



**HAL**  
open science

# Standardized syndromic and virologic surveillance facilitating unbiased signal detection of emerging and re-emerging infectious diseases

Patrick Erich Obermeier

► **To cite this version:**

Patrick Erich Obermeier. Standardized syndromic and virologic surveillance facilitating unbiased signal detection of emerging and re-emerging infectious diseases. Microbiology and Parasitology. Université Bourgogne Franche-Comté, 2021. English. NNT : 2021UBFCE019 . tel-04348098

**HAL Id: tel-04348098**

**<https://theses.hal.science/tel-04348098v1>**

Submitted on 16 Dec 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## **THÈSE DE DOCTORAT DE L'ÉTABLISSEMENT UNIVERSITÉ BOURGOGNE FRANCHE-COMTÉ**

Ecole doctorale n° 554

ÉCOLE DOCTORALE Environnements - Santé

Doctorat de médecine, microbiologie et maladies transmissibles

DOCTORAT EUROPÉEN - DOCTOR EUROPAEUS

Par

M. OBERMEIER Patrick Erich

### **Surveillance syndromique et virologique standardisée facilitant la détection non-biaisée des maladies infectieuses émergentes et ré-émergentes**

---

### **Standardized syndromic and virologic surveillance facilitating unbiased signal detection of emerging and re-emerging infectious diseases**

Thèse présentée et soutenue à Besançon, le 15 décembre 2021

Composition du Jury :

M. CHRISTMANN Daniel	Professeur émérite de maladies infectieuses, Université de Strasbourg	Président/Rapporteur
Mme BERGER Annemarie	Professeur de virologie médicale, Université Goethe, Francfort-sur-le-Main	Rapporteur
M. LEPILLER Quentin	Maître de Conférence, Praticien Hospitalier, UBFC, CHU, Besançon	Examineur
M. EBERLE Josef	Professeur de virologie, Université Louis-et-Maximilien de Munich	Examineur
Mme RATH Barbara	Directrice de recherche, UBFC, Besançon	Directrice de thèse
M. OPOTA Onya	Privat-Docteur et Maître d'enseignement et de recherche, Lausanne University Hospital, Lausanne	Codirecteur de thèse
Mme VUITTON Dominique	Professeur émérite d'immunologie clinique, UBFC, Besançon	Invité

**Titre :** Surveillance syndromique et virologique standardisée facilitant la détection non-biaisée des maladies infectieuses émergentes et ré-émergentes

**Mots clés :** Surveillance syndromique, virus, gravité de la maladie, séquençage de nouvelle génération, reconnaissance des formes, apprentissage automatique

**Résumé :** Les maladies infectieuses constituent une menace pour la santé mondiale, les infections respiratoires et du système nerveux central (SNC) étant parmi les principales causes de morbidité et de mortalité, en particulier chez les enfants. Nous surveillons généralement les maladies infectieuses (ré) émergentes sur la base de bases de données de santé publique ou de réseaux de surveillance régionaux, nationaux et internationaux, qui reposent tous sur des rapports de prestataires de soins de systèmes de vigilance sentinelles ou des patients eux-mêmes. La plupart des données de santé publique ne sont pas collectées au niveau individuel des patients et les parties prenantes ont exprimé à plusieurs reprises des inquiétudes concernant l'opportunité et l'exactitude des estimations de la morbidité, en particulier dans les situations d'épidémie/pandémie. Sans des critères standardisés et un diagnostic des maladies infectieuses au lit des patients, l'association entre une maladie particulière et un agent pathogène particulier peut être faussée et retardée. Pour étudier ce sujet, la Vienna Vaccine Safety Initiative, une organisation de recherche internationale à but non lucratif, a conçu deux programmes qualité entre 2009 et 2015 ayant pour but d'améliorer le contrôle des infections et la surveillance des maladies infectieuses dans l'un des plus grands centres de soins tertiaires pédiatriques d'Europe à la Charité Centre médical universitaire de Berlin, en Allemagne, en collaboration avec l'Institut Robert Koch : les présentations cliniques et la gravité de la maladie chez les patients atteints de syndrome grippal ont été évaluées en temps réel à l'aide de la technologie de santé mobile. Les applications mobiles ont permis l'application de critères standardisés pour la classification des cas, le risque et la gravité. L'exhaustivité des données conformément aux procédures opérationnelles standard a contribué à réduire les biais de vérification. Un programme similaire existait pour les patients suspects d'infection du SNC. Dans le présent projet de thèse de doctorat, nous nous appuyons sur les enseignements tirés de ces deux programmes d'amélioration.

Nous allons maintenant regrouper deux ensembles de données cliniques hautement standardisées et précises de plusieurs centaines de patients pédiatriques présentant une infection respiratoire et du SNC aux biobanques correspondantes pour une analyse virologique approfondie. Les analyses par PCR conventionnelle seront complétées par des analyses de métagénomique virale. En détail, nous aborderons :

- 1) L'utilité de la classification automatique des cas et de la métagénomique pour combler les « lacunes diagnostique » dans les infections complexes du SNC.
- 2) L'utilisation de la technologie de santé mobile liée au diagnostic PCR et à la métagénomique pour améliorer la gestion clinique des infections respiratoires.
- 3) L'évolution génétique de l'adénovirus C basée sur le séquençage du génome entier.
- 4) Une approche de médecine de précision pour étudier la gravité de la maladie et les schémas cliniques d'infection à adénovirus respiratoire à l'aide d'une méthodologie d'apprentissage automatique.

Les résultats des études discutées ici démontrent que la combinaison de :

- (a) L'évaluations de cas cliniques standardisées, aidées par la technologie de santé mobile, permettent la reconnaissance des formes, l'apprentissage automatique et le calcul fiable de mesures de gravité de la maladie et
- (b) combinant (a) avec des tests virologiques complets, y compris des techniques de PCR et de séquençage de nouvelle génération agnostiques facilitera la détection non biaisés des maladies infectieuses (ré) émergentes, permettant ainsi une amélioration de la gestion des maladies infectieuses dans les situations d'infections aiguës, de meilleurs systèmes de surveillance et un soutien à la recherche.

**Title:** Standardized syndromic and virologic surveillance facilitating unbiased signal detection of emerging and re-emerging infectious diseases

**Keywords:** Syndromic surveillance, viruses, disease severity, next-generation sequencing, pattern recognition, machine learning

**Abstract:** Infectious diseases pose a global health threat with respiratory infections and infections of the central nervous system (CNS) ranking among the leading causes of morbidity and mortality, especially in children. Usually, we monitor (re-)emerging infectious diseases based on public health databases or regional, national, and international surveillance networks, all of which depend on reports from vigilant healthcare providers or patients themselves. The majority of public health data are not obtained at the individual patient level and stakeholders have repeatedly raised concerns about the timeliness and accuracy of disease burden estimates, especially in epi-/pandemic situations. Unless standardized criteria are applied and infectious diseases are recognized and diagnosed at the point-of-care, any association between a certain disease and a specific pathogen may be subject to bias and delays.

To investigate this topic, the Vienna Vaccine Safety Initiative, an international non-profit research organization, designed two quality improvement programs between 2009 and 2015 for the purpose of enhanced infection control and infectious disease surveillance in one of Europe's largest pediatric tertiary care centers at the Charité University Medical Center in Berlin, Germany in collaboration with the Robert Koch Institute: clinical presentations and disease severity in patients with flu-like illness were assessed in real-time using mobile health technology. Mobile apps allowed for the application of standardized criteria for case classification, risk, and severity. Completeness of data according to standard operating procedures helped to diminish ascertainment bias. A similar program existed for patients with suspected CNS infection.

In the present PhD thesis project, we build on the lessons learned from these two quality improvement programs.

We will now link two highly standardized and granular clinical datasets from several hundred pediatric patients presenting with respiratory and CNS infection to a matching biobank for in-depth virological analysis and pattern recognition. Conventional PCR will be complemented by agnostic viral metagenomics. In detail, we will discuss:

- 1) The utility of combining digital case classification with metagenomics in closing the 'diagnostic gap' in complex CNS infectious disease
- 2) The use of mobile health technology linked to PCR diagnostics and metagenomics to improve the clinical management of respiratory infections.
- 3) Molecular evolution of adenovirus species C based on whole-genome sequencing.
- 4) A precision medicine approach to investigating disease severity and clinical patterns of respiratory adenovirus infection using machine learning methodology.

The results of the studies discussed herein demonstrate that

(a) standardized clinical case assessments, aided by mobile health technology, enable pattern recognition, machine learning, and the reliable computation of meta-analyzable disease severity measures

and that

(b) combining (a) with comprehensive virological testing, such as PCR and agnostic next-generation sequencing, will facilitate unbiased signal detection of (re-)emerging infectious diseases, thus allowing for improved infectious disease management in acute patient care, surveillance, and research.

---

*“Infectious disease is one of the few genuine adventures left in the world.  
The dragons are all dead and the lance grows rusty in the chimney corner.”*

(Hans Zinsser, 1935)

---

## REMERCIEMENTS

To my mentor and PhD thesis supervisor Barbara Rath, I owe my deepest gratitude. Not only has she agreed to supervise my PhD project at the Université Bourgogne - Franche-Comté but she has also taught me that integrity and perseverance be cornerstones of scientific conduct. Besides, I want to express my special thanks to Barbara Rath for introducing me to Brunhilde Schweiger, Dominique Vuitton, and Onya Opota.

I thank Brunhilde Schweiger whom I first met at the 5<sup>th</sup> German Influenza Meeting in Erfurt, Germany in 2012 for giving prudent advice on both technical and inter-personal matters ever since.

Without Dominique Vuitton, my PhD project would have been impossible: At the crack of dawn or deep in the night, she provided invaluable support and helped to interconnect with experts in France and beyond. I also want to thank Dominique Vuitton for her heartfelt hospitality.

To Onya Opota, I am grateful because he agreed to support my PhD thesis as co-supervisor by bringing in his expertise in clinical microbiology, molecular diagnostics, and genomics.

At UBFC, many others have supported me with fruitful discussions and food for thought – Among those, I want to acknowledge Didier Hocquet and Quentin Lepiller, in particular.

Also, I thank Martine Gautheron who has helped me a lot over the last four years. Merci beaucoup !

Among our collaborators, I want to express my gratitude to Albert Heim whose commitment to accuracy impressed me and whose support has been a key component of the PhD project. Moreover, thanks go to our long-time companion Tim Conrad, who all too often added the icing on the cake to our research manuscripts with ingenious computations.

In Berlin, there are just as many people whom I want to acknowledge: Needless to say, I appreciate that Alex Rothhaar, Jürgen Acht, and Rami Hamscho have facilitated my PhD studies. Ultimately, it is my family who went through this adventure with me – Thank you, Anna-Maria, Johannes, Christina, Dieter, Andreas, and Jeffrey!

## LISTE DES PUBLICATIONS

### Publications scientifiques et communications liées à la thèse

- ◆ Dhingra A, Hage E, Ganzenmueller T et al. Molecular Evolution of Human Adenovirus (HAdV) Species C. *Sci Rep* 2019; 9:1039.
- ◆ **Obermeier PE**, Alchikh M, Ma X et al. Leveraging digital surveillance to study emerging pathogens in influenza-like illness in children. *J Clin Virol Plus* (submitted April 2021).
- ◆ **Obermeier PE**, Ma X, Heim A et al. Avenues towards closing the diagnostic gap in encephalitis and acute disseminated encephalomyelitis in children: insight from a digital surveillance program. *Sci Rep* (submitted August 2021).
- ◆ **Obermeier PE**, Heim A, Biere B et al. Linking digital surveillance and in-depth virology to study clinical patterns of viral respiratory infections in vulnerable patient populations. *iScience* (submitted October 2021).

