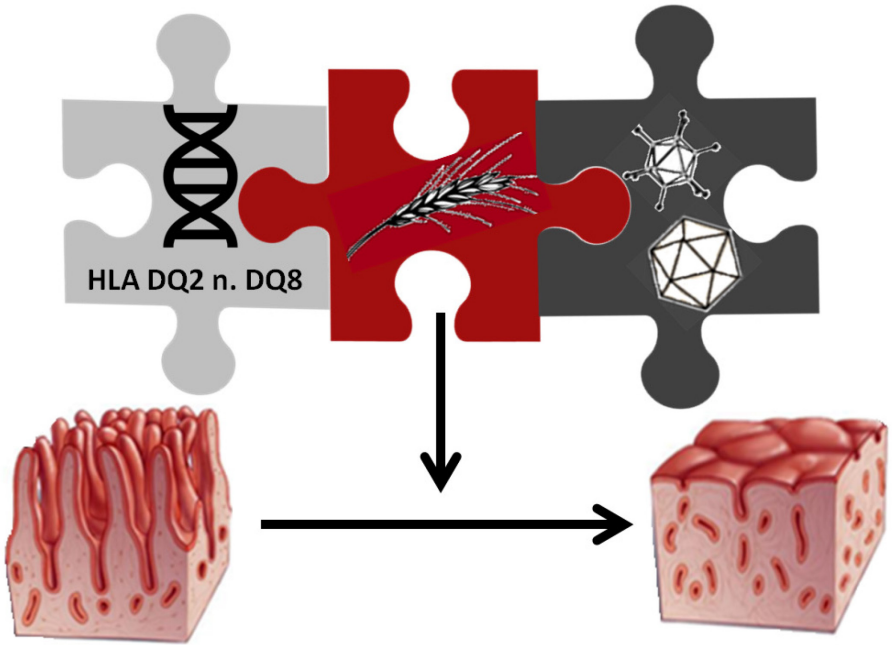


Figure 1. ***Scheme of the celiac disease etiology.*** Celiac disease is an immune-mediated gluten-sensitive enteropathy that develops in genetically predisposed individuals. The overwhelming majority of cases are carriers of HLA allele variants HLA-DQ2 and/or HLA-DQ8. Generally accepted model supposes that CD is a result of an interaction and a specific combination of genetic and environmental factors conditioned by the consumption of gluten. In the small intestine, the pathology is manifested by lymphocyte infiltration and villous atrophy, which can lead to the malabsorption of nutrients and chronic diarrhea causing malnutrition, failure to thrive, anemia, osteoporosis etc. The figure is inspired by the illustration of the dissertation defence presentation of Ch. R. Kahrs; MD, PhD. Drawn by using the image from www.mayoclinic.org, 2. 11. 2018.

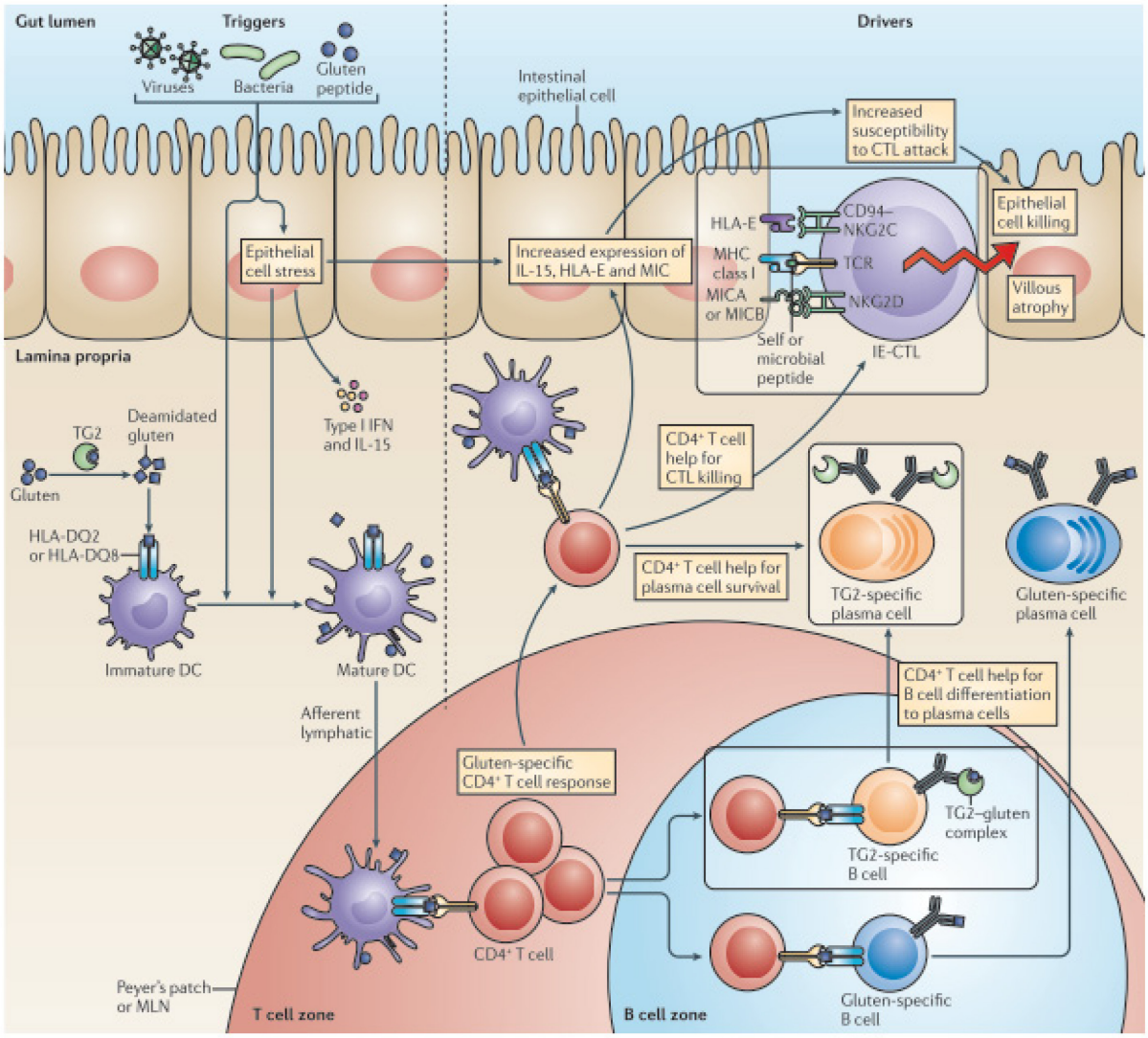


***Figure 2*. Model of celiac disease immunopathology**. Immunopathologic processes are driven by the consumption of gluten. Bacteria, viruses or even gluten itself might figure as triggers and promote the activation of antigen-presenting cells (APCs). Gluten is deamidated by an enzyme called tissue transglutaminase 2 (TG2). Impaired epithelial cells can serve as a source of interferon type I (IFN I) and interleukin 15 (IL-15), which are requires for the maturation of APCs. Mature APCs bind deamidated gluten peptides by their HLA receptors and migrate to lymph nodes (MLN = mesenteric lymph nodes) or Peyer´s patches and present gliadin peptides to CD4+ T cells. In addition, B cells can present the complex of TG2 and gluten on their HLA-DQ2 or HLA-DQ8. T cells encourage gluten-specific and TG2-specific B cells to differentiate into plasma cells producing anti-gluten and anti-TG2 specific antibodies. Activated T cells also confer to the expression of IL-15 and non-classical HLA class I molecules (HLA-E, MIC) in epithelial cells. Consequently, intraepithelial cytotoxic T lymphocytes (IE-CTLs) can cause destruction of stressed epithelial cells. From Sollid et Jabri, 2013.

2.1 Genetics

The higher incidence of CD in monozygotic compared to dizygotic twins and the higher incidence of CD in first-degree relatives compared to the CD incidence in general population show the importance of genetic components influencing the CD susceptibility (Tye-Din et al. 2018).

The main genetic predisposition for CD is determined by carrying specific variants of HLA-DQ alleles. The presence of HLA-DQ2 and/or HLA-DQ8 molecules is the strongest genetic risk factor associated with CD. Almost all CD cases carry either or both of these HLA variants (Megiorni et Pizzuti 2012). These alleles are very common, their joint prevalence in general population is around 40 %, but only 1 % of all people develop CD (Tighe et Beattie 2017) as shown in ***Figure 3***. Also HLA-DQ7 can confer slightly increased risk of CD (Tinto et al. 2015).

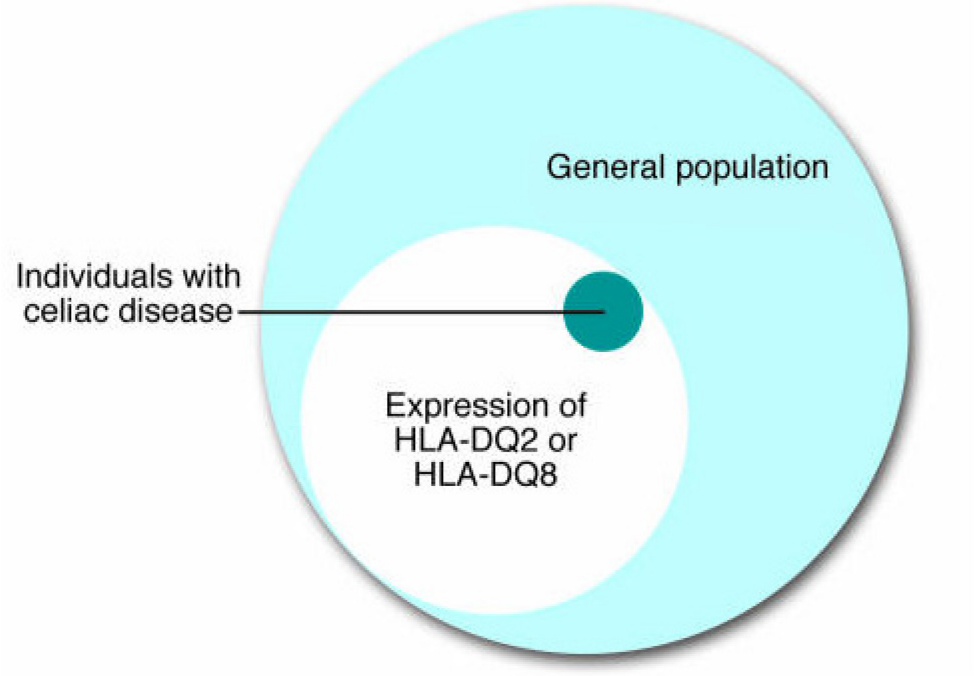


Figure 3. ***Illustration of the HLA-DQ2 and HLA-DQ8 distribution in the general population and among CD patients.*** Almost all CD cases carry these HLA variants. From Kagnoff, 2007.

Genes *HLA-DQA1* and *HLA-DQB1* encode the *alpha* and *beta* chain of heterodimer proteins HLA-DQ that is expressed on the surface of antigen presenting cells (APCs). A heterozygous

individual can produce two *alpha* and two *beta* chains, which means four isoforms can theoretically be formed on the APC surface (Abadie et al. 2011).

The HLA-DQ variants are in close genetic linkage disequilibrium to HLA-DR variants (as shown in ***Figure 4***), which means that their alleles are mainly inherited together as a haplotype.

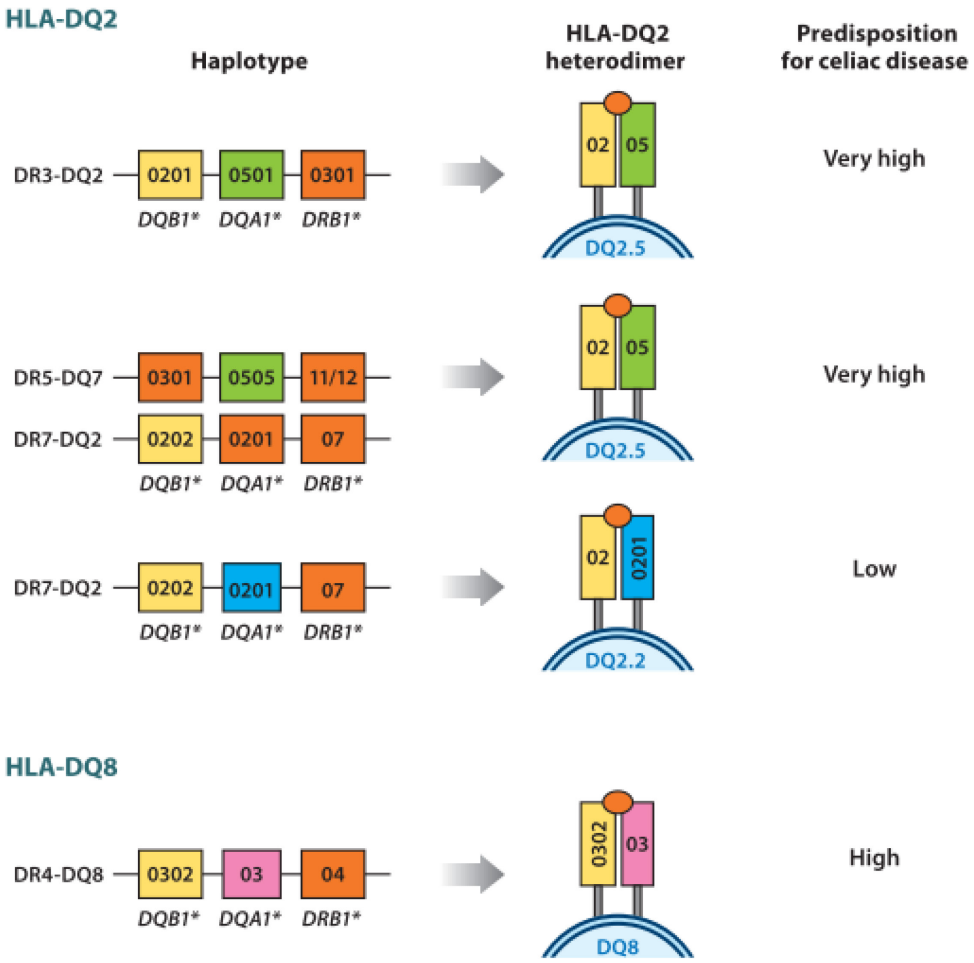


Figure 4. Associations of CD with HLA haplotypes. DQ antigen is made up of two different subunits forming an αβ-heterodimer protein on the APC surface. The DQ α subunit is encoded by the *HLA-DQA1* gene and the DQ β subunit is encoded by the *HLA-DQB1* gene. The most associated with CD are haplotypes resulting in the expression of HLA-DQ2.5 antigen. Patients can also express HLA-DQ2.5 heterodimer, which is encoded by DR3-DQ2 in *cis* form or by DR5-DQ7 combined with DR7-DQ2 in *trans* form in a heterozygote. The surface molecule DQ2.2, which is encoded by the haplotype DR7-DQ2 is only mildly associated with CD. Another CD-related HLA-DQ variant is DQ8 resulting from the haplotype DR4-DQ8. From Abadie et al. 2011.

HLA variants are estimated to be responsible for 40-50 % of estimated genetic risk (Sollid et Lie 2005). Genome wide association studies (GWAS) revealed several non-HLA genes that contribute to the CD susceptibility. About 40 non-MHC loci related to CD are known. They are responsible for approximately 14 % of the genetic susceptibility. These genes mainly encode proteins with functions in the immune system (Trynka et al. 2012). Noteworthy, variants of viral response genes are related to CD (Dubois et al. 2010, Kumar et al. 2015). Lastly, large mutations in genes, especially in genes regulating immune responses, could confer to the CD susceptibility (Trynka et al. 2012). Also, genes coding proteins of tight junction between enterocytes are considered (Visser et al. 2009).

2.2 Gluten

The consumption of gluten-containing products is crucial for the manifestation of CD. Gluten consists of alcohol-insoluble components named glutenins and alcohol-soluble gliadins as is illustrated in ***Figure 5*** (Abadie et al. 2011).

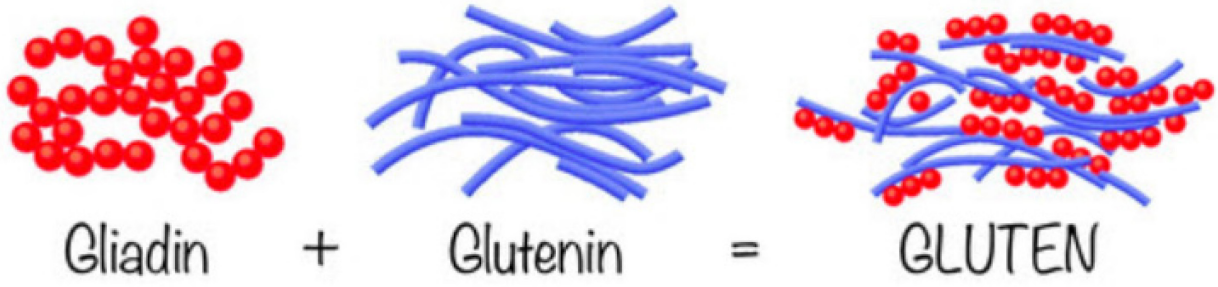
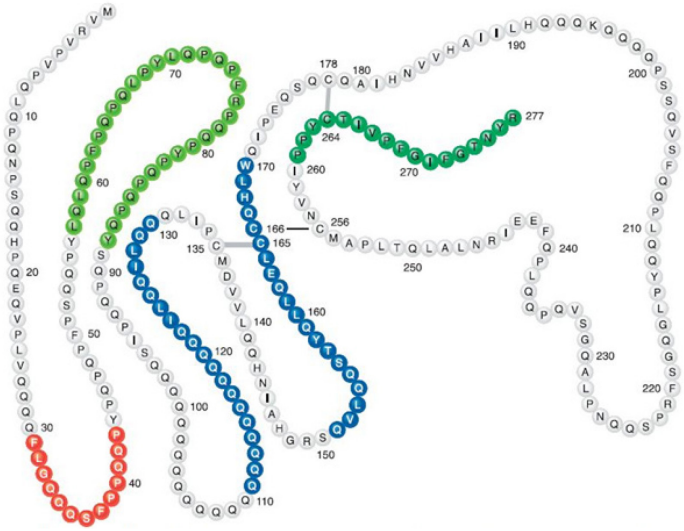


Figure 5. ***Composition of gluten.*** Gluten consists of glutenin and gliadin, which is rich in proline and glutamine. From https://cwsimons.com/gluten-chemistry-and-functionality/, 1 9.7.2019.

Some gliadin peptides are partly resistant to intestinal digestion (hydrolysis by gastric, pancreatic or intestinal proteases) due to their rich content of prolines. Undigested gliadin fragments can figure as drivers of immune responses under certain conditions. Gliadin, rich in glutamine and proline (its structure is shown in ***Figure 6***), is a good substate for tissue transglutaminase. Tissue transglutaminase 2 (TG2) is a multifunctional enzyme. In CD, TG2

changes glutamine to glutamate through deamidation (***Figure 7***). The immunogenicity of gliadin is higher after this reaction, deamidated gliadin fits better than non-deamidated gliadin in the grooves of the DQ2 and DQ8 molecules, which present deamidated gliadin peptides to CD4+ T cells (Mejías et al. 2014).



<u>Figure 6.</u> ***Structure of an alpha-gliadin.*** Parts highlighted in color have specific biologic activities. The evident abundancy of glutamine (Q) and proline (P) and the presence of disulfide bonds can lead to incomplete cleavage by digestive enzymes. From https://glutagen.com/whats-the-deal-with-gluten/, 15.7.2019.

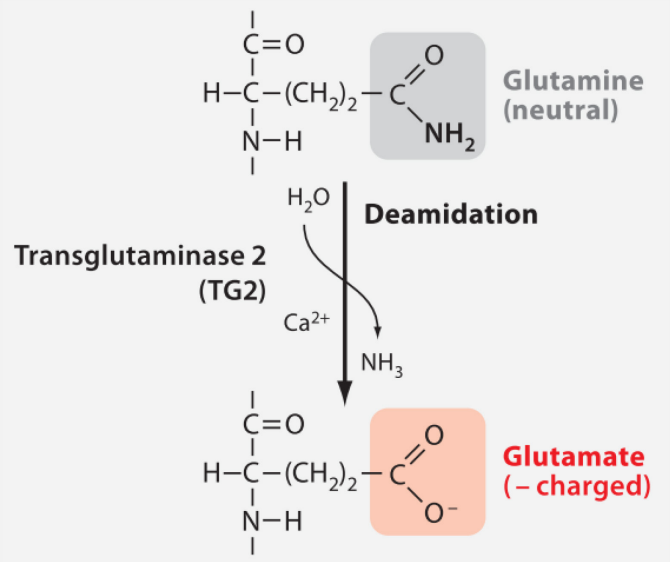


Figure 7. Deamidation of gliadin. The enzyme TG2 (protein glutamyl γ-glutamyltransferase) converts gliadin´s glutamines to glutamates and increases its immunogenicity. From Abadie et al. 2011.

2.3 Environment

Genetic background is necessary but not sufficient for CD manifestation. The CD pathology is induced and influenced by environmental conditions. As mentioned above, the prevalence of CD differs among countries. It appears to be relatively rare in sub-Saharan Africa and southeast Asia, whereas the highest prevalence of CD is observed in Scandinavian countries (Singh et al. 2018). The observed distribution of CD cannot be explained only by genetic predispositions or by the amount of wheat product intake. A good illustration is the significantly different prevalence in neighboring Finland and Russian Karelia (Singh et al. 2018). It could be caused by different recommendation guidelines for infant feeding, differences in type of daycare, the type of community in which individuals live (city/village, large multigenerational families, large number of siblings, the presence of farm animals), hygienic standards or intake of industrial prepared food. It is supposed that the specific risk combination of environmental stimuli can also influence the composition of intestinal microbiota, which in turn could precipitate the intestinal tissue for the increased susceptibility of CD development (Cenit et al. 2015, Abadie et al. 2011).

Population studies have suggested that infectious agents can act as triggers of CD (Mårild et al. 2015, Tjernberg et Ludvigsson 2014, Canova et al. 2014). It is generally accepted hypothesis that autoimmunity can be initiated by an infection through a variety of mechanisms. However, the exact mechanism of CD development remains elusive and its studies are difficult due to the lack of proper animal models (Marietta et al. 2012). Suggested mechanisms how infections can trigger CD are summarized by Lerner et al. (Lerner et al. 2017). Infectious agents could trigger autoimmune processes e.g. by:

- the mechanism of molecular mimicry (Dolcino et al. 2013, Kagnoff et al. 1987),
- increased gut permeability (e. g. caused by the damage of tight junction) enabling the deamidation of gluten by TG2 and its presentation to immune cells (Visser et al. 2009, Vorobjova et al. 2017),
- indirectly, the microbiota modification caused by an infection, the microbiota imbalance could induce the intestine tissue damage, e. g. by increased amount of produced proteolytic molecules or generation of immunogenic peptides (Olivares et Sanz 2015, Lerner et al. 2019),
- initiation of the release of danger signals leading to the immunopathogenic reaction resulting in loss of gluten tolerance (Sollid et Jabri 2013, Bouziat et al. 2017, Matzinger 2002),
- induction of inflammatory conditions leading to the activation of gluten-specific CD4+ T cells, which have been inactive until that time point (Sollid et Jabri 2013).

3 VIRUSES IN THE PATHOGENESIS OF CELIAC DISEASE

Previous epidemiological and immunological studies imply that infections, mainly gastrointestinal viruses and bacteria, may be involved in the CD development (Bouziat et al. 2017, Mårild et al. 2015, Lerner et al. 2017). Viral infections have been linked to the initiation of several immune disorders such as systemic lupus erythematosus (Schmitt et al. 2016), multiple sclerosis or type 1 diabetes (T1D) (Ercolini et Miller 2009).

Focusing on viral infections of the intestine, several potential candidates were suggested including e. g. adenovirus, astrovirus, enterovirus, rotavirus or orthoreovirus.

Astrovirus, relatively rare virus that can cause gastroenteritis in children, was detected in a four-year old boy diagnosed with CD. This case report initiated the testing of the astrovirus presence in 328 patients suffering from gastroenteritis. However, nobody was astrovirus-positive (Smits et al. 2010), which documents astrovirus´ low prevalence and thus the difficulty to study its involvement in CD pathogenesis.

As a very common intestinal virus, rotavirus was studied in the context of CD pathogenesis. Stene et al. found out a significant association of rotavirus infections and CD. Almost 2 000 of CD high-risk children were prospectively followed and regularly blood samples were collected, in which TGA and anti-rotavirus antibodies were measured. Children had four times higher risk of later CD autoimmunity, if they had more than two rotavirus infections (Stene et al. 2006). In 2013, the cross-sectional study was performed and showed the significant difference in the presence of anti-rotavirus antibodies in 21/26 CD patients compared to 10/37 controls (Dolcino et al. 2013). Non-significant results were obtained by other large cross-sectional studies comparing rotavirus infections between 118 CD patients and 60 controls (Bouziat et al. 2017) and between 118 CD cases and 214 healthy controls (Ziberna et al. 2016). Interestingly, reduced risk of later celiac disease in children, who were vaccinated against rotavirus, was observed by Kemppainen et al. (Kemppainen et al. 2016).

Concerning unspecified infections, a prospective study performed in Sweden has indicated no association between infections and CD onsets. Unfortunately, infections were only parent-reported, which decreases the reliability and do not catch asymptomatic infections.

Also, infection types and the sites where infections occur, are limited to determine by parent-reporting questionnaires. In addition, if we assume only slight contribution of infections in triggering CD, we need more sensitive tools for evaluating the rate of their contribution. The advantage of the Swedish study is its large cohort comprising data from more than 9 000 children. However, at the time of investigation, only 44 children had CD based on duodenal biopsy (Welander et al. 2019).

In spite of a number of hints coming from epidemiological studies, mechanistic studies researching the infections´ involvement in CD pathogenesis are relatively rare. A recent study conducted by Bouziat (Bouziat et al. 2017) investigated orthoreovirus infections in mouse models. Orthoreovirus is a dsRNA virus belonging to the *Reoviridae* family. In human, orthoreoviruses can cause gastroenteritis or respiratory diseases or can be asymptomatic. Bouziat et al. showed that an avirulent orthoreovirus can induce the immune response to gluten, promote the generation of Th1 cells and participate in intestinal damage (Bouziat et al. 2017). These findings support the hypothesis of virus-induced CD.

Because the diploma thesis is focused on adenovirus and enterovirus infections, these viruses are described in detail below.

3.1 Adenovirus

Adenovirus is a non-enveloped, double stranded DNA virus with an icosahedral capsid. The structure of the virion is shown in ***Figure 8***. The adenovirus genome is linear and monopartite. Adenovirus infections are very frequent, its prevalence is approximately 14 % among young children under three years of age (Kahrs et al. 2019) (the virus distribution during early childhood is depicted in ***Figure 11***). Infections can occur symptomatically or can be clinically silent. Adenovirus infections can predominantly cause respiratory or gastrointestinal illnesses (for instance common cold, diarrhea). Adenoviruses are spread through fecal-oral route, respiratory droplets or urine, depending on virus type. Family *Adenoviridae* consists of five genera, in which genus *Mastadenovirus* includes all human adenoviruses. The hypervariable region of the hexon gene (shown in ***Figure 9***) are used to classify individual adenovirus genotypes. Serotyping has been partly replaced by genotyping.

In humans, there are more than 50 distinct adenoviral serotypes (Human adenovirus 1-57; HAdV-1 to HAdV-57) ordered in seven species (*Human adenovirus A, B, C, D, E, F* and *G*). Certain adenovirus serotypes are predominantly associated with specific pathology. For instance, HAdV-40 and HAdV-41 is related to gastroenteritis in children (https://talk.ictvonline.org/ictv-reports/ictv_9th_report/dsdna-viruses-2011/w/dsdna_viruses/93/adenoviridae, 12.7.2019).

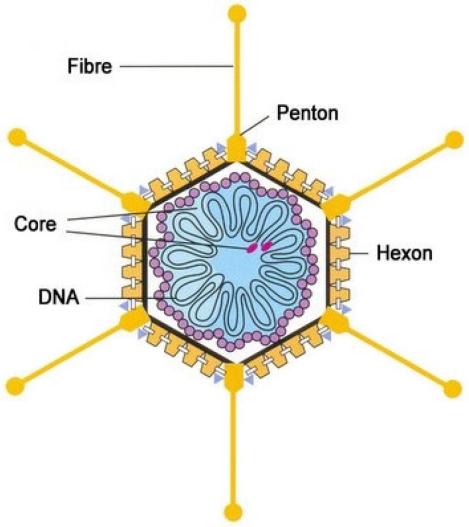


Figure 8. ***The structure of the adenovirus virion.*** The particle has 70–90 nm in diameter. Surface structures (hexon, penton and fiber), which determine serotypes, have the most variable genome regions among adenovirus species.
From https://sites.google.com/site/dsfoster/themolecularbiology, 12.7.2019.

Virus attaches via the fiber to different receptors on the susceptible cells and is internalized. Subsequently, virions are released from endosomes and are transported on microtubules to nuclear pores. Adenovirus replicates in host nucleus. (https://talk.ictvonline.org/ictv-reports/ictv_9th_report/dsdna-viruses-2011/w/dsdna_viruses/93/adenoviridae, 12.7.2019)

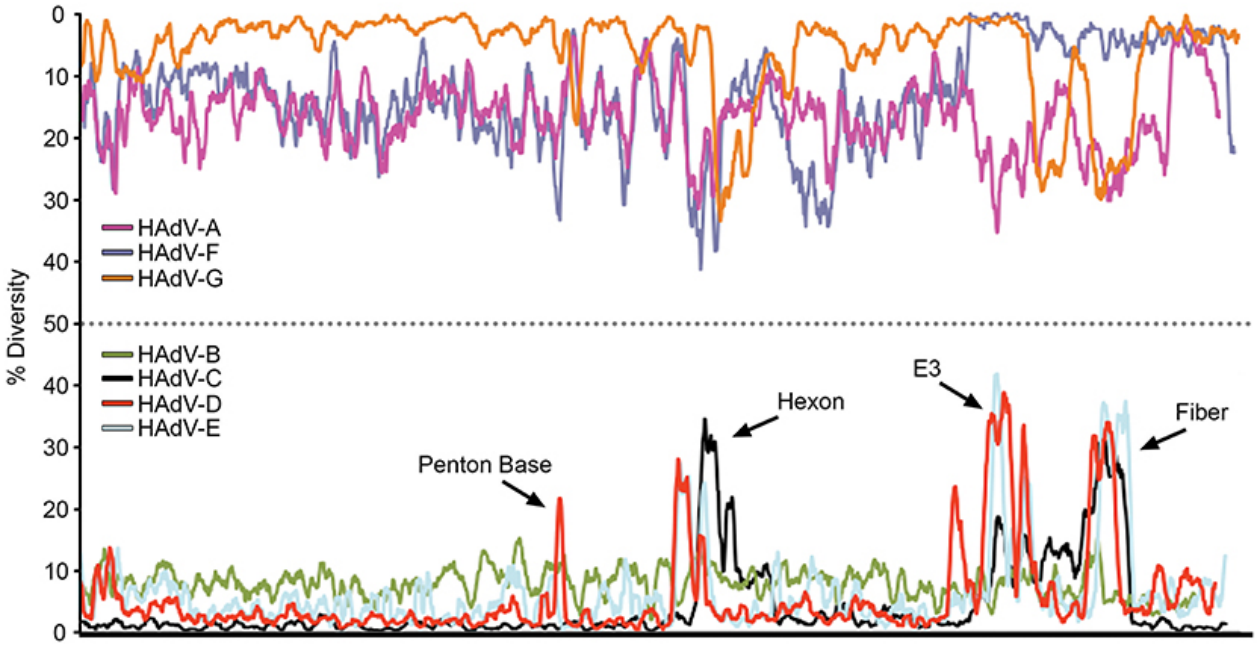


Figure 9. The comparison of nucleotide diversity among genomes of HAdV species. The x-axis depicts the position in genomes. On the y-axis, the % diversity is shown. The adenovirus genomes are relatively conserved apart from regions coding the penton base, E3 proteins, fiber or hexon. The hexon region can be used to genotyping. From Robinson et al. 2013.

3.1.1 Evidence of adenovirus infections in the pathogenesis of celiac disease

Adenovirus was suggested as a trigger of celiac disease in 1984, which means it is the longest studied virus in the CD pathogenesis. In that year, Kagnoff et al. came up with the hypothesis of molecular mimicry and demonstrated that alpha-gliadin protein shares a homologous sequence with a virus protein E1B which figures as a blocker of cell apoptosis through stabilizing the cell protein p53. Authors also showed that antibodies against E1B protein bind to alpha-gliadin (Kagnoff et al. 1984). Three years later, the same authors investigated the prevalence of anti-adenovirus antibodies in 80 CD cases and 135 healthy controls and documented significantly higher prevalence of anti-adenovirus antibodies in CD patients compared to controls. The further study performed by Lahdeaho et al. supported the hypothesis of association between adenovirus infections and CD. Authors measured anti-adenovirus antibodies in 41 cases and 57 controls (Lahdeaho et al. 1993). However, other studies did not confirm the association. By PCR evaluating of adenovirus presence in

duodenal biopsies of 18 CD cases and 24 controls, authors obtained non-significant results (Lawler et al. 1994). Similarly, Mahon et al. tested intestinal biopsies by PCR methods in 27 CD patients and 23 controls and did not detect any differences in the prevalence (Mahon et al. 1991). A larger cross-sectional study including 80 CD patients and 80 healthy controls also was conducted. However, the prevalence of blood anti-HAdV antibodies between two groups statistically did not differ (Tarish et al. 2016).

The main obvious disadvantage of the above mentioned studies is that all of them were cross-sectional, thus we cannot exclude reverse causality. There is the possibility of that CD patients are more susceptible to virus infections than healthy individuals and the measurement just catches this effect. Also, some of the studies are limited by its sample size and they did not take into consideration potential confounding factors increasing e.g. the risk of adenovirus infections (e. g. number of siblings etc.).

3.2 Enterovirus

Enterovirus is a small, spherical ssRNA virus. Its genome is approximately 7 500 bases long. Enterovirus, belonging to the family *Picornaviridae,* is a genus including 15 species. Species Enterovirus A-D and Rhinovirus A-C are common in humans. Enteroviruses are known to have a high mutation rate. Enteroviruses are very common, e. g. their occurrence in feces of young children can reach up to 17 % (Kahrs, Chuda et al. 2019) (enterovirus time course during infancy is shown by ***Figure 12)***. Infections by different enterovirus types can cause different diseases depends on their tropism or the state of host´s immune system. Enteroviruses typically multiply in the gastrointestinal tract or in the respiratory system. Diseases caused by enteroviruses can range from rashes (e. g. hand, foot and mouth disease), through common cold, gastroenteritis to encephalitis. Enterovirus infections may frequently be asymptomatic (https://talk.ictvonline.org/ictv-reports/ictv_online_report/positive-sense-rna-viruses/picornavirales/w/picornaviridae/681/genus-enterovirus, 12.7.2019).

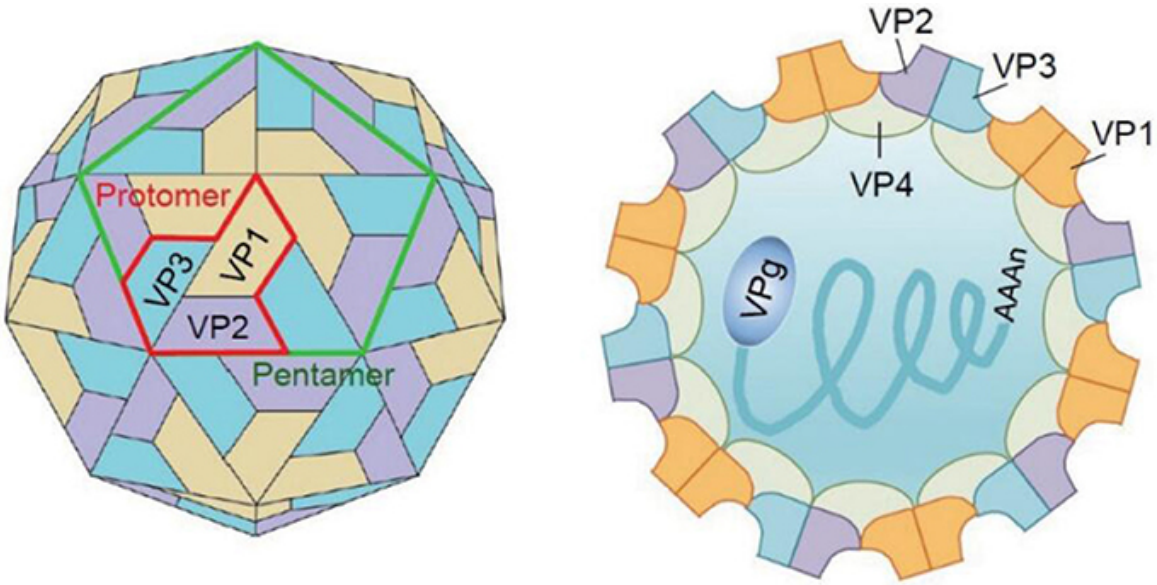


***Figure 10*. The structure of the enterovirus particle.** The virion has about 30 nm in diameter. The capsid consists of four types of proteins VP1, VP2 and VP3, which form the surface structures, and VP4 protein without any epitopes exposed on the capsid surface. Structural proteins form a protomer, five protomers are folded into a pentamer, twelve of which can self-associate to a virion particle. The enterovirus genome is one single RNA chain. From Yi et al. 2017.

The viral particle (its structure is shown in ***Figure 10***) binds to a cell surface receptor. Subsequently, a pore in the cell membrane is formed and RNA is injected into the cell. The RNA genome is uncoated, VPg protein is removed and then RNA is translated into a single polyprotein, which is cleaved into capsid proteins and proteins involving in the virus replication. A double-strand RNA is synthesized from the single strand RNA with positive polarity (+ssRNA). The RNA chain with negative polarity from the dsRNA is then replicated producing viral +ssRNA genomes, which are packed into capsids and leave the cell (https://talk.ictvonline.org/ictv-reports/ictv_online_report/positive-sense-rna-viruses/picornavirales/w/picornaviridae/681/genus-enterovirus, 12.7.2019).

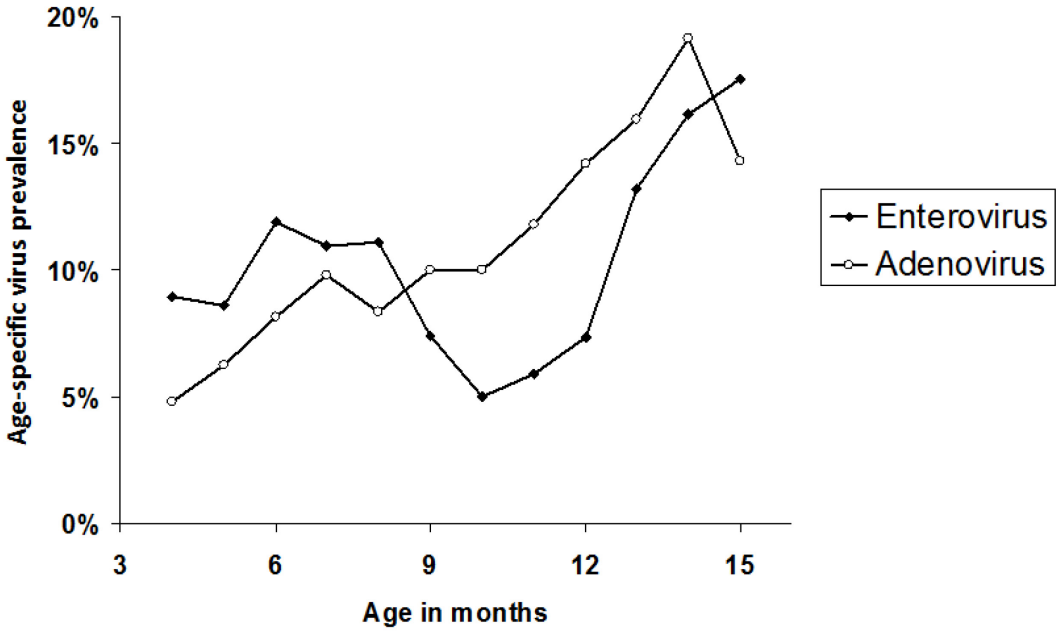


Figure 11. ***The observed prevalence of adenovirus and enterovirus in feces of infants***. As can be seen, the prevalence of both viruses is relatively high, ranging between 5 to 20 % in toddlers. Data obtained from the MIDIA project, Cinek et al. 2006.

3.2.1 Evidence of enterovirus infections in the pathogenesis of celiac disease

In 2012, Sarmiento and colleagues investigated anti-TG2 antibodies in patients with previously confirmed HEV infection. The prevalence of anti-TG2 antibodies (TGA) in these patients was 55 % (11/20 patients). Thus, authors have demonstrated that the presence of TGA is significantly associated with HEV infection. Patients had confirmed HEV infection by PCR and by the detection of enterovirus-specific IgG (serotype echovirus 16). Although relatively small sample size, this observation supports the hypothesis indicating the involvement of enteroviruses in the pathogenesis of CD. Authors also compared TGA prevalence (55 %) in these 20 HEV positive patients with the TGA prevalence in 21 CMV (cytomegalovirus) positive patients (9,5 % patients were TGA positive), 20 EBV (Epstein-Barr virus) positive patients (25 % TGA-positive patients) and 21 HCV (Hepatitis C virus) positive

patients (21 % TGA-positive patients), which implies HEV infections can be associated with the generation of TG2-specific antibodies.

Enterovirus has not been thoroughly studied in the context of CD as yet. In 2012, Mercalli et al. have investigated enterovirus infections in 27 CD patients and 21 healthy controls. They performed enterovirus-specific RT-PCR of small intestine biopsies and also detected anti-VP1 IgG levels in plasma samples. However, the study was cross-sectional and found no increased risk of CD (Mercalli et al. 2012). As suggested by the previous studies (Vorobjova et al. 2017, Li et Atkinson, 2015), HEVs could be sufficient to increase the intestinal permeability, which can enable the contact of TG2 with immune cells resulting in production of TGA (Sarmiento et al. 2012).

Enterovirus has been hypothesized to be involved in the development of autoimmune diseases such as type 1 diabetes (Tapia et al. 2011). Presumably, T1D and CD can share some environmental factors (Hagopian 2017, Frisk et al. 2008). Although enterovirus has been studied in the pathogenesis of T1D for a long time (Craig et al. 2013), insufficient evidence is available regarding its possible association with CD. Thus, investigation of enterovirus infections in the pathogenesis of CD is warranted.

3.3 Implications from the evidence of viruses in the context of CD

Taken aforementioned studies together, even contradictory results were noticed. It is required to pay attention because of several research weaknesses in designs of the studies.

Generally, there is the lack of prospective studies. Only one study was prospective (Stene et al. 2006).

Also, as seen from the evidence presented above, the most frequent restriction was the study size, which can result in the disability to detect just mild effects of infections operating in CD pathogenesis.

A common disadvantage is the retrospective character, which increases the risk of biases. Also, infections were sometimes parent-reported or unspecified, which could decrease the reliability. Another limitation is sometimes unstandardized or unclear diagnosis of CD.

4 MATERIALS AND METHODS

The steps of the workflow were:

- Formation of the study cohort MIDIA
- Implementation of the celiac disease sub-study (Recruitment of study participants – serology, selection of matched controls)
- Processing of stool samples
    - Extraction of nucleic acids
    - Detection of viruses by PCR methods
    - Virus genotyping by massive parallel amplicon sequencing
- Statistical analysis

4.1 The MIDIA cohort

MIDIA, which is the Norwegian acronym for Environmental Triggers of Type 1 Diabetes Study, is a prospective population-based cohort of children carrying the HLA-DRB1*0401-DQA1*03-DQB1*0302/DRB1*0301-DQA1*05-DQB1*02 (often referred to as DR4-DQ8/DR3-DQ2) genotype, which confers a high-risk for both T1D and celiac disease. It is thus a homogeneous cohort of children defined by a single HLA genotype conferring the highest risk of T1D, and also high risk of CD.

Study participants were born during 2001-2007 in Norway. Up to 50 000 newborns from general population throughout Norway were screened for the above HLA genotype. The time distribution of recruitment is depicted in ***Figure 12***. Originally, the MIDIA project aims to identify environmental causes of T1D (Stene et al. 2007). Parents were invited to participate when babies were 1-2 weeks old. Babies provided mouth brushes for DNA extraction and were HLA typed. In total, 912 (1.9 % of HLA typed newborns) were carriers of the high-risk genotype and were included to the study cohort. These children were followed.

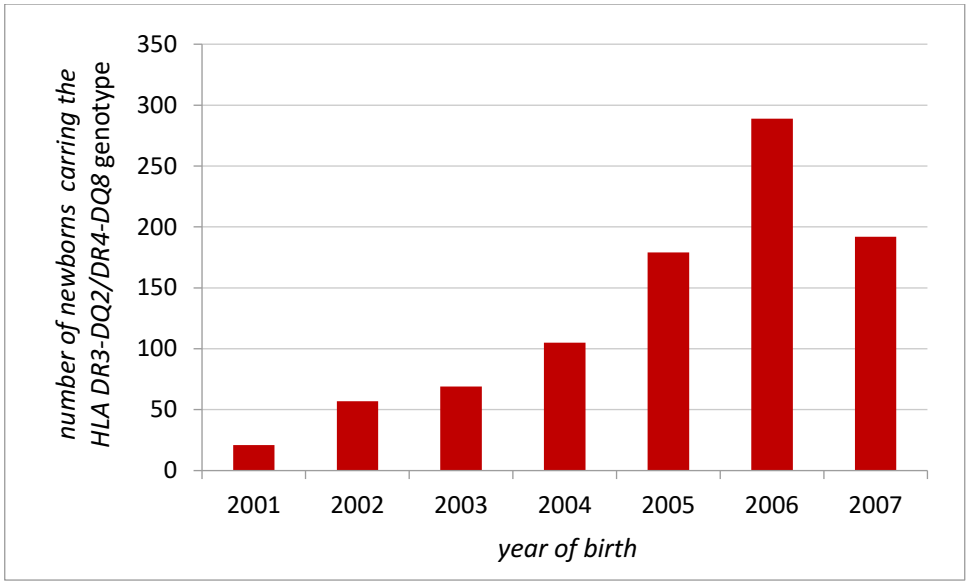


***Figure 12*. The distribution of birthdates of followed subjects carring the genotype HLA DR3-DQ2/DR4-DQ8.** Data from Kahrs, Chuda et al. 2019.

Blood samples were scheduled at ages 3, 6, 9 and 12 months and annually thereafter. Questionnaires investigating the infants´ thrive and well-being were required at the same time periods. T1D specific autoantibodies to glutamic acid decarboxylase, insulin, and insulinoma-associated antigen-2 were measured, and aliquots of plasma were archived at -80 °C.

Parents of all infants were requested to provide stool samples from the diapers or pot every month from the children´s age of 3 months to 36 months. Containers with a spoon under the lid were used. Specimens were then mailed by the post to the laboratory (Cinek et al. 2006).

4.2 Implementation of the celiac disease sub-study

The large biobank of the MIDIA project made it possible to conduct a nested case-control sub-study investigating celiac disease. During 2014-2016, participants remaining within the follow-up of the MIDIA study were invited to the celiac disease sub-study. As illustrated in ***Figure 13***, 220 of 501 invited children consented to participation. ***Table 1*** contains detailed characteristics of participants. Within the cohort, 27 children with CD were identified, 17 of them were diagnosed before screening. Subsequently, two of CD cases were excluded due to improper sampling or unreliable diagnosis. Each of 25 cases was matched to two similar controls by age, number of siblings and the place of residence.

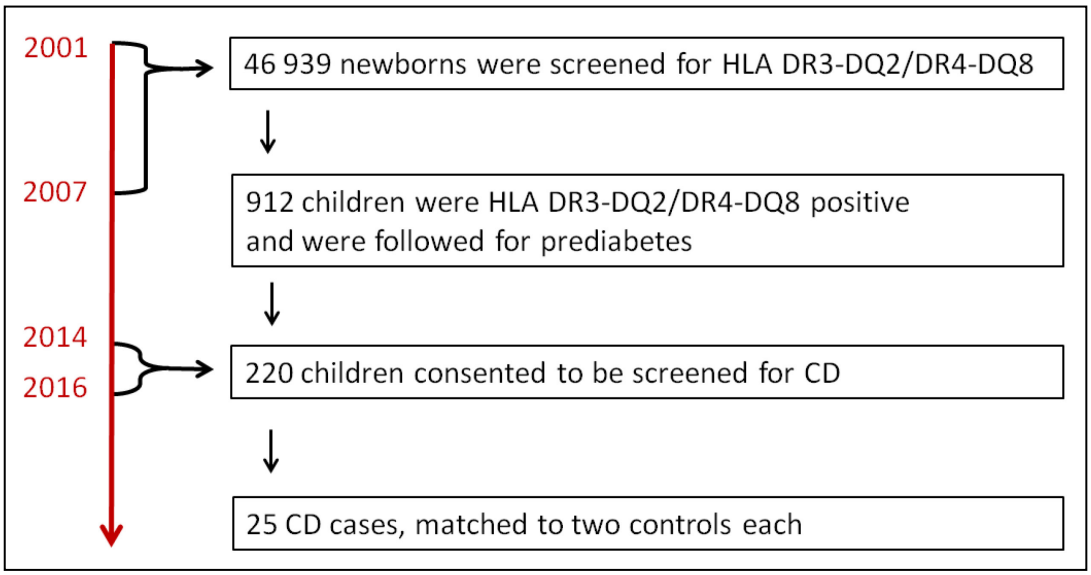


Figure 13. ***The scheme illustrates the study timeline and the enrolment of participants to the celiac disease sub-study.*** In the fall 2001, recruitment of participants started and was continuing to 2007. During this period, almost 50 000 babies were genotyped. Individuals carrying the HLA DR3-DQ2/DR4-DQ8 genotype were included and followed for T1D islet autoantibodies. Of these 912 children, 501 remained in the follow-up as of 2014, and of these, 220 consented to participate in the CD study. From 2014 to 2016, these children were screened and 25 CD cases were defined and matched to 50 healthy controls from the same cohort.

Characteristic	Babies with the chosen risk HLA genotype (n=912)	Participants in the CD study (n=220)
Female sex (n (%))	445 (49)	107 (49)
Other children in household (n (%))		
None	319 (35)	76 (35)
1	367 (40)	102 (46)
≥2	226 (25)	42 (19)
Family history of celiac disease at the time of enrolment (n (%))	19 (2)	7 (3)

Table 1. ***Characteristics of the followed children.*** The participants tended to have fewer siblings, higher prevalence of CD and higher prevalence of T1D in their families. Published in Kahrs, Chuda et al. 2019.

4.2.1 Celiac disease diagnosis

All CD cases were diagnosed according to generally accepted ESPGHAN 2012 criteria (The European Society for Pediatric Gastroenterology Hepatology and Nutrition) (Frühauf et al. 2012). During 2014-2016, 220 children were screened for the presence of specific CD antibodies in their most recent plasma specimens (IgA anti-TG2 autoantibodies and in addition, IgG anti-deamidated gliadin antibodies in order to catch cases with IgA deficiency). Some children within the cohort had already been diagnosed as CD cases, other were identified by screening. If the test was positive, the child was recommended to further testing according to the ESPGHAN diagnostic procedures. Only reliable confirmed cases were included. The detailed scheme of the case-control recruitment is depicted in ***Figure 14***.

The diagnosis was based on biopsy (Marsh ≥3) in 14 cases, further 9 children were diagnosed by serology and two children with unclear Marsh grade on biopsy, were diagnosed on repeated anti-TG2 antibodies >10 times cut-off level and symptoms.

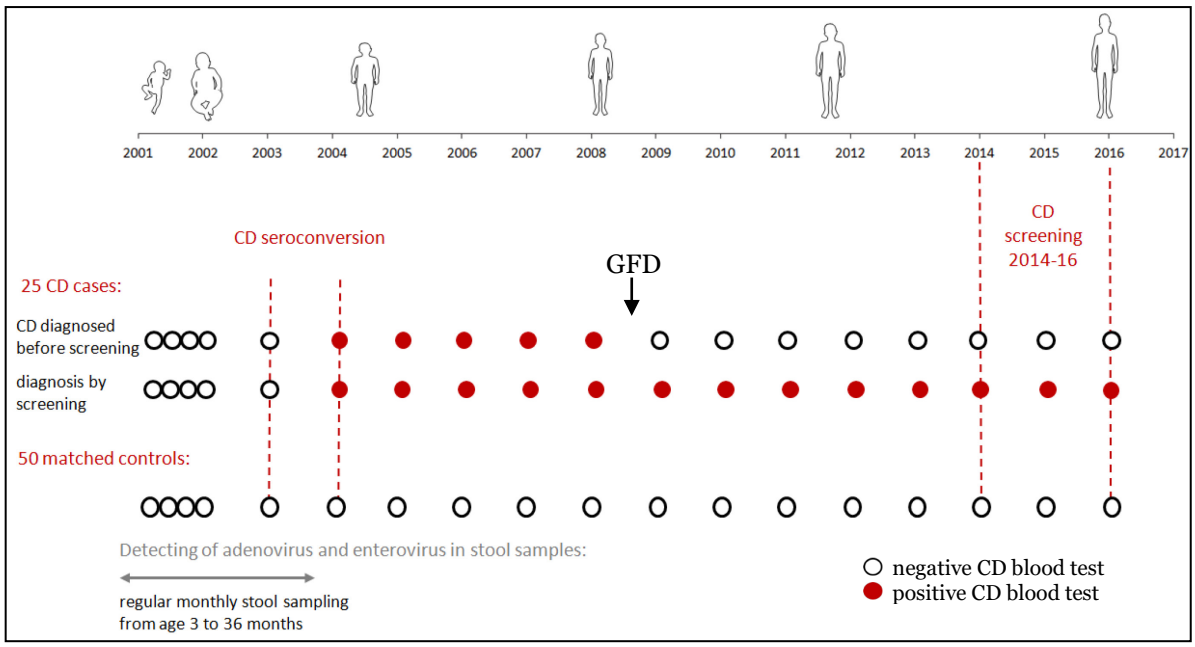


Figure 14. The sampling scheme. A total of 220 children were enrolled to the celiac disease sub-study and underwent celiac disease screening from 2014 to 2016. Some of them have already been diagnosed with celiac disease at that time. In those cases, we can see disappearance of CD specific antibodies in blood samples after the introduction of GFD. In celiac disease cases, retrospective testing of longitudinally collected blood samples was performed in order to identify the window of seroconversion. Blood samples were taken regularly every three months up to the age of one year and then annually. All matched healthy controls were negative for CD biomarkers throughout the whole observational period. Monthly collected fecal samples of all 75 enrolled children were subjected to PCR testing for the presence of enteroviruses and adenoviruses. Positive samples were genotyped. Rearranged from materials kindly provided by C. R. Kahrs; MD, PhD.

4.3 Processing of stool samples

In the central laboratory in Norway, fecal samples were suspended in phosphate buffered saline with the addition of 2.5 µg/ml fungizone, 50 IU/ml penicillin, 0.5% bovine serum albumin, 50 µg/ml streptomycin, 50 IU/ml penicilin, 50 µg/ml streptomycin. Consequently, samples were vortexed and centrifuged at 4000×g for 30 min. Supernatants were frozen at -80°C until further processing. Testing of stool supernatant samples by PCR methods began in 2016. Samples were sent from the Norwegian biobank to prof. Cinek´s Laboratory of Molecular genetics in Prague, where we have tested them.

4.3.1 Extraction of nucleic acids

Then optimized protocol for increasing the throughput was developed using the QIAamp Viral RNA Mini Kit (Qiagen, Hilden, Germany) for co-purification of RNA and DNA, with a modification for 96-well plate format instead of individual columns. Procedures were performed in biosafety cabinets whenever possible, and samples were processed blinded to the case-control status.

Samples were thawed at room temperature, vortexed and centrifuged at 22 °C 3 min at 15 000 g to sediment debris. Negative controls were incorporated in this step. Within each 96-well plate, at least six negative and one positive control randomly placed in 96-well format were used. Negative controls were 140 µl of TE buffer and positive controls were 2 µl of Enterovirus Armored RNA (Ambion Diagnostics, Austin, TX, USA) dissolved in 138 µl saline.

4.3.1.1 Lysis

The lysis was done in individual 2.0ml tubes (Biosphere SafeSeal 2.0 ml), which were prefilled by 560 µl of AVL lysis buffer at room temperature. The AVL solution contained 1 µg/ml carrier RNA and a low amount of West Nile Virus Armored RNA that served as an exogenous internal control because of the potential presence of PCR inhibitors in feces. The stool supernatant samples, each at the volume of 140 µl, were added into their respective 2.0ml tubes. After 10min incubation, 560 µl of absolute ethanol was added.

4.3.1.2 Binding

A QiaAMP filter plate was inserted into a 96-well plate (1.2 ml, Eppendorf). Subsequently, the filter plate was placed under a small transparent plastic bridge ***(Figure 15)*** with one hole of the size of one well in the plate. All other at a moment unused wells apart from a working well were covered and protected from contamination by the transparent surface of the bridge. The volume of 800 µl of lysed liquid was applied onto the QiaAMP filter plate through the hole. By shifting the plastic bridge, all wells were filled. After filling all

positions, the filter plate was covered by a microporous tape sheet (AirPore Tape sheets, Qiagen).

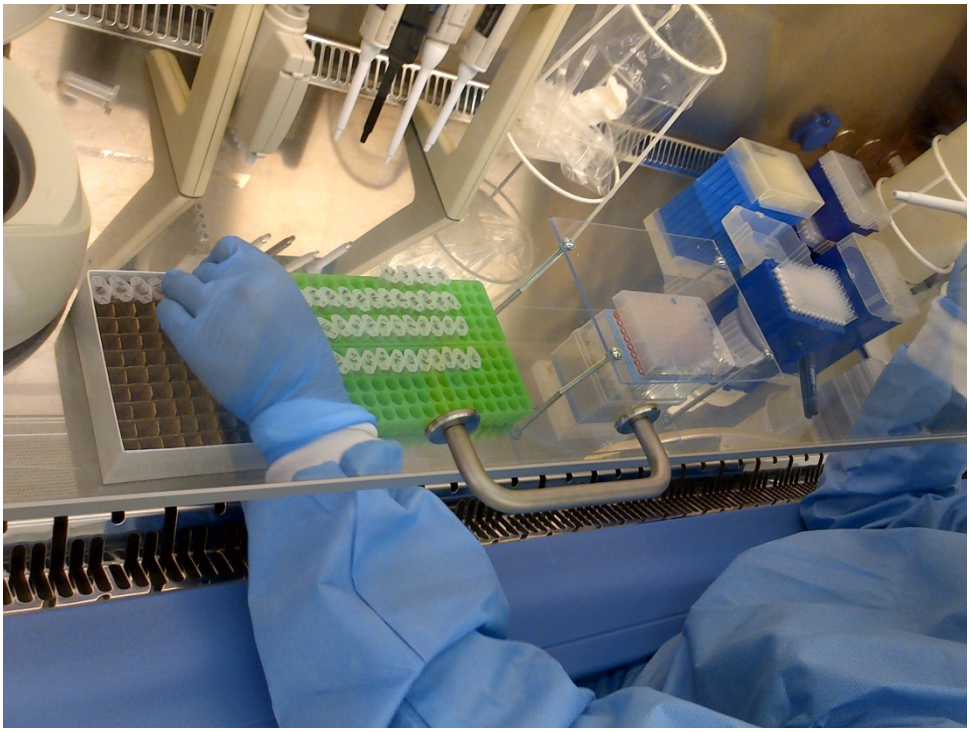


Figure 15. ***Using of the transparent plastic bridge.*** The collection plate with the filter plate is under the transparent surface of the bridge and the hole is set above the currently used sample column. The photo was provided by prof. Cinek.

4.3.1.3 Washing

The filter with bound nucleic acids was moved on a new collection plate and 400 µl of AW1 buffer was applied into each position of the filter plate. After further centrifugation, 400 µl of AW2 buffer was added and the plate was centrifuged again. Then, the filter was placed on a new 96-well plate.

4.3.1.4 Elution

After application of 110 µl of AE solution (with AVE already inside) into each position, the plate was centrifuged to get the first elution. The eluted plate was covered by an aluminium

foil and put into the fridge until further manipulation (within two days). The second elution was prepared similarly as the first elution and frozen.

4.3.2 Detection of viruses by PCR methods

Testing for enterovirus RNA was performed in one-step real-time PCR and for adenovirus DNA in one-step PCR described in detail earlier (Cinek et al. 2006). Enterovirus was tested in 15µl volume (containing 2 µl of tested nucleic acid elution) using the QuantiTect Probe RT-PCR Kit (Qiagen, Hilden, Germany) and 900 nmol/l primers and 300 nmol/l probes designed for enterovirus. The amplified genome section was conserved 5´ UTR region. These primers and probes react with an equal sensitivity to *Enterovirus A—D* species and do not react with human rhinoviruses (Honkanen et al. 2013). The amplification had to occur before the PCR cycle 35 and the positivity cut-off was set to 10 copies/µl. Amplification of WNV (the layout of the PCR plate is shown in ***Figure 16***) was carried out in 8µl volume without duplicates as was described before (Cinek et al. 2006), sequences of primers and probes were taken from Briese et al. (Briese et al., 2002). Adenovirus was tested in a 15µl volume (containing 2 µl of each tested sample). Amplification was performed using primer sequences designed by Claas et al. (Claas et al. 2005) and Echavarria et al. (Echavarria et al. 1998) targeting the 3´end of the hexon gene (Claas et al. 2005). The positivity threshold was the same like in the case of enterovirus. Reactions were performed on Biorad CFX384 cycler.

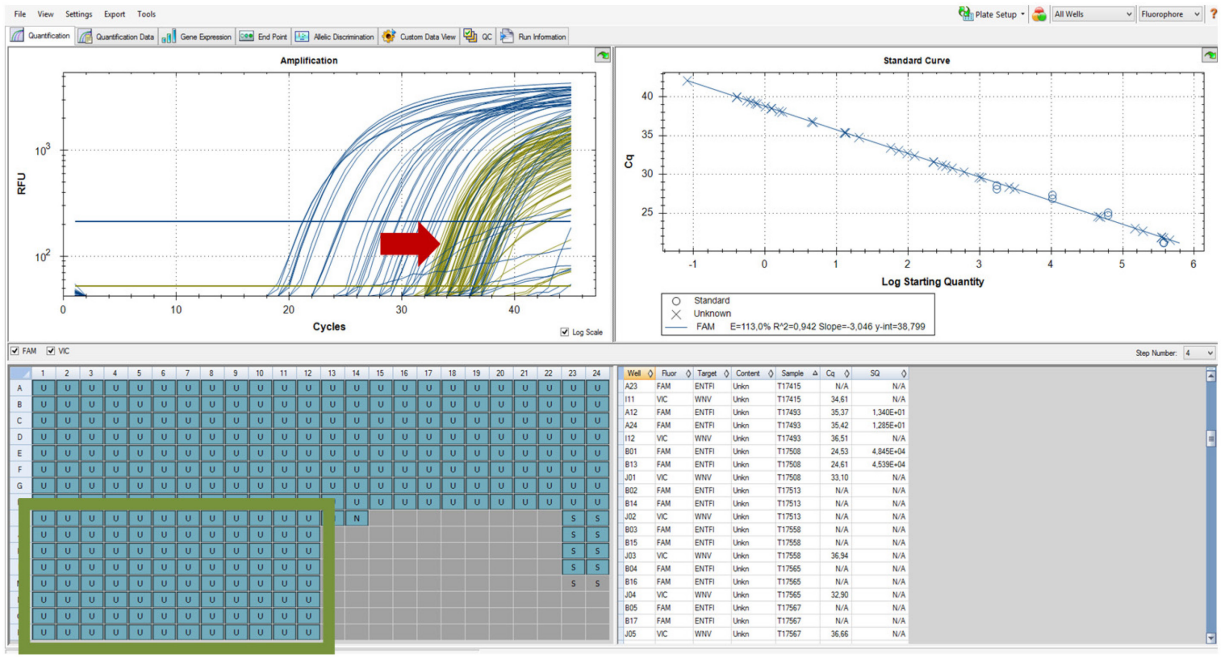


Figure 16. ***Layout of a 384-well plate for real-time (RT-)PCR testing.*** Enterovirus and adenovirus were tested in duplicates. In addition, the quality of extraction was evaluated by the PCR quantification of an internal exogenous control, West Nile Virus (WNV), highlighted in green. As seen in the first quadrant, WNV nucleic acid from the most of samples was extracted equally, thus the extraction was mostly successful and not inhibited. From one of our analyses.

4.3.3 Virus genotyping

Genotyping was an essential tool to investigate whether specific virus types are associated with celiac disease. Also, sequencing was crucial in deciding on whether the infection is new or prolonged within individual clusters of serial samples. It is necessary, because adenoviruses and enteroviruses can persist in stool and be shed for a time period longer than one month. Thus, we need to distinguish successive infections caused by different serotypes.

Compared to the protocol for adeno- and enterovirus detection, which was already pre-established (Cinek et al. 2006) and is mentioned above, the method for virus genotyping was optimized during my lab work. Therefore, the procedure is described in more detail. In addition, our paper published by Cinek et al. in 2018 (Cinek et al. 2018) and detailed supplementary protocols of described methods are attached in the last chapter "Appendix, Paper II, Supplementary materials of paper II".

The procedure consisted of several steps:

1. amplification of the informative region, using tailed primers
2. PCR products pooling
3. indexing, which provides the fragments with indices and adapters
4. equalization of library quantities
5. massive parallel amplicon sequencing
6. computational analysis

Firstly, the presented protocol was evaluated on 44 isolates of adenoviruses and 30 isolates of enteroviruses having known genotypes. To verify the ability to distinguish multiple infections, specimens containing two different adenovirus or enterovirus genotypes were prepared in several different ratios. The protocol was then applied on fecal samples originating from the MIDIA study.

4.3.3.1 Adenovirus

Adenovirus genotyping was carried out by sequencing of the seventh hypervariable region of its hexon gene using a mixture of ten forward and six reverse primers (depicted in ***Figure 17***), which were designed in program Genious 8.1.8 (https://www.geneious.com) as illustrated by ***Figure 18***. Virus specimens subjected to genotyping had the minimum quantity threshold set as 10 copies/μl.

A. Adenovirus (hypervariable region 7 of the hexon gene)	
Designation	**Sequence 5' to 3'**
Forward primers[1]	
Adv_1243_F1_for5	TCGTCGGCAGCGTC**AGATGTGTATAAGAGACAG**GAACCAGATACTTTAGCATGTGGAACTCTG
Adv_1243_F2_for5	TCGTCGGCAGCGTC**AGATGTGTATAAGAGACAG**GAACCAGATACTTYAGCATGTGGAATCAGG
Adv_1243_F3_for5	TCGTCGGCAGCGTC**AGATGTGTATAAGAGACAG**GAACCAGATAYTTTTCYATGTGGAATCAGG
Adv_1243_F4_for5	TCGTCGGCAGCGTC**AGATGTGTATAAGAGACAG**GAACCCGGTATTTCAGTATGTGGAATCARG
Adv_1243_F5_for5	TCGTCGGCAGCGTC**AGATGTGTATAAGAGACAG**GGAGTCGATACTTCTCCATGTGGAACCAGG
Adv_1243_F6_for5	TCGTCGGCAGCGTC**AGATGTGTATAAGAGACAG**GAACACGGTACTTTTCCTTGTGGAATTCCG
Adv_1243_F7_for5	TCGTCGGCAGCGTC**AGATGTGTATAAGAGACAG**GAACAAGATACTTTTCAATGTGGAATCAAG
Adv_1243_F8_for5	TCGTCGGCAGCGTC**AGATGTGTATAAGAGACAG**GAAGTCGCTACTTTTCTATGTGGAATCAAG
Adv_1243_F9_for5	TCGTCGGCAGCGTC**AGATGTGTATAAGAGACAG**GAAGCCGTTATTTTTCCATGTGGAATTCTG
Adv_1243_F10_for5	TCGTCGGCAGCGTC**AGATGTGTATAAGAGACAG**GCAGTCGGTATTTTTCTATGTGGAACTCAG
Reverse primers[2]	
Adv_1611_R1_rev7	GTCTCGTGGGCTCGG**AGATGTGTATAAGAGACAG**CCGTTCATGTACTCGTAGGTGTTGGTRTT
Adv_1611_R2_rev7	GTCTCGTGGGCTCGG**AGATGTGTATAAGAGACAG**CCGTTCATGTASTCGTAGGTGTTYTTGTT
Adv_1611_R3_rev7	GTCTCGTGGGCTCGG**AGATGTGTATAAGAGACAG**TTGTTCATGTAGTCGTAGGTGTTGGGGTT
Adv_1611_R4_rev7	GTCTCGTGGGCTCGG**AGATGTGTATAAGAGACAG**CCATTCATGTAGTCATAAGTGTTGGTGTT
Adv_1611_R5_rev7	GTCTCGTGGGCTCGG**AGATGTGTATAAGAGACAG**TTGTTCATATAATCGTATGAATTAGGGTT
Adv_1611_R6_rev7	GTCTCGTGGGCTCGG**AGATGTGTATAAGAGACAG**CCGTTCATGTAGGCATAKGTRTTITTGTT

Figure 17. ***Design of adenovirus primers.*** Plain typeface nucleotides of primers represent the specific parts for individual adenovirus genotypes. Used primer sequences surround the seventh hypervariable region of adenovirus hexon gene. PCR products are all in the same frame. Published in Cinek et al. 2019.

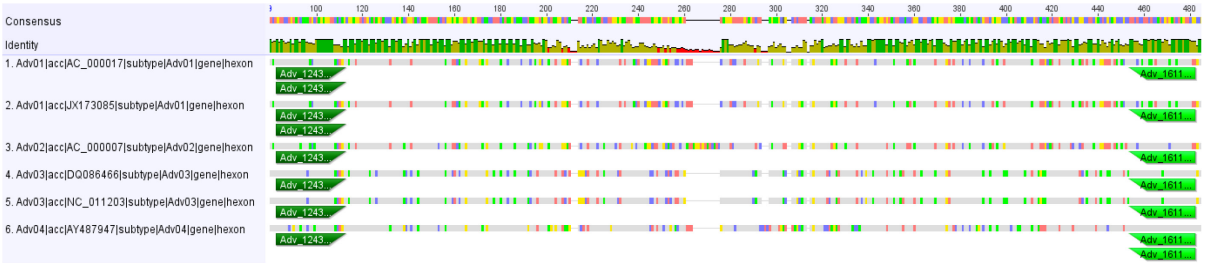


Figure 18. ***Creation of adenovirus primer designs in program Genious 8.1.8***. Primers surround the hypervariable region of the hexon gene. Each primer set targets a specific adenovirus subtype.

4.3.3.1.1 PCR amplification of the seventh hypervariable portion of hexon gene

The reaction of 20 µl contained:

- 2 µl of extracted nucleic acid
- 1 µM of forward primers (10 primers in an equimolar ratio)
- 1 µM reverse primers (6 primers in an equimolar ratio)
- 2.5mM $MgCl_2$
- 0.2mM each dNTP
- 1 U HotStar Taq polymerase (Qiagen, Hilden, Germany).