### Contribution à d'autres publications scientifiques dans le domaine

- ◆ Wang X, Li Y, O'Brien KL et al. Global burden of respiratory infections associated with seasonal influenza in children under 5 years in 2018: a systematic review and modelling study. *Lancet Glob Health* 2020; 8:e497-e510.
- ◆ Wang X, Li Y, Deloria-Knoll M et al. Global burden of acute lower respiratory infection associated with human metapneumovirus in children under five years for 2018: a systematic review and modelling study. *Lancet Glob Health* 2021; 9(1): e33-e43.
- ◆ Wang X, Li Y, Deloria-Knoll M et al. Global burden of acute lower respiratory infection associated with human parainfluenza virus in children under five years for 2018: a systematic review and meta-analysis. *Lancet Glob Health* 2021; 9(8): e1077-e87.

## ABREVIATIONS

(S)ARI	-	(Severe) acute respiratory infection
ADEM	-	Acute disseminated encephalomyelitis
CDC	-	Centers for Disease Control and Prevention
CE	-	Conformité Européenne
CLI	-	COVID-19-like illness
CNS	-	Central nervous system
COVID-19	-	Coronavirus disease 2019
DSC	-	Disease Severity with Signs and Symptoms of Complicated Disease
DSS	-	Disease Severity Score
DSU	-	Disease Severity with Signs and Symptoms of Uncomplicated Disease
ECDC	-	European Centre for Disease Prevention and Control
ECMO	-	Extracorporeal Membrane Oxygenation
eHealth	-	Electronic Health
ELISA	-	Enzyme-linked immunosorbent assay
FDA	-	Food and Drug Administration
FDA-ARGOS	-	FDA dAtabase for Regulatory-Grade micrObial Sequences
GBD	-	Global Burden of Disease
GDPR	.	General Data Protection Regulation
GISRS	-	Global Influenza Surveillance and Response System
GPEI	-	Global Polio Eradication Initiative
GROG	-	Groupes Régionaux d'Observation de la Grippe
HAdV	-	Human adenovirus
HBoV	-	Human bocavirus
HCoV	-	Human coronavirus
HIPAA	-	Health Insurance Portability and Accountability Act
HIV	-	Human immunodeficiency virus
HMPV	-	Human metapneumovirus
HPeV	-	Human parechovirus
HPIV	-	Human parainfluenza virus
HRV	-	Human rhinovirus
IBM	-	International Business Machines Corporation
ICD	-	International Classification of Diseases
ICTV	-	International Committee on Taxonomy of Viruses
ICU	-	Intensive Care Unit
ILI	-	Influenza-like illness
mHealth	-	Mobile Health
mNGS	-	metagenomic next-generation sequencing
NGS	-	Next-generation sequencing
PCR	-	Polymerase chain reaction



PEDSIDEA	- Partnering for Enhanced Digital Surveillance of Influenza-like Disease and the Effect of Antivirals and vaccines
PHEIC	- Public Health Emergency of International Concern
PISA	- Pandemic Influenza Severity Assessment
RNA	- Ribonucleic acid
RSV	- Respiratory syncytial virus
SARS-CoV-2	- Severe acute respiratory syndrome coronavirus 2
SNP	- Single nucleotide polymorphism
UCSF	- University of California, San Francisco
US	- United States (of America)
VACC-Tool	- Vienna Vaccine Safety Initiative Automated Case Classification Tool
ViVI	- Vienna Vaccine Safety Initiative
WHO	- World Health Organization

## TABLE DE MATIÈRES – TABLE OF CONTENTS

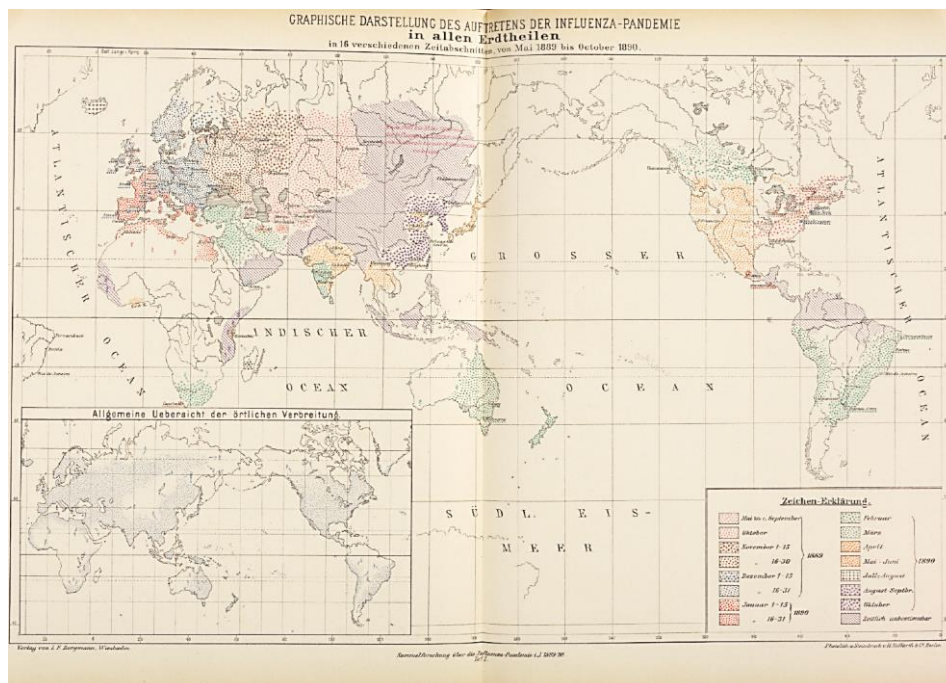
<b>RÉSUMÉ EN FRANÇAIS.....</b>	<b>II</b>
<b>RÉSUMÉ EN ANGLAIS.....</b>	<b>III</b>
<b>REMERCIEMENTS.....</b>	<b>V</b>
<b>LISTE DES PUBLICATIONS.....</b>	<b>VI</b>
<b>ABRÉVIATIONS.....</b>	<b>VII</b>
<b>TABLE DE MATIÈRES – TABLE OF CONTENTS.....</b>	<b>IX</b>
<b>1 INTRODUCTION .....</b>	<b>- 1 -</b>
1.1 Viral Infectious Diseases – A Genuine Adventure .....	- 1 -
1.2 Infectious Disease Surveillance – Breaking New Ground .....	- 4 -
1.3 Aims and Objectives of the Thesis Project .....	- 7 -
<b>2 THE PURSUIT OF UNBIASED SIGNAL DETECTION .....</b>	<b>- 8 -</b>
2.1 CNS Infectious/ Inflammatory Disease.....	- 8 -
2.1.1 Facilitating Precision Medicine using Mobile Health Technology .....	- 10 -
2.1.2 Exploring Viruses in Encephalitis and ADEM .....	- 13 -
2.1.3 Manuscript A .....	- 20 -
2.1.4 Poster A .....	- 39 -
2.2 Acute Respiratory Infections .....	- 40 -
2.2.1 Agnostic Identification of (Re-)Emerging Viruses.....	- 46 -
2.2.2 Manuscript B .....	- 50 -
<b>3 MINING THE BIG PICTURE .....</b>	<b>- 68 -</b>
3.1 Whole-Genome Analysis – Leveraging Big Data in Virology.....	- 68 -
3.2 Manuscript C.....	- 70 -
3.3 Machine Learning Analysis for Pattern Recognition of Adenovirus Infection.....	- 84 -
3.4 Manuscript D .....	- 87 -
<b>4 CONCLUSIONS AND PERSPECTIVES.....</b>	<b>- 118 -</b>
<b>RÉFÉRENCES.....</b>	<b>X</b>

# 1 INTRODUCTION

## 1.1 Viral Infectious Diseases – A Genuine Adventure

“Infectious disease is one of the few genuine adventures left in the world.”<sup>1</sup>, stated US American microbiologist Hans Zinsser (1878–1940) in his best-selling popular science book “Rats, Lice, and History”, first published in 1935<sup>1</sup> – and history has proved him right ever since:

It was only at the beginning of the 20<sup>th</sup> century when medical microbiologists, an avant-garde group of physician-scientists knowledgeable about parasites, fungi, and bacteria were finally realizing that in man and animals a yet unknown group of “ferocious little fellow creatures”<sup>1</sup> fully invisible to the human eye would cause disease, *videlicet*, viruses. A well-recorded chapter of this infectious disease adventure opened in the early winter days of 1889 in Saint Petersburg, capital of the Russian Empire at that time when an unforeseen respiratory disease pandemic proposed a conundrum to the world<sup>2,3</sup>. It was the Russian Flu or *La Grippe* that very rapidly spread around the globe, claiming >1 million lives in the 1890’s, thus being considered one of the first major disease outbreaks to happen in the globalizing modern world<sup>3</sup> (Figure 1<sup>4</sup>).



**Figure 1.** Historical map illustrating the dates of first Russian Flu cases reported worldwide between 1889 and 1890. Here, first influenza cases were depicted to happen in Central Asia between May and September 1889, then spreading towards east and west reaching Western Europe, Africa, Australia, and the American continent by the beginning of 1890. According to the map at issue, cases were detected worldwide except for the Arctic, Antarctic, and Central Africa<sup>4</sup>.