The PCR program:

95°C	15 min	initial denaturation
45 cycles of:		
94°C	45 min	denaturation
55°C	1 min	annealing
68°C	1 min	synthesis
68°C	5 min	final synthesis

4.3.3.2 Enterovirus

Genotyping of enterovirus was done by reverse transcriptase PCR of VP1 gene. The protocol was optimalized according to the previously published one (Nix et al. 2006). Primers were redesigned in order to increase the variability spectrum of detected enteroviruses (depicted in ***Figure 19***) (Cinek et al. 2018). Only samples with the virus load over 100 copies/µl were genotyped.

B. Enterovirus (fragment of the VP1 gene)	
Second round of PCR[3]	Sequence 5' to 3'
Forward primers[4]	
E2F1_for5	TCGTCGGCAGCGTC**AGATGTGTATAAGAGACAG**GCNGYTGAGACAGG
E2F2_for5	TCGTCGGCAGCGTC**AGATGTGTATAAGAGACAG**GCNGYRGAGACTGG
E2F3_for5	TCGTCGGCAGCGTC**AGATGTGTATAAGAGACAG**GCWGYNGAAACTGG
E2F4_for5	TCGTCGGCAGCGTC**AGATGTGTATAAGAGACAG**GCWGYNGAAACGGG
Reverse primers[5]	
E2R1_rev7	GTCTCGTGGGCTCGG**AGATGTGTATAAGAGACAG**GGNAYRWACATRATTTG
E2R2_rev7	GTCTCGTGGGCTCGG**AGATGTGTATAAGAGACAG**GGNAYRWACATGTACTG
E2R3_rev7	GTCTCGTGGGCTCGG**AGATGTGTATAAGAGACAG**GGNAYRTACATTATCTG
E2R4_rev7	GTCTCGTGGGCTCGG**AGATGTGTATAAGAGACAG**GGDAYRTACATSASCTG
E2R5_rev7	GTCTCGTGGGCTCGG**AGATGTGTATAAGAGACAG**GGWAYRTACATNAMTTG

Figure 19. ***Design of enterovirus primers.*** Plain typeface nucleotides of primers represent the specific parts for individual enteroviral genotypes. Used primer sequences surround the proximal part of enterovirus VP1 region. PCR products are all in the same frame. Published in Cinek et al. 2019.

4.3.3.2.1 PCR amplification of the proximal portion of VP1 gene

Firstly, **reverse transcriptase PCR** was performed. The reaction mixture of 10 µl for each sample consists of:

- denatured primer mix Nix-RT, each primer at final concentration of 1 pmol
- 1x First-strand buffer
- 0.005 M DTT
- 20 U per reaction of RNAsin
- 100 U per reaction of reverse transcriptase
- dNTP with 250µM final concentration of each nucleotide type
- 5 µl of extracted RNA

Primer mix Nix-RT was denatured for 5 min. at 65°C and then chilled on ice. Still on ice, remaining components were added. The mixture was vortexed. Extracted 5 µl of samples were pipetted into their respective positions of a 96-well plate. The plate was incubated at 65°C for 5 min. and promptly then chill on ice. Previously prepared RT was added à 5 µl.

The RT-PCR program was:

22°C	15 min	primer annealing
50°C	45 min	reverse transcription
70°C	15 min	inactivation of reverse transcriptase
10°C	forever	storage

The first round of PCR was almost performed as in the previously published protocol (Nix et al. 2006). The reaction mixture of 25 µl for each sample contains:

- 15.25 µl of water
- 1x buffer
- 25 mM $MgCl_2$
- 20 µM of NIX224 primer
- 20 µM of NIX222 primer (Nix. et al. 2006)
- dNTP with 200µM final concentration of each nucleotide type
- 1.25 U per reaction of HotStar polymerase
- 2.5 µl of cDNA from the previous step

The 96-well plate was subjected to the PCR program:

95°C	15:00 min	initial denaturation
40 cycles of:		
94°C	0:30 min	denaturation
42°C	0:30 min	annealing
60°C	1:00 min	synthesis
60°C	2:00 min	final synthesis
15°C	forever	storage

After PCR, 15 µl of samples were diluted with 5 µl of a blue dye and the PCR product was checked by electrophoresis. Remaining 10 µl was used in the second round of PCR.

The second round of PCR was more optimalized in comparison with the protocol published by Nix et al. Promega GoTaq G2 Hot Start polymerase chemistry was used.

The reaction volume consists of:

- 2 µl of the first round product that was diluted with water in a ratio 1:5
- 1 µM of forward primers (four primers in an equimolar ratio)
- 1 µM of reverse primers (5 primers in an equimolar ratio)
- 2.5mM $MgCl_2$
- 0.2mM of each dNTP
- 1X Green Go Taq Flexi Buffer
- 0.5 U Taq polymerase

The second round PCR was performed at:

95°C	2:00 min	denaturation
40 cycles of:		
95°C	0:15 min	denaturation
55°C	1:00 min	annealing
72°C	0:45 min	synthesis
72°C	5:00 min	final synthesis
10°C	forever	storage

4.3.3.3 PCR products pooling

The PCR amplified fragments of adenovirus and enterovirus were controlled by electrophoresis. If a sample was positive for both adenovirus and enterovirus, the sample was pooled together. Purification was performer by Ampure XT (Beckman, Brea, CA). To wash primer dimers out and to promote binding of longer chains, the ratio of the magnetic particles to PCR product ratio was set to 0.8:1.

4.3.3.4 Indexing

By a short PCR reaction of eight cycles, indices and adapters were added to fragments.

The volume of 15 µl consists of:

- 1.5 µl primer 1, with either of 24 indices N701-N729
- 1.5 µl of primer 2, with either of 16 indices S502-S522 (Nextera XT Index Kit, Illumina; USA)
- 3 µl of the purified product from the previous round
- 1X KAPA HiFi HotStart Ready Mix (Sigma-Aldrich, Roche, Basel, Switzerland)

The PCR program of indexing reaction contained:

95°C	3:00 min	initial denaturation
8 cycles of:		
95 °C	0:30 min	denaturation
55 °C	1:00 min	annealing
72 °C	0:30 min	synthesis

Purification was performer by using a 1:1 ratio of PCR products and magnetic beads (Ampure XT, Beckman, USA) to eliminate artifacts.

4.3.3.5 Equalization of library quantities

Using the KAPA library quantification kit (Roche), PCR products were quantified by real-time PCR. Then, the samples were equalized to the same concentration each and pooled. The total final concentration was 2 nmol/l. Finally, the library was checked for concentration on a Bioanalyzer High Sensitivity DNA chip (Agilent, Santa Clara, CA).

4.3.3.6 Massive parallel amplicon sequencing

An Illumina MiSeq instrument was used for sequencing (Illumina; Germany).

4.3.3.7 Computational analysis

The pipeline was derived from VIPIE pipeline (Lin et. al 2017). The sequencing reads of low-quality were eliminated and trimmed to remove their low-quality ends. The maximum length of a read could be 250 bp. The left and right reads were merged. Overlapped left and right enterovirus reads produced a 347bp long amplicon. Adenovirus amplicon had 321 bp. Subsequently, amplicons were mapped to a reference panel of adenovirus or enterovirus known genotypes. The proportion of amplicons that were caught by an individual reference were expressed in percents of the total signal. The multiple infections were defined as a presence of more viruses with the positivity threshold set as 3 % of the total signal.

4.4 Statistical analysis

Firstly, the association of viruses with celiac disease was evaluated by a mixed effects logistic regression model regarding the clustered case-control matched design with longitudinal data containing repeated measurements. Logistic regression serves to model binary outcome variables (healthy/unhealthy), thus may be used to predict the risk of developing a disease on the basis of characteristics of the patient (age, sex, number of siblings, exposure to risk factors…). It is a kind of generalized estimated models.

The mixed models refer to a model regarding both random effects and fixed effects (Katz 2006). The data were analyzed in Stata program version 12 (https://www.stata.com/, 2.2.2019) by German Tapia, PhD. from our collaborating Norwegian Institute of Public Health. The adenovirus and enterovirus data were analyzed independently on each other. Virus positivity of stool samples was included as the dependent variable and case-control status was evaluated as an independent variable. The odds ratio for celiac disease status is calculated as the odds of a fecal sample being positive for virus given that it came from a future case, compared with the same odds for virus positivity for samples from matched controls. The calculation was made with regard to the potential predictors of CD and viral

infection and was adjusted for relevant variables such as family history of CD, season of sample collection, number of siblings, age, age squared and sex.

In order to avoid reverse causality, only stool samples before the CD antibody seroconversion in matched clusters were included in the primary analysis as was pre-planned. A sequence of consecutively virus positive fecal samples was considered as a single episode, which means a negative stool sample was required before a new episode. Further analyzing investigated the link between these viruses of interest in stool samples and the time periods during or after the CD seroconversion time period. Besides, infections with reported symptoms or long infectious episodes or infections of high virus quantities (≥100 000 copies per one µl of nucleic acid) were investigated. In addition, documented infectious symptoms from regular parental questionnaires were assessed and their association with the development of CD serologic markers was investigated.

Furthermore, the analysis was adjusted for weaning, the time of gluten introduction and also for T1D. The adjustment was performed by restriction of the analysis only on those infections occurred after the end of weaning or by restriction only on infections after the gluten introduction. In the case of adjustment for T1D, calculated results coming from children with T1D were compared with those without T1D and the difference was smoothed.

Exploring virus types, enteroviruses were clustered into *Enterovirus A-D* species and adenovirus into specific types like human adenovirus C1, C2, C3, because almost all positive samples were identified as *Human adenovirus C*.

4.5 Materials

4.5.1 Chemicals

	Manufacturer
10% Armored WNV	Ambion Diagnostics; USA
Agencourt® AMPure® PCR Purification System	Beckman Coulter Genomics; USA
DNA marker	Biotech; Czechia
dNTP (5 mM of each)	Qiagen; Germany
Enterovirus Armored RNA	Ambion Diagnostics; USA
GelRed™ Nucleic Acid Stain 10 000x in water	Biotium; USA
GoTaq® DNA Polymerase Kit	Promega; USA
HotStarTaq DNA Polymerase Kit	Qiagen; Germany
KAPA Library Quantification Kit	Roche; USA
Nextera XT Index Kit v2 Set A, set D	Illumina; CA
QIAamp Viral RNA Mini Kit	Qiagen; Germany
QuantiTect Probe RT-PCR Kit	Qiagen, Hilden, Germany
PCR (injection) water	Braun; Germany
RNasin® Ribonuclease Inhibitors	Promega; USA
SeaKem® LE Agarose	Lonza; USA
SuperScript III Reverse Transcriptase	Invitrogen; USA

4.5.2 Laboratory equipments

	Manufacturer
1,5ml test tubes	Eppendorf; Germany
2.0ml tubes	Biosphere SafeSeal 2.0 ml; Germany
96-well PCR plates	Applied Biosystems; USA
Adhesive PCR Plate Foils	Termofisher; Germany
Agencourt® SPRIPlate 96R ring magnetic plate	Beckman Coulter Genomics; USA
Hard-Shell® 384-Well PCR Plates	Biorad
LightCycler® 480 Sealing Foils	Roche
LightCycler® 480 Multiwell Plates 384	Roche
pipette filter tips (DNAase, RNAase, ATP free)	Sarstedt; Germany
Plate Deepwell 96 wells 2000 µl	Eppendorf
Sealing foil, for 96-well plates	Eppendorf

	Manufacturer
2100 Bioanalyzer Instrument	Agilent, USA
3130xl Genetic Analyzer	Applied Biosystems; USA
7300 Real Time PCR system	Applied Biosystems; USA
AccuBlock Digital Dry Bath	Labnet International; USA
Biohazard box EF/S 6	Clean Air Techniek; Netherlands
Bioanalyzer High Sensitivity DNA Chip	Agilent; CA
Camera OLYMPUS SP500uz	Olympus; Japan
Centrifuge B4i	Jouan; Francie
Centrifuge MiniSpin plus	Eppendorf; Germany
Centrifuge Universal 320R	Hettich Zentrifugen; Germany
CFX 384 Real-Time Cycler	Biorad; Germany
Electrophoresis Power Pac P25 T	Biometra; Germany
Illumina MiSeq instrument	Illumina; Germany
LightCycler® 480 Instrument II	Roche, Basel; Switzerland
MiSeq sequencer	Ilumina; USA
pippettes	Eppendorf; Germany
	Thermo – Finnpipette; Finland
	Thermo – Matrix; USA
Robot Biomek 3000	Beckman Coulter; USA
Robot Biomek 4000	Beckman Coulter; USA
Termocycler Labcycler	Sensoquest; Germany
Termocycler Mastercycler pro S	Eppendorf; Germany
Transilluminator TFX-26MC	Vilber Lourmat; France
Vortex MS1 Minishaker	IKA Werke & Co; Germany

6 RESULTS

6.1 Enterovirus

Enterovirus was present in 370/2135 (17 %) stool samples (***Table 2, Table 3***). The overview of all enterovirus-positive samples is in ***Figure 20***. Prior to the development of celiac disease markers, there were 84/429 (20 %) enterovirus positive stool samples in cases and 129/855 (15 %) in matched controls, which resulted in adjusted odds ratio (aOR) 1.49 (95% CI 1.07 to 2.06; P=0.02). The adjusted odds ratio for prolonged infections (more than two months) was 2.16 (95% CI 1.16 to 4.04; P=0.02) and 2.11 (95% CI 1.24 to 3.60; P=0.01) for specimens of high virus load (>100 000 copies/µl) (***Table 4)***. There was no significant difference in enterovirus frequency during or after development of celiac disease markers in comparison between cases and controls.

Enterovirus frequency was dependent on age and season (***Figure 23***). Almost all followed children (73/75 individuals) have at least one enterovirus positive stool sample.

In total, 309/370 (83.5 %) enterovirus positive samples had data on enterovirus types. The genotyping failed in 20 specimens. The most frequent enterovirus species was *Enterovirus A*, which was significantly associated with the later development of celiac disease antibodies (aOR = 1.62, 95% CI 1.04 to 2.53; P=0.03), similarly *Enterovirus B* (OR = 2.27, 95% CI 1.33 to 3.88; P=0.003). In total, the most often observed genotype was Coxsackievirus A2 (CV-A2) (13 % of genotyped specimens), which was detected also as dual infection with CV-A5 or E-25. The second the most frequent genotype was CV-A4 that was found in 10 % of enterovirus positive samples**.** Enterovirus infections occurred most often as infections of a single enterovirus type. Coinfections of more enterovirus types within the same genus was documented in 5 % of enterovirus-positive samples. Examples of succession of dual infection episodes in children are shown in ***Figure 24.***

Enterovirus positivity during weaning was not associated with the later celiac disease (aOR 0.78, 95% CI 0.34 to 1.79; P=0.56). However, enterovirus infections after the end of weaning were associated (aOR 1.80, 95% CI 1.17 to 2.78; P=0.01). Enterovirus infections before introduction of gluten consumption were not associated with celiac disease (aOR 0.75, 95%

CI 0.34 to 1.79; P=0.65), whereas infections after the gluten introduction were related (aOR 1.52, 95% CI 1.05 to 2.20; P=0.03) **(*Table 5*).** Detection of enterovirus infections in combination with recorded fever was associated with the subsequent development of celiac disease (aOR 2.12, 95% CI 1.16 to 3.85; P=0.01), as well as enterovirus infections in combination with reported diarrhea (aOR 3.62, 95% CI 1.14 to 11.49; P=0.03) **(*Table 6*)**.

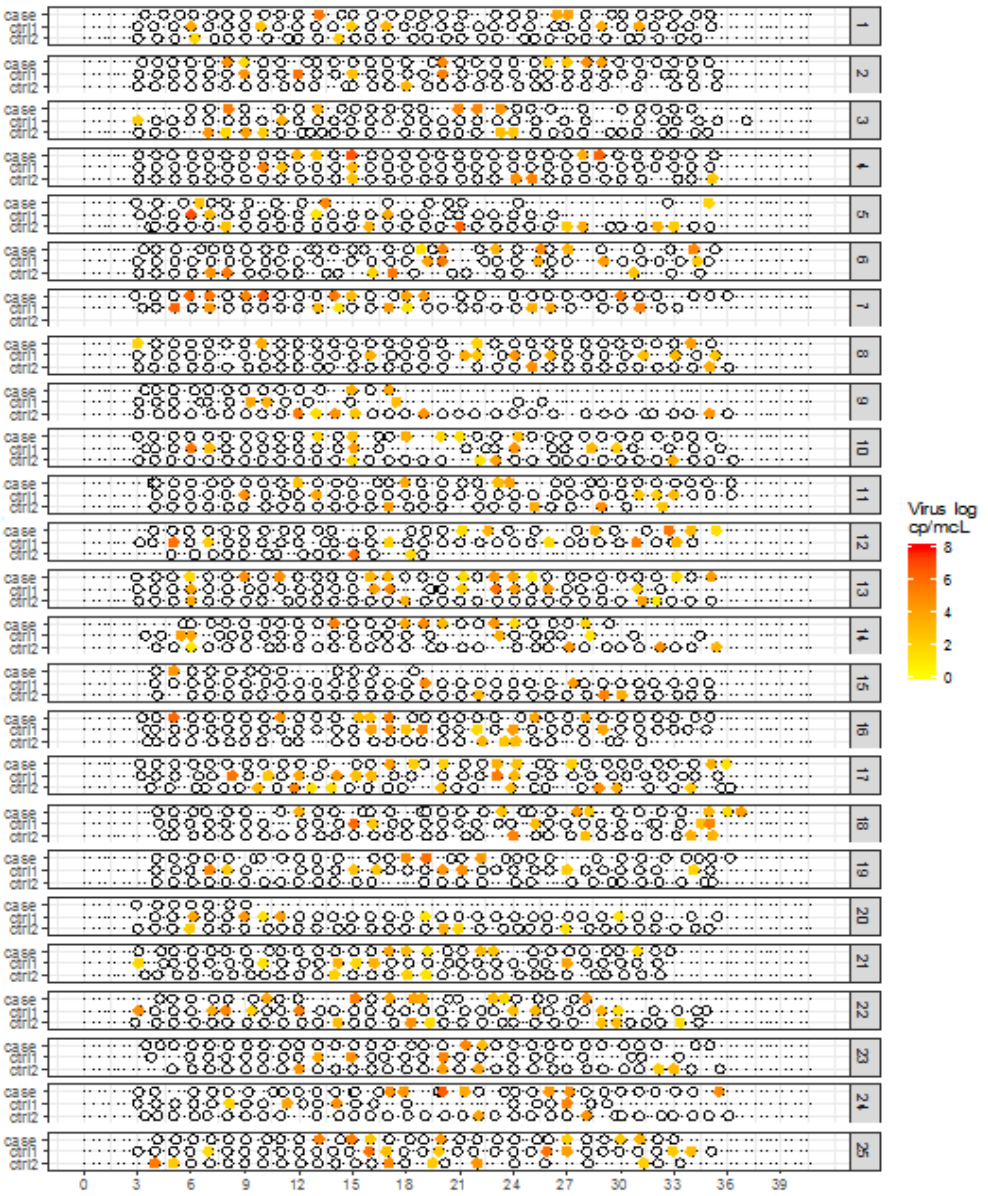


***Figure 20.* The graph illustrating enterovirus positivity of all studied stool samples.** On the x-axis, the age of children in months is plotted. The y-axis shows 25 clusters of a matched case and two controls. The virus quantity is distinguished by shades of orange.

	Cases (n=25)	Controls (n=49)	Adjusted Odds ratio (95% CI)	P value
	Positives (n)/Total (N) samples (%)			
All HEV BEFORE	84/429 (20)	129/855 (15)		0.02
Enterovirus A	42/387 (11)	60/787 (8)		0.03
CV-A2	10/355 (3)	10/737 (1)		0.09
CV-A4	5/350 (1)	11/738 (1)		0.90
CV-A6	7/352 (2)	9/736 (1)		0.39
CV-A10	4/349 (1)	10/737 (1)		0.96
CV-A16	4/349 (1)	7/734 (1)		0.63
Enterovirus B	34/379 (9)	43/770 (6)		0.003
All HEV DURING	30/140 (21)	58/332 (17)		0.48
All HEV AFTER	21/121 (17)	48/258 (19)		0.54

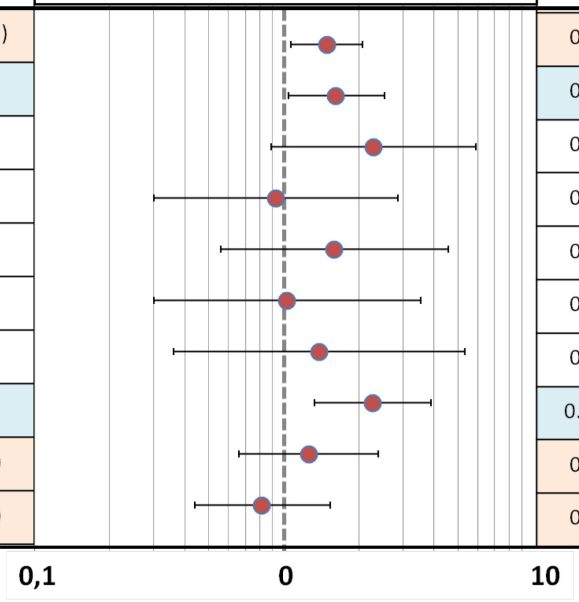


Table 2. ***The number of enterovirus positive stool samples in cases and controls before, during and after seroconversion.*** "Before" means the period prior the last celiac disease negative blood sample. *Enterovirus* species *A* and *B* are mentioned. *Enterovirus C* and *Enterovirus D* were not frequent enough, thus they are not depicted in this table. "During" covers the period after the last CDA negative sample and the first CDA positive sample. "After" represents the sampling period after the first CDA positive blood sample. Graphically, aORs are showed. Significant results in enterovirus frequency between cases and controls were obtained before the seroconversion period, with the P value 0.02 for all enterovirus infections, 0.03 for species *Enterovirus A* and 0.003 for *Enterovirus B.* Adapted from Kahrs, Chuda et al. 2019.

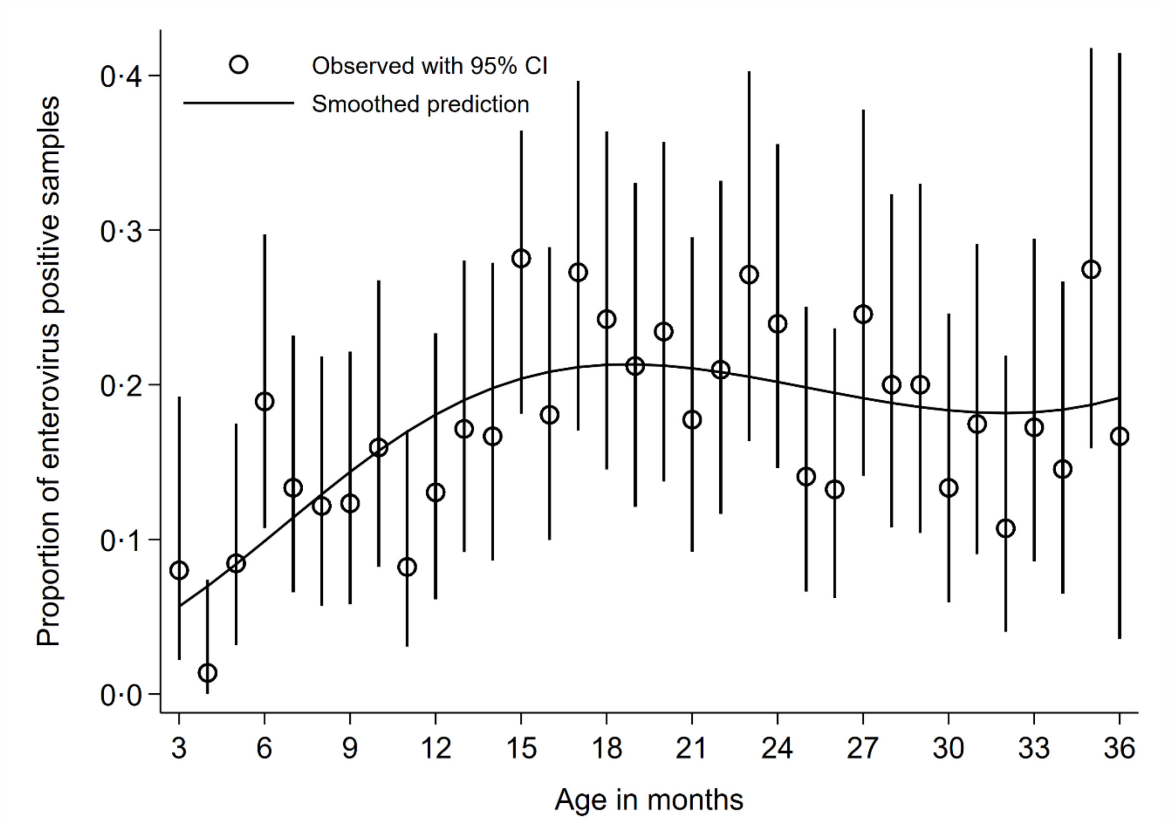


Figure 4. ***The proportion of enterovirus positive samples depending on a child's age.*** Published in Kahrs, Chuda et al. 2019.

Serotypes	Samples (N)	Percent of total samples (2 135 samples)	Percent of all enterovirus infections (a cut-off of 100 copies/µl)
Enterovirus A	172	8.1	46.5
CV-A2	41	1.9	11.1
CV-A4	30	1.4	8.1
CV-A5	16	0.7	4.3
CV-A6	24	1.1	6.5
CV-A8	2	0.1	0.5
CV-A10	22	1.0	5.9
CV-A12	1	0.0	0.3

CV-A14	1	0.0	0.3
CV-A16	22	1.0	5.9
EV-A71	15	0.7	4.1
Enterovirus B	139	6.5	37.6
CV-B1	12	0.6	3.2
CV-B2	19	0.9	5.1
CV-B3	11	0.5	3.0
CV-B4	9	0.4	2.4
CV-B5	12	0.6	3.2
CV-A9	17	0.8	4.6
E-3	3	0.1	0.8
E-6	5	0.2	1.4
E-7	2	0.1	0.5
E-9	10	0.5	2.7
E-11	8	0.4	2.2
E-13	1	0.0	0.3
E-18	9	0.4	2.4
E-25	19	0.9	5.1
E-30	6	0.3	1.6
Enterovirus C	4	0.2	1.1
CV-A1	3	0.1	0.8
CV-A22	1	0.0	0.3
PV-2	1	0.0	0.3
Enterovirus D	0	0	0

Table 3. Enterovirus serotypes detected within the sample set. In total, 370/2135 (17 %) stool samples were positive, from which 309 provide genotyping data. Coinfections were noticed in 5.8 % positive samples. The infection by poliovirus PV-2 is highlighted in yellow. Likely, the poliovirus positivity is an artifact from the vaccination. Published in Kahrs, Chuda et al. 2019.

	Cases	Controls	Adjusted	
	Positives (n)/Total (N) samples (%)		Odds ratio (95% CI)	P value
Main analysis	84/429 (20)	129/856 (15)	1.49 (1.07 to 2.06)	0.02
Long infections (≥2 positive sequent samples)	22/367 (6)	27/755 (4)	2.16 (1.16 to 4.04)	0.02
High-quantity infections (≥100 000 copies)	28/429 (7)	33/856 (4)	2.11 (1.24 to 3.60)	0.01
Infectious episodes	55/400 (14)	95/822 (12)	1.27 (0.87 to 1.86)	0.21

Table 4. ***Overview of the number of enterovirus infectious episodes, high-quantity infections and long infections prior to the development of celiac disease antibodies.*** Adjusted for sex, age, number of siblings, celiac disease family history, age squared and season of sample collection. Published in Kahrs, Chuda et al. 2019.

	Cases	Controls	Adjusted	
	Positives (n)/Total (N) samples (%)		Odds ratio (95% CI)	P value
Main analysis	84/429 (20)	129/856 (15)	1.49 (1.07 to 2.06)	0.02
Samples 3 to 6 months of age	7/87 (8)	18/174 (10)	0.44 (0.14 to 1.37)	0.16
Samples 6-12 months of age	17/155 (11)	42/302 (14)	0.66 (0.30 to 1.46)	0.30
Samples ≥12 months of age	67/231 (29)	83/463 (18)	1.97 (1.33 to 2.93)	0.001
Prior to gluten introduction	7/64 (11)	9/113 (8)	0.75 (0.21 to 2.63)	0.65
At gluten introduction (samples ±1 month of gluten introduction)	4/37 (11)	13/70 (19)	0.42 (0.11 to 1.61)	0.21
After gluten introduction	66/298 (22)	99/587 (17)	1.52 (1.05 to 2.20)	0.03
While breastfed	18/163 (11)	42/369 (11)	0.78 (0.34 to 1.79)	0.56
After end of breastfeeding	55/199 (28)	83/455 (18)	1.80 (1.17 to 2.78)	0.01
After end of breastfeeding and gluten introduction	55/195 (28)	71/383 (19)	1.83 (1.19 to 2.81)	0.01

Table 5. Data on enterovirus positivity, breastfeeding and gluten introduction. Presented samples belong to the period before CDA seroconversion. Adjusted for sex, age, number of siblings, family history of celiac disease, age squared and season of sample collection. The significant results are highlighted. Published in Kahrs, Chuda et al. 2019.

Association enterovirus and simultaneous symptoms	Reported symptoms/ Total samples (N)	Adjusted odds ratio (95% CI)	P value
Any symptom	696/2119	0.97 (0.88 to 2.52)	0.14
Fever	418/2136	1.10 (1.16 to 3.85)	0.01
Diarrhea	163/2120	3.62 (1.14 to 11.49)	0.03
Common cold	422/2121	1.00 (0.51 to 1.97)	0.99

Table 6. Summary of detected enterovirus infections and reported symptoms. Analyzed by logistic regression. All symptoms of infections were taken into consideration and evaluated for the period starting two weeks prior to and ending two weeks after the monthly stool sampling. The significant association of infections and fever is highlighted as well as the association of infections and reported diarrhea and. Published in Kahrs, Chuda et al. 2019.

6.2 Adenovirus

Adenovirus was detected in 258 of 2006 (13 %) fecal samples. The frequency of adenovirus infections depended on age. The most common serotypes were human adenovirus C1 and C2 (***Table 7, Table 8***). In total, 61/75 (81 %) children had at least one positive stool sample during the first three years of life. Adenovirus infections were associated with fever but no other common symptoms (***Table 9***). Seasonal pattern was not evident (***Figure 23***).

Adenovirus infections were not associated with celiac disease. Adjusted OR was 0.82 (95% CI 0.49 to 1.38; P=0.46) for adenovirus positive stool samples before the celiac disease development. No significant differences in adenovirus frequency were observed during or after the CD development. No adenovirus types were associated with the development of celiac disease ***(Table 7).***

In total, 301 samples containing adenovirus DNA were subjected to genotyping by massive parallel amplicon sequencing. Genotyping failed in nine samples. Detailed commented overview is given in ***Table 8.*** Examples of individual succession of infections in children are shown by ***Figure 24.***

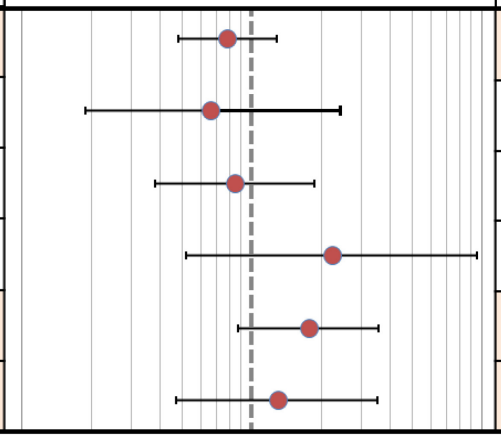
	Cases (n=25)	Controls (n=49)	Adjusted Odds ratio (95% CI)	P value
	Positives (n)/Total (N). samples (%)			
All HAdV BEFORE	47/390 (12)	111/775 (14)		0.32
HAdV-C1	17/360 (5)	60/724 (8)		0.51
HAdV-C2	20/363 (6)	43/707 (6)		0.67
HAdV-C5	5/348 (1)	7/671 (1)		0.28
All HAdV DURING	26/139 (19)	36/325 (11)		0.12
All HAdV AFTER	12/120 (10)	26/257 (10)		0.62

Table 7. ***The proportion of adenovirus positive stool samples in cases and controls before, during and after CD seroconversion.*** The numbers of positive/total stool samples are shown. The results from the mixed effects logistic regression (depicted as adjusted odds ratios) are given for the cases and controls at three different time periods. “Before” means the period prior to the last celiac disease negative blood sample. Adenovirus species HAdV-C1, HAdV-C2 and HAdV-C5 are depicted in detail. HAdV-C2 dominated in our sample collection. “During” covers the period after the last CDA negative sample and the first CDA positive sample. “After” represents the sampling period after the first CDA positive blood sample. Graphically, adjusted odds ratios (aORs) are showed. The aORs were calculated from the model of mixed effects logistic regression adjusted for relevant variables such as sex, age, number of siblings, celiac disease family history, age squared and season of sample collection. Evidently, there are no significant differences in adenovirus frequency between cases and controls comparing three, previously defined, time periods. Adapted from Kahrs, Chuda et al. 2019.

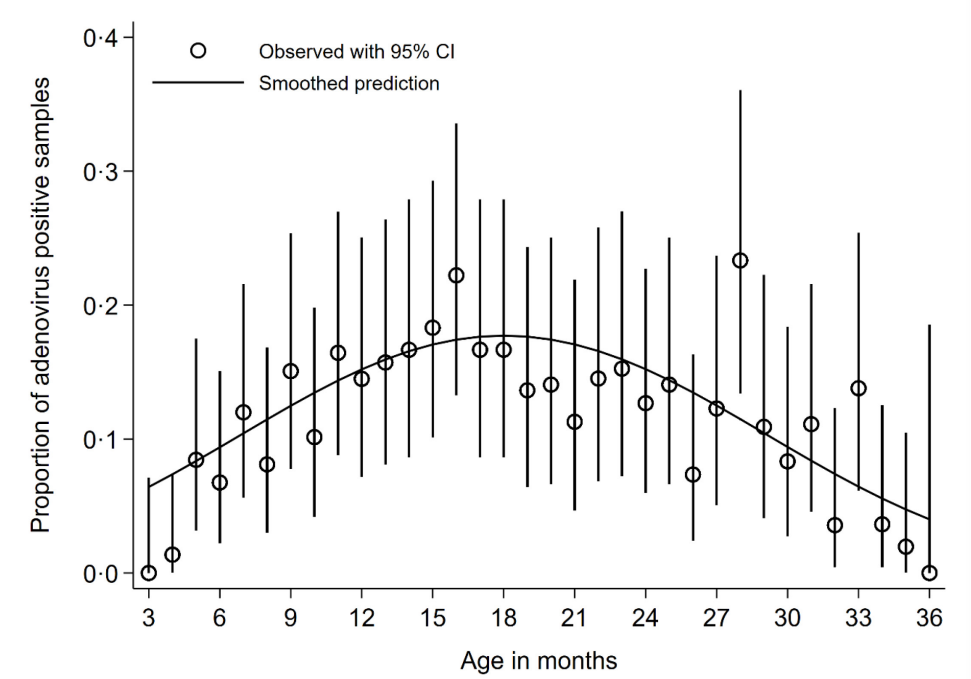


Figure 5. Distribution of adenovirus positive samples depending on a child´s age. Published in Kahrs, Chuda et al. 2019.

Serotypes	Samples (N)	Percent of total samples (2006 samples)	Percent of all adenovirus infections (a cut-off of 10 copies/ µl)
HAdV-C1	102	5.1	39.5
HAdV-C2	109	5.4	42.2
HAdV-B3	17	0.8	6.6
HAdV-A31	8	0.4	3.1
HAdV-F41	9	0.4	3.5
HAdV-C5	24	1.2	9.0
HAdV-C57	1	0.0	0.4

Table 8. Adenovirus serotypes detected in the whole sample set. Coinfections with two HAdV serotypes were observed in 18 adenovirus positive samples, which means 0.9% of positive samples. Published in Kahrs, Chuda et al. 2019.

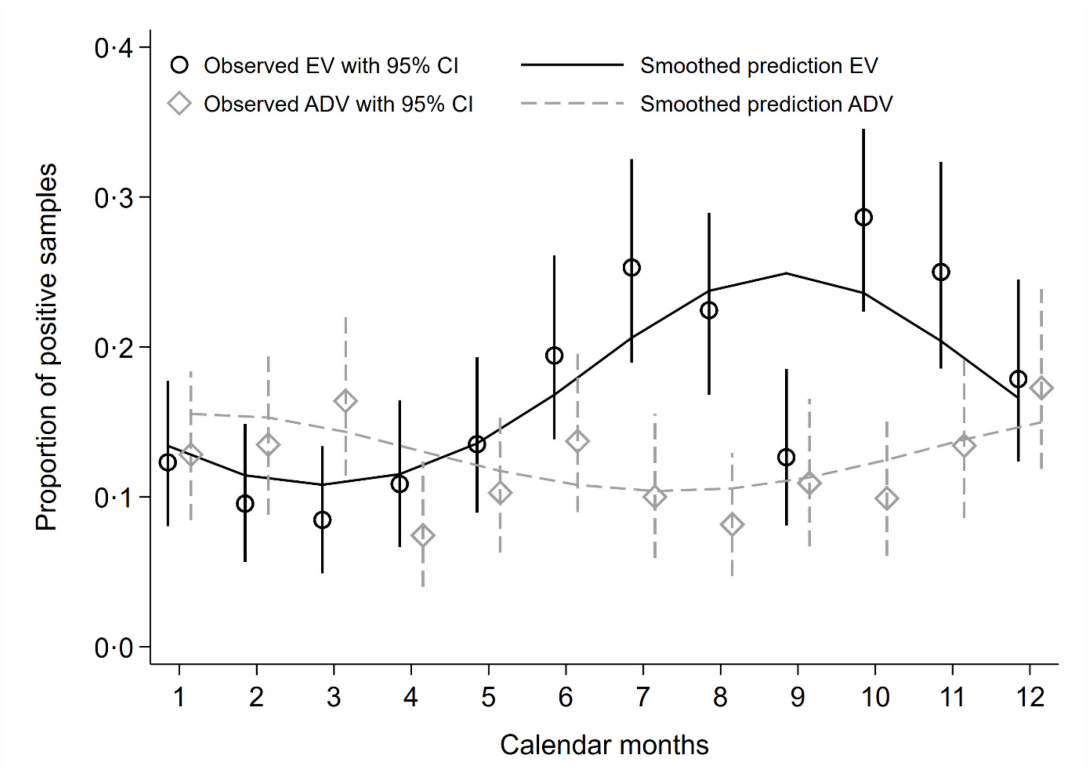
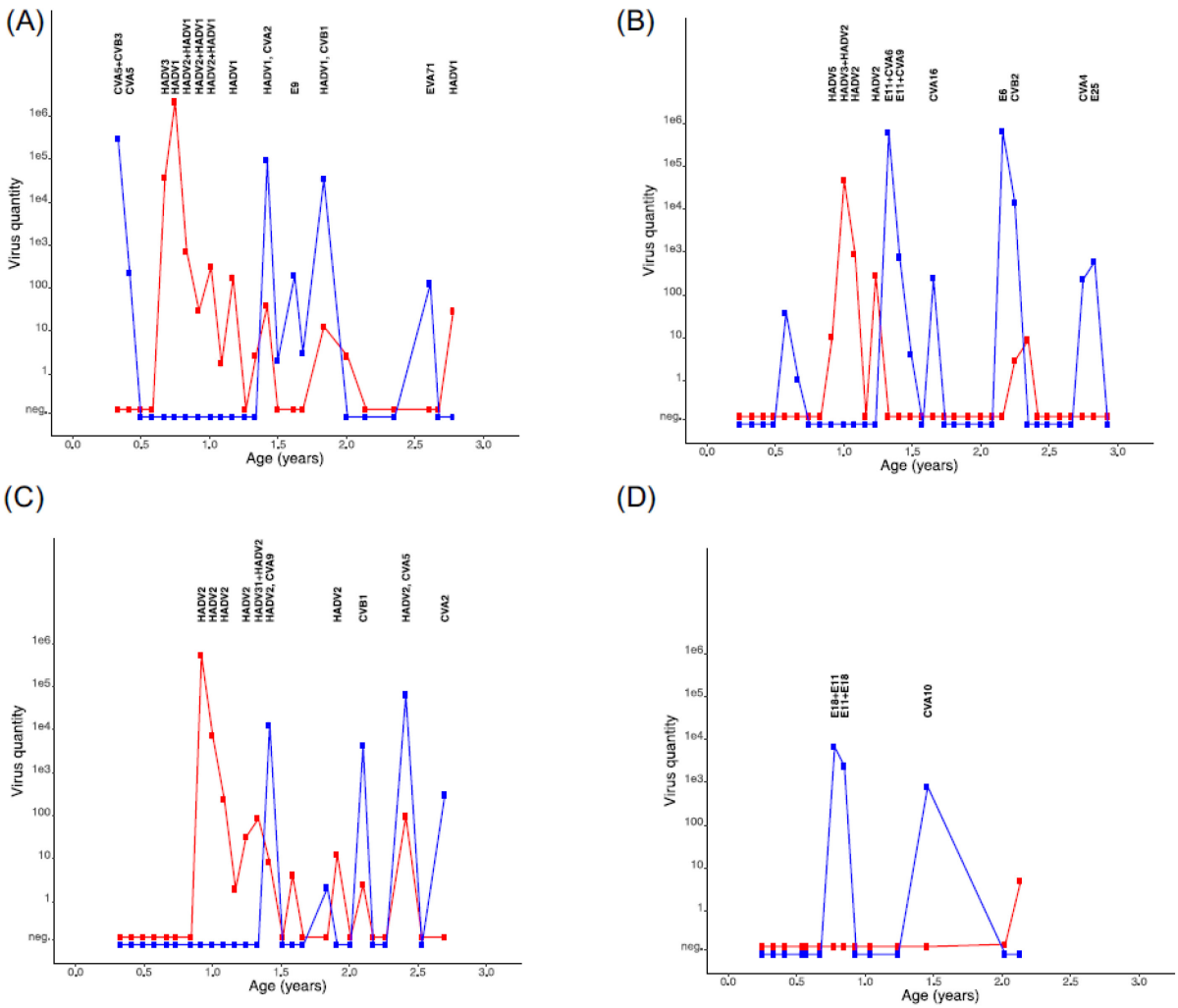


Figure 23. ***Seasonality of enterovirus and adenovirus infections.*** There is an obvious seasonal variation with a peak in autumn for enterovirus infections, whereas adenovirus infections have no clear seasonality. Published in Kahrs, Chuda et al. 2019.

Association adenovirus and simultaneous symptoms			
Any symptom	644/1992	1.26 (0.93 to 1.70)	0.13
Fever	396/2006	1.62 (1.22 to 2.16)	<0.01
Diarrhea	149/1992	1.11 (0.67 to 1.85)	0.68
Common cold	388/1994	1.11 (0.77 to 1.61)	0.58

Table 9. ***Summary of detected adenovirus infections and reported symptoms.*** Analyzed by logistic regression. All symptoms of infections were taken into consideration and evaluated for the period starting two weeks prior to and ending two weeks after the monthly stool sampling. The significant association between infections and fever is highlighted. Published in Kahrs, Chuda et al. 2019.



<u>Figure 24.</u> Four examples of children with simultaneous or prolonged infections in their consecutively taken stool specimens. Enterovirus is depicted in blue, adenovirus in red. **<u>A:</u>** two long-term and dual adenovirus and also enterovirus positivity in a child. **<u>B:</u>** Another example of simultaneous long-term infections. **<u>C:</u>** A long-term HAdV-2 followed by HAdV-31infection. **<u>D:</u>** Dual echovirus (abbreviated as "E") infections in two serial samples. Published in Cinek et al. 2018.

Overall, 50 samples were positive for both adenovirus and enterovirus. In total, 28 % of followed children under three years of age have gone through at least one dual infection of virus types belonging to the same genus.

7 DISCUSSION

7.1 Enterovirus and adenovirus in the pathogenesis of celiac disease

We found a significant association between enterovirus exposure and subsequent increased risk of celiac disease.

Previously published studies had a suboptimal design. Concerning the association of enteroviruses and CD, there was the lack of studies in general. Although enterovirus has been speculated to be involved in the development of immune-mediated diseases such as type 1diabetes (Tapia et al. 2011), its pathogenic effects in CD was not investigated. In 2012, Sarmiento et al. investigated the significant association of the presence of TGA and HEV infection and supports its involvement in the development of TGA, a CD biomarker. However, the study was cross-sectional and limited by its small sample size of 20 individuals (Sarmiento et al. 2012). In 2012, Mercalli performed enterovirus-specific RT-PCR of small intestine biopsies and measured anti-enterovirus antibodies in blood of 27 CD cases and 21 controls. However, the study was cross-sectional and found no increased risk of CD (Mercalli et al. 2012). Enterovirus infections could be able to increase the gut permeability (Vorobjova et al. 2017, Li et Atkinson, 2015), that is likely at the beginning of CD pathogenic processes (Visser et al. 2009).

Our finding is supported by the gradient in the effect: high titer of enterovirus and prolonged infections showed stronger association with CD. Temporal sequence of events was also clear: no association was observed for infections after the seroconversion period - this speaks against reverse causality. It is therefore likely that enterovirus infection is truly linked to CD, either causally, or through yet unknown confounder.

Our data do not support any strong associations between adenoviruses and celiac disease, which is coherent with the previous studies. The recent larger investigation of 80 CD patients and 8 healthy controls did not detect any association of adenoviruses and CD (Tarish et al. 2016), which is in line with previous studies (Mahon et al. 1991, Lawler et al. 1994). All previously performed studies were cross-sectional, without evaluation of potential

confounders and we have to keep in mind the significant observed involvement of adenoviruses in CD by Kagnoff et al. (Kagnoff et al. 1987) that was not confirmed by others.

Interestingly, adenovirus positivity was observed in 13 % of all stool samples of cohort children under the three years of age, whereas dual infections were detected in around 5 % of all adenovirus-positive stool samples, although the assumption should be 13 % again. The lower prevalence of the second infections is explainable by the antiviral milieu of already infected tissue, which is commonly observed effect of virus interference. However, we can assume slightly higher real frequency of multiple infections than observed 5 %, due to MIDIA sampling design and due to potential loss of some viral subtypes during lab processing (PCR amplification with a certain rate of randomness) or potential throwing some reads out during bioinformatic analysis. Noteworthy, children´s gut with simultaneous infections might serve as a milieu for recombinant events between viruses within the same genus. At minimum one dual infection was confirmed at more than a fourth of followed children up to their 3 years of age (Cinek et al. 2019).

7.1.1 Strengths of the study

We were the first to use a correct longitudinal design of the study, and to adjust for known confounders.

The study is unique in its robustness and longitudinal design with precise regular sampling, which provides complex information. Large, complex design and regularly taken samples represents extraordinarily valuable data. Only longitudinal studies can provide reliable information on environmental risk factors and can eliminate the chance of reporting bias. Its prospective nature enables to diminish the risk of reverse causality and careful documentation of children´s health characteristics in the combination with uniquely complex study design provide the adjustment for potential high reliability of results.

Also, generally accepted diagnostic criteria ensure the correct determination of CD status and no control subject is expected to be an undiagnosed case.

Because the viruses of interest can be shed for approximately three weeks, monthly sampling enables to detect most of these infections. The PCR methods used for detection of viruses provides the ability to uncover the vast majority of otherwise asymptomatic infections of the gut and also provide information on virus load. Observed virus prevalence and distribution are in line with the previous investigations (Cinek et al. 2006).

Because Sanger sequencing cannot distinguish simultaneously present virus types in one sample and requires the presence of high virus quantity, next generation amplicon sequencing was performed. By massive parallel amplicon sequencing, we genotyped viruses directly from clinical samples without confusion of non-viral stool nucleic acids. We were able to detect simultaneous virus infections within one phylogenetic genus.

7.1.2 Limitations of the study

First of all, only participants with genotype HLA DR3-DQ2/DR4-DQ8 were included. These heterozygotes might react differently to the studied infections than carriers of other predisposing genotypes, notably of the HLA-DR3-DQ2/DR3-DQ2. Thus, we cannot outright generalize to all CD cases. In future studies, inclusion of further risk haplotypes could be considered (Tinto et al. 2015). Nevertheless, the cost-effectiveness would be likely lower due to the lower overall CD risk when more genotypes will be used for defining the cohort. In addition, the present study was focused on children, and the pathogenesis of CD developing in adulthood may be different.

Another limitation could be that children were followed by sampling from the three months of age, thus the earliest infections during the first three months of life is not covered. Also, some CD seroconversions occurred after the stool sampling period of three years, which implies that we have not covered by stool sampling the period of "during" and "after" development of CDA in blood.

Despite the high robustness of the study, the results are restricted by a relatively small sample size, which was used within some specific sub-analyses. Also, the reliability is more likely diminished because infections with a shorter shedding period could be undetected by used sampling scheme.

We have to keep in mind that the observed association is mild, thus the reverse causality cannot be absolutely excluded in spite of the significance of observed effects. However, owing to the prospective nature of the survey, the reverse causality is very unlikely.

The celiac disease study consists of children with fewer siblings and higher prevalence of T1D compared to the whole MIDIA cohort. Nevertheless, it most likely does not influence the exposure-outcome associations.

We cannot exclude that future CD cases had mucosal barrier damage prior to CD development, which increase their susceptibility to enterovirus infections as well as the susceptibility to initiate the production of anti-TG2 autoantibodies. However, in that case, a similar amount of enterovirus infections would also occur after the CD development.

In addition, increased anti-TG2 levels are not totally specific signs of CD (Elitsur et al. 2016) and can result from other physiological changes, for instance, some infections can be accompanied by increased anti-TG2 levels (Farrace et al. 2001, Garcia-Peris et al. 2018). However, the detection of high anti-TG2 levels in several consecutively taken blood samples likely is not caused by this reason.

7.2 Interpretation

Presented findings can have several explanations. One of the explanations could be that children with more infections have an increased chance of later CD because of a stronger immune system stimulation can cause more easily the induction of autoimmune processes (Tighe et Beattie 2019).

Generally, we can hypothesize that enteroviruses may provide a danger signal that induce immune reactions such as activation of CD4+ gluten reactive T-cells or promotion of APCs to present available antigens (gliadins) to T-cells. Mouse models using reovirus infections aspired to describe the mechanistic pathways by which a virus can cause the loss of oral tolerance and initiate Th1 immunity to gluten (Bouziat et al. 2017). Thus, enterovirus may follow a similar pattern of pathogenesis.

Another explanation could be that increased frequency of enteroviruses is just a sign of a yet unknown risk factor (a confounder) with a true impact on the CD pathogenesis. However, it is unlikely because of the enterovirus frequencies between cases and controls after the development of CD autoimmunity are not significantly different.

Another remark is related to the potential temporal separation between the disease initiating event and the appearance of CD markers in blood. Our data shows that such a period, during which the risk infections are accumulating, and which is followed by a slow precipitation of intestinal tissue by immune processes, exists. The potential latent period between the environmental triggers and the CD manifestation can be illustrated by phenomenon observed by Mårild et al. (Mårild et al. 2015): children later diagnosed with CD had a significantly impaired growth relatively long before the CD diagnosis. That can be explained by a variety of reasons and we cannot be sure about their true causality. In CD patients, there could be the unique factor that influence the slower growth as well as the higher risk of later CD (depicted in ***Figure 25***). We have to keep in mind this potentially uncovered CD physiology.

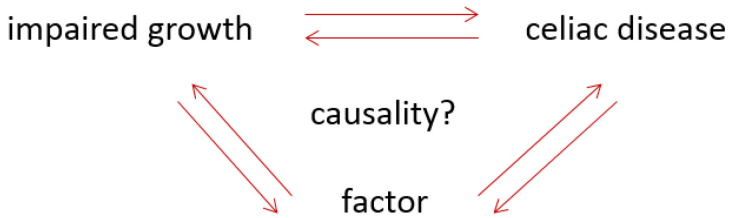


Figure 25. The scheme depicting causality. Inspired by the illustration made by C.R. Kahrs, MD, PhD.

7.3 Future prospects

Further investigations in other cohort studies are required to corroborate our observations. Challenge of further investigations could be the proper detection of infections with shorter shedding period or lower frequency such as rotavirus, which has been suggested as a CD risk factor (Stene et al. 2006).

Appropriate CD animal model is currently lacking. Available animal models cannot develop CD with the same phenotype as humans. For instance, transgenic mice carrying the human HLA DQ2 or -DQ8 alleles do not develop full CD, which could be explained by the lack of further genetic or environmental factors. Thus, our knowledge on the immune processes is restricted and there is an essential requirement of functional studies (Tye-Din et al. 2018). Nonetheless, mice have been used for studying potential therapies (Rossi et al. 2017). Additionally, it has been shown in mice that CVB vaccine can protect against T1D development (Hyöty et al. 2018). In 2014, investigators have published the study in mouse models, in which they have tested vaccine´s safety and efficacy in prevention of T1D. Focus on enterovirus serotypes associated with CD can lead to the development of the appropriate vaccine for CD prevention intervention studies. Nowadays, no commercially available non-polio enterovirus vaccine exists. Vaccination against enterovirus in children in high risk of celiac disease could figure as the option in the future (Kahrs, Chuda et al. 2019, Hyöty et al. 2018). But first, the role of enteroviruses in the CD pathogenesis need to be elucidated by further studies. Currently, we are expecting the results of the same associations "enteroviruses-CD" from the large neonatal TEDDY cohort , which is the largest study in the world investigating environmental triggers of CD and T1D with over 26 000 participants (Lernmark 2016, Kim et al. 2019).

If our findings will be confirmed by others, intervention studies could take place. Up to now, there are no conclusive recommendation for CD risk groups of individuals. However, the data from the MIDIA cohort have given valuable hints.

The MIDIA collection of samples have been also used to study the association of CD and parechovirus and anellovirus infections, which are able to be reliably caught and evaluated

by the MIDIA study design, because of their relatively long shedding period. Unfortunately, rotavirus infections, which are considered as a relevant candidate in CD triggering, cannot be properly evaluated within the MIDIA study. Rotavirus has a short shedding period, thus there is no chance to precisely detect this virus due to monthly stool sampling scheme.

Regarding NGS methods, we can expect the ongoing trend of their price reduction and more affordable sequencing of CD risk candidate viruses. Because we have detected dual infections within one genus in relatively high frequency (6 % of all available samples), we can aim to elucidate their potential recombination events. Our future attention will be also focused on the whole-genome sequencing of identified enteroviruses within the MIDIA study. Because we have got the unique collection of samples, we can also aim to identify the specific genome regions responsible for the triggering of the immune pathology. As described, different enterovirus serotypes can have different tropism and can be associated with different clinical manifestation (Jing-Yi et Shin-Ru 2014). Thus, we can hypothesize there may be a specific enterovirus type or its part responsible for triggering of CD pathology.

8 CONCLUSION

Exposure to enteroviruses in early childhood may increase the risk of later celiac disease, while adenovirus does not. This is the first prospective population-based cohort study investigating the associations between celiac disease and two most frequent human gut viruses, adeno- and enteroviruses. Further studies are warranted to evaluate the clinical significance and corroborate our findings. New insights into celiac disease pathogenesis could provide the opportunity to intervene in CD risk patients and could help in order to prevent celiac disease.

9 REFERENCES

ABADIE, V., SOLLID, L. M., BARREIRO, L. B. and JABRI, B., 2011. Integration of Genetic and Immunological Insights into a Model of Celiac Disease Pathogenesis. *Annual Review of Immunology*. 29(1), 493–525. ISSN 0732-0582. Available at: doi:10.1146/annurev-immunol-040210-092915

ALHASSAN, E., SERENA, G., KELLY, C. P. and FASANO, A., 2019. Nondietary Therapies for Celiac Disease. *Gastroenterology Clinics of North America*. 48(1), 145–163. ISSN 1558-1942. Available at: doi:10.1016/j.gtc.2018.09.011

BOUZIAT, R., HINTERLEITNER, R., BROWN, J. J., STENCEL-BAERENWALD, J. E., IKIZLER, M., MAYASSI, T., MEISEL, M., KIM, S. M., DISCEPOLO, V., PRUIJSSERS, A. J., ERNEST, J. D., ISKARPATYOTI, J. A., COSTES, L. M. M., LAWRENCE, I., PALANSKI, B. A., VARMA, M., ZURENSKI, M. A., KHOMANDIAK, S., SAMSOM, J. N., REINECKER, H. and KUPFER, S. S., 2017. Reovirus infection triggers inflammatory responses to dietary antigens and development of celiac disease. *Science*. 356(4), 44–50. ISSN 0036-8075. Available at: doi:10.1126/science.aah5298

CANOVA, C., ZABEO, V., PITTER, G., ROMOR, P., BALDOVIN, T., ZANOTTI, R. and SIMONATO, L., 2014. Association of Maternal Education, Early Infections, and Antibiotic Use With Celiac Disease: A Population-Based Birth Cohort Study in Northeastern Italy. *American Journal of Epidemiology*. 180(1), 76–85. Available at: doi:10.1093/aje/kwu101

CENIT, M. C., OLIVARES, M., CODONER-FRANCH, P. and SANZ, Y., 2015. Intestinal microbiota and celiac disease: Cause, consequence or co-evolution? *Nutrients*. 7(8), 6900–6923. ISSN 2072-6643. Available at: doi:10.3390/nu7085314

CINEK, O., KRAMNA, L., MAZANKOVA, K., KUNTEOVÁ, K., CHUDÁ, K., C. J. CLAAS, E., STENE, L. C. and TAPIA, G., 2018. Virus genotyping by massive parallel amplicon sequencing: adenovirus and enterovirus in the Norwegian MIDIA study. *Journal of Medical Virology*. 91(4), 606–614. ISSN 1096-9071. Available at: doi:10.1002/jmv.25361

CINEK, O., WITSØ, E., JEANSSON, S., RASMUSSEN, T., DREVINEK, P., WETLESEN, T., VAVRINEC, J., GRINDE, B. and RØNNINGEN, K. S., 2006. Longitudinal observation of enterovirus and adenovirus in stool samples from Norwegian infants with the highest genetic risk of type 1 diabetes. *Journal of Clinical Virology*. 35(1), 33–40. ISSN 1386-6532. Available at: doi:10.1016/j.jcv.2005.03.007

CRAIG, M. E., NAIR, S., STEIN, H. and RAWLINSON, W. D., 2013. Viruses and type 1 diabetes: A new look at an old story. *Pediatric Diabetes*. 14(3), 149–158. ISSN 1399-543X. Available at: doi:10.1111/pedi.12033

CROCKER, H., JENKINSON, C. and PETERS, M., 2018. Quality of life in coeliac disease: item reduction, scale development and psychometric evaluation of the Coeliac Disease Assessment Questionnaire (CDAQ). *Alimentary Pharmacology and Therapeutics*. 48(8), 852–862. ISSN 1365-2036. Available at: doi:10.1111/apt.14942

DOLCINO, M., ZANONI, G., BASON, C., TINAZZI, E., BOCCOLA, E., VALLETTA, E., CONTREAS, G., LUNARDI, C. and PUCCETTI, A., 2013. A subset of anti-rotavirus antibodies directed against the viral protein VP7 predicts the onset of celiac disease and induces typical features of the disease in the intestinal epithelial cell line T84. *Immunologic Research*. 56(3), 465–476. ISSN 0257-277X. Available at: doi:10.1007/s12026-013-8420-0

DUBOIS, P. C. A., TRYNKA, G., FRANKE, L., HUNT, K. A., ROMANOS, J., CURTOTTI, A., ZHERNAKOVA, A., HEAP, G. A. R., ÁDÁNY, R., AROMAA, A., BARDELLA, M. T., VAN DEN BERG, L. H., BOCKETT, N. A., DE LA CONCHA, E. G., DEMA, B., FEHRMANN, R. S. N., FERNÁNDEZ-ARQUERO, M., FIATAL, S., GRANDONE, E., GREEN, P. M., GROEN, H. J. M., GWILLIAM, R., HOUWEN, R. H. J., HUNT, S. E., KAUKINEN, K., KELLEHER, D., KORPONAY-SZABO, I., KURPPA, K., MACMATHUNA, P., MÄKI, M., MAZZILLI, M. C., MCCANN, O. T., MEARIN, M. L., MEIN, C. A., MIRZA, M. M., MISTRY, V., MORA, B., MORLEY, K. I., MULDER, C. J., MURRAY, J. A., ÑEZ, C., OOSTEROM, E., OPHOFF, R. A., POLANCO, I., PELTONEN, L., PLATTEEL, M., RYBAK, A., SALOMAA, V., SCHWEIZER, J. J., SPERANDEO, M. P., TACK,

G. J., TURNER, G., VELDINK, J. H., VERBEEK, W. H. M., WEERSMA, R. K., WOLTERS, V. M., URCELAY, E., CUKROWSKA, B., GRECO, L., NEUHAUSEN, S. L., MCMANUS, R., BARISANI, D., DELOUKAS, P., BARRETT, J. C., SAAVALAINEN, P., WIJMENGA, C. and VAN HEEL, D. A., 2010. Multiple common variants for celiac disease influencing immune gene expression. *Nature Genetics*. 42(4), 295–302. ISSN 1546-1718. Available at: doi:10.1038/ng.543

ELITSUR, Y., SIGMAN, T., WATKINS, R., PORTO, A. F., LEONARD PUPPA, E. L., FOGLIO, E. J. and PRESTON, D. L., 2016. Tissue Transglutaminase Levels Are Not Sufficient to Diagnose Celiac Disease in North American Practices Without Intestinal Biopsies. *Digestive Diseases and Sciences*. 62(1), 1–5. ISSN 1573-2568. Available at: doi:10.1007/s10620-016-4354-4

ERCOLINI, A. M. and MILLER, S. D., 2009. The role of infections in autoimmune disease. *Clinical and Experimental Immunology*. 155(1), 1–15. ISSN 0009-9104. Available at: doi:10.1111/j.1365-2249.2008.03834.x

FARRACE, M., A PICARELLI, TOLA, M. Di, SABBATELLA, L., MARCHIONE, O., IPPOLITO, G. and PIACENTINI, M., 2001. Presence of anti-'tissue' transglutaminase antibodies in inflammatory intestinal diseases: An apoptosis-associated event? *Cell Death and Differentiation*. 8(7), 767–770. ISSN 1350-9047. Available at: doi:10.1038/sj.cdd.4400880

FRISK, G., HANSSON, T., DAHLBOM, I. and TUVEMO, T., 2008. A unifying hypothesis on the development of type 1 diabetes and celiac disease: Gluten consumption may be a shared causative factor. *Medical Hypotheses*. 70(6), 1207–1209. ISSN 0306-9877. Available at: doi:10.1016/j.mehy.2007.05.058

FRÜHAUF, P., SZITÁNYI, P. and VYHNÁNEK, R., 2012. Nové doporučení ESPGHAN pro diagnostiku celiakie. *Pediatrie pro Praxi*. 13(3), 211–213. ISSN 1213-0494.

GARCIA-PERIS, M., DONAT ALIAGA, E., ROCA LLORENS, M., MASIP SIMÓ, E., POLO MIQUEL, B. and RIBES KONINCKX, C., 2018. Anti-tissue transglutaminase antibodies not related to gluten intake. *Anales de Pediatria*. 89(5), 279–285. ISSN 1695-9531. Available at: doi:10.1016/j.anpedi.2018.01.013

HAGOPIAN, W., LEE, H.-S., LIU, E., REWERS, M., SHE, J.-W., ZIEGLER, A., LERNMARK, A., TOPPARI, J., RICH, S., KRISCHER, J., ERLICH, H., AKOLKAR, B. and AGARDH, D., 2017. Co-occurrence of Type 1 Diabetes and Celiac Disease Autoimmunity. *Pediatrics.* 140(5). Available at: doi:10.1542/peds.2017-1305

HEIKKILÄ, K., KOSKINEN, O. A., AGARWAL, A., TIKKINEN, K. A. O. and MÄKI, M., 2015. Nutrition , Metabolism & Cardiovascular Diseases Associations of coeliac disease with coronary heart disease and cerebrovascular disease: A systematic review and meta-analysis. *Nutrition, Metabolism and Cardiovascular Diseases*. 25(9), 816–831. ISSN 0939-4753. Available at: doi:10.1016/j.numecd.2015.05.004

HYÖTY, H., LEON, F. and KNIP, M., 2018. Developing a vaccine for type 1 diabetes by targeting coxsackievirus B. *Expert Review of Vaccines*. 17(12), 1071–1083. ISSN 1744-8395. Available at: doi:10.1080/14760584.2018.1548281

JING-YI, L. and SHIN-RU, S., 2014. Cell and tissue tropism of enterovirus 71 and other enteroviruses infections. *Journal of Biomedical Science*. 21(18). Available at: http://www.jbiomedsci.com/content/pdf/1423-0127-21-18.pdf

KAGNOFF, B. Y. M. F., AUSTIN, R. K., HUBERT, J. J., BERNARDIN, J. E. and KASARDA, D. D., 1984. Possible role for a human adenovirus in the pathogenesis of celiac disease. *Gut*. 160, 1544–1557.

KAGNOFF, M. F., 2007. Celiac disease : pathogenesis of a model immunogenetic disease Find the latest version: Review series Celiac disease : pathogenesis of a model immunogenetic disease. *The journal of clinical Investigation*. 117(1), 41–49. Available at: doi:10.1172/JCI30253.severe

KAGNOFF, M. F., PATERSON, Y. J., KUMAR, P. J., KASARDA, D. D., CARBONE, F. R., UNSWORTH, D. J. and AUSTIN, R. K., 1987. Evidence for the role of a human intestinal adenovirus in the pathogenesis of coeliac disease. *Gut*. 28, 995–1001. Available at: doi:10.1136/gut.28.8.995

KAHRS, C. R., CHUDA, K., TAPIA, G., STENE, L. C., MÅRILD, K., RASMUSSEN, T., RØNNINGEN, K. S., LUNDIN, K. E. A., KRAMNA, L., CINEK, O. and STØRDAL, K., 2019. Enterovirus as trigger of coeliac disease: nested case-control study within prospective birth cohort. *BMJ*. 364, ISSN 1756-1833. Available at: doi:10.1136/bmj.l231

KATZ, M., 2006. *Multivariable analysis: a practical guide for clinicians.* 2nd ed. Cambridge University Press. ISBN 978-0521840514.