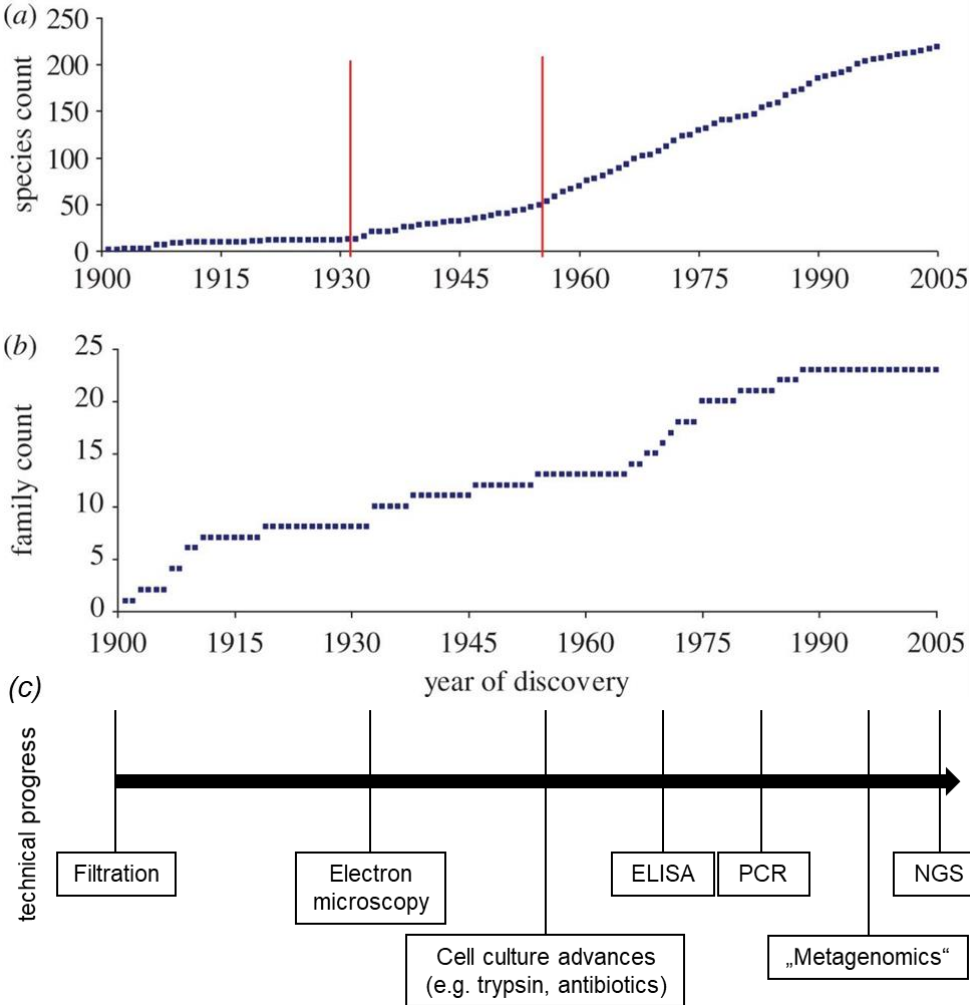
*Post hoc* serological analyses in the mid-20<sup>th</sup> century suggested that the Russian Flu was caused by influenza A virus H2N2<sup>5</sup> or H3N8<sup>6,7</sup>. Back in those days of the late 19<sup>th</sup> century however, the pandemic was a “*genuine adventure*” from the scientific perspective, indeed. Besides ideas of the Russian Flu being a somatopsychic disease of the “*overstrained brain*”<sup>8</sup> in the industrializing world or the corollary of miasmatic influences, literally giving the disease its name *influenza*<sup>9</sup>, bacteria were *a priori* considered the culprit. Meeting the spirit of that golden age of thriving bacteriology, renowned microbiologists Richard Pfeiffer (1858–1945) and Shibasaburo Kitasato (1852–1931) at Robert Koch’s laboratory in Berlin, Germany claimed that Pfeiffer’s bacillus (nowadays known as *Haemophilus influenzae*) was the cause of the late 19<sup>th</sup>-century pandemic<sup>10</sup>. During the next devastating influenza pandemic starting in 1918 however, Peter Olitsky (1886–1964) and Frederic Gates (1886–1933) at The Rockefeller Institute for Medical Research in New York City indicated that another germ might have been the causative agent: a so called “*filterable virus*”, i.e. the still infectious filtrate (*vīrus*, slime or stench) of ultrafine earthenware filters<sup>11</sup>. With technological innovation such as “*Übermikroskopie*” (i.e. electron microscopy) introduced for virological purposes by Helmut Ruska (1908–1973) and Gustav-Adolf Kausche (1901–1960) in the 1930’s as a world premiere in Berlin, Germany, the invisible “*ferocious little fellow creatures*” were finally exposed<sup>12</sup>. Among the very first *filterable viruses* uncovered was the influenza virus, merely being visualized as 2D filamentous structures<sup>13,14</sup>.

By the end of the 20<sup>th</sup> century, modern means of direct and indirect virus detection methods including viral cultures, serological analyses, and polymerase chain reaction (PCR) led to the discovery and in-depth characterization of ≈200 virus species accounting for human disease<sup>15</sup>.

Towards the turn of the millennia, virus discovery was again revolutionized by a groundbreaking approach termed “*metagenomics*”. Metagenomic analysis, first mentioned by Jo Handelsman in 1998<sup>16</sup>, denotes the exploration of the “*collective genomes*”<sup>16</sup> directly from a sample, i.e. without prior culturing methods. Standard operating procedures of metagenomic testing usually include 1) sample preparation (nucleic acid extraction, indexing, etc.), 2) sequence-independent amplification of nucleic acids, 3) sequencing, and 4) bioinformatics analysis<sup>17,18</sup>. With the advent of next-generation sequencing (NGS) allowing for cost-effective, high-throughput agnostic sequence amplification and identification, the real potential of metagenomic testing has been unleashed in that it can provide an unbiased answer to the question “*what is in the sample?*” with reasonable effort. When the first NGS systems became commercially available in

the 2000's, metagenomic next-generation sequencing (mNGS) has ushered in a new era of virus discovery and understanding of the pathogenesis of viral disease in humans. This being said, modern molecular virology is gradually experiencing a shift from mere laboratory to bioinformatic and technical as well as interpretative challenges when it comes to the real-world application of next-generation virologic testing, thus warranting continuous technical optimization as well as integration of comprehensive clinical data.

Refer to **Figure 2**<sup>15</sup> for a brief virology timeline.



**Figure 2.** Brief overview of virus discovery curves depicting (a) species count and (b) family count of viruses infecting humans and reported to cause disease from 1900-2005. In (a), vertical red lines highlight significant increases in species count numbers. In synopsis with (a) and (b), (c) illustrates key technical advances in virology in the same period of time. Adapted from Woolhouse et al.<sup>15</sup>

## 1.2 Infectious Disease Surveillance – Breaking New Ground

To the clinical end, our understanding of infectious diseases has improved with refined observation of illnesses, *videlicet*, by surveillance which literally means “to watch [diseases]” (French *surveiller*).

Written records of early infectious disease surveillance conducted in Europe are as old as 700 years, e.g. originating from bubonic plague outbreaks in medieval Marseille, France<sup>19</sup>. Gradually, surveillance has evolved from notarial or church registers to well-organized international structures such as the Global Influenza Surveillance and Response System (GISRS), initiated in 1947 by the newly founded World Health Organization (WHO)<sup>20</sup>. Among the main aims of such surveillance systems are the detection and characterization of (re-)emerging pathogens, the monitoring of disease burden, and public health guidance<sup>21</sup>. For the purpose of surveillance data to become meaningful and actionable, data capturing, analysis, and interpretation need to be timely, accurate, and uniform.

But for the sake of simplicity and cost effectiveness, the monitoring of (re-)emerging infectious diseases is frequently based on passive surveillance networks or secondary data analysis from pre-existing databases, e.g. international classification of diseases (ICD) registries. Consequently, surveillance data may be subject to bias as well as time delay and harmonization of terminology allowing for meta-analyzability and interoperability is hardly ever performed<sup>22</sup>. To overcome those hurdles, public health authorities and WHO advocate the use of standardized terminology, case definitions, and innovative electronic health (eHealth) technology<sup>19,21,23,24</sup>. In 2020 however, a pandemic caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has grown to a notable crisis in recent human history, declared a Public Health Emergency of International Concern (PHEIC) - despite globally acting sophisticated surveillance and response systems for respiratory infectious disease.

Sketching out new avenues towards standardized infectious disease surveillance and precision medicine, the Vienna Vaccine Safety Initiative (ViVI, [www.vi-vi.org](http://www.vi-vi.org)), an international scientific think tank and non-profit research organization, implemented two extensive quality improvement programs for the purpose of enhanced infection control and quality of care in one of Europe’s largest pediatric tertiary care centers at the Charité University Medical Center in Berlin, Germany in collaboration with the Robert Koch Institute between 2009 and 2015. In these two prospective surveillance programs, all pediatric patients presenting with pre-defined signs and symptoms of influenza-like illness (ILI) and/or infection/inflammation of the

central nervous system (CNS) were assessed by specifically trained quality improvement staff using mobile health (mHealth) technology at the patient's bedside. Clinical presentations, captured with up to 1600 data points per patient, were analyzed in real-time based on consensus case criteria, thus providing a very detailed picture of the clinical presentation of each patient<sup>22,25-38</sup>. Disease severity was assessed using the validated and published ViVI Disease Severity Score (ViVI Score, <https://score.vi-vi.org/>)<sup>30,39</sup> at the point-of-care, serving as a real-time support for clinical decision making and also to systematically analyze disease severity in specific risk groups using uniform criteria. Throughout, data were captured based on standard data formats and terminologies making them comparable across sites as mandated by regulatory authorities<sup>22,30</sup>. By 2015, a total of >6000 pediatric ILI patients and >700 pediatric patients with signs of CNS infection/inflammation had entered the standardized syndromic surveillance cohorts.

For gold-standard virological PCR testing of specimens collected in these two surveillance cohorts, a matching biobank was generated. Nasopharyngeal samples were collected from each ILI patient and stool samples from patients with CNS symptoms. In collaboration with the National Reference Centers for Influenza and for Poliomyelitis and Enteroviruses at the Robert Koch Institute, samples were tested for influenza A/B viruses, respiratory syncytial viruses (RSV), human rhinoviruses (HRV), human metapneumoviruses (HMPV), human bocavirus (HBoV-1), parainfluenza viruses (PIV1-4), human adenoviruses (HAdV), human betacoronaviruses (HCoV-NL63, HCoV-229E, HCoV-HKU1, HCoV-OC43) as well as for entero- und human parechoviruses (HPeV)<sup>30,33,36</sup>.

Upon completion, both ViVI projects demonstrated that standardized case criteria and state-of-the-art technology need to be leveraged at the bedside to generate complete surveillance data without missing values. Unless surveillance data measure the same standardized outcomes using the same criteria, data will be prone to bias and delays in generating signals to detect emerging viruses or outbreaks in a timely fashion. Only with rigorous virological testing in addition, changes in clinical presentation can be detected with sufficient certainty to generate adequate clinical management and public health response. The recent COVID-19 pandemic has shown how important it is to detect emerging viruses early, and to differentiate a disease outbreak from known clinical presentations and seasonal epidemiology<sup>40,41</sup>. In the prospect of international travel and globalization, next-generation surveillance systems should rely on large and

comprehensive datasets from all over the world, making standardization, particularly in terms of case ascertainment, disease severity measurement, and terminology a prerequisite<sup>22,28,30,34,38</sup>.



### **1.3 Aims and Objectives of the Thesis Project**

Building on the lessons learned from the above-mentioned quality improvement programs by the Vienna Vaccine Safety Initiative, this PhD thesis project aimed to investigate the merits of standardized syndromic infectious disease surveillance aided by innovative technology in combination with in-depth, state-of-the-art virological testing at the interface of clinical care and virology.