KEMPPAINEN, K. M., LYNCH, K. F., LIU, E., LÖNNROT, M., SIMELL, V., BRIESE, T., KOLETZKO, S., HAGOPIAN, W., REWERS, M., SHE, J.-X., SIMELL, O., TOPPARI, J., ZIEGLER, A.-G., AKOLKAR, B., KRISCHER, J. P., LERNMARK, Å., HYÖTY, H., TRIPLETT, E. W. and AGARDH, D., 2016. Factors That Increase Risk of Celiac Disease Autoimmunity After a Gastrointestinal Infection in Early Life. *Clinical Gastroenterology and Hepatology*. 15(5), 694–702. ISSN 1542-3565. Available at: doi:10.1016/j.cgh.2016.10.033

KIM, K. W., HORTON, J. L., PANG, C. N. I., JAIN, K., LEUNG, P., ISAACS, S. R., BULL, R. A., LUCIANI, F., WILKINS, M. R., CATTEAU, J., LIPKIN, W. I., RAWLINSON, W. D., BRIESE, T. and CRAIG, M. E., 2019. Higher abundance of enterovirus A species in the gut of children with islet autoimmunity. *Scientific Reports*. 9(1), 1–8. ISSN 2045-2322. Available at: doi:10.1038/s41598-018-38368-8

KUMAR, V., GUTIERREZ-ACHURY, J., KANDURI, K., ALMEIDA, R., HRDLICKOVA, B., ZHERNAKOVA, D. V., WESTRA, H. J., KARJALAINEN, J., RICAÑO-PONCE, I., LI, Y., STACHURSKA, A., TIGCHELAAR, E. F., ABDULAHAD, W. H., LÄHDESMÄKI, H., HOFKER, M. H., ZHERNAKOVA, A., FRANKE, L., LAHESMAA, R., WIJMENGA, C. and WITHOFF, S., 2015. Systematic annotation of celiac disease loci refines pathological pathways and suggests a genetic explanation for increased interferon-gamma levels. *Human Molecular Genetics*. 24(2), 397–409. ISSN 1460-2083. Available at: doi:10.1093/hmg/ddu453

LAHDEAHO, M. L., PARKKONEN, P., REUNALA, T., MAKI, M. and LEHTINEN, M., 1993. Antibodies to E1b Protein-Derived Peptides of Enteric Adenovirus Type 40 Are Associated with Celiac Disease and Dermatitis Herpetiformis. *Clinical Immunology and Immunopathology*. 69(3), 300–305. ISSN 0090-1229. Available at: doi:10.1006/clin.1993.1184

LAWLER, M., HUMPHRIES, P., O'FARRELLY, C., HOEY, H., SHEILS, O., JEFFERS, M., O'BRIAIN, D. S. and KELLEHER, D., 1994. Adenovirus 12 E1A gene detection by polymerase chain reaction in both the normal and coeliac duodenum. *Gut*. 35(9), 1226–1232. ISSN 0017-5749. Available at: doi:10.1136/gut.35.9.1226 Updated

LERNER, A., ARLEEVSKAYA, M., SCHMIEDL, A. and MATTHIAS, T., 2017. Microbes and Viruses Are Bugging the Gut in Celiac Disease. Are They Friends or Foes?. *Frontiers in Microbiology.* 8(5), 1–16. Available at: doi:10.3389/fmicb.2017.01392

LERNER, A., RAMESH, A. and MATTHIAS, T., 2019. The Revival of the Battle between David and Goliath in the Enteric Viruses and Microbiota Struggle: Potential Implication for Celiac Disease. *Microorganisms*. 7(6), 173. ISSN 2076-2607. Available at: doi:10.3390/microorganisms7060173

LERNMARK, Å., 2016. Environmental factors in the etiology of type 1 diabetes, celiac disease, and narcolepsy. *Pediatric Diabetes*. 17, 65–72. ISSN 1399-5448. Available at: doi:10.1111/pedi.12390

LOHI, S., MUSTALAHTI, K., KAUKINEN, K., LAURILA, K., COLLIN, P., RISSANEN, H., LOHI, O., BRAVI, E., GASPARIN, M., REUNANEN, A. and MÄKI, M., 2007. Increasing prevalence of coeliac disease over time. *Alimentary Pharmacology and Therapeutics*. 26(9), 1217–1225. ISSN 0269-2813. Available at: doi:10.1111/j.1365-2036.2007.03502.x

MAHON, J., BLAIR, G. E., WOOD, G. M., SCOTT, B. B., LOSOWSKY, M. S. and HOWDLE, P. D., 1991. Is persistent adenovirus 12 infection involved in coeliac disease? A search for viral DNA using the polymerase chain reaction. *Gut*. 32(10), 1114–6. ISSN 0017-5749. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1955164

MARIETTA, E., DAVID, C. and MURRAY, J., 2012. Important Lessons Derived From Animal Models of Celiac Disease. *Int Rev Immunology.* 30(4), 197–206. ISSN 1563-5244. Available at: doi:10.3109/08830185.2011.598978

MÅRILD, K., KAHRS, C., TAPIA, G. and STENE, L., STØRDAL, K., 2015. Infections and risk of celiac disease in childhood: a prospective nationwide cohort study. *The American Journal of Gastroenterology*. 110(10), 1475–1484. Available at: doi:10.1038/ajg.2015.287

MATZINGER, P., 2002. The Danger Model: A Renewed Sense of Self. *Science*. 296(5566), 301–306. Available at: doi:10.1126/science.1071059

MEGIORNI, F. and PIZZUTI, A., 2012. HLA-DQA1 and HLA-DQB1 in Celiac disease predisposition: Practical implications of the HLA molecular typing. *Journal of Biomedical Science*. 19(1). ISSN 1021-7770. Available at: doi:10.1186/1423-0127-19-88

MEJÍAS, J. H., LU, X., OSORIO, C., ULLMAN, J. L., VON WETTSTEIN, D. and RUSTGI, S., 2014. Analysis of wheat prolamins, the causative agents of celiac sprue, using reversed phase high performance liquid chromatography (RP-HPLC) and matrix-assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOF-MS). *Nutrients*. 6(4), 1578–1597. ISSN 2072-6643. Available at: doi:10.3390/nu6041578

NIX, W. A., OBERSTE, M. S. and PALLANSCH, M. A., 2006. Sensitive, seminested PCR amplification of VP1 sequences for direct identification of all enterovirus serotypes from original clinical specimens. *Journal of Clinical Microbiology*. 44(8), 2698–2704. ISSN 0095-1137. Available at: doi:10.1128/JCM.00542-06

OLIVARES, M. and SANZ, Y., 2015. Intestinal Microbiota and Celiac Disease. *Omnia Science.* 193–221. Available at: doi:10.3926/oms.253

REWERS, M., HYÖTY, H., LERNMARK, Å., SHE, J.-X., REWERS, M., HAGOPIAN, W., ZIEGLER, A.-G., AKOLKAR, B., KRISCHER, J., SCHATZ, D. and TOPPARI, J., 2018. The Environmental Determinants of Diabetes in the Young (TEDDY) Study: 2018 Update. *Current Diabetes Reports*. 18(12). ISSN 1534-4827. Available at: doi:10.1007/s11892-018-1113-2

ROBINSON, C. M., SINGH, G., LEE, J. Y., DEHGHAN, S., RAJAIYA, J., LIU, E. B., YOUSUF, M. A., BETENSKY, R. A., JONES, M. S., DYER, D. W., SETO, D. and CHODOSH, J., 2013. Molecular evolution of human adenoviruses. *Scientific Reports*. 3, 1812-1815. ISSN 2045-2322. Available at: doi:10.1038/srep01812

ROSSI, S., LUONGO, D., MAURANO, F., BERGAMO, P. and ROSSI, M., 2017. Immunomodulatory activity of recombinant α-gliadin conjugated to cholera toxin in DQ8 transgenic mice. *Immunology Letters*. 187(5), 47–52. ISSN 1879-0542. Available at: doi:10.1016/j.imlet.2017.05.004

SARMIENTO, L., GALVAN, J.A., CABRERA-RODE, E., AIRA, L., CORREA, C., SARIEGO, S., FONSECA, M.,CUBAS-DUEN, C., 2012. Type 1 diabetes associated and Tissue Transglutaminase Autoantibodies in Patients Without Type 1 Diabetes and Coeliac Disease With Confirmed Viral Infections. *Journal of Medical Virology*. 30(12), 4799–4804. ISSN 02507005. Available at: doi:10.1002/jmv

SCHMITT, H., SELL, S., KOCH, J., SEEFRIED, M., SONNEWALD, S., DANIEL, C., WINKLER, T. H. and NITSCHKE, L., 2016. Siglec-H protects from virus-triggered severe systemic autoimmunity. *The Journal of Experimental Medicine*. 213(8), 1627–1644. ISSN 0022-1007. Available at: doi:10.1084/jem.20160189

SINGH, P., ARORA, A., STRAND, T. A., LEFFLER, D. A., CATASSI, C., GREEN, P. H., KELLY, C. P., AHUJA, V. and MAKHARIA, G. K., 2018. Global Prevalence of Celiac Disease: Systematic Review and Meta-analysis. *Clinical Gastroenterology and Hepatology*. 16(6), 823–836. ISSN 1542-7714. Available at: doi:10.1016/j.cgh.2017.06.037

SMITS, S. L., VAN LEEUWEN, M., VAN DER EIJK, A. A., FRAAIJ, P. L. A., ESCHER, J. C., SIMON, J. H. and OSTERHAUS, A. D. M. E., 2010. Human astrovirus infection in a patient with new-onset celiac disease. *Journal of Clinical Microbiology*. 48(9), 3416–3418. ISSN 0095-1137. Available at: doi:10.1128/JCM.01164-10

SOLLID, L. M. and JABRI, B., 2013. Triggers and drivers of autoimmunity: lessons from coeliac disease. *Nature Reviews Immunology*. 13(4), 294–302. ISSN 1474-1733. Available at: doi:10.1038/nri3407

SOLLID, L. M. and LIE, B. A., 2005. Celiac disease genetics: Current concepts and practical applications. *Clinical Gastroenterology and Hepatology*. 3(9), 843–851. ISSN 1542-3565. Available at: doi:10.1016/S1542-3565(05)00532-X

STENE, L. C., HONEYMAN, M. C., HOFFENBERG, E. J., HAAS, J. E., SOKOL, R. J., EMERY, L., TAKI, I., NORRIS, J. M., ERLICH, H. A., EISENBARTH, G. S. and REWERS, M., 2006. Rotavirus infection frequency and risk of celiac disease autoimmunity in early childhood: A longitudinal study. *American Journal of Gastroenterology*. 101(10), 2333–2340. ISSN 0002-9270. Available at: doi:10.1111/j.1572-0241.2006.00741.x

STENE, L. C., WITSØ, E., TORJESEN, P. A., RASMUSSEN, T., MAGNUS, P., CINEK, O., WETLESEN, T. and RØNNINGEN, K. S., 2007. Islet autoantibody development during follow-up of high-risk children from the general Norwegian population from three months of age: Design and early results from the MIDIA study. *Journal of Autoimmunity*. 8., 29(1), 44–51. ISSN 0896-8411. Available at: doi:10.1016/j.jaut.2007.04.003

TAPIA, G., CINEK, O., RASMUSSEN, T., WITSØ, E., GRINDE, B., STENE, L. C. and RØNNINGEN, K. S., 2011. Human enterovirus RNA in monthly fecal samples and islet autoimmunity in Norwegian children with high genetic risk for type 1 diabetes: The MIDIA study. *Diabetes Care*. ISSN 0149-5992. Available at: doi:10.2337/dc10-1413

TARISH, H. R., HAMEED, W. S., ABDUL-MEHDI, R. J., ALI, H. and ALSHEREES, A., 2016. Role of Previous Adenovirus Infection and Its Association with IFN- α in Occurrence of Celiac Disease in Iraqi Patients Authors. *Journal of Medical Science And Clinical research*. 04(04), 10326–10330. ISSN 2455-0450. Available at: doi:10.18535/jmscr/v4i4.58

TIGHE, M. P. and BEATTIE, R. M., 2017. Is HLA testing for coeliac disease helpful in associated autoimmune conditions?. *BMJ.* 102(7), 595–596. Available at: doi:10.1136/archdischild-2017-312714

TIGHE, M. P. and BEATTIE, R. M., 2019. Searching for the missing link in coeliac disease. *BMJ.* 696(2), 1–2. Available at: doi:10.1136/bmj.l696

TINTO, N., COLA, A., PISCOPO, C., CAPUANO, M., GALATOLA, M., GRECO, L. and SACCHETTI, L., 2015. High frequency of haplotype HLA-DQ7 in celiac disease patients from south Italy: Retrospective evaluation of 5.535 subjects at risk of celiac disease. *PLoS ONE*. 10(9), 1–7. ISSN 1932-6203. Available at: doi:10.1371/journal.pone.0138324

TJERNBERG, A. R. and LUDVIGSSON, J. F., 2014. Children with celiac disease are more likely to have attended hospital for prior respiratory syncytial virus infection. *Digestive Diseases and Sciences*. ISSN 1573-2568. Available at: doi:10.1007/s10620-014-3046-1

TRYNKA, G., HUNT, K. A., BOCKETT, N. A., ROMANOS, J., CASTILLEJO, G., CONCHA, E. G. De and ALMEIDA, R. C. De, 2012. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nature Genetics.* 43(12), 1193–1201. Available at: doi:10.1038/ng.998.Dense

TYE-DIN, J. A., GALIPEAU, H. J. and AGARDH, D., 2018. Celiac Disease: A Review of Current Concepts in Pathogenesis, Prevention, and Novel Therapies. *Frontiers in Pediatrics*. 6, 350-356. Available at: doi:10.3389/fped.2018.00350

VISSER, J., ROZING, J., SAPONE, A., LAMMERS, K. and FASANO, A., 2009. Tight Junctions, Intestinal Permeability, and Autoimmunity Celiac Disease and Type 1 Diabetes Paradigms. *New York Academy of Sciences*. Available at: doi:10.1111/j.1749-6632.2009.04037.x

VOROBJOVA, T., RAIKKERUS, H., KADAJA, L., TALJA, I., UIBO, O., HEILMAN, K. and UIBO, R., 2017. Circulating Zonulin Correlates with Density of Enteroviruses and Tolerogenic Dendritic Cells in the Small Bowel Mucosa of Celiac Disease Patients. *Digestive Diseases and Sciences*. 62(2), 358–371. ISSN 1573-2568. Available at: doi:10.1007/s10620-016-4403-z

YI, E.-J., SHIN, Y.-J., KIM, J.-H., KIM, T.-G. and CHANG, S.-Y., 2017. Enterovirus 71 infection and vaccines. *Clinical and Experimental Vaccine Research*. 6(1). ISSN 2287-3651. Available at: doi:10.7774/cevr.2017.6.1.4

ZIBERNA, F., DE LORENZO, G., SCHIAVON, V., ARNOLDI, F., QUAGLIA, S., LEO, L., VATTA, S., MARTELOSSI, S., BURRONE, O. R., VENTURA, A. and NOT, T., 2016. Lack of Evidence of Rotavirus-Dependent Molecular Mimicry as a Trigger of Celiac Disease. *Clinical and experimental immunology*. 186(3). 356–363. ISSN 1365-2249. Available at: doi:10.1111/cei.12855

Internet sources

Adenoviridae - dsDNA Viruses - dsDNA Viruses - International Committee on Taxonomy of Viruses (ICTV) [online] [accessed. 2019-07-12]. Available at: https://talk.ictvonline.org/ictv-reports/ictv_9th_report/dsdna-viruses-2011/w/dsdna_viruses/93/adenoviridae

Geneious | Bioinformatics Software for Sequence Data Analysis [online] [accessed. 2019-08-11]. Available at: https://www.geneious.com/

Genus: Enterovirus - Picornaviridae - Picornavirales - International Committee on Taxonomy of Viruses (ICTV) [online] [accessed. 2019-07-12]. Available at: https://talk.ictvonline.org/ictv-reports/ictv_online_report/positive-sense-rna-viruses/picornavirales/w/picornaviridae/681/genus-enterovirus

Gluten Chemistry and Functionality – Food Science Toolbox [online] [accessed. 2019-07-19]. Available at: https://cwsimons.com/gluten-chemistry-and-functionality/

Mayo Clinic [online] [accessed. 2018-11-02]. Available at: www.mayoclinic.org

Stata: Software for Statistics and Data Science [accessed. 2019-02-02]. Available at: https://www.stata.com/

The molecular biology [online] [accessed. 2019-07-12]. Available at: https://sites.google.com/site/dsfoster/themolecularbiology

What's the deal with gluten? - Glutagen [online] [accessed. 2019-07-15]. Available at: https://glutagen.com/whats-the-deal-with-gluten/

World Population Clock: 7.7 Billion People - Worldometers [online] [accessed. 2019-08-03]. Available at: https://www.worldometers.info/world-population/

10 APPENDIX

10.1 Paper I

Enterovirus as trigger of coeliac disease: nested case-control study within prospective birth cohort

Christian R Kahrs,[1,2,3] Katerina Chuda,[4] German Tapia,[2] Lars C Stene,[2] Karl Mårild,[5] Trond Rasmussen,[6] Kjersti S Rønningen,[7] Knut E A Lundin,[8,9] Lenka Kramna,[4] Ondrej Cinek,[4] Ketil Størdal[1,2]

For numbered affiliation see end of the article

Correspondence to: K Størdal ketil.stordal@fhi.no (or @ketil_st on Twitter ORCID 0000-0001-7826-8646)

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2019;364:l231 http://dx.doi.org/10.1136/bmj.l231

Accepted: 14 January 2019

ABSTRACT

OBJECTIVE

To determine whether infection with human enterovirus or adenovirus, both common intestinal viruses, predicts development of coeliac disease.

DESIGN

Case-control study nested within Norwegian birth cohort recruited between 2001 and 2007 and followed to September 2016.

SETTING

Norwegian population.

PARTICIPANTS

Children carrying the HLA genotype DR4-DQ8/DR3-DQ2 conferring increased risk of coeliac disease.

EXPOSURES

Enterovirus and adenovirus detected using real time polymerase chain reaction in monthly stool samples from age 3 to 36 months.

MAIN OUTCOME MEASURE

Coeliac disease diagnosed according to standard criteria. Coeliac disease antibodies were tested in blood samples taken at age 3, 6, 9, and 12 months and then annually. Adjusted odds ratios from mixed effects logistic regression model were used to assess the relation between viral infections before development of coeliac disease antibodies and coeliac disease.

RESULTS

Among 220 children, and after a mean of 9.9 (SD 1.6) years, 25 children were diagnosed as having coeliac disease after screening and were matched to two controls each. Enterovirus was found in 370 (17%) of 2135 samples and was significantly more frequent in samples collected before development of coeliac disease antibodies in cases than in controls (adjusted odds ratio 1.49, 95% confidence interval 1.07 to 2.06; P=0.02). The association was restricted to infections after introduction of gluten. High quantity samples (›100 000 copies/µL) (adjusted odds ratio 2.11, 1.24 to 3.60; P=0.01) and long lasting infections (›2 months) (2.16, 1.16 to 4.04; P=0.02) gave higher risk estimates. Both the commonly detected enterovirus species *Enterovirus A* and *Enterovirus B* were significantly associated with coeliac disease. The association was not found for infections during or after development of coeliac disease antibodies. Adenovirus was not associated with coeliac disease.

CONCLUSIONS

In this longitudinal study, a higher frequency of enterovirus, but not adenovirus, during early childhood was associated with later coeliac disease. The finding adds new information on the role of viral infections in the aetiology of coeliac disease.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Recent evidence suggests that viral infections are involved in an animal model of coeliac disease

Prospective studies have shown a higher prevalence of infections in children before diagnosis of coeliac disease

Previous studies of adenoviruses and enteroviruses have been limited to cross sectional designs, and reverse causality is a possible explanation for these observations

WHAT THIS STUDY ADDS

Children with increased genetic risk of coeliac disease had a higher frequency of enteroviruses in stool samples before development of coeliac disease than did healthy controls

Adenovirus was not associated with coeliac disease

Identification of specific viruses as triggers of coeliac disease may have implications for preventive strategies and justify future studies to clarify mechanisms

Introduction

Coeliac disease is an immune mediated disease believed to result from gluten intake and unknown environmental trigger factors in genetically susceptible individuals.[1] Coeliac disease develops almost exclusively in people with the HLA-DQ2 (*DQA1*05:01-DQB1*02:01*) or HLA-DQ8 (*DQA1*03-DQB1*03:02*) haplotype. The HLA-DQ2 and HLA-DQ8 haplotypes also occur in about 40% of the general population. Although non-HLA genetic variants have also been identified, the predictive value of these variants, even in combination, is limited.[2 3] As most children are exposed to gluten, this suggests that additional genetic and environmental triggers are involved in development of coeliac disease.[4] The autoimmune process may start months or years before clinical manifestations,[5] and the identification and confirmation of environmental triggers remain a major challenge.[6]

Both experimental studies and epidemiological studies based on parental reporting of illness suggest a role for infections in the development of coeliac disease, particularly gastrointestinal infections.[6-9] Gastrointestinal infections are common in childhood and may impair the mucosal barrier for transfer of dietary proteins as gluten regardless of the presence of clinical symptoms.[10 11] The only prospective study on coeliac disease and viral infection suggested that frequent rotavirus infections might increase the risk of development of coeliac disease antibodies in a cohort at high risk.[12] Other studies with retrospective

designs have studied adenovirus, enterovirus, and orthoreovirus as potential triggers of coeliac disease with conflicting or inconclusive results.[9 13-15] Study designs including patients diagnosed as having coeliac disease and controls may be subject to bias due to reverse causality. Therefore, long term follow-up of longitudinal birth cohorts is needed.

In this study, we approached the question of potential gastrointestinal triggers by using a longitudinal birth cohort analysis of the most frequently occurring viruses: enterovirus and adenovirus. We chose these two viruses on the basis of a pilot screening of the stool virome in the cohort members. We aimed to test whether the presence of human enterovirus and adenovirus in monthly faecal samples was more common before development of coeliac disease antibodies in cases subsequently diagnosed as having coeliac disease compared with children not developing the disease.

Methods

Participants and study design

We designed a nested case-control study of coeliac disease within a birth cohort of children with the HLA-DQ2/DQ8 genotype, with monthly stool samples tested for nucleic acid of enterovirus and adenovirus. Figure 1 outlines the enrolment of the study sample, and figure 2 gives an overview of the study design. The MIDIA study, which was originally designed to study type 1 diabetes, is described in detail by Stene et al.[16]

Briefly, during 2001-07, 46 939 newborns throughout Norway were screened for the HLA-DQ2/DQ8 genotype conferring an increased risk of both coeliac disease and type 1 diabetes (HLA-*DRB1*04:01-DQA1*03-DQB1*03:02/DRB1*03-DQA1*05:01-DQB1*02:01*). This risk genotype was identified in 912 (1.9%) children, who were followed with repeated blood and faecal samples from the age of 3 months. We collected plasma samples at age 3, 6, 9, and 12 months and annually thereafter. Monthly stool samples collected at age 3-36 months were diluted in preservation buffer, centrifuged, and supernatant separated. All samples were stored at −80°C until testing.[17]

Children who still actively contributed with blood samples during 2014-16 (n=501) were invited to coeliac disease screening, of whom 220 consented to participate. These participants tended to have a higher prevalence of type 1 diabetes, fewer siblings, and slightly higher prevalence of family history of coeliac disease at initial recruitment around 3 months of age, than the whole cohort (supplementary table A).

Case definition and selection of matched controls

From the 220 consenting participants, we first identified those who had coeliac disease by measuring serological markers of coeliac disease in their most recent sample and by review of clinical documentation, followed by an invitation for a confirmatory blood sample and for standard paediatric diagnostic investigation for coeliac disease. We identified previously diagnosed cases by using a parental questionnaire followed by a review of medical files. For an overview of the coeliac screening process, see supplementary figure A. We diagnosed coeliac disease according to the European Society for Paediatric Gastroenterology Hepatology and Nutrition (ESPGHAN) 2012 criteria.[18] By this procedure, we identified 27 cases of coeliac disease by February 2016 and confirmed that the remaining 193 participants were free from coeliac disease. Mean age at end of follow-up was 9.9 (SD 1.6) years. After evaluation for eligibility, we matched 25 cases (16 girls and 9 boys; 11% of participants) with coeliac disease to two controls each. Matching was done for duration of follow-up, date of birth, and county of residence. We excluded one matched control from analysis owing to missing stool samples. See table 1 for characteristics of the nested cases and controls. For more details about case definition, see supplementary methods.

We then determined the time interval when cases seroconverted for coeliac disease markers by retrospectively analysing biobanked plasma samples that had been collected longitudinally since age 3 months, searching for the last sample that was negative for the serological markers and the first sample indicative of coeliac disease. Of the 25 case-control groups, 15 had this seroconversion period covered by monthly stool sampling, whereas the remaining 10 seroconverted after the collection of stools was terminated (after 36 months of age). For characteristics of the children with coeliac disease, see supplementary table B.

Serological testing for coeliac disease

IgA anti-tissue transglutaminase and IgG anti-deamidated gliadin peptide (EliA Celikey IgA/EliA GliadinDP IgG, Thermo Fischer Scientific, Phadia AB; Uppsala, Sweden) were measured in the most recent blood sample from each participant at the laboratory of Akershus University Hospital, Lørenskog, Norway. Mean age at screening was 8.6 (SD 2.0) years. Screening with IgA anti-tissue transglutaminase in combination with IgG anti-deamidated gliadin peptide should better capture people with IgA deficiency and samples in which tissue transglutaminase concentrations may have been influenced by haemolysis.[19-21] In accordance with a previous screening study and to increase sensitivity,[22] we defined IgA anti-tissue transglutaminase values of 3 U/mL or higher as positive. For IgG anti-deamidated gliadin peptide, we used 7 U/mL as the cut-off value in accordance with the manufacturer's instruction. We contacted children with positive serological results (n=20) for a second blood sample (Varelisa tissue transglutaminase IgA, Phadia, cut-off <3 U/mL (<5 U/mL until 2015); IgG anti-deamidated gliadin peptide, QUANTA Lite Gliadin IgG IIInova, cut-off <20 U/mL) which was analysed at Oslo University Hospital Ullevål, Oslo, Norway.

Detection of enterovirus and adenovirus in stool samples

All available faecal samples from cases and controls were subjected to RNA and DNA extraction using

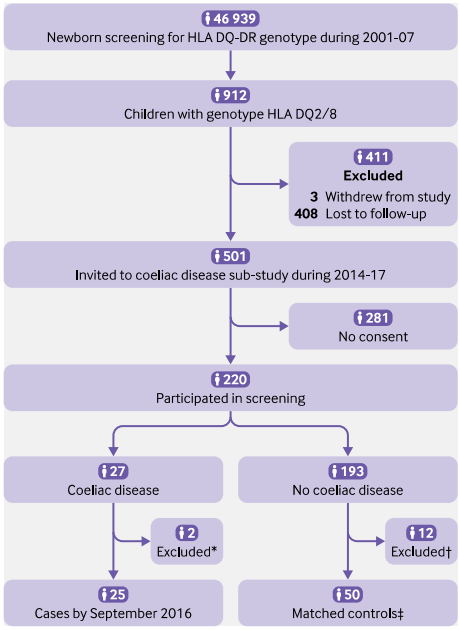


Fig 1 | Enrolment of study sample. "Invited" children are those who were still actively participating (delivering samples, questionnaires) at start of coeliac disease sub-study and were invited to participate; "participated" are those who consented to participate and were screened for coeliac disease antibodies. *Two cases were excluded from analyses, one owing to missing stool samples and one owing to incorrect use of diagnostic criteria. †Children with positive coeliac disease antibodies at first screening test but for whom follow-up did not confirm coeliac disease diagnosis (n=10) were excluded from further analyses; control children with high antibody titres (>10 times cut-off) (n=2) in old samples but with normal values at screening were excluded and replaced by new controls selected using same matching criteria. ‡One matched control was excluded from analysis owing to missing stool samples

Qiagen chemistry (Qiagen, Hilden, Germany). We tested enterovirus by quantitative real time polymerase chain reaction (PCR).[23] The primers detected with an equal sensitivity the *Enterovirus A-D* species (that is, members of species *Enterovirus A*, *Enterovirus B*, *Enterovirus C*, and *Enterovirus D* of the genus *Enterovirus*, family *Picornaviridae*, order *Picornavirales;* the former nomenclature of these species was *Human enterovirus A-D*) but did not react with human rhinoviruses (members of species *Rhinovirus A-C*). Adenovirus was tested using a previously published real time PCR assay.[24] The threshold of enterovirus and adenovirus positivity was set to 10 copies/µL nucleic acid. We tested 2161 stool samples, of which 2135 and 2006 provided data on quantity of enterovirus and adenovirus, respectively. All samples were tested blinded as to the case-control status.

All samples with more than 100 copies/µL of enterovirus were subjected to genotyping with Sanger or next generation amplicon sequencing of PCR amplicons of the VP1 gene segment informative for the virus type. Samples with more than 10 copies/µL of adenovirus were genotyped by a similar protocol amplifying the seventh hypervariable region of its hexon gene. Primers, probes, and a detailed protocol have been described in Cinek et al.[25] See also supplementary methods for more details.

Statistical analysis

We analysed the association of virus with coeliac disease primarily by using a mixed effects logistic regression model with random intercepts for each matched set and for each individual child, thus accounting for the matched design. We used faecal sample virus positivity as the dependent variable (separate models for enterovirus and adenovirus) and case-control status as an independent variable. The odds ratio for coeliac disease status is then interpreted as the odds of a faecal sample being positive for virus given that it came from a child who later developed coeliac disease, compared with the same odds for virus positivity for samples from matched controls. We included the following potential predictors of viral infection and coeliac disease as covariates in the main regression model: sex, age, age squared, season of sample collection, number of siblings (categorised as none, 1, or ≥2), and family history of coeliac disease. In the primary analysis, we included only stool samples from the matched sets collected before the case child developed coeliac disease antibodies to avoid reverse causality (fig 3).

In pre-planned sub-analyses, we explored the association between enterovirus and adenovirus and subsequent coeliac disease for stool samples collected during and after the development of coeliac disease

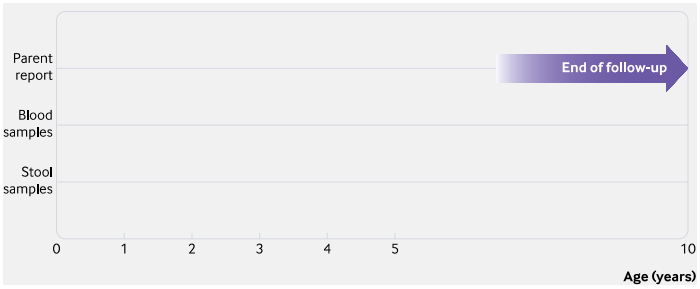


Table 1 | Characteristics of nested cases and controls. Values are numbers (percentages) unless stated otherwise

Characteristics	Cases (n=25)	Controls (n=49)
Female sex	16 (64)	26 (53)
Other children in household:		
None	7 (28)	15 (31)
1	10 (40)	28 (57)
≥2	8 (32)	6 (12)
Type 1 diabetes	4 (16)	1 (2)
Family history of coeliac disease*	7 (28)	4 (8)
Mean (SD) age (months) at antibody screening test†	99 (30)	106 (21)
Mean (SD) age (months) at end of follow-up‡	119 (19)	119 (19)
Mean (SD) age (months) at last stool sample	32 (7)	34 (3)
Stool samples§	703	1458
Enterovirus:		
Stool samples providing enterovirus data	690 (98)	1445 (99)
Enterovirus positive stool samples	135 (20)	235 (16)
Median (range) count of positive samples per child	6 (0-11)	4 (0-9)
Adenovirus:		
Stool samples providing adenovirus data	649 (92)	1357 (93)
Adenovirus positive samples	85 (13)	173 (13)
Median (range) count of positive samples per child	4 (0-9)	3 (0-11)
Blood samples§:		
Total	326	593
Samples providing coeliac disease antibody data¶	259 (79)	374 (63)
Coeliac disease antibody positive samples¶	131 (51)	7 (2)

*Known coeliac disease in first degree relative or half-sibling ascertained at coeliac disease screening 2014-16.
†Antibody screening test performed on both cases and controls at time of inclusion in coeliac disease sub-study of MIDIA.
‡At time when all 25 cases were ascertained by end of February 2016.
§Discrepancy between number of stool samples and samples providing virus data and number of blood samples and samples providing data on coeliac disease antibodies was due to missing samples or failed laboratory test.
¶Antibodies to tissue transglutaminase.

antibodies (fig 3). We analysed the samples before development of coeliac disease antibodies, defining the exposure as number of infectious episodes, counting a sequence of consecutively virus positive faecal samples as a single episode (a negative stool sample is required before a new episode is defined). We speculated that a higher quantity of enterovirus, longer duration of viral shedding, or symptomatic infections would have a greater effect on development of coeliac disease. Therefore, we analysed infections with reported symptoms and defining exposures as long infectious episodes (at least two positive consecutive monthly samples) and per infection with high quantity (≥100 000 copies per µL nucleic acid). We investigated specific virus types, using an a priori determined threshold of at least 20 positive samples to be analysed. We grouped enteroviruses into *Enterovirus A-D* species as detailed above. We grouped adenovirus types only into specific types (for example, human adenovirus C2), as the species *Adenovirus C* dominated in our samples.

Additionally, we adjusted the primary analysis for the timing of introduction of gluten and breast feeding. As exploratory analyses, we investigated time periods less than six months, six to 12 months, more than 12 months, after the end of breast feeding, and after the introduction of gluten. We also investigated whether infectious symptoms, as reported by parents in longitudinal questionnaires in early life, were associated with coeliac disease antibodies and whether specific symptoms were linked to infections. In sensitivity analyses, we also adjusted for type 1 diabetes.

Patient and public involvement

Patients were not involved in setting the research question or the outcome measures, nor were they involved in developing plans for recruitment, design, or implementation of the study. No patients were asked to advise on interpretation or writing up of results. We will disseminate the results of the research to study participants and the general public.

Results

Enterovirus for whole observational period

We detected enterovirus in 370 (17%) of 2135 stool samples (table 1), with 73 children having at least one positive sample. The distribution of enterovirus in cases and controls of the 25 matching groups is shown in supplementary figure B. Enterovirus showed variation with age and season, with a peak in autumn (supplementary figures C and D). Several different

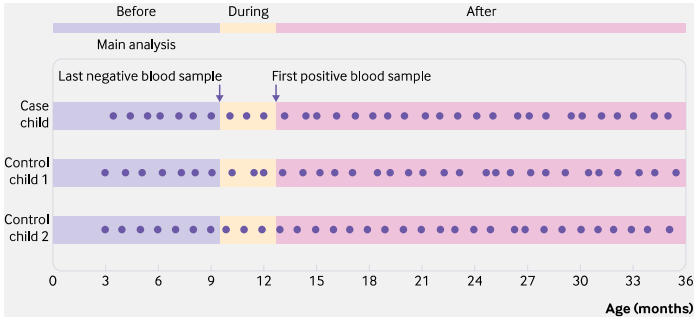


Fig 3 | Development of coeliac disease antibodies. Figure illustrating definition of before, during, and after development of coeliac disease antibodies. Dots represent faecal samples. Primary analysis included only stool samples collected up to age at last antibody negative blood sample for cases and corresponding age for matched controls. Pre-planned sub-analyses explored same association for stool samples collected during development of coeliac disease antibodies (between last negative and first antibody positive blood sample) and samples collected after first antibody positive blood sample

doi: 10.1136/bmj.l231 | *BMJ* 2019;364:l231 | the bmj

types were detected, occasionally as co-infections, with coxsackievirus A2 and A4 being most common (supplementary table C).

Enterovirus and coeliac disease

The frequency of enterovirus positive stool samples before development of coeliac disease antibodies was 84/429 (20%) in cases and 129/855 (15%) in matched controls (adjusted odds ratio 1.49, 95% confidence interval 1.07 to 2.06; P=0.02) (table 2). The adjusted odds ratio was 2.11 (1.24 to 3.60; P=0.01) for high quantity samples (>100 000 copies/µL), 2.16 (1.16 to 4.04; P=0.02) for long lasting infections (more than two months), and 1.27 (0.87 to 1.86; P=0.21) for infectious episodes (consecutive positive samples counted as a single episode) (supplementary table D). The frequency of enterovirus in stool samples during or after development of coeliac disease antibodies was not associated with coeliac disease (table 2).

Exploratory analyses of enterovirus and coeliac disease

Both the commonly identified enterovirus species *Enterovirus A* (adjusted odds ratio 1.62, 1.04 to 2.53; P=0.03) and *Enterovirus B* (2.27, 1.33 to 3.88; P=0.003) were significantly associated with later coeliac disease. *Enterovirus C* and *Enterovirus D* were detected in few or no samples (table 2).

Enterovirus infections after the first year of life showed increased estimates, whereas infections from age 3 to 6 months or from 6 to 12 months did not. Enterovirus infections after introduction of gluten were associated with coeliac disease, whereas infections before or at the time of gluten introduction were not (supplementary table E). Similarly, infections after the end of breast feeding were associated with coeliac disease, but enterovirus infections during breast feeding were not. Enterovirus infections in shorter time windows before the last sample negative for coeliac disease antibody showed borderline increased estimates (supplementary table E). We found no association between reported infectious symptoms and coeliac disease or between infectious symptoms and enterovirus positivity (supplementary tables F and G). Enterovirus positivity at the same time of reported fever or diarrhoea was associated with development of coeliac disease, but other symptoms were not (supplementary table F). Adjustment for type 1 diabetes only marginally changed the estimates (adjusted odds ratio 1.43, 1.02 to 2.02).

Adenovirus and coeliac disease

We detected adenovirus in 258 (13%) of 2006 faecal samples (table 1), with 61 children having at least one positive sample. Adenovirus showed variation with age but did not follow a seasonal pattern (supplementary figures C and D). Several different types were detected, occasionally as co-infections, with human adenovirus C2 and C1 being most common (supplementary table H).

We found no significant difference between cases and controls in adenovirus positive stool samples before (adjusted odds ratio 0.82, 0.49 to 1.38; P=0.46), during, or after development of coeliac disease antibodies. No adenovirus types were associated with development of coeliac disease (table 3). Adenovirus was associated with fever, but adenovirus positive samples with reported symptoms were not associated with development of coeliac disease (supplementary table F).

Discussion

We found a significant association between exposure to enterovirus and subsequent risk of coeliac disease. Adenovirus was not associated with coeliac disease.

Strengths and limitations of study

This is the first population based study on viruses in stool samples collected longitudinally before development of coeliac disease antibody markers in children later diagnosed as having coeliac disease. We

Table 2 | Enterovirus positive stool samples and subsequent coeliac disease*

	Positive samples/total samples (%)			Adjusted†	
	Cases (n=25)	Controls (n=49)	Unadjusted odds ratio (95% CI)	Odds ratio (95% CI)†	P value
Before development of coeliac disease antibodies‡					
All enteroviruses	84/429 (20)	129/855 (15)	1.37 (1.01 to 1.87)	1.49 (1.07 to 2.06)	0.02
Enterovirus A‡	42/387 (11)	60/787 (8)	1.47 (0.97 to 2.23)	1.62 (1.04 to 2.53)	0.03
CV-A2	10/355 (3)	10/737 (1)	2.11 (0.87 to 5.11)	2.29 (0.89 to 5.87)	0.09
CV-A4	5/350 (1)	11/738 (1)	0.96 (0.33 to 2.78)	0.93 (0.30 to 2.86)	0.90
CV-A6	7/352 (2)	9/736 (1)	1.64 (0.60 to 4.44)	1.59 (0.56 to 4.57)	0.39
CV-A10	4/349 (1)	10/737 (1)	0.84 (0.25 to 2.76)	1.03 (0.30 to 3.53)	0.96
CV-A16	4/349 (1)	7/734 (1)	1.20 (0.35 to 4.14)	1.39 (0.36 to 5.30)	0.63
Enterovirus B‡	34/379 (9)	43/770 (6)	1.74 (1.08 to 2.81)	2.27 (1.33 to 3.88)	0.003
During/after development of coeliac disease antibodies§					
All enteroviruses during	30/140 (21)	58/332 (17)	1.24 (0.74 to 2.10)	1.26 (0.66 to 2.39)	0.48
All enteroviruses after	21/121 (17)	48/258 (19)	0.92 (0.52 to 1.62)	0.82 (0.44 to 1.53)	0.54

*Number of positive/total stool samples and results from mixed effects logistic regression are given for cases and controls at different time periods. In analysis for specific genotypes, samples positive for other enteroviruses have been set to missing. *Enterovirus C* and *Enterovirus D* were too few for a meaningful analysis and are not presented.
†Adjusted for sex, age, age squared, season of sample collection, number of siblings, and family history of coeliac disease.
‡Only enterovirus types found in more than 10 samples before development of coeliac disease antibodies are presented.
§Before: before last coeliac disease antibody negative blood sample. During: between last coeliac disease antibody negative and first coeliac disease antibody positive blood sample. After: after first coeliac disease antibody positive blood sample.