For this purpose, we linked the pre-existing highly-granular, standardized clinical datasets from several thousand pediatric patients presenting with ILI and/or infectious/-inflammatory CNS disease to a matching biobank for mNGS analysis and in-depth exploration of emerging and re-emerging viruses. Specifically, the objectives of this PhD thesis are to investigate

- 1) The utility of combining digital case classification with metagenomics in closing the 'diagnostic gap' in complex CNS infectious disease.**
- 2) The use of mobile health technology linked to with PCR diagnostics and metagenomics to improve the clinical management of respiratory infections.**
- 3) Molecular evolution of adenovirus species C based on whole-genome sequencing.**
- 4) A precision medicine approach to investigating disease severity and clinical patterns of respiratory adenovirus infection using machine learning methodology.**

## 2 THE PURSUIT OF UNBIASED SIGNAL DETECTION

### 2.1 CNS Infectious/ Inflammatory Disease

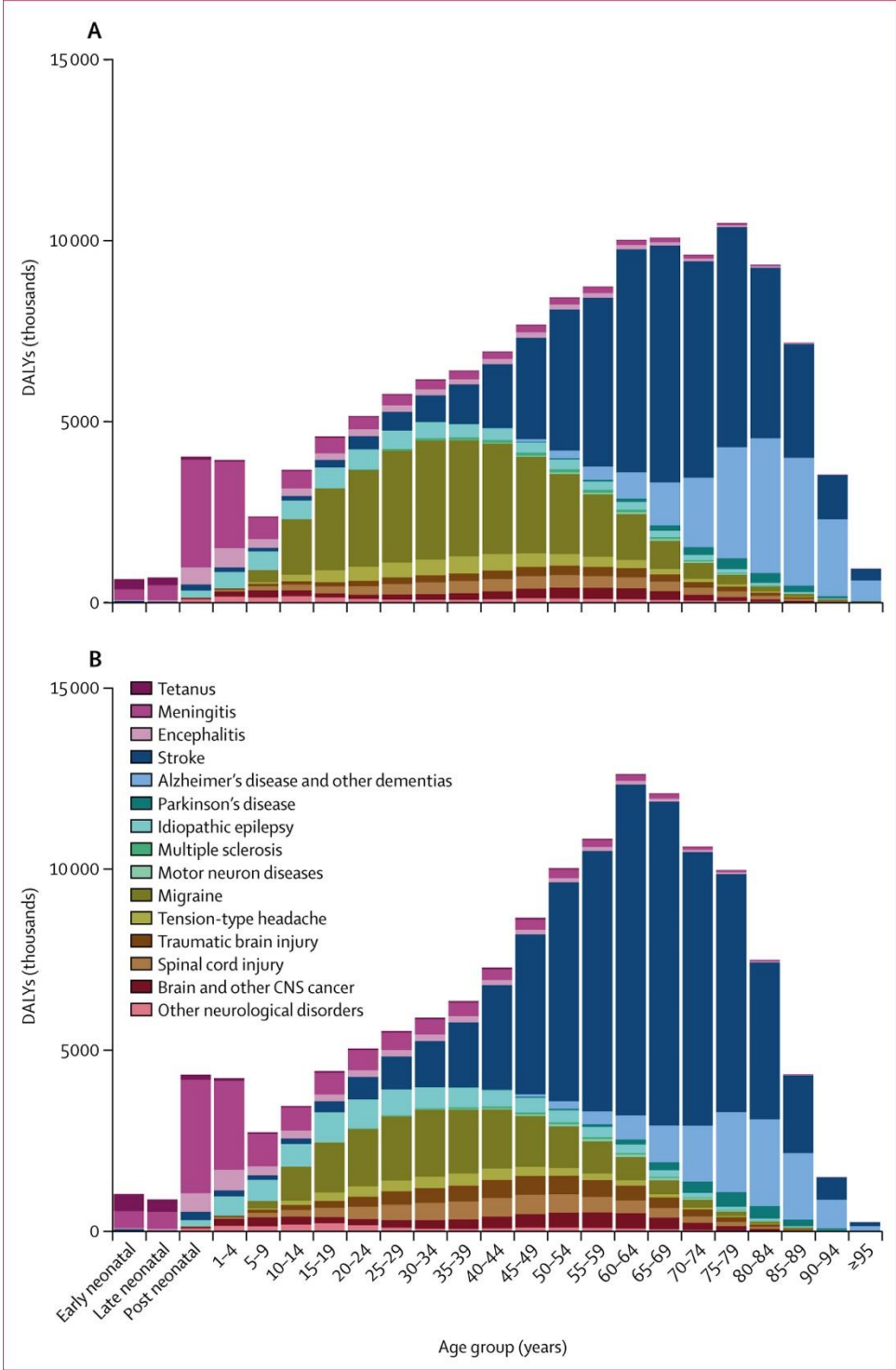
Among public health threats, WHO and the Global Burden of Disease (GBD) project declared neurological disorders the leading cause of disability and the second-leading cause of death on the globe. Depending on risk factors such as age, education, socioeconomic status, comorbidities (e.g. HIV infection) or geographic regions, CNS infectious and inflammatory disease in particular impose vital danger to human health<sup>42,43</sup>.

Certain communicable CNS diseases have almost vanished, such as poliomyelitis thanks to extensive global immunization, surveillance, and containment (Global Polio Eradication Initiative, GPEI). Other communicable CNS disease have at least become less threatening, such as bacterial meningitis with the advent of antibiotics. However, some forms of infectious CNS disease are not (yet) vaccine-preventable and therapeutic options are unavailable, as is the case with the majority of viral CNS disease. Thus, CNS infectious/ inflammatory meningitis and encephalitis have ranked among the top-10 CNS disorders accounting for disability-adjusted life years (DALY) worldwide, and especially in the developing world (see Figure 3<sup>42,44</sup>).

	Global	East Asia	Southeast Asia	Oceania	Central Asia	Central Europe	Eastern Europe	High-income Asia Pacific	Australasia	Western Europe	Southern Latin America	High-income North America	Caribbean	Andean Latin America	Central Latin America	Tropical Latin America	North Africa and Middle East	South Asia	Central sub-Saharan Africa	Eastern sub-Saharan Africa	Southern sub-Saharan Africa	Western sub-Saharan Africa	
Stroke	1	1	1	1	1	1	1	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1
Migraine	2	3	3	3	2	2	2	2	1	1	2	2	2	2	2	3	2	2	4	3	3	3	3
Alzheimer's disease and other dementias	3	2	2	2	4	3	3	3	3	3	3	3	3	3	3	2	3	4	3	4	4	4	4
Meningitis	4	11	5	4	9	12	10	14	13	13	11	13	4	9	10	8	5	3	2	2	5	2	2
Epilepsy	5	5	4	5	3	7	8	6	7	6	5	6	5	4	4	4	4	6	5	5	5	2	5
Spinal cord injury	6	7	8	9	7	6	5	4	4	4	4	4	9	8	9	9	6	9	6	7	10	9	9
Traumatic brain injury	7	6	6	7	5	4	4	7	8	8	9	8	7	7	6	7	9	7	7	8	6	7	7
Brain and other CNS cancer	8	4	9	10	6	5	6	8	5	5	6	5	8	6	7	5	8	10	9	11	9	10	10
Tension-type headache	9	8	10	8	10	8	7	5	6	7	7	7	6	5	5	6	7	8	8	9	7	6	6
Encephalitis	10	9	7	6	8	13	11	11	14	14	12	14	11	10	11	12	10	5	10	10	11	8	8
Parkinson's disease	11	10	11	12	12	9	9	10	9	10	8	9	12	11	12	11	12	13	13	13	12	13	13
Other neurological disorders	12	12	12	11	11	10	12	9	10	9	10	10	10	12	8	10	11	12	12	12	8	12	12
Tetanus	13	15	13	14	15	15	15	15	15	15	15	15	13	15	15	15	14	11	11	6	15	11	11
Multiple sclerosis	14	14	15	15	13	11	13	13	12	11	13	11	15	14	14	14	13	14	14	14	13	15	15
Motor neuron diseases	15	13	14	13	14	14	14	12	11	12	14	12	14	13	13	13	15	15	15	15	14	14	14

Figure 3. Ranking of age-standardized disability-adjusted life years for top-15 neurological disorders by region in 2016<sup>42</sup>.

Notably, infectious/ inflammatory CNS diseases, i.e. meningitis and encephalitis, are of highest impact in the pediatric age group (see **Figure 4**<sup>42</sup>) whereas in older patients, stroke, neoplasms, or dementia become more important in terms of DALY<sup>42</sup>. At the same time though, recognizing infectious/ inflammatory CNS disease can be particularly challenging in children<sup>22</sup>.



**Figure 4.** Global disability-adjusted life years for neurological disorders by sex (A for females and B for males) and age, 2016. Early neonatal is 0–7 days; late neonatal is 7–28 days; and post-neonatal is 28 days to 1 year<sup>42</sup>.

### **2.1.1 Facilitating Precision Medicine using Mobile Health Technology**

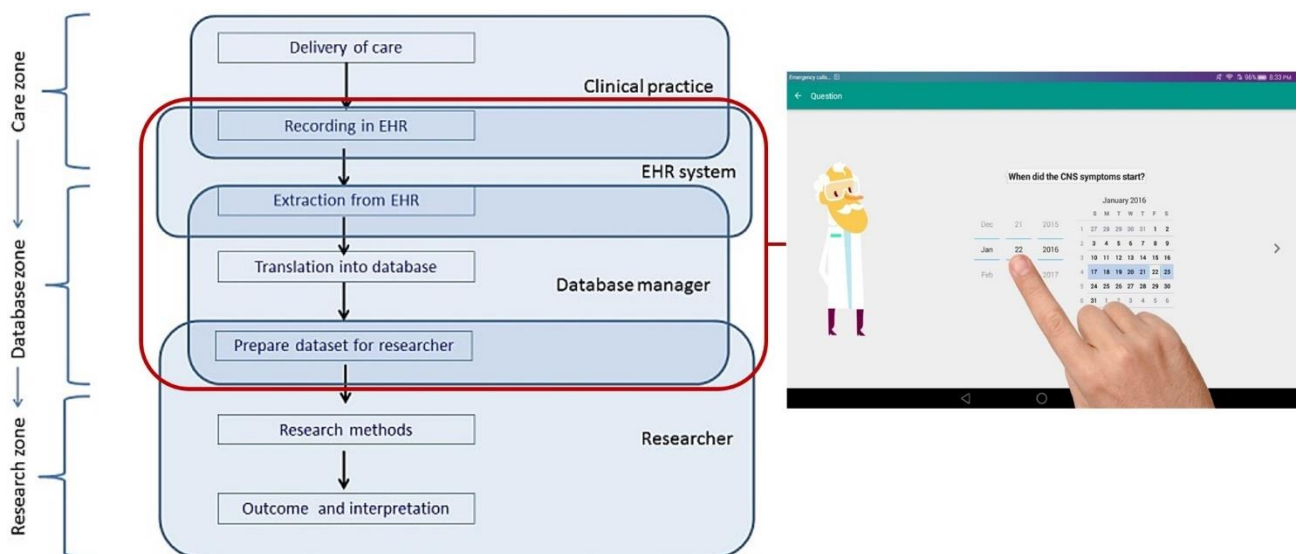
In the past, research has shown that routine care clinical diagnoses, i.e. ICD codes, tend to be prone to observer bias and inconsistency, especially in case of CNS pediatric infectious diseases<sup>22,45</sup>. To harmonize disease classification, expert panels have published validated consensus case criteria, but they are rarely put to use. The level of standardization in recently published studies on CNS infectious disease is usually limited to “clinical suspicion of CNS infectious diseases”<sup>46</sup> or “retrospective, in-depth chart review”<sup>47</sup> or else, if pre-defined case criteria are indeed implemented, they tend to be intrinsically inconclusive leaving parts of the definition open to subjective interpretation<sup>48</sup>. Not only do such methodologies lead to observer bias, but they also impede interoperability and meta-analyzability of data<sup>49</sup>. This fact will become particularly critical when (multinational) surveillance systems lack standardization, e.g. as to be seen during a recent report describing an outbreak of “aseptic meningitis [...] associated with echovirus 4 in Northern Europe in 2013–2014”<sup>50</sup>. The above and other virological surveillance reports do not use point-of-care ascertainment of endorsed case criteria but retroact chart review, based on incidentally reported symptoms such as “Fever, cough, deteriorated condition, headache”<sup>50</sup>. Pleocytosis in the CSF, the histopathological hallmark of meningeal inflammation<sup>51</sup>, was not considered, nor were other critical elements of international case definitions such as bulging fontanelle or nuchal rigidity<sup>50,51</sup>. Lumping together a large number of patients presenting with nonspecific signs/symptoms such as fever, malaise or headache under a diagnosis of “meningitis,” and correlation of this loosely defined group with a positive virus detection result may ultimately lead to unsound conclusions from surveillance systems, conveying erroneous images of epidemiology and pathogenicity.

#### *Paving the Way towards Medical Innovation*

To facilitate the implementation of international case definitions at the point of care, ViVi developed an innovative mobile application that enables healthcare professionals to perform unbiased automated case classification based on standardized case criteria at the bedside. This allows to ask all pertinent questions at the right time, i.e. when the patient is still in the physician’s reach at the hospital<sup>22</sup>.

Use of this eHealth technology called VACC-Tool (Vienna Vaccine Safety Initiative Automated Case Classification Tool, <https://www.vi-vi.org/surveillance-of-meningitis-and-encephalitis-in-children.html>) in a

proof-of-concept study published in 2016 proved to provide consistent and reliable case ascertainment when used at the point-of-care. Not surprisingly, retroactive application of the VACC-Tool (using the same algorithms for case classification) based on routine medical records was less conclusive because medical records were often missing relevant data and conclusive information on the presence or absence of hallmark criteria<sup>22,34,35,38,52</sup>. This fact highlighted the importance of immediate, prospective application of case criteria at the bedside so that all pertinent information can be obtained. Hence, the added value of eHealth technology will become actionable once implemented in routine work-flows in real time. If routine care electronic health records are leveraged for surveillance purposes, this will usually be performed as an extra, *post hoc* work step as described by Verheij et al. in 2018, thus requiring additional effort in the “database zone”<sup>53</sup> (in addition to the “care zone”<sup>53</sup>) by a “database manager”<sup>53</sup> to translate and re-process data for research purposes, implicating potential time delay and bias<sup>53</sup>. However, precision medicine warrants eHealth technology to be reasonably incorporated in routine care, thereby unifying and simplifying processes at the interface of patient care, data curation, and research (**Figure 5**<sup>53</sup>). If so, eHealth technology has the potential to aid physicians in that mobile apps or web applications such as the VACC-Tool will automatically prompt the healthcare professional to obtain all pertinent data points (e.g. lumbar puncture when clinical signs and symptoms justify the suspected diagnosis meningitis)<sup>22,38</sup>.



**Figure 5.** Work flow and parties involved when routine care electronic health records (EHR) are used for research purposes as depicted by Verheij et al.<sup>53</sup>. Right-hand side shows a snapshot of the VACC-Tool question and answer tree leading to an automated case classification result when all pertinent questions

are answered on a hand-held device, thereby amalgamating “care zone”, “database zone”, and “research zone”.

The VACC-Tool 1.0 was primarily used for four different disease entities: aseptic meningitis, encephalitis, myelitis, and acute disseminated encephalomyelitis (ADEM)<sup>22,38</sup> and then updated to ascertain up to 14 different diseases (VACC-Tool 2.0) over time if the patient had to be hospitalized, thus providing even more, dynamic information on the course of the patient’s disease, like a flip-book<sup>34</sup>. Highly granular datasets comprising up to 1600 data points per patient lead to unbiased diagnoses and were also used for downstream bioinformatics analyses and machine learning algorithms, e.g. allowing to identify and weigh key clinical characteristics of certain disease entities (see **Figure 6**<sup>22</sup>).

	Bulging fontanelle	Reduced level of consciousness	Lethargy	Cranial nerve deficit	Fever (body temperature $\geq 38.0^{\circ}\text{C}$ )	Sensory abnormalities	Cerebellar deficits (e.g. dysmetria)	Focal cortical signs (e.g. alexia)	Decreased deep tendon reflexes	Behavioral changes	Pleocytosis in cerebrospinal fluid	Gram stain	Neuroimaging
Aseptic Meningitis	0.05	0.12	0.28	0.88			1.32				8.34	4.7	
Encephalitis				0.68		0.26	0.06	0.13			0.19		
ADEM				0.88	0.54	0.29		0.04	0.05	0.32			0.77

**Figure 6.** Key clinical characteristics for VACC-Tool classification results of aseptic meningitis, encephalitis, and ADEM. Dark grey background color indicates a positive correlation (i.e. the presence of a symptom is important) between clinical sign/laboratory finding and case classification result. Light grey background color indicates a negative correlation (i.e. the absence of a symptom is important). Numbers are Kullback-Leibler-divergences, indicating increasing importance with increasing number<sup>22</sup>.

Ultimately, use of the VACC-Tool at the point-of-care proved to facilitate comprehensive and accurate, interoperable syndromic surveillance, literally representing the zeitgeist of precision medicine that will enable improved healthcare<sup>22</sup>.

### 2.1.2 Exploring Viruses in Encephalitis and ADEM

Following the case ascertainment process, be it in a standardized syndromic surveillance or acute care setting, the search for the cause of the disease will be key for both, optimal therapy and epidemiological investigations<sup>54,55</sup>. Pursuing the most probable trigger in the individual patient case, serology, antigen testing, and/or PCR testing (including multiplex panels) will be applied in the majority of instances. Depending on the laboratory (e.g. Reference Centers), viral cultures might be part of the work-up, too. For all those detection methods however, there are characteristic limitations:

- ◆ Antigen-based and serology might have varying sensitivity and specificity, with particular regards to the pathogen to be detected, the sample type used, the time-point of sampling, and/or the immune-status of the host, i.e. immunocompromization or immunization. Indirect detection methods will not allow for any detailed pathogen characterization. Yet, both antigen-based and serology testing are affordable and easily applicable, making them a good choice for point-of-care rapid testing as well as *post-hoc* (sero-) epidemiological studies.
- ◆ PCR methods usually have high to extremely high sensitivity and specificity but are more expensive and may require certain user facility requirements, e.g. biological safety cabinets, controlled ventilation system, freezers, thermocyclers, etc<sup>56</sup>. High sensitivity and specificity are, at the same time, among the most relevant disadvantages of nucleic acid amplification testing. Technically, some PCR assays will detect the tiniest amount of nucleic acid in a sample that might not be relevant in the acute phase, i.e. positive test results could come from genetic residues of non-viable pathogens following successful therapy, prior vaccination<sup>57</sup>, shedding of a bystander or from contamination<sup>55</sup>. To the clinical end, the significance of PCR testing directly depends on the physician ordering a pathogen-specific PCR test, i.e. the physician will only find or rule-out what they have on their mind.
- ◆ More recent developments allow for multiplex and/or point-of-care PCR testing, thereby broadening the window for pathogens to be detected in reasonable time. But with some of such assays, sensitivities and specificities do not approach the level of single-pathogen lab-based PCR testing for various reasons, including primer-mismatch on the subtype level or in case of newly emerging pathogens<sup>58,59</sup>. One Examples of commercial point-of-care multiplex PCR assay are, among others, the MeningoFinder® or

the BIOFIRE® FILMARRAY® Meningitis/Encephalitis panel. The latter tests for a total of 14 pathogens, including bacteria, viruses, and yeast, in a turnaround time of approximately 1 hour from CSF sampling to test result<sup>60</sup>. Several validation studies have concluded that overall, sensitivities and specificities of this multiplex panel are good<sup>61,62</sup> and obviously, turnaround time is excellent but some concerns may remain with this particular widely-used test panel, e.g.: false-negative results for human enterovirus infection caused by potential cross reactivity with human rhinovirus will not be differentiated, neither will the assay reliably detect all virus subtypes (e.g. with HPeV), nor has use of other sample types than CSF been cleared<sup>60</sup>. Especially in pediatrics however, CSF may be hard to come by and with certain CNS infectious/ inflammatory disease, the diagnostic window for successful pathogen detection in the CSF is narrow<sup>35</sup>. Also, the amount of CSF might sometimes not be sufficient for all test modalities ordered, i.e. screening microscopy in the emergency room, Gram stain, fungal cultures, virus culture, specific PCR test(s), and the multiplex panel in addition.

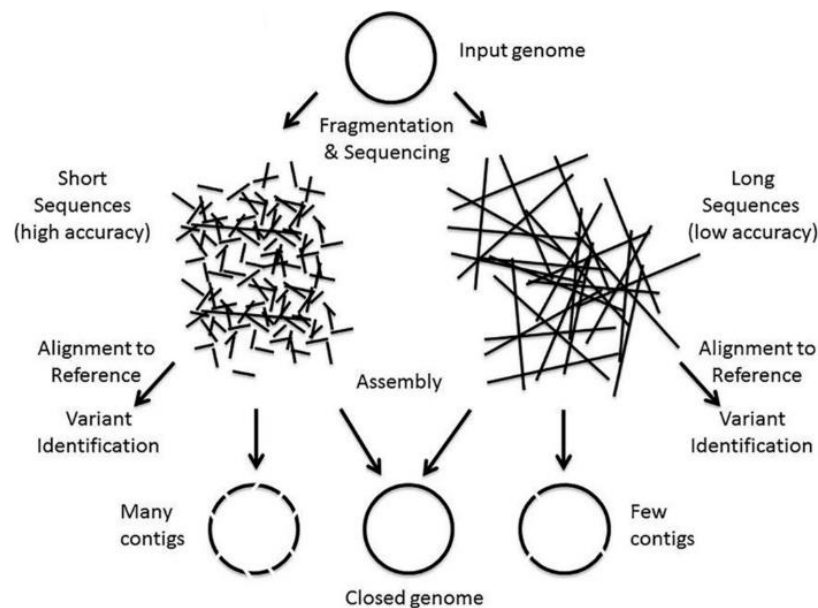
- ◆ Culture-based methods have represented the gold standard for clarification of the trigger of communicable CNS disease for decades, in particular when bacterial or fungal infection is suspected. If preanalytical sources or error are diminished, e.g. contamination during the sampling process or pre-sampling administration of antibiotics, bacterial and fungal cultures can be of immeasurable value for optimal therapeutic regimens. For viral CNS disease however, cultures tend to be less suitable as they require considerable work effort and often lack sensitivity, depending of the virus in question<sup>54,63,64</sup>.

With the aforementioned armamentarium of methodologies however, disease etiology remains unresolved in more than 50% of cases where patients present with signs and symptoms of CNS infection/ inflammation<sup>65-68</sup>. In many cases, where no pathogen has been identified, a virus that is not even on the “diagnostic radar” may have been the trigger<sup>69</sup>.

In recent years, innovative agnostic approaches such as viral metagenomic NGS (mNGS) have been increasingly leveraged to find the etiology of CNS disease where no cause could be identified prior<sup>46,47,54,70-78</sup>. Briefly, mNGS samples, e.g. stool or cerebrospinal fluid (CSF), will be subjected to nucleic acid extraction, complementary DNA (cDNA) synthesis and barcode indexing if multiplexing is performed. Following



manufacturer's instructions, input genomes are then further processed, sequenced, and analyzed (Figure 7<sup>79</sup>).