Table 3 | Adenovirus positive stool samples and subsequent coeliac disease*

	Positive samples/total samples (%)			Adjusted†	
	Cases (n=25)	Controls (n=49)	Unadjusted odds ratio (95% CI)	Odds ratio (95% CI)	P value
Before development of coeliac disease antibodies‡					
All adenoviruses	47/390 (12)	111/775 (14)	0.82 (0.50 to 1.35)	0.82 (0.49 to 1.38)	0.46
HAdV-C1	17/360 (5)	60/724 (8)	0.65 (0.19 to 2.19)	0.66 (0.19 to 2.29)	0.51
HAdV-C2	20/363 (6)	43/707 (6)	0.88 (0.43 to 1.80)	0.84 (0.38 to 1.87)	0.67
HAdV-C5	5/348 (1)	7/671 (1)	1.42 (0.29 to 6.96)	2.22 (0.52 to 9.57)	0.28
During/after development of coeliac disease antibodies‡					
All adenoviruses during	26/139 (19)	36/325 (11)	1.84 (0.89 to 3.79)	1.76 (0.87 to 3.56)	0.12
All adenoviruses after	12/120 (10)	26/257 (10)	0.99 (0.41 to 2.39)	1.29 (0.47 to 3.52)	0.62

HAdV=human adenovirus.
*Number of positive/total stool samples and results from mixed effects logistic regression are given for cases and controls at different time periods. In analysis for specific genotypes, samples positive for other adenoviruses have been set to missing. Adenovirus types were grouped into only specific genotypes (eg, HAdV-C2), as species *Adenovirus C* dominated in samples.
†Adjusted for sex, age, age squared, season of sample collection, number of siblings, and family history of coeliac disease.
‡Before: before last coeliac disease antibody negative sample. During: between last coeliac disease antibody negative and first coeliac disease antibody positive blood sample. After: after first coeliac disease antibody positive blood sample.

used accepted diagnostic criteria for coeliac disease. The development of coeliac disease antibodies, which is believed to be the first sign of disease, may start months or even years before clinical manifestations and diagnosis. Importantly, we analysed exposure before development of coeliac disease antibodies and thus avoided the potential reverse causation that may bias studies of infections at or after clinical diagnosis of coeliac disease. The long time period from development of coeliac disease antibodies to diagnosis of coeliac disease makes it highly unlikely that infections before development of coeliac disease antibodies prompted symptoms and diagnostic investigation for coeliac disease. Enterovirus positivity and type distribution in our study was in line with previous studies.[26] We used a PCR technique, which is a highly sensitive method for detection of virus locally in the gut but does not measure the immune response as serological testing does.[27] We believe that PCR was most appropriate in our study, as most infections were asymptomatic and current enterovirus serology not is designed for this kind of setting. We used a quantitative assay that enabled us to follow the dynamics of the viral load. Regular monthly stool sampling and high completeness were important strengths, because duration of viral shedding is expected to be around three to four weeks.[28 29] Administration of oral poliovirus vaccine may increase the detection of enterovirus in stool samples, but this cohort received inactivated poliovirus vaccine injections.[30 31]

Our study also had some limitations. The participants were followed up for about 10 years, and some children (particularly among those with the shortest follow-up) would be likely to be diagnosed as having coeliac disease after that age. Even though we analysed a large number of samples, the number of children with coeliac disease was limited. Furthermore, the limited sample size for some of the sub-analyses could have led to spurious associations. Loss to follow-up and modest response rate is inevitable in most cohort studies and also raises the possibility of selection bias. The nested cases and controls tended to have a higher prevalence of type 1 diabetes, fewer siblings, and a slightly higher frequency of family history of coeliac disease than the whole cohort. This could have led to non-representativeness, but it does not necessarily bias exposure-outcome associations.[32] The study was originally designed to study type 1 diabetes, with inclusion of only children with the HLA-DQ2/DQ8 genotype. Although we cannot generalise our findings beyond this genotype, we note that this genotype accommodates the two well established susceptibility haplotypes (DQ2 and DQ8) present in nearly all patients. We therefore believe that our findings are likely to apply to a sizeable proportion of patients with coeliac disease. Furthermore, the frequency of faecal enterovirus shedding did not differ significantly between people with the HLA-DQ2/DQ8 genotype and other genotypes in a previous study.[33] We did our best to adjust for potential confounding factors, and we are not aware of any obvious confounding factors that could plausibly explain our findings, but unmeasured confounding factors or residual confounding can never be entirely ruled out in non-randomised studies. Further studies in other settings would therefore be necessary to corroborate our findings.

Comparison with other studies

Coeliac disease and type 1 diabetes have several common features, and viral infections are candidate triggers in both diseases. In type 1 diabetes, enterovirus infections before islet autoimmunity, as well as persistent enterovirus infections in the gut mucosa or pancreatic tissue, have been described.[34-36] The single study on enterovirus and coeliac disease was cross sectional but could not find persistent viral infection in intestinal biopsies from untreated coeliac disease patients.[15] The study did not investigate signs of previous exposure to enterovirus. Finally, adjusting for type 1 diabetes in our analyses yielded similar results. This is consistent with the lack of association between enterovirus and diabetes autoimmunity in the same cohort.[37]

Adenoviruses have been proposed to precipitate coeliac disease via molecular mimicry.[13 38 39] However, later studies did not confirm the association.[14 40] Our negative findings are in line with the latter studies and

suggest that adenovirus in general does not influence risk of coeliac disease. An important strength of our study was the prospective design, whereas the studies mentioned above were cross sectional. Although we cannot rule out moderate effects from our data, the 95% confidence interval around the odds ratios estimated from our data suggest that strong associations with adenovirus infections are unlikely.

Interpretation

Although the effect sizes are relatively small, this study suggests that infections with enterovirus in early life could be one among several key risk factors for development of a disease with lifelong consequences. Our observations suggest that, rather than a specific enterovirus species or type driving the association, several enterovirus types, high titre, and long duration infections in the period after introduction of gluten were involved. Collectively, our results are compatible with a mechanism whereby viral infections may disrupt the mucosal barrier with increased translocation of gluten peptides into the mucosa as the initial event in the loss of tolerance.[41] We speculate that enteroviruses may provide a danger signal that activates dendritic cells acting as antigen presenting cells for CD4 positive gluten reactive T cells in the presence of transglutaminase modified gluten peptides.[9 42 43]

Patients with coeliac disease may have enteric barrier disruption before the development of autoantibodies and thus a susceptibility to enterovirus. However, we believe a more plausible explanation is that enterovirus causes impaired barrier function, which in turn increases the risk of coeliac disease. A challenge in studying the temporal association is the potential separation in time between the triggering event and disease onset as marked by appearance in the blood of coeliac disease antibodies. The results from our study suggest a certain time window between these events. As stool sampling was restricted to the period between birth and 3 years of age, we are lacking data just before the development of coeliac disease antibodies in a few cases diagnosed after 4 years of age.

Our study design is less sensitive to detection of infections occurring with lower frequency and shorter duration of viral shedding (rotavirus, orthoreovirus), so a non-specific response to several virus infections should be further investigated. If enterovirus is confirmed as a trigger factor, vaccination could reduce the risk of development of coeliac disease. Currently, except for poliovirus, non-polio enterovirus vaccines are not commercially available, but efforts are ongoing to develop such vaccines.[27 44] Further research in larger sample sets are needed to confirm our results.

Conclusion

In this longitudinal study, we found that a higher frequency of enterovirus infections was associated with increased risk of coeliac disease. Given the limited number of cases, we call for corroboration in similar studies and preferably interventional studies to reach conclusions about causality.

AUTHOR AFFILIATIONS

[1]Department of Pediatrics, Østfold Hospital Trust, Grålum, Norway
[2]Norwegian Institute of Public Health, Oslo, Norway
[3]Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway
[4]Department of Paediatrics, 2nd Faculty of Medicine, Charles University in Prague and University Hospital Motol, Prague, Czech Republic
[5]Department of Pediatrics, Institute of Clinical Sciences, The Sahlgrenska Academy at University of Gothenburg and Queen Silvia Children's Hospital, Gothenburg, Sweden
[6]Department of IT and e-health, Division of Institute Resources, Norwegian Institute of Public Health, Oslo, Norway
[7]Department of Pediatric Research, Oslo University Hospital, Oslo, Norway
[8]Department of Gastroenterology, Oslo University Hospital Rikshospitalet, Oslo, Norway
[9]K.G. Jebsen Coeliac Disease Research Centre, University of Oslo, Oslo, Norway

We express our sincere gratitude to the participants. We also especially thank Nicolai Andre Lund-Blix of Oslo University Hospital, Norway, for his help with the breastfeeding data; Lisa Merete Markussen of Akershus University Hospital for help with analysing coeliac disease antibodies in historical blood samples; Kaja Klykken Aas and Nina Stensrud from the Norwegian Institute of Public Health for handling plasma and stool samples in the MIDIA project; and the public health nurses in the MIDIA project (Asbjørg Skorge Hornseth, Liv Kjeldstadli Onsrud, and Turid Wetlesen). Adriana Lourenco Caixinha, Universidade de Lisboa, is acknowledged for expert technical help in nucleic acid extraction during her IFMSA internship.

Contributors: CRK and KS coordinated the coeliac disease sub-study in MIDIA, wrote the analysis plans, and had the primary responsibility for writing the paper. KS and LCS supervised the study, interpreted the data, and reviewed and commented on all drafts. GT did the statistical analysis, contributed with reagents/materials/analysis tools, and reviewed and commented on drafts. OC supervised the laboratory testing of virus PCR, contributed with analysis tools, and reviewed and commented on the data analysis. KC and LK did the laboratory testing of virus PCR and reviewed and commented on drafts. KSR designed the MIDIA study and reviewed and commented on the data analyses. TR, KL, and KM reviewed and contributed to the data processing. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted. KS is the guarantor.

Funding: The Research Council of Norway (grant numbers 205086/F20 to GT, 166515/V50 to KSR) and the Project for the Conceptual Development of Research Organization 00064203 (University Hospital Motol, Prague, Czech Republic). CRK received grant support from Østfold Hospital Trust, Kalnes, Norway, and KS received grant support from Oak Foundation, Geneva, Switzerland. The EliA Celikey and EliA Gliadin kits were supported by Thermo Fisher Scientific, Norway. The Norwegian Coeliac Society supported the study financially. The funders had no influence on the contents of this paper.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare: no support from any organisation for the submitted work other than that described above; 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.

Ethical approval: The study was approved by the Regional Committee for Medical Research Ethics. Written informed consent was obtained from the parents.

Data sharing: The individual raw data are provided in supplementary figure B. The Stata codes for all data handling and analyses are available on request from the corresponding author. Norwegian legislation prevents publication of the full dataset, but data supporting the presented results are available from the authors on reasonable request.

Transparency: The lead author (the manuscript's guarantor) affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

1 Green PH, Cellier C. Celiac disease. *N Engl J Med* 2007;357:1731-43. doi:10.1056/NEJMra071600
2 Romanos J, Rosén A, Kumar V, et al, PreventCD Group. Improving coeliac disease risk prediction by testing non-HLA variants additional to HLA variants. *Gut* 2014;63:415-22. doi:10.1136/gutjnl-2012-304110
3 Mårild K, Vistnes M, Tapia G, et al. Midpregnancy and cord blood immunologic biomarkers, HLA genotype, and pediatric celiac disease. *J Allergy Clin Immunol* 2017;139:1696-8. doi:10.1016/j.jaci.2016.10.016
4 Liu E, Lee HS, Aronsson CA, et al, TEDDY Study Group. Risk of pediatric celiac disease according to HLA haplotype and country. *N Engl J Med* 2014;371:42-9. doi:10.1056/NEJMoa1313977
5 Agardh D, Lee HS, Kurppa K, et al, TEDDY Study Group. Clinical features of celiac disease: a prospective birth cohort. *Pediatrics* 2015;135:627-34. doi:10.1542/peds.2014-3675
6 Abadie V, Sollid LM, Barreiro LB, Jabri B. Integration of genetic and immunological insights into a model of celiac disease pathogenesis. *Annu Rev Immunol* 2011;29:493-525. doi:10.1146/annurev-immunol-040210-092915
7 Mårild K, Kahrs CR, Tapia G, Stene LC, Størdal K. Infections and risk of celiac disease in childhood: a prospective nationwide cohort study. *Am J Gastroenterol* 2015;110:1475-84. doi:10.1038/ajg.2015.287
8 Kemppainen KM, Lynch KF, Liu E, et al, TEDDY Study Group. Factors That Increase Risk of Celiac Disease Autoimmunity After a Gastrointestinal Infection in Early Life. *Clin Gastroenterol Hepatol* 2017;15:694-702.e5. doi:10.1016/j.cgh.2016.10.033
9 Bouziat R, Hinterleitner R, Brown JJ, et al. Reovirus infection triggers inflammatory responses to dietary antigens and development of celiac disease. *Science* 2017;356:44-50. doi:10.1126/science.aah5298
10 Webb A, Starr M. Acute gastroenteritis in children. *Aust Fam Physician* 2005;34:227-31.
11 Schumann M, Siegmund B, Schulzke JD, Fromm M. Celiac Disease: Role of the Epithelial Barrier. *Cell Mol Gastroenterol Hepatol* 2017;3:150-62. doi:10.1016/j.jcmgh.2016.12.006
12 Stene LC, Honeyman MC, Hoffenberg EJ, et al. Rotavirus infection frequency and risk of celiac disease autoimmunity in early childhood: a longitudinal study. *Am J Gastroenterol* 2006;101:2333-40. doi:10.1111/j.1572-0241.2006.00741.x
13 Kagnoff MF, Paterson YJ, Kumar PJ, et al. Evidence for the role of a human intestinal adenovirus in the pathogenesis of coeliac disease. *Gut* 1987;28:995-1001. doi:10.1136/gut.28.8.995
14 Mahon J, Blair GE, Wood GM, Scott BB, Losowsky MS, Howdle PD. Is persistent adenovirus 12 infection involved in coeliac disease? A search for viral DNA using the polymerase chain reaction. *Gut* 1991;32:1114-6. doi:10.1136/gut.32.10.1114
15 Mercalli A, Lampasona V, Klingel K, et al. No evidence of enteroviruses in the intestine of patients with type 1 diabetes. *Diabetologia* 2012;55:2479-88. doi:10.1007/s00125-012-2591-4
16 Stene LC, Witsø E, Torjesen PA, et al. Islet autoantibody development during follow-up of high-risk children from the general Norwegian population from three months of age: design and early results from the MIDIA study. *J Autoimmun* 2007;29:44-51. doi:10.1016/j.jaut.2007.04.003
17 Cinek O, Witsø E, Jeansson S, et al. Longitudinal observation of enterovirus and adenovirus in stool samples from Norwegian infants with the highest genetic risk of type 1 diabetes. *J Clin Virol* 2006;35:33-40. doi:10.1016/j.jcv.2005.03.007
18 Husby S, Koletzko S, Korponay-Szabó IR, et al, ESPGHAN Working Group on Coeliac Disease Diagnosis, ESPGHAN Gastroenterology Committee. European Society for Pediatric Gastroenterology, Hepatology, and Nutrition. European Society for Pediatric Gastroenterology, Hepatology, and Nutrition guidelines for the diagnosis of coeliac disease. *J Pediatr Gastroenterol Nutr* 2012;54:136-60. doi:10.1097/MPG.0b013e31821a23d0
19 Bürgin-Wolff A, Gaze H, Hadziselimovic F, et al. Antigliadin and antiendomysium antibody determination for coeliac disease. *Arch Dis Child* 1991;66:941-7. doi:10.1136/adc.66.8.941
20 Bürgin-Wolff A, Dahlbom I, Hadziselimovic F, Petersson CJ. Antibodies against human tissue transglutaminase and endomysium in diagnosing and monitoring coeliac disease. *Scand J Gastroenterol* 2002;37:685-91. doi:10.1080/00365520212496
21 Arguelles-Grande C, Norman GL, Bhagat G, Green PH. Hemolysis interferes with the detection of anti-tissue transglutaminase antibodies in celiac disease. *Clin Chem* 2010;56:1034-6. doi:10.1373/clinchem.2009.141242
22 Myléus A, Ivarsson A, Webb C, et al. Celiac disease revealed in 3% of Swedish 12-year-olds born during an epidemic. *J Pediatr Gastroenterol Nutr* 2009;49:170-6. doi:10.1097/MPG.0b013e31818c52cc
23 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;56:348-53. doi:10.1016/j.jcv.2012.11.018
24 Claas EC, Schilham MW, de Brouwer CS, et al. Internally controlled real-time PCR monitoring of adenovirus DNA load in serum or plasma of transplant recipients. *J Clin Microbiol* 2005;43:1738-44. doi:10.1128/JCM.43.4.1738-1744.2005
25 Cinek O, Kramna L, Mazankova K, et al. Virus genotyping by massive parallel amplicon sequencing: adenovirus and enterovirus in the Norwegian MIDIA study. *J Med Virol* 2018;10: [Epub ahead of print]. doi:10.1002/jmv.25361
26 Tapia G, Cinek O, Rasmussen T, et al. Human enterovirus RNA in monthly fecal samples and islet autoimmunity in Norwegian children with high genetic risk for type 1 diabetes: the MIDIA study. *Diabetes Care* 2011;34:151-5. doi:10.2337/dc10-1413
27 Baggen J, Thibaut HJ, Strating JRPM, van Kuppeveld FJM. Publisher Correction: The life cycle of non-polio enteroviruses and how to target it. *Nat Rev Microbiol* 2018;16:391. doi:10.1038/s41579-018-0022-3
28 Pallansch MA, Roos RP. Enteroviruses: polioviruses, coxsackieviruses, echoviruses, and newer enteroviruses. In: Knipe DM, Howley PM, eds. *Fields Virology*. 7th ed. Lippincott Williams &Wilkins, 2007: 839-94.
29 Witsø E, Palacios G, Cinek O, et al. High prevalence of human enterovirus a infections in natural circulation of human enteroviruses. *J Clin Microbiol* 2006;44:4095-100. doi:10.1128/JCM.00653-06
30 Viskari H, Oikarinen S, Hoppu S, et al. Live attenuated enterovirus vaccine (OPV) is not associated with islet autoimmunity in children with genetic susceptibility to type 1 diabetes: prospective cohort study. *Diabetologia* 2018;61:203-9. doi:10.1007/s00125-017-4410-4
31 Norwegian Institute of Public Health. Vaksinasjonsveilederen for helsepersonell (information about the norwegian vaccine program). 2017 [cited 6 February 2018]. https://www.fhi.no/nettpub/vaksinasjonsveilederen-for-helsepersonell/vaksiner-mot-de-enkelte-sykdommene/poliovaksinasjon-poliomyelitt---vei/.
32 Nilsen RM, Vollset SE, Gjessing HK, et al. Self-selection and bias in a large prospective pregnancy cohort in Norway. *Paediatr Perinat Epidemiol* 2009;23:597-608. doi:10.1111/j.1365-3016.2009.01062.x
33 Witsø E, Cinek O, Aldrin M, et al. Predictors of sub-clinical enterovirus infections in infants: a prospective cohort study. *Int J Epidemiol* 2010;39:459-68. doi:10.1093/ije/dyp333
34 Oikarinen M, Tauriainen S, Oikarinen S, et al. Type 1 diabetes is associated with enterovirus infection in gut mucosa. *Diabetes* 2012;61:687-91. doi:10.2337/db11-1157
35 Honkanen H, Oikarinen S, Nurminen N, 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;60:424-31. doi:10.1007/s00125-016-4177-z
36 Krogvold L, Edwin B, Buanes T, et al. Detection of a low-grade enteroviral infection in the islets of langerhans of living patients newly diagnosed with type 1 diabetes. *Diabetes* 2015;64:1682-7. doi:10.2337/db14-1370
37 Tapia G, Cinek O, Rasmussen T, et al. Human enterovirus RNA in monthly fecal samples and islet autoimmunity in Norwegian children with high genetic risk for type 1 diabetes: the MIDIA study. *Diabetes Care* 2011;34:151-5. doi:10.2337/dc10-1413
38 Kagnoff MF, Austin RK, Hubert JJ, Bernardin JE, Kasarda DD. Possible role for a human adenovirus in the pathogenesis of celiac disease. *J Exp Med* 1984;160:1544-57. doi:10.1084/jem.160.5.1544
39 Lähdeaho ML, Parkkonen P, Reunala T, Mäki M, Lehtinen M. Antibodies to E1b protein-derived peptides of enteric adenovirus type 40 are associated with celiac disease and dermatitis herpetiformis. *Clin Immunol Immunopathol* 1993;69:300-5. doi:10.1006/clin.1993.1184
40 Lawler M, Humphries P, O'Farrelly C, et al. Adenovirus 12 E1A gene detection by polymerase chain reaction in both the normal and coeliac duodenum. *Gut* 1994;35:1226-32. doi:10.1136/gut.35.9.1226
41 Schumann M, Siegmund B, Schulzke JD, Fromm M. Celiac disease: role of the epithelial barrier. *Cell Mol Gastroenterol Hepatol* 2017;3:150-62. doi:10.1016/j.jcmgh.2016.12.006
42 Sollid LM, Jabri B. Triggers and drivers of autoimmunity: lessons from coeliac disease. *Nat Rev Immunol* 2013;13:294-302. doi:10.1038/nri3407
43 Matzinger P. The danger model: a renewed sense of self. *Science* 2002;296:301-5. doi:10.1126/science.1071059
44 Stone VM, Hankaniemi MM, Svedin E, et al. A Coxsackievirus B vaccine protects against virus-induced diabetes in an experimental mouse model of type 1 diabetes. *Diabetologia* 2018;61:476-81. doi:10.1007/s00125-017-4492-z

Supplementary materials

10.1.1 Supplementary materials of paper I

Supplementary material

Kahrs CR, Chuda K, Tapia G, et.al. Enterovirus as trigger of coeliac disease: nested case-control study within prospective birth cohort.

Table of Contents

Supplementary methods 2
Supplementary references 2
Supplementary figures 3
Supplementary figure A. Coeliac disease screening process 3
Supplementary figure B. Enterovirus distribution over the whole period 4
Supplementary figure C. Age variation of enterovirus and adenovirus 6
Supplementary figure D. Seasonal variation of entero- and adenovirus 7
Supplementary tables 8
Supplementary table A. Characteristics of the whole cohort 8
Supplementary table B. Characteristics of the children with coeliac disease 9
Supplementary table C. Enterovirus types observed in this study 10
Supplementary table D. Enterovirus infectious episodes, long infections and high-quantity infections 11
Supplementary table E. Enterovirus, gluten introduction, breastfeeding and age-groups 12
Supplementary table F. Reported symptoms, infections with symptoms and coeliac disease 13
Supplementary table G. Association between reported symptoms and viral infections independent of case status 14
Supplementary table H: Adenovirus types observed in this study 15

This supplementary material has been provided by the authors to give readers additional information about their work.

Supplementary methods

Case definition and exclusion criteria

Out of the 220 children who consented to participate in our study, 17 had been diagnosed with coeliac disease before inclusion. Of the remaining 203 children, 20 had elevated coeliac disease antibodies and were retested, with 11 children having positive antibodies in the confirmatory blood sample. These were referred to a hospital for diagnostic work-up, with 10 children being confirmed as having a coeliac disease diagnosis. The one child where the diagnosis was not confirmed and the nine children with negative antibodies in the confirmatory blood sample were excluded from the current analysis. All these children were only DGP positive in the first sample, none had tTG antibodies.

Case validation and confirmation

In the parental questionnaire at inclusion of the sub-study on coeliac disease the following covariates were included; first degree relatives with coeliac disease, type of symptoms, age at onset of symptoms, age at coeliac disease diagnosis, whether diagnosis was detected by screening or clinical symptoms, biopsy and serology status. In the validation scheme sent to hospitals after end of follow up we asked for additional diagnostic information including symptoms and coeliac disease antibody levels at diagnosis, biopsy status and Marsh grade, and date of diagnosis.

Analysis of fecal samples

All available fecal supernatant samples from cases and controls were subjected to RNA and DNA co-purification using Qiagen spin columns (Qiagen, Hilden, Germany). We co-purified the total RNA and DNA using a procedure derived from the QiaAmp Viral RNA Mini kit protocol (Qiagen, Hilden, Germany). For increasing the throughput, we utilized a 96-well format with the "QiaAmp 96" plates (Qiagen) instead of individual columns, and to minimize the risk of contamination we utilized centrifugation instead of applying vacuum. West Nile virus Armored RNA as an exogeneous internal control (Asuragen, Austin, TX) and carrier RNA (Qiagen, Hilden, Germany) were added in a constant quantity to the lysis buffer, which was used in the first step of the protocol. Each extraction plate contained a minimum of six negative controls (TE buffer instead of a sample) in irregular intervals. No negative control was positive for entero- or adenovirus.

We tested enterovirus by reverse-transcriptase real-time PCR with the QuantiTect Probe RT-PCR Kit (Qiagen), using 900 nmol/L primers and 300 nmol/L probes (sequences described by Honkanen et al).[1] The combination of primers and probes reacts with an equal sensitivity to Enterovirus A—D species (i.e. members of species *Enterovirus A*, *Enterovirus B*, *Enterovirus C* and *Enterovirus D* of the genus *Enterovirus*, family *Picornaviridae*, order *Picornavirales*) but does not react with human rhinoviruses (i.e. members of species *Rhinovirus A-C*, genus *Enterovirus*). Enterovirus positivity cut-off was 10 copies / µL, and the amplification had to occur before the cycle 35. Adenovirus was tested using a previously published real-time PCR assay,[2] and the positivity threshold was likewise 10 copies / µL.

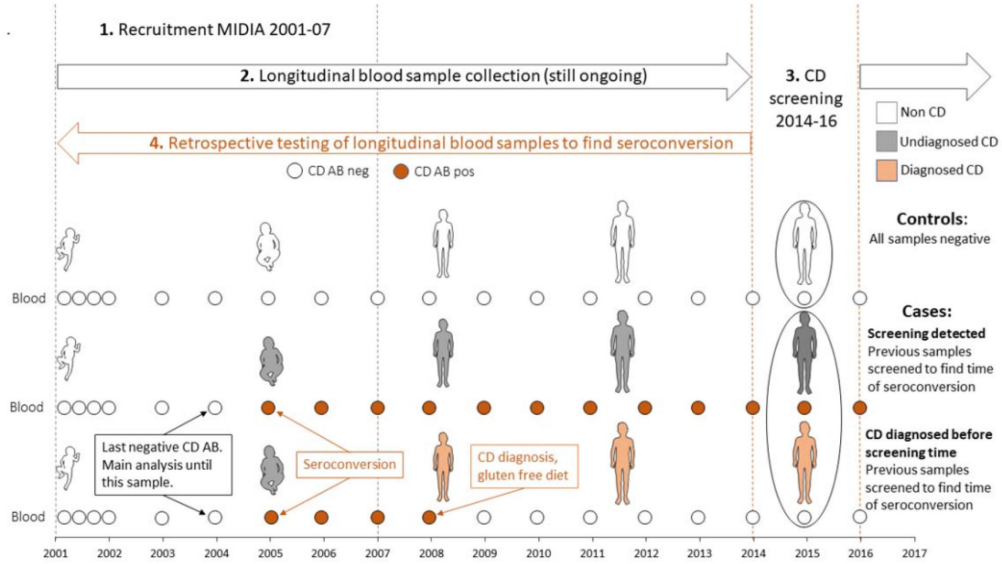
Enterovirus was genotyped using a nested reverse transcriptase PCR targeted to the VP1 gene. Primers were redesigned from the work by Nix et al to accommodate for higher variability of the genotypes,[3] and to make the reaction truly nested as compared to the semi-nested design of the original method. Genotyping of adenovirus was performed by sequencing of hexon gene spanning its hypervariable region 7. The product was amplified using a mixture of 10 forward and 6 reverse primers.

Supplementary references

1. 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;56:348-53.
2. Claas EC, Schilham MW, de Brouwer CS, et al. Internally controlled real-time PCR monitoring of adenovirus DNA load in serum or plasma of transplant recipients. J Clin Microbiol 2005;43:1738-44.
3. Nix WA, Oberste MS, Pallansch MA. Sensitive, seminested PCR amplification of VP1 sequences for direct identification of all enterovirus serotypes from original clinical specimens. J Clin Microbiol 2006;44:2698-2704.

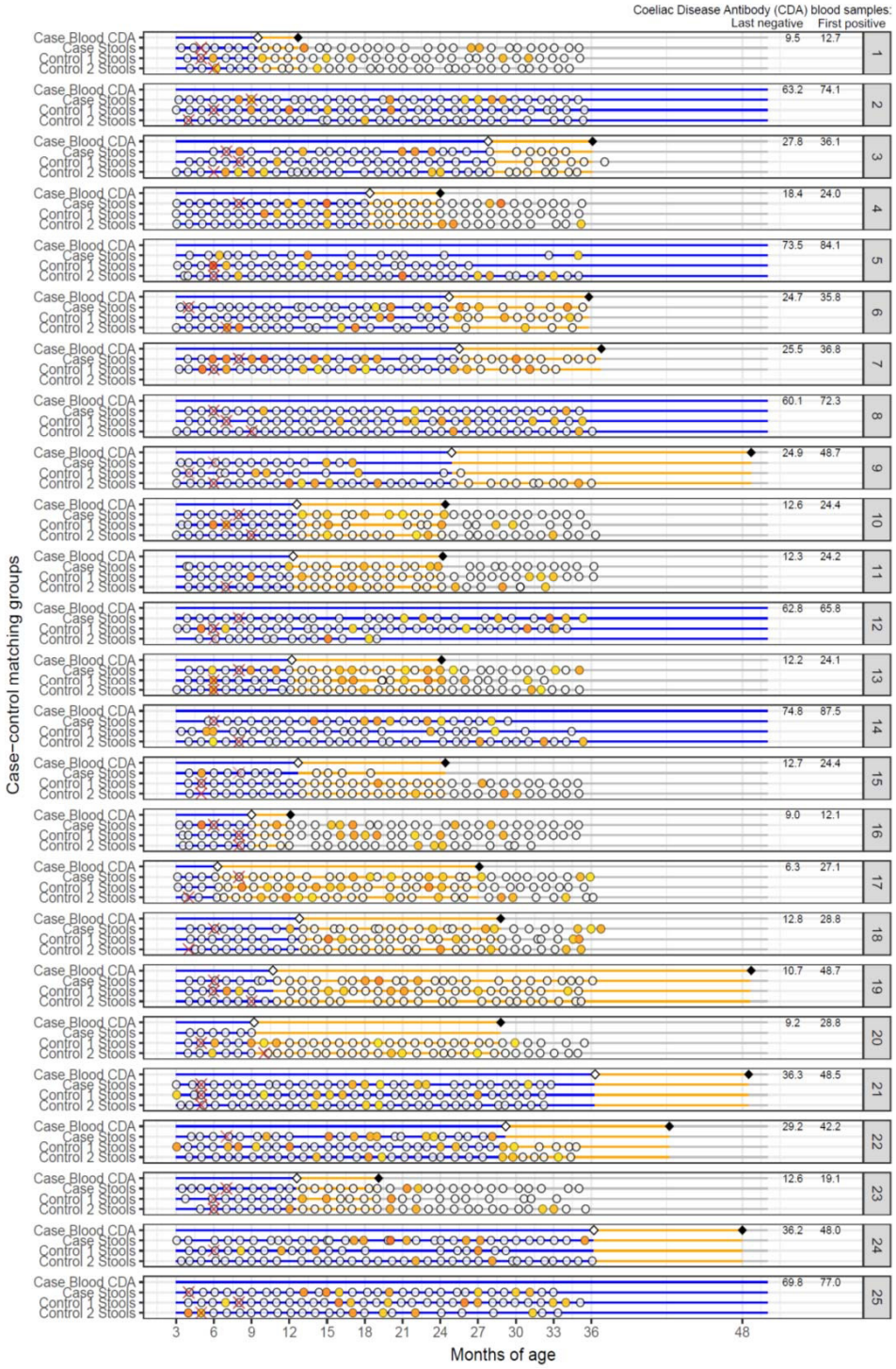
Supplementary figures

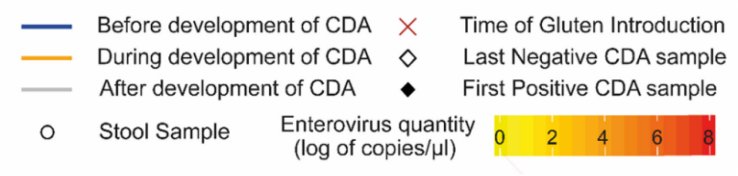
Supplementary figure A. Coeliac disease screening process



CD denotes coeliac disease; AB, antibody. Coeliac disease antibody screening was performed in samples collected from consenting participants during 2014-2016. Those diagnosed after being positive for coeliac disease antibodies on screening in addition to those with known coeliac disease were then screened retrospectively for coeliac disease antibodies in stored samples longitudinally collected from birth to identify time of seroconversion. Children negative at coeliac disease screening during 2014-16 were eligible as potential controls, and confirmed as controls if all previous samples were negative (if previous sample(s) were positive, they were not included as controls). We diagnosed coeliac disease according to the ESPGHAN (The European Society for Paediatric Gastroenterology Hepatology and Nutrition) 2012 criteria.

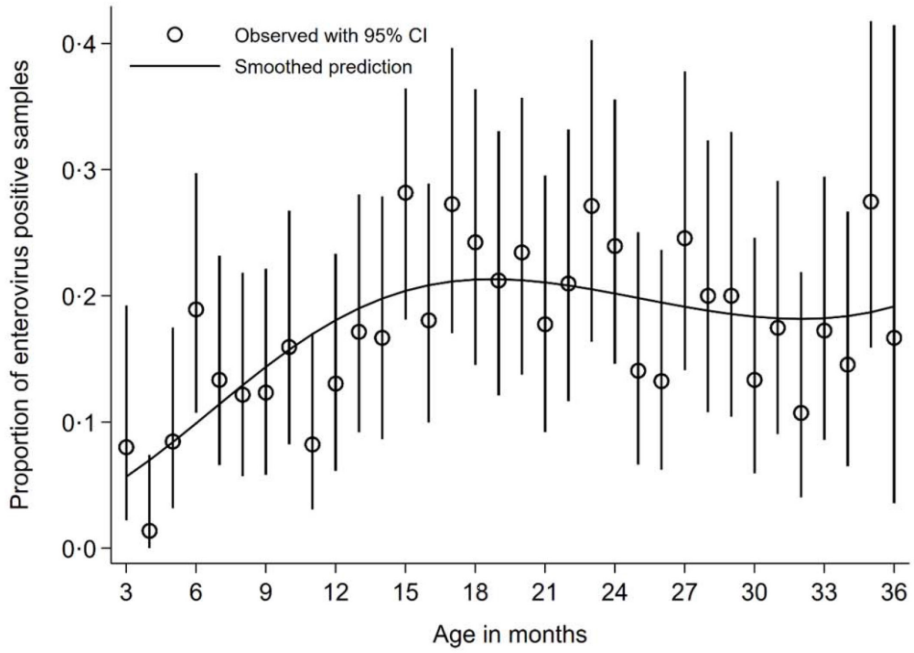
Supplementary figure B. Enterovirus distribution over the whole period

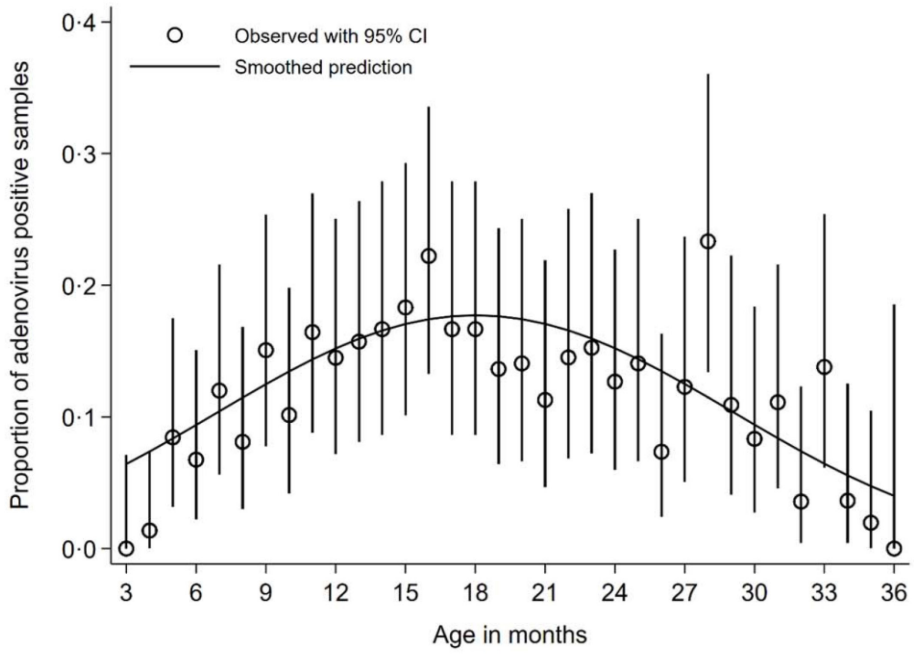




Supplementary figure 1 shows samples from the participants in the study, by month of age and matching set. One control (in matching group 7) was excluded due to missing stool samples. Top line shows the last negative and first positive coeliac disease antibody (CDA) blood sample (marked by a diamond; last negative = white, first positive = black) in cases, followed by stool samples (Enterovirus quantity denoted by colour) from cases and controls. The age in months at last negative and first positive CDA blood sample is shown in columns at the right of the plot. The coloured lines denote the time before (blue line; the main analysis), during (orange line) and after (grey line) development of CDA in cases and the corresponding age in the matched controls. A red X marks time of gluten introduction.