**Figure 7.** Simplified key steps in current metagenomic next-generation sequencing (mNGS) work-flows. Note that with mNGS (in contrast to NGS in general), input genomes directly come from the sample without culturing or other pre-selection of singular nucleic acid<sup>79</sup>.

There are varying protocols and operating procedures for mNGS (incl. the bioinformatics pipeline *post hoc*), making standardization one of the key challenges on the technical level. Thus, health authorities and national guidelines recommend use of (m)NGS in case of unresolved CNS disease as auxiliary means, if at all. The German guideline on “viral meningoencephalitis”, for example, was released in 2018 (remaining valid until 2023) when commercial NGS had been in use for 18 years already but the only molecular virological detection method mentioned therein is PCR<sup>80</sup>. Interestingly, the 2017 French guideline<sup>81</sup>, the 2018 revised British guideline, and the 2019 Italian guideline<sup>82</sup> on encephalitis at least correspondingly mention next-generation sequencing where it “may provide a diagnosis for more patients in the future”<sup>83</sup>. Accordingly does a 2019 systematic review of “clinical guidelines on the management of acute, community-acquired CNS infections”<sup>84</sup> not recognize NGS as a diagnostic approach for causality assessment<sup>84</sup>.

Hesitancy towards the use of (m)NGS in national guidelines for CNS diagnostics may be due to technical challenges, lack of standardization, and interpretative difficulty with mNGS results. Although with traditional PCR the same question about the clinical significance of a positive result may arise, the matter is incomparably more apparent with mNGS where the absolute entirety of viruses, the virome, in a sample

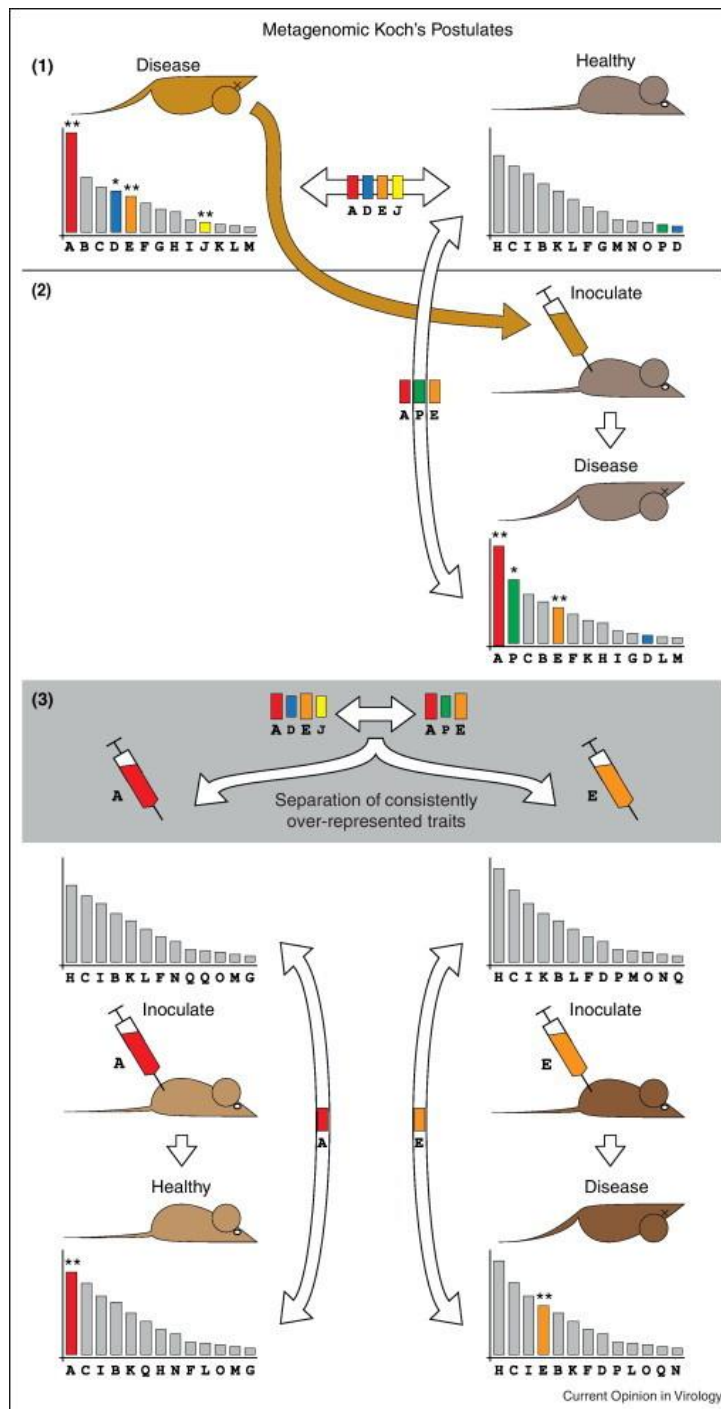
will be revealed. On multiple layers, it will be difficult to constitute an mNGS hit the cause of a certain disease<sup>54</sup>.

#### *Causality Assessment*

Once a technically false-positive test result caused by contamination or computational artifacts can be excluded, any detected virus needs to be categorized as bystander (or viral commensal) or trigger of the disease in the causality assessment process.

In an attempt to facilitate causality assessment, “metagenomic Koch’s postulates” have been introduced by Mokili et al. in 2012 as a consequence of the implications inherent with agnostic detection methods such as mNGS<sup>85</sup>. Unlike traditional Koch’s postulates originating from the 19<sup>th</sup> century, metagenomic Koch’s postulates from the 21<sup>st</sup> century do not require propagation of the potential pathogen in culture<sup>85,86</sup> but formulate three conditions, i.e.

- I. The metagenome of diseased individuals differs from healthy controls and suspicious traits (i.e. viruses or genomic traits) are found more frequently in diseased individuals (**Figure 8. (1)**<sup>85</sup>)
- II. Suspicious traits from diseased individuals cause disease in healthy controls (**Figure 8. (2)**<sup>85</sup>)
- III. Selected traits alone cause the disease state in previously healthy controls (**Figure 8. (3)**<sup>85</sup>)



**Figure 8.** Metagenomic Koch's postulates by Mokili et al.<sup>85</sup>.

However, causality assessment will sometimes involve multiple factors impeding straightforward deduction based on metagenomic Koch's postulates. As an example, the role of so-called bystanders (in terms of metagenomic Koch's postulates "metagenomic traits") may be more complex upon closer examination as some viral co-infections might interfere with each other, either beneficially<sup>87</sup> or in an aggravating manner<sup>88</sup>.

Considering other causation criteria, e.g. Bradford-Hill criteria<sup>89</sup>, the complexity becomes more obvious<sup>90</sup>. In case of peri- or post-infectious CNS inflammatory disease such as ADEM, for example, the fourth Bradford-Hill criterion – temporality<sup>89</sup> - will turn out to be the *crux* of the matter. Neurological deficits coherent with ADEM will usually only appear once the acute infection has already been cleared from the body compartment that had primarily been affected, i.e. the CSF compartment. That is, timing of sampling *and also* the sample type will be key to facilitate sound causality assessment by means of agnostic detection methods<sup>69</sup>. Accounting for narrow windows of detection, especially in the CSF, paired sampling from multiple body compartments will help to shed light on a potentially causative pathogen<sup>91,92</sup>. Among well-known pathogens underlying CNS infections (such as meningitis) are enteroviruses, a large group of viruses that is only shortly and therefore inconsistently detectable in the CSF of encephalitis/ADEM patients<sup>92</sup>. In such cases, stool samples can provide a good opportunity for virus detection due to prolonged shedding<sup>33,35,92,93</sup>. To the practical end, stool samples are also easy to come by, as opposed to CSF samples or even brain biopsy which is hardly ever performed in patients.

Host-related factors will contribute to manifestation of a disease state and disease severity, as well. From research on Crohn's disease, a concept termed "'virus-plus-susceptibility gene' effect"<sup>94</sup> has been established that aims to explain why norovirus infection may be more likely to cause chronic inflammatory bowel disease in individuals with a certain autophagy gene polymorphism<sup>95</sup>. Phenotypic variability in the host, either on the large scale (e.g. primary immunodeficiency with Down, DiGeorge, or Turner syndrome)<sup>96</sup> or small-scale, i.e. single nucleotide polymorphisms (SNPs) have been shown to determine the impact of infections<sup>94</sup>. In acute infectious encephalitis, in particular, multiple predisposing factors for symptomatic (vs. asymptomatic) and severe disease have been identified, e.g. SNPs in CCL-2 and CCR-2 for Japanese encephalitis<sup>97</sup>, or in TLR3 for chikungunya and herpes simplex virus encephalitis<sup>98</sup>, or in RFC1, SCN1A, and ANPEP for West Nile encephalitis<sup>99</sup>.

Thus, drawing generic conclusions from singular mNGS findings is hardly possible. Any finding will have to be evaluated on a case-by-case basis and in due consideration of the pre-existing scientific evidence before an 'association' can be named 'causation'. While conventional PCR detection methods have led to isolated discoveries of novel associations between disease and potential trigger in the last years, e.g. ADEM and human parechovirus infection<sup>35</sup>, the introduction of mNGS has pushed boundaries. A good example for this

fact are orthobunyaviruses: As of July 2019, there were 88 different species within the orthobunyavirus genus, according to the Virus Taxonomy: 2019 Release by the International Committee on Taxonomy of Viruses (ICTV). Those small, vector-borne RNA viruses can cause febrile illness and encephalitis, one of which is the long-known California encephalitis orthobunyavirus<sup>100</sup>. Lately, mNGS has enabled the discovery of multiple novel orthobunyaviruses that most probably account for CNS infectious/ inflammatory disease, e.g. Ntvetwe virus from a Ugandan child with fatal encephalitis in 2016<sup>101</sup> or Cristoli virus (named after the Parisian suburb of Créteil) from a French woman with fatal encephalitis in 2018<sup>77</sup>. Of note, the French case of Cristoli virus-associated fatal encephalitis published by Rodriguez et al.<sup>77</sup> *post hoc* fulfills consensus case criteria of encephalitis with the highest level of diagnostic certainty<sup>102</sup>, making this singular case a worthwhile example of thorough clinical examination combined with up-to-date unbiased virological testing.

Hence, mNGS has evidently been leveraged extensively for the purpose of exploring unresolved CNS infectious/ inflammatory disease.<sup>46,47,70-76,78,103-106</sup> However, the nested approach using agnostic mNGS within a standardized digital surveillance system described in **Study 1 (Manuscript A)** of the PhD thesis project at issue is, thus far, unique. The following manuscript depicts an innovative avenue towards diminishing bias to both, the clinical and virological side of a patient's history, thus allowing for the detection of novel disease-pathogen-associations such as the first published case of ADEM linked to rotavirus infection (**Poster A**).

### **2.1.3 Manuscript A**

Obermeier PE, Ma X, Heim A, Rath B. **Avenues towards closing the diagnostic gap in encephalitis and acute disseminated encephalomyelitis in children: insight from a digital surveillance program.** Sci Rep (submitted August 2021).