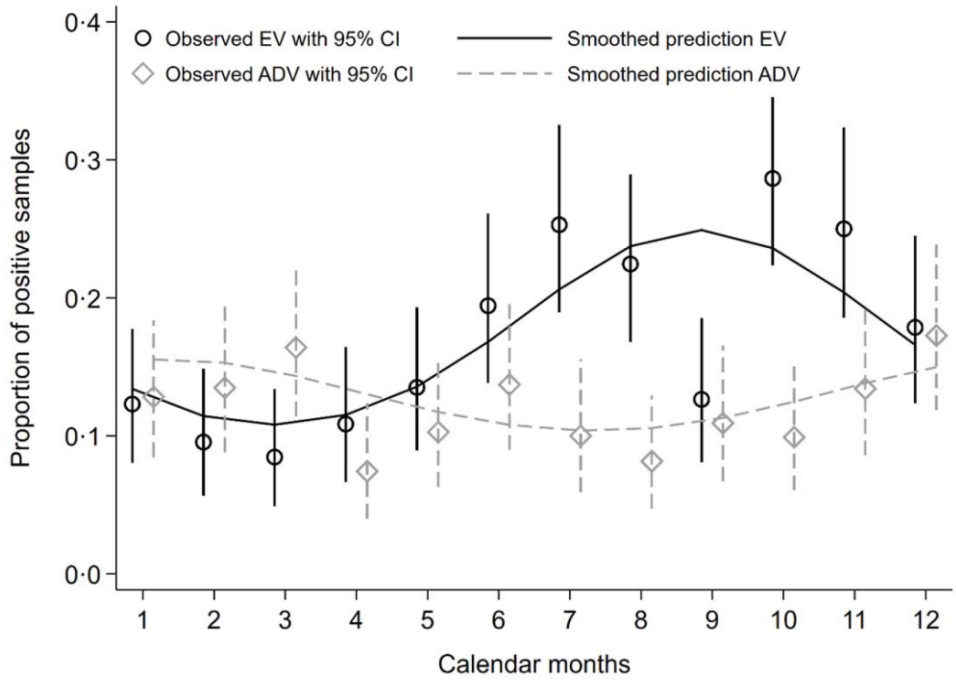
Supplementary figure C. Age variation of enterovirus and adenovirus





The solid line is predicted values from third degree polynomial logistic regression.

Supplementary figure D. Seasonal variation of entero- and adenovirus



EV denotes enterovirus; ADV, adenovirus. Lines are smoothed predictions from regression model with sine and cosine terms for month of sample collection (cosinor model). There was a significant seasonal variation for E (2 degree of freedom likelihood ratio test; P<0.001), but not for adenovirus (P=0.11).

Supplementary tables

Supplementary table A. Characteristics of the whole cohort, children invited to coeliac disease autoantibody screening, and those who consented to screening (participated). Values are numbers (percentages).

Characteristics	Whole cohort (n=912)	Invited (n=501)	Participated (n=220)
Year of birth* 2001 – 2005	431 (47)	241 (48)	103 (47)
Year of birth* 2006 – 2007	481 (53)	260 (52)	117 (53)
Female sex	445 (49)	250 (50)	107 (49)
Other children in household			
None	319 (35)	172 (34)	76 (35)
1	367 (40)	220 (44)	102 (46)
2	226 (25)	109 (22)	42 (19)
Family history of coeliac disease†	19 (2)	12 (2)	7 (3)

The "whole cohort" includes children of all consenting parents who were followed up at least once. The "invited" are those who were still actively participating (delivering samples, questionnaires) at the start of the coeliac disease sub-study and were invited to participate in the coeliac disease sub-study, and the "participated" are those who consented to participate.

*Year of birth was distributed as follows in the whole cohort (n, %): 2001 (21, 2.3%); 2002 (57, 6.3%); 2003 (69, 7.6%); 2004 (105, 11.5%); 2005 (179, 19.6%); 2006 (289, 31.7%); 2007 (192, 21.1%)

†Ascertained at time of enrolment. The nested case-control study with screened subjects that were cases and their matched controls, were asked again and have thus more affected first-degree relatives and half-siblings as shown in table 1.

Supplementary table B. Characteristics of the children with coeliac disease (n=25). Values are numbers (percentages) unless stated otherwise.

Ages	Mean (SD) age at development of coeliac disease antibodies, months	
	Last negative sample	30 (23)
	First positive sample	42 (23)
	Mean (SD) age at debut of symptoms, months[*]	73 (38)
	Mean (SD) age at diagnosis, months[†]	87 (34)
Coeliac disease antibodies	Coeliac disease antibody concentration at first positive sample[§]	
	IgA anti-tTG	
	<3 U/mL	5 (22)
	3-70 U/mL	10 (40)
	>70 U/mL	10 (40)
	IgG anti-DGP[‖]	
	<7 U/mL	4 (17)
	7-70 U/mL	15 (65)
	>70 U/mL	4 (17)
	IgA anti-tTG max values	
	<7 U/mL	1 (4)
	7-70 U/mL	3 (12)
	>70 U/mL	21 (84)
	IgG anti-DGP max values	
	<7 U/mL	3 (12)
	7-70 U/mL	12 (48)
	>70 U/mL	10 (40)
Basis for diagnosis	Biopsy (Marsh grade 3)	14 (56)
	Serology according to ESPGHAN 2012 criteria	9 (36)
	Serology and biopsy (uncertain Marsh classification)[‡]	2 (8)
Symptoms	Symptoms before diagnosis[¶]	
	Intestinal	16 (64)
	Extra-intestinal	2 (8)
	None	7 (28)

SD denotes standard deviation; tTG, tissue transglutaminase; DGP, deamidated gliadin peptide.
[*]Missing: 10.
[†]Date of the biopsy taking or date of consultation if diagnosis based on serology.
[‡]Two children with unknown/unclear Marsh-grading were diagnosed based on symptoms and repeated anti-tTG-levels >10 times cut-off level before the ESPGHAN criteria were established in 2012.
[§]IgA anti-tTG >3 or IgG DGP >7.
[‖]Missing: 2.
[¶]All participants with a coeliac disease diagnosis (both those diagnosed prior to the study screening and those diagnosed after the study screening) were asked in retrospect if they noted any symptoms prior to coeliac disease diagnosis and when these symptoms debuted.

Supplementary table C. Enterovirus types observed in this study

	Samples (n)	Percent of total samples*	Percent of infections†
Enterovirus type not determined‡	**62**	**-**	**16.8**
Enterovirus type determined§	**308**	**14.4**	**83.2**
Enterovirus A	172	8.1	46.5
CV-A2	41	1.9	11.1
CV-A4	30	1.4	8.1
CV-A5	16	0.7	4.3
CV-A6	24	1.1	6.5
CV-A8	2	0.1	0.5
CV-A10	22	1.0	5.9
CV-A12	1	0.0	0.3
CV-A14	1	0.0	0.3
CV-A16	22	1.0	5.9
EV-A71	15	0.7	4.1
Enterovirus B	139	6.5	37.6
CV-B1	12	0.6	3.2
CV-B2	19	0.9	5.1
CV-B3	11	0.5	3.0
CV-B4	9	0.4	2.4
CV-B5	12	0.6	3.2
CV-A9	17	0.8	4.6
E-3	3	0.1	0.8
E-6	5	0.2	1.4
E-7	2	0.1	0.5
E-9	10	0.5	2.7
E-11	8	0.4	2.2
E-13	1	0.0	0.3
E-18	9	0.4	2.4
E-25	19	0.9	5.1
E-30	6	0.3	1.6
Enterovirus C	4	0.2	1.1
CV-A1	3	0.1	0.8
CV-A22	1	0.0	0.3
PV-2¶	1	0.0	0.3
Enterovirus D	0	0	0

*2135 stool samples provided enterovirus data of which 370 (17%) were enterovirus positive with 309 (83.5%) having genotype information. Coinfections were common (5.8%).

†Number of samples with this genotype divided by number of samples positive for enterovirus (using a cut-off of at least 10 copies/µl).

‡Low quantity.

§CV denotes coxsackievirus; EV, enterovirus; E, echovirus, PV poliovirus.

¶Poliovirus vaccine strain Sabin.

Supplementary table D. Enterovirus infectious episodes, long infections and high-quantity infections*

	Cases (n=25)	Controls (n=49)	Unadjusted	Adjusted†	
	Positives (n)/ total (N) samples (%)		Odds ratio (95% CI)	Odds ratio (95% CI)	P value
Main analysis	*84/429 (20)*	*129/856 (15)*	*1.37 (1.01 to 1.87)*	*1.49 (1.07 to 2.06)*	*0.02*
Long infections‡	22/367 (6)	27/755 (4)	1.72 (0.96 to 3.06)	2.16 (1.16 to 4.04)	0.02
High-quantity infections§	28/429 (7)	33/856 (4)	1.73 (0.97 to 3.06)	2.11 (1.24 to 3.60)	0.01
Infectious episodes¶	55/400 (14)	95/822 (12)	1.22 (0.85 to 1.74)	1.27 (0.87 to 1.86)	0.21

*Before development of coeliac disease antibodies (prior to last negative sample).
†Adjusted for sex, age, age squared, season of sample collection, number of siblings, and family history of coeliac disease.
‡≥2 positive consecutive monthly samples.
§Per infection with high quantity (≥100 000 copies).
¶A sequence of consecutively virus-positive fecal samples (a negative stool sample was demanded before defining a new episode).

Supplementary table E. Enterovirus, gluten introduction, breastfeeding and age-groups*

	Cases (n=25)	Controls (n=49)	Adjusted†	
	Positives (n)/ total (N) samples (%)		Odds ratio (95% CI)	P value
Main analysis	*84/429 (20)*	*129/856 (15)*	*1.49 (1.07 to 2.06)*	*0.02*
Samples 3 to 6 months of age	7/87 (8)	18/174 (10)	0.44 (0.14 to 1.37)	0.16
Samples 6-12 months of age	17/155 (11)	42/302 (14)	0.66 (0.30 to 1.46)	0.30
Samples 12 months of age	67/231 (29)	83/463 (18)	1.97 (1.33 to 2.93)	0.001
Prior to gluten introduction	7/64 (11)	9/113 (8)	0.75 (0.21 to 2.63)	0.65
At gluten introduction‡	4/37 (11)	13/70 (19)	0.42 (0.11 to 1.61)	0.21
After gluten introduction§	66/298 (22)	99/587 (17)	1.52 (1.05 to 2.20)	0.03
While breastfed	18/163 (11)	42/369 (11)	0.78 (0.34 to 1.79)	0.56
After end of breastfeeding‖	55/199 (28)	83/455 (18)	1.80 (1.17 to 2.78)	0.01
After end of breastfeeding and gluten introduction	55/195 (28)	71/383 (19)	1.83 (1.19 to 2.81)	0.01

*Samples before development of coeliac disease antibodies (prior to last negative sample) independent of case/control/status.
†Adjusted for sex, age, age squared, season of sample collection, number of siblings, and family history of coeliac disease.
‡Samples at ± 1month of gluten introduction.
§Median age 6 months, range 2 to 10.
‖Median age 12 months, range 2 to 23.

Supplementary table F. Reported symptoms, infections with symptoms and coeliac disease*

	Cases (n=25)	Controls (n=49)	Adjusted†	
	Positives (n)/ total (N) samples (%)		Odds ratio (95% CI)	P value
Reported symptoms‡				
Any symptom	157/430 (37)	302/858 (35)	1.07 (0.68 to 1.67)	0.77
Fever reported	104/440 (24)	181/865 (21)	1.21 (0.77 to 1.90)	0.41
Diarrhoea reported	30/431 (7)	61/858 (7)	1.07 (0.66 to 1.72)	0.79
Common cold reported	93/432 (22)	195/858 (23)	0.99 (0.58 to 1.68)	0.97
Enterovirus positive samples with reported symptoms				
Enterovirus + any symptom	26/419 (6)	44/849 (5)	1.49 (0.88 to 2.52)	0.14
Enterovirus + fever	24/429 (6)	28/856 (3)	2.12 (1.16 to 3.85)	0.01
Enterovirus + diarrhoea	7/420 (2)	6/849 (1)	3.62 (1.14 to 11.49)	0.03
Enterovirus + common cold	13/421 (3)	32/849 (4)	1.00 (0.51 to 1.97)	0.99
Adenovirus positive samples with reported symptoms				
Adenovirus + any symptom	19/383 (5)	48/768 (6)	0.94 (0.44 to 2.01)	0.87
Adenovirus + fever	15/390 (4)	36/775 (5)	0.88 (0.43 to 1.83)	0.74
Adenovirus + diarrhoea	3/383 (1)	9/768 (1)	0.89 (0.23 to 3.49)	0.87
Adenovirus + common cold	10/385 (3)	30/768 (4)	0.78 (0.26 to 2.30)	0.65

*Analysed as main analysis (mixed model), before development of coeliac disease antibodies (prior to last negative sample). Adjusted for sex, age, age squared, season of sample collection, number of siblings, and family history of coeliac disease.

†Self-reported symptoms of the infants as reported by their mothers, coded into yes/no per monthly sample (tolerating ± 15 days from reported date).

Supplementary table G. Association between reported symptoms and viral infections independent of case status[*]

	Reported symptoms[†] / total samples (n)	Odds ratio (95% CI)	P value
Association enterovirus and simultaneous symptoms[‡]			
Any symptom	696/2119	0.97 (0.77 to 1.22)	0.81
Fever	418/2136	1.10 (0.86 to 1.40)	0.46
Diarrhoea	163/2120	0.90 (0.57 to 1.41)	0.63
Common cold	422/2121	0.97 (0.77 to 1.23)	0.82
Association adenovirus and simultaneous symptoms[§]			
Any symptom	644/1992	1.26 (0.93 to 1.70)	0.13
Fever[¶]	396/2006	1.62 (1.22 to 2.16)	<0.01
Diarrhoea	149/1992	1.11 (0.67 to 1.85)	0.68
Common cold	388/1994	1.11 (0.77 to 1.61)	0.58

[*] All samples analysed using logistic regression.
[†] Number of samples with positive symptoms, tolerating 15 days difference between a sample and date of reported symptoms.
[‡] 368 Enterovirus samples with valid symptom data, with the exception of fever (370 samples).
[§] 258 Adenovirus samples with valid symptom data.
[¶] Sequencing revealed a possible association between human adenovirus (HAdV)-C5 and fever (OR 2.95 (95% CI 1.45 to 6.03); p<0.01, and HAdV-C2 and fever (OR 1.59 (1.08 to 2.34); p=0.02.

Supplementary table H: Adenovirus types observed in this study*

	N	Percent of total samples†	Percent of infections‡
No adenovirus type determined	**6**	**-**	**2.3**
Adenovirus type determined	**252**	**12.6**	**97.7**
HAdV-C1	102	5.1	39.5
HAdV-C2	109	5.4	42.2
HAdV-B3	17	0.8	6.6
HAdV-A31	8	0.4	3.1
HAdV-F41	9	0.4	3.5
HAdV-C5	24	1.2	9.
HAdV-C57	1	0.0	0.4

*HAdV denotes *human adenovirus*. Adenovirus types were grouped into only specific types (e.g. HAdV-C2), as species *Adenovirus C* dominated in our samples.

†Percent with genotype information of all 2006 samples. Coinfections with two HAdV types were observed in 18 (0.9%) positive samples.

‡Number of samples with this genotype, divided by number of positive samples (using a cut-off of 10 copies/µl).

10.2 Paper II

Received: 7 August 2018 | Accepted: 4 November 2018
DOI: 10.1002/jmv.25361

RESEARCH ARTICLE

WILEY JOURNAL OF MEDICAL VIROLOGY

Virus genotyping by massive parallel amplicon sequencing: adenovirus and enterovirus in the Norwegian MIDIA study

Ondrej Cinek[1] | Lenka Kramna[1] | Karla Mazankova[1] | Kateřina Kunteová[1] | Kateřina Chudá[1] | Eric C. J. Claas[3] | Lars C. Stene[2] | German Tapia[2]

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

[2]Division of Mental and Physical Health, Department of Chronic Diseases and Ageing, Norwegian Institute of Public Health, Oslo, Norway

[3]Department of Medical Microbiology, Leiden University Medical Center, Leiden, The Netherlands

Correspondence
Ondrej Cinek, Department of Pediatrics, 2nd Faculty of Medicine, Charles University in Prague and University Hospital Motol, Prague, V Uvalu 84, Prague 5, Czech Republic.
Email: ondrej.cinek@fmotol.cuni.cz

Funding information
Ministerstvo Zdravotnictví Ceské Republiky, Grant/Award Number: 15□29078A

Objectives: Direct genotyping of adenovirus or enterovirus from clinical material using polymerase chain reaction (PCR) followed by Sanger sequencing is often difficult due to the presence of multiple virus types in a sample, or due to varying efficacy of PCR amplifying the capsid gene on the background of foreign nucleic acids. Here we present a simple protocol for virus genotyping using massive parallel amplicon sequencing.

Methods: The protocol utilized a set of 16 tailed degenerate primers flanking the seventh hypervariable region of the adenovirus hexon gene and 9 tailed degenerate primers targeted to the proximal portion of the enterovirus VP1 gene. Subsequent addition of dual indices enabled simultaneous sequencing of 384 different samples on an Illumina MiSeq instrument. Downstream bioinformatic analysis was based on remapping to a set of references representative of the presently known repertoire of virus types.

Results: After validation with known virus types, the sequencing method was applied on 301 adenovirus□positive samples and 350 enterovirus□positive samples from a longitudinally collected series of stools from 83 children aged 3 to 36 months. We detected 7 different adenovirus types and 27 different enterovirus types. There were 37 (6.2%) samples containing more than one genotype of the same viral genus. At least one dual infection was experienced by 23 of 83 (28%) of the children observed over the 3 years' observation period.

Conclusions: Amplicon sequencing with a multiplex set of degenerate primers seems to be a rapid and reliable technical solution for genotyping of large collections of samples where simultaneous infections with multiple strains can be expected.

KEYWORDS
adenovirus, enterovirus, genotype, infants, massive parallel sequencing, virus type

1 | INTRODUCTION

Virus genotyping of adenoviruses and enteroviruses is most often performed by sequencing of informative variable regions of capsid genes (the hexon gene of adenovirus, and the VP1 gene of enterovirus). Degenerate primers and relaxed polymerase chain reaction (PCR) stringency help to achieve amplification of a broad spectrum of virus types. However, in complex samples as are nucleic acids from stools, this often leads to the generation of spurious PCR by□products from the foreign nonviral nucleic acid. These by□products complicate or thwart the evaluation of genotyping products on Sanger electrophoretograms. The same problem arises when two or more viral strains are present in one sample—in Sanger sequencing, such signals can be dissected only by arduous subcloning

of the PCR product or by using an extensive panel of type- or group-specific primer pairs.

Massive parallel sequencing, also termed next-generation sequencing, is revolutionizing the field of molecular virology. Its metagenomic applications have been utilized, for example, for pathogen discovery,[1] in the investigation of outbreaks,[2] and in research of primarily noninfectious diseases such as type 1 diabetes.[3,4] Amplicon sequencing, in contrast to the above mentioned metagenomic sequencing, targets only a limited set of PCR products defined by the flanking primers—therefore it is noticeably cheaper, easier to perform, has higher sample throughput and higher reproducibility.

Here we describe the use of massive parallel amplicon sequencing in genotyping of enterovirus and adenovirus. As massive parallel sequencing provides one pair of reads for each sequenced amplicon molecule, any complex PCR product can be dissected in detail: this is in fundamental contrast to the summary signal ensuing from Sanger sequencing. Dissecting mixed samples is useful whenever a specimen contains more than one virus types, or when the desired PCR product is compounded with numerous by-products. Such by-products are especially frequent in genotyping protocols utilizing highly degenerate primers directly on clinical materials.

The aim of the present study was to set up a simple and reliable protocol for genotyping by massive parallel sequencing of two of the most prevalent gut viruses found in infancy—adenovirus and enterovirus—and to perform a genotyping study on a set of entero- and adenovirus-positive stools of healthy infants and toddlers from a Norwegian birth cohort.

2 | SUBJECTS AND METHODS

2.1 | The amplification and sequencing procedure

The protocol consisted of: (a) PCR amplification of informative virus fragments using tailed primers; (b) pooling of PCR products by sample when more than one product was amplified from each sample; (c) indexing reactions that provisioned the amplified fragments with indices and sequencing adapters; (d) equalization of library quantities across the index space; (e) massive parallel sequencing; (f) bioinformatic analysis. For a detailed protocol, please refer to the Supplementary Materials.

2.1.1 | PCR amplification of the virus fragments

Adenovirus hexon amplification

The amplification of adenovirus hexon fragment spanning its hypervariable region 7 was accomplished using a mixture of 10 forward and 6 reverse primers (Table 1). The amplification reaction had a volume of 20 µl and contained the HotStar Taq chemistry (Qiagen, Hilden, Germany) with 2.5 mM $MgCl_2$, 0.2 mM each dNTP, 1 µM total concentration of forward primer pool (10 primers in an equimolar ratio), 1 µM reverse primer pool (6 primers in an equimolar ratio), 1 U HotStar Taq polymerase, and 2 µl of extracted nucleic acid. The PCR program consisted of an initial denaturation of 15 minutes at 95°C, followed by 45 cycles of denaturation of 15 seconds at 94°C, annealing step of 1 minute at 55°C, and synthesis of 1 minute at 68°C. The final synthesis at 68°C lasted 5 minutes.

Enterovirus VP1 amplification

Enterovirus was genotyped using a nested reverse transcriptase PCR modified from the work by Nix et al.[5] The reverse transcription and the first round of PCR were performed with minor modifications only. In contrast, the primers for the second PCR round have been redesigned (Table 1): the forward primer AN-89 has been split into four variants and extended to increase the annealing temperature, whereas the reverse primer (AN-88) has been split into five variants, repositioned in the 5′ direction by six bases, and extended—this has made the reaction truly nested as compared to the semi-nested design of the original method.[5] The second round of PCR was performed with Promega GoTaq G2 Hot Start polymerase chemistry, with 1X Green Go Taq Flexi Buffer, 2.5 mM $MgCl_2$, 0.2 mM each dNTP, 1 µM total concentration of forward primer pool (four primers in an equimolar ratio), 1 µM total of reverse primer pool (5 primers in an equimolar ratio), 0.5 U Taq polymerase, and 2 µl of the first round product as a template, previously diluted 1:5 with sterile PCR-grade water.

2.1.2 | Purification and pooling

The PCR products were then checked on an agarose gel, and—if a sample was tested for both viruses—pooled by the sample. Then the product was purified with Ampure XT (Beckman, Brea, CA) according to the manufacturer's instruction with the exception of the magnetic beads to PCR product ratio, which was decreased to 0.8:1 to prioritize binding of longer PCR fragments and to eliminate primer dimers and other PCR artifacts.

2.1.3 | Indexing reaction

The indexing reaction extended the tailed PCR products with indices and sequencing adapters—it was a short PCR with eight cycles. The reaction of 15 µl contained 1X KAPA HiFi HotStart Ready Mix (cat. number 07958935001; Sigma-Aldrich, Roche, Basel, Switzerland), 3 µl of the purified product from the previous round, 1.5 µl primer 1 (with either of 24 indices N701-N729), and 1.5 µl of primer 2 (with either of 16 indices S502-S522). The indexed adapter primers were from the Nextera XT Index Kit v2 Set A (FC-131–2001) and set D (FC-131–2004, both Illumina, San Diego, CA), which together could create 384 unique index combinations (16 rows × 24 columns of the index space), the maximum sample count in one run. The PCR program of the indexing reaction consisted of an initial denaturation of 3 minutes at 95°C, followed by 8 cycles of 30 seconds at 95°C, 1 minute at 55°C, and 30 seconds at 72°C. A PCR purification step followed, using a 1:1 concentration of the Ampure purification solution (Beckman) to retain longer fragments and effectively remove shorter artifacts.

TABLE 1 Primers utilized in the amplification of sequenced fragments

A. Adenovirus (hypervariable region 7 of the hexon gene)	
Designation	**Sequence 5' to 3'**
Forward primers[1]	
Adv_1243_F1_for5	TCGTCGGCAGCGTC**AGATGTGTATAAGAGACAG**GAACCAGATACTTTAGCATGTGGAACTCTG
Adv_1243_F2_for5	TCGTCGGCAGCGTC**AGATGTGTATAAGAGACAG**GAACCAGATACTTYAGCATGTGGAATCAGG
Adv_1243_F3_for5	TCGTCGGCAGCGTC**AGATGTGTATAAGAGACAG**GAACCAGATAYTTTTCYATGTGGAATCAGG
Adv_1243_F4_for5	TCGTCGGCAGCGTC**AGATGTGTATAAGAGACAG**GAACCCGGTATTTCAGTATGTGGAATCARG
Adv_1243_F5_for5	TCGTCGGCAGCGTC**AGATGTGTATAAGAGACAG**GGAGTCGATACTTCTCCATGTGGAACCAGG
Adv_1243_F6_for5	TCGTCGGCAGCGTC**AGATGTGTATAAGAGACAG**GAACACGGTACTTTTCCTTGTGGAATTCCG
Adv_1243_F7_for5	TCGTCGGCAGCGTC**AGATGTGTATAAGAGACAG**GAACAAGATACTTTTCAATGTGGAATCAAG
Adv_1243_F8_for5	TCGTCGGCAGCGTC**AGATGTGTATAAGAGACAG**GAAGTCGCTACTTTTCTATGTGGAATCAAG
Adv_1243_F9_for5	TCGTCGGCAGCGTC**AGATGTGTATAAGAGACAG**GAAGCCGTTATTTTTCCATGTGGAATTCTG
Adv_1243_F10_for5	TCGTCGGCAGCGTC**AGATGTGTATAAGAGACAG**GCAGTCGGTATTTTTCTATGTGGAACTCAG
Reverse primers[2]	
Adv_1611_R1_rev7	GTCTCGTGGGCTCGG**AGATGTGTATAAGAGACAG**CCGTTCATGTACTCGTAGGTGTTGGTRTT
Adv_1611_R2_rev7	GTCTCGTGGGCTCGG**AGATGTGTATAAGAGACAG**CCGTTCATGTASTCGTAGGTGTTYTTGTT
Adv_1611_R3_rev7	GTCTCGTGGGCTCGG**AGATGTGTATAAGAGACAG**TTGTTCATGTAGTCGTAGGTGTTGGGGTT
Adv_1611_R4_rev7	GTCTCGTGGGCTCGG**AGATGTGTATAAGAGACAG**CCATTCATGTAGTCATAAGTGTTGGTGTT
Adv_1611_R5_rev7	GTCTCGTGGGCTCGG**AGATGTGTATAAGAGACAG**TTGTTCATATAATCGTATGAATTAGGGTT
Adv_1611_R6_rev7	GTCTCGTGGGCTCGG**AGATGTGTATAAGAGACAG**CCGTTCATGTAGGCATAKGTRTTITTGTT
B. Enterovirus (fragment of the VP1 gene)	
Second round of PCR[3]	**Sequence 5' to 3'**
Forward primers[4]	
E2F1_for5	TCGTCGGCAGCGTC**AGATGTGTATAAGAGACAG**GCNGYTGAGACAGG
E2F2_for5	TCGTCGGCAGCGTC**AGATGTGTATAAGAGACAG**GCNGYRGAGACTGG
E2F3_for5	TCGTCGGCAGCGTC**AGATGTGTATAAGAGACAG**GCWGYNGAAACTGG
E2F4_for5	TCGTCGGCAGCGTC**AGATGTGTATAAGAGACAG**GCWGYNGAAACGGG
Reverse primers[5]	
E2R1_rev7	GTCTCGTGGGCTCGG**AGATGTGTATAAGAGACAG**GGNAYRWACATRATTTG
E2R2_rev7	GTCTCGTGGGCTCGG**AGATGTGTATAAGAGACAG**GGNAYRWACATGTACTG
E2R3_rev7	GTCTCGTGGGCTCGG**AGATGTGTATAAGAGACAG**GGNAYRTACATTATCTG
E2R4_rev7	GTCTCGTGGGCTCGG**AGATGTGTATAAGAGACAG**GGDAYRTACATSASCTG
E2R5_rev7	GTCTCGTGGGCTCGG**AGATGTGTATAAGAGACAG**GGWAYRTACATNAMTTG

Underlined is the portion of primers that serves as an adaptor for subsequent adding of dual indices (underlined, not bold), and a pad (underlined, bold). The specific portion of the primers is in plain typeface. All primers were purified using HPLC.
All primers anneal to the same place in the alignment of types. i.e. the products of these primers are in phase.
(1) Position 20,011-20,041 relative to the genome of Human adenovirus 2 NCBI reference sequence AC_000007.1.
(2) Reverse complement to the position 20,402-20,373 in the above genome.
(3) The primers for reverse transcription and for the first round of PCR were identical to the original publication[5].
(4) Position 2,615-2,628 relative to the genome of Human poliovirus 1 Mahoney, NC_002058.3.
(5) Reverse complement to the position 2945-2961 in the above genome.

2.1.4 | Equalization of libraries

The double-indexed tailed products were then quantified by real-time PCR using the KAPA library quantification kit (cat. number 07960140001; Roche). Products were equalized across samples to an identical concentration and pooled to an overall concentration of 2 nmol/L by a custom set of calculation scripts written in Python linked to programs for the Biomek 3000 or 4000 robotic platform (Beckman). The pooled library was then checked for size and for concentration on a Bioanalyzer High Sensitivity DNA chip (Agilent, Santa Clara, CA).

2.1.5 | Sequencing

Sequencing was performed on an Illumina MiSeq instrument with a 2 × 250 bp protocol using the 500 cycle kit (cat. number MS-102–2003; Illumina).

2.1.6 | Bioinformatic analysis

The whole bioinformatic procedure has been automated into an application driven by a graphical user interface and is available from the authors upon reasonable request. It can be run on a desktop computer as an instance within the Oracle VM Virtual machine. The pipeline consists of a series of scripts derived from our previously published VIPIE pipeline.[6]

The sequencing reads of individual samples, demultiplexed by the sequencer software, were first filtered for quality and trimmed to remove low-quality ends. Then the left and right sequencing reads were merged using FLASH,[7] as the reads overlapped only partially, being each 250 bp long, whereas the enterovirus and adenovirus amplicons had 347 and 321 bases, respectively. The merged reads were then remapped to a custom reference set of adenovirus hexon sequences or enterovirus VP1 sequences, representative of the known types. This remapping step was accomplished by the BWA program.[8] The reads that remapped to individual reference sequences were counted and expressed as percents of the total signal. Virus types were then inferred from the distribution of the reads among the reference sequences. The coinfection of two virus types (referred to as a dual infection hereinafter) was defined as a positivity of more than one genotype of the same viral genus, with the less abundant signal exceeding 3% of the total signal. Finally, the result was exported into a Microsoft Excel xlsx table with one dimension being the sample, the other dimension being the subtypes, and the percentage of matched reads shown in the individual cells.

2.2 | The possibility of Sanger sequencing

If needed, individual PCR products (before or after the indexing step) can be sequenced by a conventional Sanger protocol using the outer portion of the primers tails (for one direction TCGTCGGCAGCGT CAGA, for the opposite direction GTCTCGTGGGCTCGGA). Of note, Sanger sequencing fails if multiple types are present in the sample, and also in case of low virus quantities when other background by-products dominate the reaction. For obvious reasons, no Sanger sequencing is feasible after multiple targets of one sample, or if multiple samples are pooled.

2.3 | Validation of the method

The method was first tested on a set of nucleic acids from viral isolates (30 types of enterovirus and 44 types of adenovirus). The genotypes of these strains were previously determined by Sanger sequencing of a wider informative capsid gene region.[5,9] To verify the ability to distinguish dual infections, we generated and sequenced mock samples consisting of two types of enterovirus or adenovirus mixed in varying mutual ratios.

2.4 | Sequencing of samples from the Norwegian birth cohort MIDIA

The method was then piloted in the Norwegian study MIDIA, a newborn cohort of children who had been selected at birth by neonatal screening for a single genotype conferring the highest risk of type 1 diabetes.[10] These children are then followed-up for serologic signs of islet autoimmunity preceding type 1 diabetes, and for antibody markers of celiac disease. A set of stool samples collected monthly between the age of 3 to 36 months is available for most subjects, and is presently being investigated for multiple viral agents in connection to the development of the aforementioned diseases (eg,[11]). Written parental consent was obtained. and the study was approved by The Regional Committee for Medical Research.

In the present study, a nested case-control set was investigated by testing 2400 archived serial stool samples. Adenoviruses were tested by quantitative real-time PCR primarily designed to detect species *Human mastadenovirus A*, *B*, and *C* but reacting also with other human adenoviruses (HAdVs).[12] Enterovirus was tested by quantitative one-tube real-time reverse transcription-polymerase chain reaction.[13] The primers detected with an equal sensitivity the *Enterovirus A-D* species (ie, members of species *Enterovirus A*, *Enterovirus B*, *Enterovirus C* and *Enterovirus D* of the genus *Enterovirus*, family *Picornaviridae*, order *Picornavirales;* the former nomenclature of these species was *Human enterovirus A-D*) but did not react with human rhinoviruses (ie, members of species *Rhinovirus A-C*). In this study, we determined genotypes in 301 samples that contained more than 10 copies of HAdV per microlitre DNA, and in 350 samples that contained more than 100 copies of human enterovirus per microlitre DNA. Of these, 50 samples were positive for both viruses.

3 | RESULTS

3.1 | Validation of sequencing on isolates of known virus types and on mock mixed samples

The sequencing protocol was validated using a set of samples with adenoviruses and enteroviruses of various types and quantities. First, we tested the performance using DNA from 43 HAdV types previously determined by Sanger methods.[9] All tested concordantly in next-generation sequencing. Then the enterovirus protocol was tested in 30 samples of various enterovirus types that had been determined first by serotyping and later by Sanger sequencing, also with a complete concordance.