**Avenues towards closing the diagnostic gap in encephalitis and acute disseminated  
encephalomyelitis in children: insight from a digital surveillance program**

Patrick E Obermeier MD<sup>1,2,3</sup>, Xiaolin Ma MD<sup>1</sup>, Albert Heim MD<sup>4</sup>, Barbara A Rath MD, PhD<sup>1,3\*</sup>

1 Vienna Vaccine Safety Initiative, Berlin, Germany

2 Charité University Medical Center, Berlin, Germany

3 Laboratoire Chrono-Environnement, Université Bourgogne Franche-Comté, Besançon, France

4 National Reference Laboratory for Adenoviruses, Hannover Medical School, Hannover, Germany

\*Corresponding author: Barbara Rath, MD PhD, Vienna Vaccine Safety Initiative, Berlin, Germany,  
F: +49-30-69088616, [Barbara.rath@vi-vi.org](mailto:Barbara.rath@vi-vi.org)

Word count: 1956/4500 (excl. Methods Section)

Title word count: 20/20

Abstract word count: 200/200

Number of references: 22/60

Number of tables: 1

Number of figures: 2

Search terms: CNS infection, ADEM, encephalitis, surveillance, metagenomic NGS

Funding: Funding was provided in kind.

Conflict of interest: All other authors report no potential conflicts.

**Abstract:**

We combined digital case classification with metagenomic next generation sequencing (mNGS) to increase diagnostic yield in pediatric encephalitis and/or acute disseminated encephalomyelitis (ADEM). Overall, 444 pediatric patients (mean age 7 years, range 0.03-18; 56% male) fulfilling predefined criteria of infection/inflammation of the central nervous system entered our hospital-based digital surveillance program from 11/2010-12/2013. Of these, 100 patients (mean age 11 years, range 0.15-17.85; 49% male) met consensus criteria for encephalitis and/or ADEM using a validated mobile app for automated case classification at the bedside. Stool samples were analyzed using viral mNGS and confirmatory PCR. Medical records were screened to identify infectious disease diagnosis codes (ICD-10) to compare routine care with digital surveillance.

In routine care, 13 of 100 patients with confirmed encephalitis/ADEM were diagnosed as 'herpesvirus infection' (n=8) or 'enterovirus infection' (n=5). Combining automated case classification with mNGS, one enterovirus infection diagnosed in routine care was confirmed and additional viruses were found in six cases. Nine viruses were newly identified by mNGS.

For encephalitis and ADEM, the combination of digital case classification and mNGS may provide novel insight, helping to resolve complex cases. Linking two agnostic methods minimizes bias and allows to identify emerging pathogens. Confirmatory testing is still required.



## Introduction

Infectious/inflammatory diseases of the central nervous system (CNS), e.g. encephalitis and acute disseminated encephalomyelitis (ADEM), are serious health conditions requiring early and accurate diagnosis<sup>1-3</sup>. However, neurological signs and symptoms can be subtle or atypical, especially in children, thus implicating diagnostic challenges<sup>4</sup>.

To help clinicians and researchers establish a diagnosis with minimal bias, health authorities advocate use of consensus case criteria, including those for encephalitis and ADEM<sup>5</sup>. Facilitating compliance with these case definitions, we have developed the VACC-Tool (Vienna Vaccine Safety Automated Case Classification Tool), a mobile precision medicine application comparing the patient's clinical presentation to published case definitions in realtime. The VACC-Tool was prospectively validated in the context of a surveillance program where it proved to efficiently harmonize diagnoses compared to retroactive chart review and ICD coding, which are prone to bias or rather serve for billing purposes<sup>4</sup>.

Once a precision diagnosis has been established and ascertained based on standardized case criteria, causality assessments constitute the next logical step in order to initiate adequate therapy or infection control measures. However, the etiologic agent remains unresolved in the majority of encephalitis and/or ADEM cases when only a limited number of pathogens, including viruses, are tested by serology and conventional PCR<sup>1</sup>. Within the past two decades, novel molecular pathogen detection techniques have been developed, including metagenomic next-generation sequencing (mNGS) which allows to identify literally any genomic sequence present in a sample. It is hoped that use of agnostic mNGS, in contrast to sequence-dependent conventional detection methods, will help to diminish bias to the laboratory end<sup>6</sup>. However, mNGS remains second-line diagnostics at this time, e.g. due to variations in sensitivity and methodology as well as the complex interpretation of mNGS data and the significance of the clinical impact of sequences detected in a specimen<sup>6,7</sup>. It is understood that mNGS will be key to advancing the discovery of the impact of infectious disease pathogens, and research but ways to deploy it best are still under investigation<sup>7</sup>. However, the detailed information conveyed by metagenomics analyses and virus diagnostics is not always matched with equally detailed and standardized clinical data. This appears to be an obstacle to advancing knowledge of the role of specific pathogens (or combinations thereof) in human disease. Oftentimes, critical clinical variables are not to be found *post hoc* in medical records, and can only be obtained reliably at the point of care.

In this study, we implemented a nested digital surveillance approach for the comprehensive investigation of pediatric encephalitis and ADEM, *combining* two agnostic methodologies as means of precision medicine: automated case classification using the VACC-Tool at the patient's bedside and viral mNGS, attempting to help close the diagnostic gap in complex infectious/inflammatory CNS diseases.

CONFIDENTIAL

## Methods

We conducted the study within a digital surveillance and quality improvement program for children with CNS infection/inflammation at the Charité University Hospital in Berlin, Germany<sup>8</sup>. From 11/2010-12/2013, all patients (aged 0-18 years) meeting entry criteria of suspected CNS infection/inflammation (fever and CNS symptoms and/or lumbar puncture)<sup>8</sup> participated in the program. During the precision screening process, all patients underwent assessments using the VACC-Tool for automated case classification at the point of care performed by specifically trained staff, independent from routine care<sup>4,8</sup>. The VACC-Tool 1.0 is a validated mobile application allowing clinicians to assess a patient while comparing the disease presentation with published case definitions for aseptic meningitis, encephalitis, myelitis, and ADEM<sup>4,5</sup>.

Exclusion criteria were known seizure disorder, CNS lesion/tumor, intoxication, traumatic head injury or acute diarrhea/dehydration<sup>8</sup>.

Stool samples from all patients with confirmed encephalitis/ADEM underwent viral mNGS. Viral mNGS results were reassessed by specific PCR assays.

Hospital discharge letters were screened for infectious disease diagnostic codes to compare with independent digital surveillance data.

### *Ethics declarations*

The prospective non-interventional surveillance program was approved by the Charité Institutional Review Board (IRB number: EA2/161/11). The study was performed in compliance with the ethical standards of the Declaration of Helsinki and according to German laws. Written informed consent was waived for the purpose of quality improvement and infection control given the observational nature of the study. Verbal informed consent was obtained from patients above 16 years of age or parents or caretakers of underage patients<sup>8</sup>.

### *Viral mNGS – sample processing*

Stool suspensions were mixed with phosphate-buffered saline (0.75 ml) and zirconia beads (0.2 g), vortexed, and spun at 12.000 rounds per minute (rpm) in a tabletop microfuge for 10 min and supernatants were transferred into Eppendorf tubes. Supernatants (200 µl) were filtered through a 0.45 µm pores filter to exclude cells and large particles. The filtrates were then digested with a combination of DNase and RNase nucleases to reduce background of host and bacterial genetic material and enrich

for viral nucleic acids protected from nuclease digestion within their capsids<sup>9,10</sup>. Nucleic acid extraction was performed using the Qiagen Viral RNA mini kit according to manufacturer's instructions. Viral cDNA synthesis was performed separately on each individual sample using viral nucleic acids (10 µl) with random hexamers (100 pmol) at 72°C for 2 min. 200U SuperScript III reverse transcriptase (Invitrogen), 0.5 mM of each deoxynucleoside triphosphate (dNTP), 10 mM dithiothreitol, and 1× first-strand extension buffer were added to the mixture and incubated at 25°C for 10 min, followed by 50°C incubation for 1 h and 70°C for 15 min. Second-strand cDNA synthesis was performed by incubation of reversely transcribed products with 5U of Klenow Fragment DNA polymerase (New England Biolabs) at 37°C for 1 h followed by 75°C for 20 min. The resulting double-stranded cDNAs from groups of five stool samples were then pooled. The transposon based Nextera XT DNA sample preparation kit (Illumina) was then used followed by PCR using unique pairs of index barcodes for each pool. Each library was quantified using Kapa Library Quant kit (Kapa Biosystems) following manufacturer's instructions and equimolar DNA quantities pooled for sequencing using one flow cell on the Illumina HiSeq4000 instrument with 150 paired-end sequencing.

#### *Viral mNGS - bioinformatics pipeline*

Paired-end sequencing reads were debarcoded by Illumina vendor software. Using an in-house analysis pipeline running on a 36-nodes Linux cluster, bacterial reads were subtracted by mapping to nucleotide sequences from GenBank nucleotide database using bowtie2 (version 2.2.4)<sup>11</sup>.

Reads were considered duplicates if base positions 5 to 55 were identical. One random copy of duplicates was kept. Low sequencing quality tails were trimmed using Phred quality score 20. Adaptor and primer sequences were trimmed using VecScreen11 default parameters. Cleaned reads were assembled *de novo* using Ensemble Assembler<sup>12</sup>. Assembled contigs, along with the remaining singlets, were aligned to an in-house viral proteome database using BLASTx. Matches to virus proteins were aligned to an in-house non-virus-non-redundant (NVNR) universal proteome database using BLASTx. The human and bacterial nucleotide database was compiled as follows: Human reference genome sequences and mRNA sequences (hg38) were concatenated. Bacterial nucleotide sequences were extracted from NCBI nr fasta files based on NCBI taxonomy. Human and bacterial nucleotide sequences were compiled into bowtie2 databases<sup>11</sup> for cellular sequences subtraction. Two databases were constructed: 1) a virus BLASTx database using the NCBI virus reference proteome to which viral protein sequences from NCBI nr fasta files were added; and 2) a NVNR database using sequences extracted

from NCBI nr fasta file. Repeats and low-complexity regions were masked using segmasker from blast+ suite (version 2.2.7). Hits with lower adjusted E-value to NVNR than to viral proteins were removed.

#### *Confirmatory PCR testing*

Aliquots of pooled samples were reassessed using specific confirmatory PCR to narrow down pooled mNGS results to the individual patient level for sapo-, rota-, entero-, boca-, adeno-, and saffold virus<sup>8,13-</sup>

<sup>16</sup>.