To test the ability to detect multiple virus genotypes mixed in one sample in different quantities, we designed the following experiments: three different pairs of adenovirus types were chosen varying in their phylogenetic distance (HAdv1 and HAdv2; HAdv3 and HAdv5; HAdv31 and HAdv41). Forty mutual dilutions were then prepared over five orders of magnitude. In all these mock samples, the method was able to detect

TABLE 2 Types of detected viruses

A. Adenovirus				
Virus type	Total occurrence (n = 292)	As the only adenovirus type in the sample	Accompanied by another (less abundant) type in the sample	As the less prevalent virus type in the sample
HAdV-C1	**106**	93	5 (3x with HAdV-C2; 1x with HADv-C5; 1x with HAdV-A31)	8
HAdV-C2	**133**	119	5 (4x with HAdV-C1; 1x with HAdV-C5)	9
HAdV-B3	**20**	18	2 (1x with HAdV-C1; 1x with HAdV-C2)	
HAdV-C5	**29**	23	4 (2x with HAdV-C1; 2x with HAdV-C2)	2
HAdV-A31	**11**	7	3 (with HAdV-C2)	1
HAdV-F41	**12**	11	1 (with HAdV-C1)	
HAdV-C57	**1**	1		

B. Enterovirus				
Virus type	Total occurrence (n = 330)	As the only enterovirus type in the sample	Accompanied by another (less abundant) type in the sample	As the less prevalent virus type in the sample
CVA1	**2**	1		1
CVA2	**44**	42	1 (with E25)	1
CVA4	**34**	33	1 (with CVA6)	
CVA5	**15**	13	2 (1x w. CVA2, 1x w. CVB3)	
CVA6	**24**	20	1 (CVA1)	3
CVA8	**2**	2		
CVA9	**20**	19		1
CVA10	**24**	22	2 (1x w. CVB3, 1x w. CVB5)	
CVA14	**1**	1		
CVA16	**23**	22		1
CVA22	**1**	1		
CVB1	**15**	14		1
CVB2	**20**	19		1
CVB3	**12**	10		2
CVB4	**10**	9	1 (with EV-A71)	
CVB5	**12**	10	1 (with CVB2)	1
E3	**4**	4		
E6	**7**	7		
E7	**2**	2		
E9	**10**	9	1 (with CVA16)	

(Continues)

TABLE 2 (Continued)

B. Enterovirus				
Virus type	Total occurrence (n = 330)	As the only enterovirus type in the sample	Accompanied by another (less abundant) type in the sample	As the less prevalent virus type in the sample
E11	**10**	6	3 (with 1x CVA6; 1xCVA9; 1x E18)	1
E13	**1**	1		
E18	**9**	7	1 (with E11)	1
E25	**20**	18	1 (with CVA6)	1
E30	**8**	8		
EV-A71	**14**	13		1
PV2 - Sabin	**1**	1		

n (genotyped samples) = 292; the genotyping failed in remaining 9 of the 301 adenovirus-positive samples.
n (genotyped samples) = 330; the genotyping failed in remaining 20 of the 350 adenovirus-positive samples.
The "Total occurrence" columns is the count of samples where the virus type occurred either alone, or in combination, either as the more abundant or less abundant virus type.
CVA, coxsackievirus A; CVB, coxsackievirus B; E, echovirus; EV, enterovirus

and distinguish both types. The quantity ratio between the major and minor type in the mock sample was distorted from the mixture ratio in favor of the less represented type—the most pronounced difference was for a dilution of 1:1000, where the ensuing ratio of sequencing reads per type was 3:97. Analogical tests were performed on mixtures of enterovirus types (echovirus 18 and echovirus 5; echovirus 30 and echovirus 4; echovirus 2 and echovirus 15; echovirus 31 and echovirus11; echovirus 26 and echovirus 31; echovirus 30 and echovirus 15) for which we made three quantity ratios. All types from the mixtures were detected, again with substantial differences of the sequencing signal to the mock sample dilution ratio.

3.2 | Adenovirus and enterovirus genotyping in the MIDIA newborn cohort study

The genotyping protocol was then applied to 301 adenovirus-positive and to 350 enterovirus-positive samples of nucleic acids extracted from stool supernatants. The virus type assignment was based on a mean sequencing depth of 22 676 reads (median 19 057, interquartile range 13 653-23 806 reads).

In 292 of 301 adenovirus-positive samples (97%), the partial hexon PCR products were successfully sequenced and the genotype (s) determined. The most frequent type was HAdv2 present in 133 of 290 (45.8%) adenovirus-genotyped samples, mostly alone, but also in combination with HAdv1 (7; 2.4%), HAdv31 (3; 1.0%), and HAdv5 (3; 1.0%). The second most common adenovirus type was the HAdv1 type, present in 106 (36.5%) adenovirus-genotyped samples, again predominantly alone, but 13 samples carried an additional adenovirus type (Table 2A).

Of the 350 enterovirus-positive samples, the partial VP1 sequence was determined and genotype assigned in 330 (94.3%). The most often observed type was CV-A2 in 44 of 330 (13.3%) enterovirus-genotyped samples, predominantly as the sole genotype, but also in combination with E-25 or CV-A5 (one sample each). Other frequent enterovirus genotypes were CV-A4 (10.3%), CV-A10, and CV-A6 (both 7.3% of the enterovirus-genotyped samples), Table 2B.

3.3 | Detection of dual infections

Simultaneous infections with more than one type of the same genus was noted in 20 of 292 (6.8%) adenovirus-positive samples and 17 of 330 (5.2%) enterovirus-positive samples. The longitudinal character of the observation allowed us to detect individuals who shed adenovirus continuously over several months: some of them harboring the same genotype of the virus, some contracting various genotypes over time (Figure 1). At least one infection with two types of the same viral genus was experienced by 23 of 83 (28%) of the children observed over the 3 years' observation period: nine had dual infections only with enterovirus strains, nine only with adenovirus strains, and five experienced dual infections with both adenoviruses and an enteroviruses.

In formal statistical analyses by regression modeling, the likelihood of a dual virus type infection in a sample was not associated with either of age, gender, or the viral load. The probability of a dual infection with two types of the same genus was, however, lower ($P_{difference} < 0.01$) than would be expected under the assumption of independence of infections.

4 | DISCUSSION

We developed a sequencing protocol for genotyping adenovirus and enterovirus directly from clinical material, using massive parallel

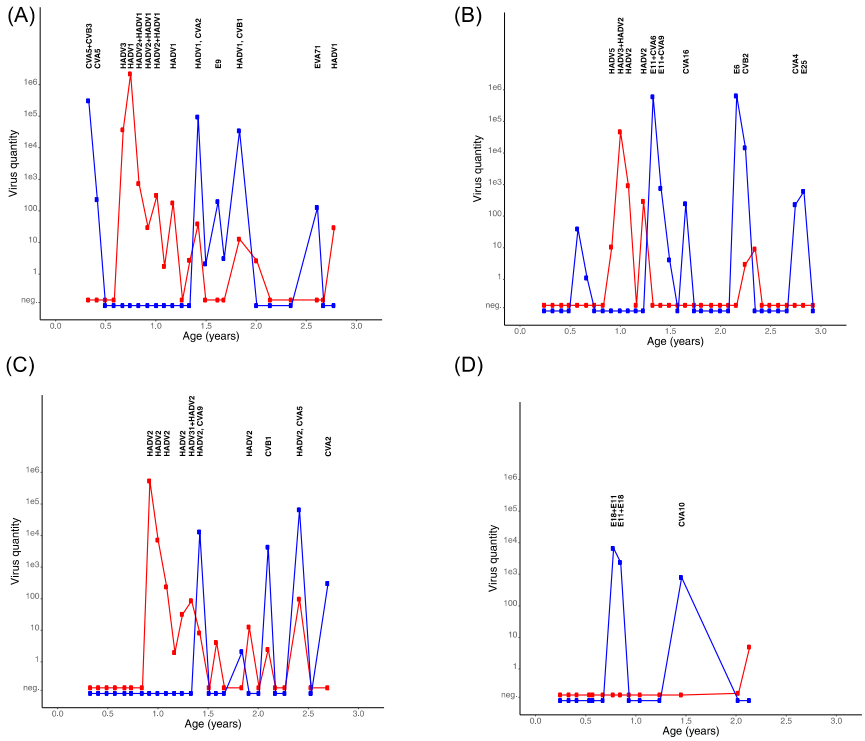


FIGURE 1 Examples of individuals with prolonged infections, and with multiple simultaneous infections. Virus quantity (blue, enterovirus; red, adenovirus) in monthly taken stool samples. A, A dual infection with two coxsackieviruses followed by a prolonged infection with two adenovirus types. B, A dual infection with adenovirus, followed by two monthly samples with dual enterovirus infection. C, A prolonged infection with HAdV-C2 superimposed on an infection with HAdV-A31. D, A 2-months' positivity for two echovirus types.

amplicon sequencing with degenerate primers. The protocol proved its ability to detect a wide repertoire of virus types alone or in combination and was resistant to foreign DNA and RNA in stool samples.

4.1 | Choice of primers

For adenovirus genotyping, we used the seventh hypervariable region of the hexon gene. This region is known to correlate with the virus type as it contains serotype-specific epitopes, and it also correlates with the adenovirus species (with the exception of HAdV-E4 that clustered with HAdV-B types). Enterovirus primers targeting the proximal portion of VP1 were redesigned from the highly cited work by Nix et al[5] to make the assay fully nested. Our choice was to design more variants of better-targeted primers, rather than utilize a higher degree of degeneracy. By this more specific design, we aimed to limit the amplification of spurious products. Testing of the protocol in genotyping of mock mixtures, as well as in real samples, indicated its capability of detecting multiplex infections and showed resistance to concomitant nonviral DNA.

4.2 | The repertoire of virus types in healthy Norwegian children

The study of stools longitudinally collected from Norwegian infants and young children demonstrated a high success rate in determining the virus type. The virus type repertoire was similar to what had been previously found in other children of the same cohort, including the surprisingly high occurrence of EV-A71.[14-16] Interestingly, although the detection primers anneal perfectly to deposited sequences of enterovirus D68, this type was not observed in our sample set—this may be related not only to the rarity of HEV-D68 in stool,[17] but also to the time of collection of the samples—most samples were collected between 2006 and 2009, long before the 2014-2015 outbreaks in the USA and Europe.[18]

The repertoire of found adenovirus types was skewed by the targeting of the adenovirus real-time PCR detection assay identifying positive samples that were later sequenced. This PCR assay primarily detected HAdVs of species *A*-*C*[12] and its sensitivity to other species was lower. Thus, the adenovirus results are not entirely representative because the occurrence of HAdV species *D* to *G* is underestimated.

4.3 | Dual infections

Dual infections (ie, simultaneous infections with two virus types of the same genus) offer an opportunity for homologous recombination of viral genomes. Evidence for past recombination is available both in adenovirus and enterovirus.[19-22] Children with dual infections may, thus, be an important reservoir for creating new recombinant strains: we observed two types of adenovirus or enterovirus in more than 1 of 20 of virus-positive stool samples taken from mostly asymptomatic Norwegian infants and young children. Over a fourth of these children had at least one documented dual type infection with adenovirus or enterovirus over the 3 years of observation.

The true dual infection rate is probably even higher, as some infections of short duration may have been missed because of the monthly interval between samples. In addition, some dual infections may have been missed owing to technical reasons. First, we imposed a relatively strict threshold of 3% for the minor virus type signal. This level is still at least one decadic order of magnitude higher than the level of false signals that may arise from the known imperfections in sample indexing on the Illumina platform. Second, truly minor virus subpopulation may not be efficiently amplified due to PCR bias, being overshadowed by the signal from the major virus type population. The latter technical shortcoming cannot be presently remedied, as even unbiased metagenomic sequencing or virus sequence capture methods [23] rely on nucleic acid amplification at some stages of their protocols. Any amplification can induce bias, be it random PCR[24] or multiple displacements amplification.[25]

The frequency of dual infections was indeed significantly lower than would be expected under the assumption of independence. With the overall frequency of adenovirus 12.5% (301 of 2400), upon the assumption of independence and after a slight simplification, approximately 11% of positive samples should contain two types of adenovirus; this is roughly two times more than was actually observed. Apart from the above-described reasons hampering the detection of all dual infections, this observation may also reflect the phenomenon of in vivo viral interference: a second viral infection is less likely because of the antiviral state of the already infected tissues, as reported previously in the clinical setting.[26,27]

Figure 1 exemplifies several subjects having dual infections with adenovirus or enterovirus types. Subject A shed two adenovirus types for three months, within seven successive months of adenovirus positivity. Subject B experienced both adenovirus and enterovirus dual infections. Subject C had a long infection with adenovirus C2, with an episode of adenovirus A31 superimposed. Whether this reflects their general propensity towards infections, long-term persistence, or rich external sources of infection, is unclear. Finally, subject D is selected as an example of short-term excretion of two types of enterovirus.

4.4 | Strengths and limitations

The main strength of the present approach is the ability to distinguish individual reads among the complex PCR signal. This brings not only high discriminatory power for distinguishing simultaneous infections with two or more genotypes of the same viral genus, but also increases sensitivity. Most PCR protocols for genotyping of viruses, and especially of RNA viruses of high mutation rate, use primers with degenerate bases or inosines, low annealing temperatures, increased magnesium concentrations in the PCR mix, and other means of relaxing stringency of PCR to amplify as many different viral variants as possible. This often causes unspecific amplification—especially in a complex material like stool with high abundance of foreign nonviral nucleic acids. This unspecific amplification was not resolvable by Sanger sequencing without subcloning of products or the use of multiple type-specific primer combinations, but now massive parallel sequencing dissects the signal into individual read pairs, so any unwanted amplicons are eliminated by bioinformatic steps, namely by the remapping to the set of sequences of reference virus types.

A technical limitation is the maximum sequenced amplicon length. The Illumina MiSeq instrument is the only one capable of relatively high degree of multiplexing (384 reactions per sequencing run), with sequencing length in one direction of 250 bp; recently this has increased to 300 bp. The combination of forward and reverse sequencing reads allowed us to cover the seventh hypervariable region (ie, loop 2) of the adenovirus hexon and an informative portion of the enterovirus VP1 gene. However, two pairs of presently known adenovirus types cannot be reliably distinguished (HAdV-D22 vs HAdV-D53 and HAdV-D29 vs HAdV-D56).

Another limitation of the method is connected to the balance between capacity and costs. The need for accumulating high enough count of samples for the sake of cost-effectiveness of the massive parallel sequencing makes the protocol suitable chiefly for retrospective studies or in applications where the time to result is not critical.

4.5 | Conclusions

In conclusion, we here present a simple and high-throughput method for adenovirus and enterovirus typing, which is resistant to PCR by-products and dissects coinfections. We demonstrate that a considerable proportion of adenovirus and enterovirus infections in healthy infants are coinfections with two virus types that belong to the same genus.

ACKNOWLEDGMENTS

This work is supported by Ministry of Health of the Czech Republic: grant number 15-29078A to Charles University in Prague. The MIDIA study from which the samples originated, was conceived by Kjersti Skjold Rønningen and funded by the Research Council of Norway. We would like to thank Christian R. Kahrs and Ketil Størdal for the permission to work with the samples, and Kaja Klykken Aas and Nina Stensrud from Norwegian Institute of Public Health for handling stool samples in the MIDIA project. The work of public health nurses in the MIDIA project is also acknowledged (Asbjørg Skorge, Liv Kjeldstadli, and Turid Wetlesen). Professor Stig Jeansson, University of Oslo, Norway, is gratefully acknowledged for a

generous gift of a collection of enterovirus reference types donated to our laboratory in 1999.

ORCID

Ondrej Cinek http://orcid.org/0000-0002-0526-8714

REFERENCES

1. Li L, Victoria J, Kapoor A, et al. A novel picornavirus associated with gastroenteritis. *J Virol*. 2009;83:12002-12006.
2. Zhu Y, Zembower TR, Metzger KE, Lei Z, Green SJ, Qi C. Investigation of respiratory syncytial virus outbreak on an adult stem cell transplant unit by use of whole-genome sequencing. *J Clin Microbiol*. 2017;55:2956-2963.
3. Pérez-Brocal V, García-López R, Nos P, Beltrán B, Moret I, Moya A. Metagenomic analysis of Crohn's disease patients identifies changes in the virome and microbiome related to disease status and therapy, and detects potential interactions and biomarkers. *Inflamm Bowel Dis*. 2015;21:2515-2532.
4. Kramná L, Kolářová K, Oikarinen S, et al. Gut virome sequencing in children with early islet autoimmunity. *Diabetes Care*. 2015;38:930-933.
5. Nix WA, Oberste MS, Pallansch MA. Sensitive, seminested PCR amplification of VP1 sequences for direct identification of all enterovirus serotypes from original clinical specimens. *J Clin Microbiol*. 2006;44:2698-2704.
6. Lin J, Kramna L, Autio R, Hyoty H, Nykter M, OC. Vipie: web pipeline for parallel characterization of viral populations from multiple NGS samples. *BMC Genomics*. 2017;18:378.
7. Magoc T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*. 2011;27:2957-2963.
8. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754-1760.
9. Lu X, Erdman DD. Molecular typing of human adenoviruses by PCR and sequencing of a partial region of the hexon gene. *Arch Virol*. 2006;151:1587-1602.
10. Stene LC, Witsø E, Torjesen PA, et al. Islet autoantibody development during follow-up of high-risk children from the general Norwegian population from three months of age: Design and early results from the MIDIA study. *J Autoimmun*. 2007;29:44-51.
11. Tapia G, Cinek O, Rasmussen T, et al. Human enterovirus RNA in monthly fecal samples and islet autoimmunity in Norwegian children with high genetic risk for type 1 diabetes: the MIDIA study. *Diabetes Care*. 2011;34:151-155.
12. Claas ECJ, Schilham MW, de Brouwer CS, et al. Internally controlled real-time PCR monitoring of adenovirus DNA load in serum or plasma of transplant recipients. *J Clin Microbiol*. 2005;43:1738-1744.
13. 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.
14. Cinek O, Witso E, Jeansson S, et al. Longitudinal observation of enterovirus and adenovirus in stool samples from Norwegian infants with the highest genetic risk of type 1 diabetes. *J Clin Virol*. 2006;35:33-40.
15. Witso E, Palacios G, Cinek O, et al. High prevalence of human enterovirus A infections in natural circulation of human enteroviruses. *J Clin Microbiol*. 2006;44:4095-4100.
16. Witsø E, Palacios G, Rønningen KS, et al. Asymptomatic circulation of HEV71 in Norway. *Virus Res*. 2007;123:19-29.
17. Harvala H, Broberg E, Benschop K, et al. Recommendations for enterovirus diagnostics and characterization within and beyond Europe. *J Clin Virol*. 2018;101:11-17.
18. Holm-Hansen CC, Midgley SE, Fischer TK. Global emergence of enterovirus D68: a systematic review. *Lancet Infect Dis*. 2016;16:e64-e75.
19. Robinson CM, Singh G, Lee JY, et al. Molecular evolution of human adenoviruses. *Sci Rep*. 2013;3:1812.
20. Lukashev AN. Role of recombination in the evolution of enteroviruses. *Rev Med Virol*. 2005;15:157-167.
21. Lukashev AN, Ivanova OE, Eremeeva TP, Iggo RD. Evidence of frequent recombination among human adenoviruses. *J Gen Virol*. 2008;89:380-388.
22. Wang Y, Li Y, Lu R, et al. Phylogenetic evidence for intratypic recombinant events in a novel human adenovirus C that causes severe acute respiratory infection in children. *Sci Rep*. 2016;6:23014.
23. Briese T, Kapoor A, Mishra N, et al. Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis. *mBio*. 2015;6:e01491.
24. Kramná L, Cinek O. Virome sequencing of stool samples. *Methods Mol Biol*. 2018;1838:59-83.
25. Li L, Delwart E. From orphan virus to pathogen: the path to the clinical lab. *Curr Opin Virol*. 2011;1:282-288.
26. Seppälä E, Viskari H, Hoppu S, et al. Viral interference induced by live attenuated virus vaccine (OPV) can prevent otitis media. *Vaccine*. 2011;29:8615-8618.
27. Wang Z, Malanoski AP, Lin B, et al. Broad spectrum respiratory pathogen analysis of throat swabs from military recruits reveals interference between rhinoviruses and adenoviruses. *Microb Ecol*. 2010;59:623-634.

SUPPORTING INFORMATION

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

How to cite this article: Cinek O, Kramna L, Mazankova K, et al. Virus genotyping by massive parallel amplicon sequencing: adenovirus and enterovirus in the Norwegian MIDIA study. *J Med Virol*. 2019;91:606-614. https://doi.org/10.1002/jmv.25361

10.2.1 Supplementary materials of paper II

Supplemental material:
Protocol for genotyping of entero- and adenoviruses using NGS

Ondrej.Cinek@Lfmotol.cuni.cz

Name of experiment / test _____

Step	*When*	*Who*	*Notes*
Selection of samples			
Selection of tests/primers			
Protocol written			
First round of PCR			
Pooling of tests of the same sample			*Confirm that ratios are noted in the table:* ____
Cleanup 1			Ampure:sample ratio: _____
Indexing PCR			
Cleanup 2			Ampure:sample ratio: _____
Quantification of libraries			
Normalization and pooling			
Bioanalyzer control NGS run			
Data analysis			

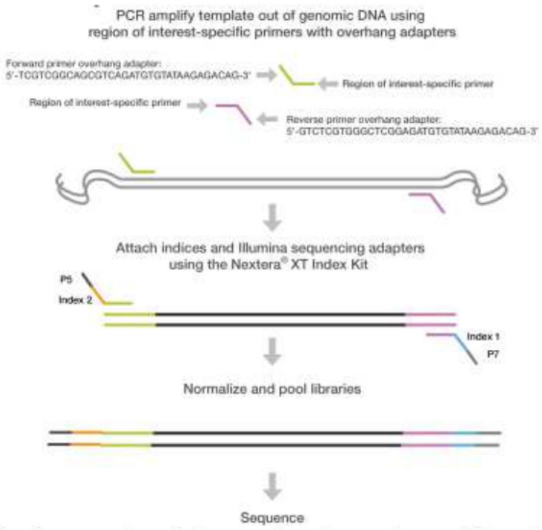
Supplemental Material: Protocol

The principle

Before starting with this protocol, please refer to the document on 16S profiling by Illumina (document 15044223 Rev. B, http://support.illumina.com/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf

Our method is designed with the "alternative amplicon primers" as mentioned on page 3 of this document.

Please see the diagram by Illumina for our primers (green and violet).



User-defined forward and reverse primers that are complementary upstream and downstream of the region of interest are designed with overhang adapters, and used to amplify templates from genomic DNA. A subsequent limited-cycle amplification step is performed to add multiplexing indices and Illumina sequencing adapters. Libraries are normalized and pooled, and sequenced on the MiSeq system using v3 reagents.

Supplemental Material: Protocol

Outline of the procedure:

(1) PCR amplification of the target

- uses pathogen-specific primers tailed with Illumina adapters.
- if the PCR is nested (like in enterovirus), the Illumina tails are used in the *last round* of amplification.

(2) Pooling of the PCR products (optional)

- pooling by sample is done is more than one amplicon is prepared from a given sample

(3) Size-limiting PCR clean-up

- here the PCR products are purified of the unused primers and dNTP
- depending on the Ampure concentration, it also limits the minimum retained size of the product - this is useful in removing primer dimers and short by-products.
- As a control of the procedure, the purified products can be sequenced by a Sanger protocol before they have been pooled.

(4) Index PCR (to attach indices and sequencing adaptors)

- Here we add indices for the Illumina sequencing. The indices distinguish samples and enable analysis of multiple samples in the same sequencing run.
- The addition is based on the complementarity of the tails from the specific PCR, and the primers from the Illumina index set.

(5) Library quantification, normalization, and pooling

- This is done using normal workflow as for any other library for the MiSeq machine.
- Specified in a separate document

(6) Sequencing and analysis

- as per a separate protocol, use the "Generate FASTQ Workflow"
- this is an amplicon sequencing, therefore **PhiX** must be added at a concentration from 5% (v3 kit) to 20% (v2 kit)

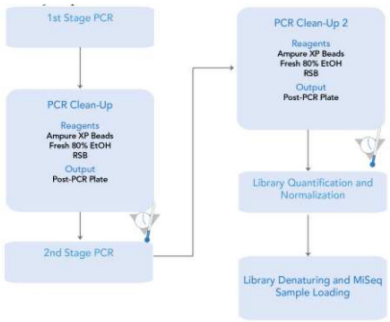


Image by Illumina

PLATE LAYOUT. Name of experiment / test _____

Plate A	Plate C
1) Amplification: when _____ who _____ stored in: _____ 2) Pooling: when _____ who _____ stored in: _____ 3) Size-limiting clean-up: when _____ who _____ stored in: _____ 4) Indexing: when_____ who _____ stored in: _____ purified after indexing: _____ 5) KAPA quantification: Normalized pool:	1) Amplification: when _____ who _____ stored in: _____ 2) Pooling: when _____ who _____ stored in: _____ 3) Size-limiting clean-up: when _____ who _____ stored in: _____ 4) Indexing: when_____ who _____ stored in: _____ purified after indexing: _____ 5) KAPA quantification: Normalized pool:
Plate B	**Plate D**
1) Amplification: when _____ who _____ stored in: _____ 2) Pooling: when _____ who _____ stored in: _____ 3) Size-limiting clean-up: when _____ who _____ stored in: _____ 4) Indexing: when_____ who _____ stored in: _____ purified after indexing: _____ 5) KAPA quantification: Normalized pool:	1) Amplification: when _____ who _____ stored in: _____ 2) Pooling: when _____ who _____ stored in: _____ 3) Size-limiting clean-up: when _____ who _____ stored in: _____ 4) Indexing: when_____ who _____ stored in: _____ purified after indexing: _____ 5) KAPA quantification: Normalized pool:

Supplemental Material: Protocol

PLATE LAYOUT. Name of experiment / test _____

Layout of MiSeq indices in the whole sequencing run. One index combination per sample, 384 samples maximum.

Plates: A | C / B | D

The indices come from the FC-131-2001 (set A) and FC-131-2004 (set D) kits.

	1 N701	2 N702	3 N703	4 N704	5 N705	6 N706	7 N707	8 N710	9 N711	10 N712	11 N714	12 N715	13 N716	14 N718	15 N719	16 N720	17 N721	18 N722	19 N723	20 N724	21 N726	22 N727	23 N728	24 N729	
A S502																									A
B S503																									B
C S505																									C
D S506																									D
E S507																									E
F S508																									F
G S510																									G
H S511																									H
I S513																									I
J S515																									J
K S516																									K
L S517																									L
M S518																									M
N S520																									N
O S521																									O
P S522																									P
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	

Supplemental Material: Protocol

(1) Amplification

List of available virus tests and protocols for amplification: **Tick tests which will be performed in the present set of samples.**

[ ] ADENOVIRUS - use the *Amplification_sheet_ADENOVIRUS.docx* (in the Annex)

[ ] ENTEROVIRUS - use the *Amplification_sheet_ENTEROVIRUS.docx* (in the Annex)

File the amplification protocol and gel photos with this document!

(2) Pooling of individual PCR products

(**Not needed if only one target is amplified** for each sample.)

This pooling must be done BEFORE clean-up and indexing. The products can be pooled only if they **belong to the same sample**, and differ **only** by the amplified target.

What is the total calculated volume of the pooled PCR products?
(a) over 50 µl: use a deep well plate as the target plate.

- deep well plate 500 µl, Eppendorf cat. no. 30503104 Deepwell Plate 96, DNA Lo Bind (0,5 ml).
- Sealed with 0030 127.552 Sealing mat (for DWP 96/1000)

A fraction of 50 µl will be then taken for downstream procedures: purification, indexing, quantification and sequencing.

(b) under 50 µl:
Use a PCR plate for pooling.

Procedure

- **Label** the target plate
- Using a multichannel pipette, transfer all content of the source plates into the target plate
- **Mix thoroughly** by pipetting
- If a deep well was used for pooling, transfer 50 µl into a new PCR plate for downstream procedures.

(3) PCR clean-up

The concentration of Ampure in the clean-up depends on the size of the **SMALLEST correct band in the pool.**

- If less than 250 bp, use 1.8x Ampure to product;
- if 250-500 bp, use 1x,
- if more than 500 bp, use 0.6 x

Smallest band size in the pool is _____bp

Ampure-to-product ratio will be: _____ : 1

Pooling and Ampure[1] purification record:

Plate	Pooled by:	Date / Time:	Purified by:	Date / Time:	On robot / Manually:
A					
B					
C					
D					

[1] See Annexes for options

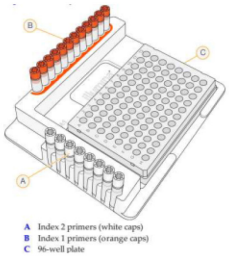
(4) Indexing PCR

Done by:_____ **Date:** _____ **Notes:** _____

Use original primers from the
Illumina kits FC-131-2001 and FC-131-2004.

Rows A-H: index 2 primers
("S5..." primers) - here shown with white caps

Columns 1-12: index 1 primers
("N7..." primers) - here shown with red caps



Master mix:

	µl per sample for 15µl	µl per sample for 10µl	Master mix	(final conc.)	Lot #, open on
PCR water	1,50	1			
2x KAPA HiFi HotStart Mix	7,50	5		1x	
primer 1 (7..)	1,50	1	-		
primer 2 (5..)	1,50	1	-		
Purified first round product[2]	3,00	**2**	-	20%	
TOTAL VOLUME	**15**	**10**			

Index PCR plate sample layout

7.. index:

5.. index	1	2	3	4	5	6	7	8	9	10	11	12
A:												
B:												
C:												
D:												
E:												
F:												
G:												
H:												

Index PCR profile:

denat	95°C	3:00
c. 8x[3]	95°C	0:30
	55°C	1:00
	72°C	0:30
final synth.	72°C	5:00
hold	10°C	forever

Cycler: _____ **Date / time:** _____
You may check the result on a gel electrophoresis.

Proceed with cleanup, normalization, pooling, addition of *PhiX* and sequencing (separate documents) **as with any other NGS amplicon library.**

[2] Can be also balanced (by eluting to a volume higher than 50µl) to avoid the need of preparing master mix.
[3] It is this short, but it works. The count of cycles may be adjusted.

(5) Sequencing

- NGS on MiSeq

Use the v2; 2x250 cycles kit; Illumina catalogue number MS-102-2003
(MiSeq Reagent kit v.2, 500-cycle, 15M, 8 Gb)

- Sanger sequencing.

Use only for individual products, does not work if different products were pooled.

Sequencing primers: The following primers anneal to the adaptors of MiSeq. Bi-directional sequencing requires one reaction with the "for5" primer, and another with the "rev7" primer.

for5_sanger1	TCGTCGGCAGCGTCaga
rev7_sanger1	GTCTCGTGGGCTCGGa

Annexes

PCR clean-up, manual procedure

All steps are done at room temperature! Before starting, also Ampure should reach room temperature.
Chemicals: 75% ethanol (400 µl per position), injection water (100 µl per position)
Material: magnetic stand, 100 µl and 200 µl pipette tips, preferably with filter.

1. Mix Ampure thoroughly in order to fully resuspend magnetic particles.

2. Outside the magnetic stand, add Ampure to PCR product following the ratios:

PCR pool volume µl	**Ampure µl** for products of **100 bp - 250 bp** (1.8x)	**Ampure µl** for products of **250 bp - 500 bp**(1x)	**Ampure µl** for products of **~500 bp** and longer (0.6x)
10	18	10	6*
15	27	15	9*
20	36	20	12
50	90	50	30

*) this is likely to fail - the total volume will not reach the magnet, and the product will not form a ring.

3. Mix the product with Ampure 10 times by pipetting.

4. Incubate 5 min - still off the magnetic stand.

5. Place the plate **on the magnetic stand**, and incubate for 2-3 min or until the suspense clears.

6. While still on the stand, aspirate and remove the liquid[4]. Take care not to scratch off the ring of magnetic beads.

7. Add 200 ul 75% ethanol to each well and incubate 30 s. Remove ethanol by aspiration.

8. Repeat step 7.

9. Let the particles air-dry for 5-10 minutes. Do not over-dry, the film of particles should not crack.

10. Remove the plate from the magnetic stand.

11. Add 55 µl PCR-grade water into each position, and mix by pipetting 10 times.

12. Incubate at room temperature for 2 minutes

13. Back on the magnetic stand, incubate for 2-3 min or until the suspense clears.

14. Transfer 50 µl purified product into a new plate. DO NOT transfer the magnetic particles, otherwise the purified product cannot be frozen and stored.

[4] This can be done **either** by a multichannel pipette, **or** by a vacuum manifold (with slender pipette tips attached) connected to a collection bottle.

ANNEX: Amplification_sheet_ADENOVIRUS

Amplification of adenovirus for NGS genotyping

Amplification of adenovirus hexon HVR7 with primers flanked by sequencing adaptors.

Done by:_____ **Date:** _____

[5] or 35 cycles in program MULTI_c35 if the samples have a high adenovirus quantity

ANNEX: Amplification_sheet_ENTEROVIRUS

Amplification of enterovirus for NGS genotyping

Amplification of enterovirus VP1 fragment with primers flanked by sequencing adaptors.
Adapted from Nix et al 2006 JCM 2698-2704

1) Reverse transcription with specific primers

Who:_____ **Date:** _____

Chemicals	Conc.	Open date
SuperScript III Reverse Transcriptase III, Invitrogen, cat no 18080-093, -0444, -085	buffer 5x enzyme 200U/ul 0,1 M DTT	
RNasin RNA inhibitor, Promega	40 U /ul	
dNTP 5 mM each primers: NIX-RT[6]	4 x 5 mM 4x 1µM	

RT Mix:

Component	Per one reaction	Master mix	Final concentration
denatured)* primer mix **Nix-RT**	**1**		1 pmol (abs.) each
5x First-strand buffer	**2**		1x
0,1 M DTT	**0,5**		0,005 M
RNAsin 40U/ul	**0,5**		20U / r.
reverse transcriptase	**0,5**		100U / r.
dNTP 5 mM each	**0,5**		250 uM each
RNA[7]	*5*	-	50% vol
Total	**10**		

*) the primer denaturation is carried out as the first step of RT mix preparation

Procedure:

a) **Prepare RT mix**: prefill into a 1,5 ml tube the Nix-RT primer mix, incubate at 65°C for 5 min and put on thawing ice. While the tube is still kept on ice, add the remaining components (buffer, DTT...). Mix by vortexing, briefly spin and keep on ice.
b) Pipette 5 µl RNA samples into sterile RNAse-free PCR tubes. Close carefully, briefly spin.
c) Incubate RNA samples in the PCR tubes in a **pre-PCR** cycler or a heat block at 65°C for 5 min
d) **Snap chill** on thawing ice (directly from 65°C).
e) While still on ice: add 5µl **RT mix** and mix by pipetting. Briefly spin.
f) Run on a **pre-PCR cycler** using program ***RT-NIX2***

22°C	15 min	primer annealing
50°C	45 min	reverse transcription
70°C	15 min	reverse transcriptase inactivation
10°C	forever	storage

g) This single-strand cDNA should be immediately used as a PCR template or stored below -20°C at the pre-PCR room.

Cycler: when? From _____ to _____

Stored: when and where: _____

[6] The **Nix-RT** primer mix consists of 1 µM each of NIX-AN32, NIX-AN33, NIX-AN34, and NIX-AN35. To be prepared from 100 µM stock solutions as follows: 5 µl each of 4 primers are added to 480 µl PCR-grade water, and mixed.
[7] Do not forget to run a positive and negative control along with each batch of samples.

ANNEX: Amplification_sheet_ENTEROVIRUS

2) First round of PCR

Who:_____ **Date:** _____

ed culture) or NIX-R1c25 (culture with >10^8 particles/µl).

ANNEX: Amplification_sheet_ENTEROVIRUS

3) Second round of PCR

Who:_____ **Date:** _____

c35; or 40 cycles in MULTI55_c40