CONFIDENTIAL

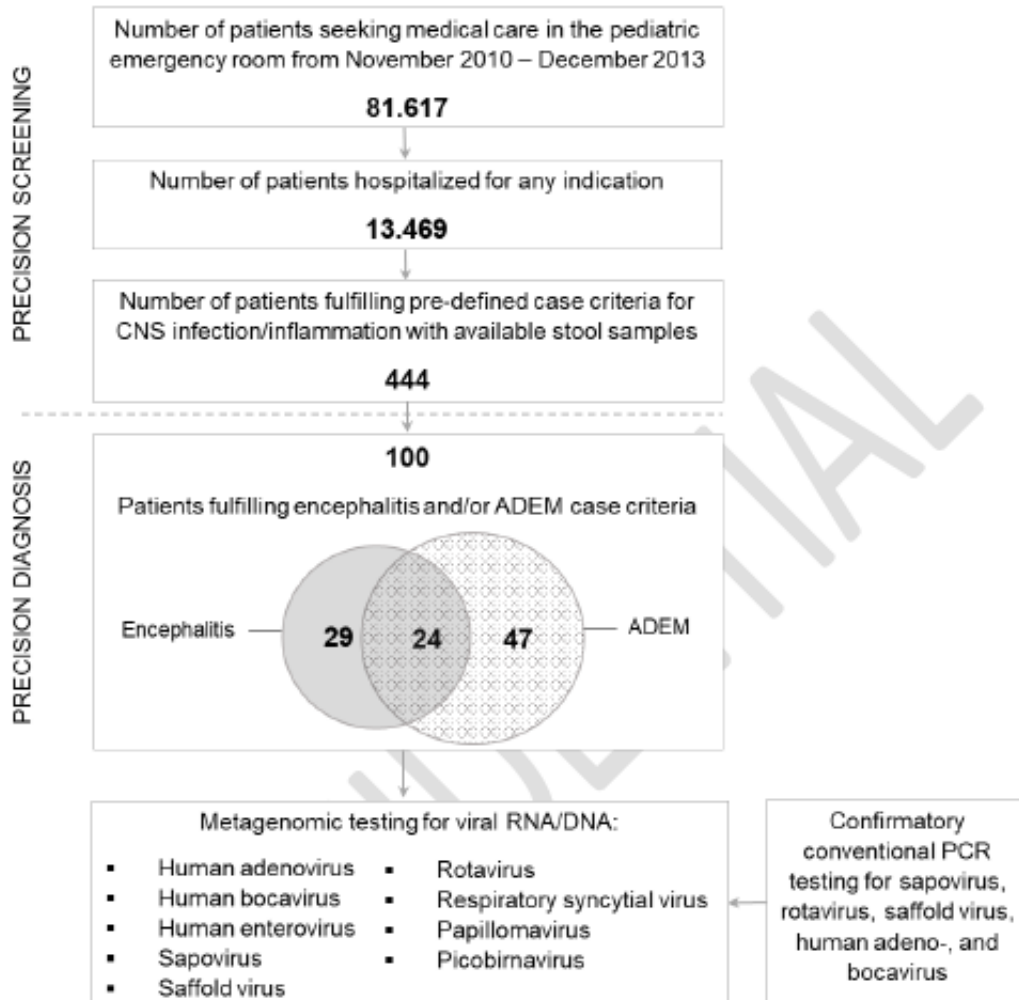
## Results

Overall, 444 patients entered the digital surveillance program (Table 1).

**Table 1.** Patient characteristics.

	Total (N=444)	Patients selected for metagenomic testing (n=100)	Encephalitis cases (n=29)	ADEM cases (n=47)	Cases meeting both encephalitis and ADEM case criteria (n=24)
Mean age in years (range)	7 (0-18)	11 (0.1-17.8)	9 (0.7-17.3)	11 (0.1-17.8)	11 (0.9-17.3)
0-28 days	11 (2%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
28-365 days	91 (20%)	4 (4%)	1 (3%)	2 (4%)	1 (4%)
1-2 years	51 (11%)	6 (6%)	5 (17%)	0 (0%)	1 (4%)
3-5 years	70 (16%)	10 (10%)	5 (17%)	3 (6%)	2 (8%)
6-18 years	221 (50%)	80 (80%)	18 (62%)	42 (89%)	20 (83%)
Gender					
Male	247 (56%)	49 (49%)	11 (38%)	24 (51%)	14 (58%)
Female	197 (44%)	51 (51%)	18 (62%)	23 (49%)	10 (42%)

Their clinical presentations were classified using the VACC-Tool 1.0<sup>4</sup>, yielding 100 cases meeting case criteria for encephalitis and/or ADEM. Figure 1 depicts the precision screening process and classification results, i.e. precision diagnosis.

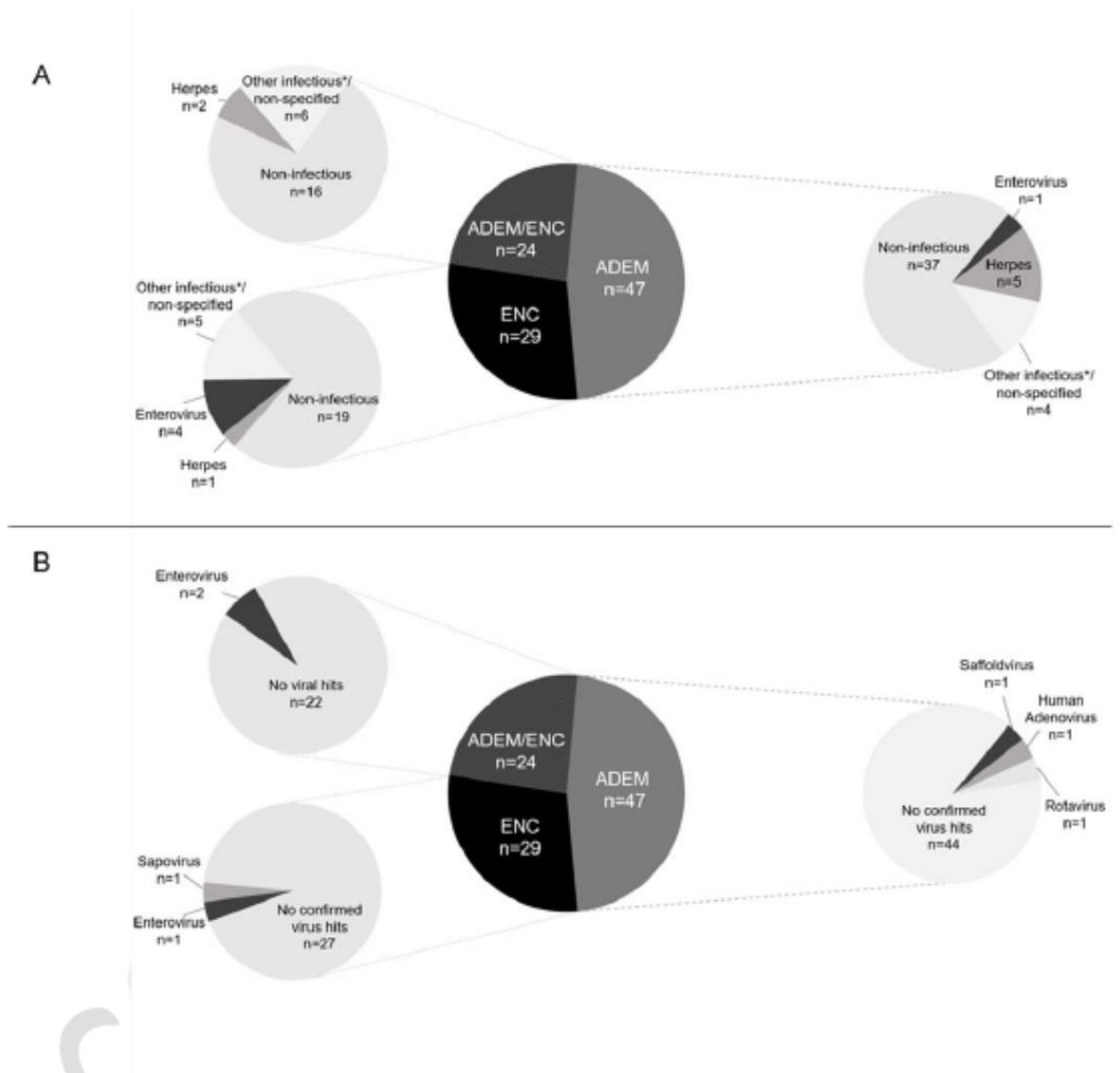
**Figure 1.** Precision screening and diagnosis process

Of 100 patients with confirmed encephalitis and/or ADEM, 13 held a virus-specific diagnosis in their routine care discharge letter. Of those, 8 were 'herpesvirus infection of the CNS', 5 of which tested positive by PCR from cerebrospinal fluid (CSF), 2 by antibody testing from CSF, and 1 by serology.

Another 5 were 'enterovirus infection of the CNS', 2 of which tested positive by PCR from CSF, 1 by PCR from saliva, 1 by PCR from serum, and 1 based on clinical suspicion only.

Another 6 cases were 'other/unspecified (viral) infection', 9 more cases 'borrelia infection'. The remainder (72%) did not receive any virus/infectious disease diagnosis.

Figure 2 depicts routine care infectious disease diagnoses compared with confirmed viral mNGS results.

**Figure 2.** Routine care infectious disease diagnoses vs. viral metagenomic test results

(A) showing virus/ infectious disease diagnoses from routine care among patients with confirmed encephalitis (ENC) and/or acute disseminated encephalomyelitis (ADEM). (B) showing PCR-confirmed viral metagenomic next-generation sequencing test results among the same patients.

\*including "other spirochetal infection", i.e. International Statistical Classification of Diseases and Related Health Problems 10<sup>th</sup> Revision (ICD-10) code A69.x and "other infectious diseases", i.e. ICD-10 code B99.



#### *Viral mNGS and confirmatory PCR testing*

Pooled viral mNGS revealed 9 distinct viral species (Figure 1). Identification of individual cases by confirmatory PCR testing was possible for sapovirus, rotavirus, human adenovirus, and saffold virus. Detection of human bocavirus reads by viral mNGS could not be confirmed by PCR.

#### *Sapovirus in a 2-year-old with encephalitis*

A 2-year-old female was hospitalized with fever, cough, decreased levels of consciousness, acute personality changes, phonophobia, and bilateral motor weakness of the limbs. Broad-spectrum antibiotics, oseltamivir, and intravenous immunoglobulins were administered upon admission.

Blood analysis showed elevated C-reactive protein and interleukin-6 levels and respiratory acidosis. CSF analysis was normal except for elevated glucose and protein levels, i.e. cytoalbuminologic dissociation. Routine care bacterial testing from CSF and blood were all negative. Routine care viral multiplex PCR testing from respiratory secretions was positive for RSV, enterovirus, rhinovirus, and Epstein-Barr virus on day 2. A stool sample obtained on day 19 tested positive for rotavirus by PCR in routine care. Cranial magnet resonance imaging revealed parenchymal signal enhancement.

The patient fully recovered and was discharged after 26 days.

Encephalitis case criteria were met, as confirmed by the VACC-Tool. The stool sample obtained within the quality improvement program on hospital day 2 tested positive for sapovirus by viral mNGS as confirmed by real-time RT-PCR (Cp value: 23.5).

#### *Rotavirus in a 2-year-old with ADEM*

A 2-year-old female presented with an afebrile convulsive seizure and ataxia. No gastroenteritis or other symptoms were present. On hospital day 7, a maximum body temperature of 38.8°C was measured. CSF analysis showed elevated total protein levels and normal white blood cell counts. In routine care, microbiology testing was all negative. The patient fully recovered without specific treatment and was discharged after 9 days.

ADEM case criteria, including cerebellar dysfunction, were met and confirmed using the VACC-Tool. The stool sample tested positive for rotavirus A by viral mNGS as confirmed by real-time RT-PCR (Cp value: 13.24).

*Saffold virus in a 4-year-old with ADEM*

A 4-year-old male presented with acute afebrile seizures, behavioral changes, ongoing fatigue, motor weakness of the limbs, and bladder dysfunction in addition to mild respiratory symptoms and conjunctivitis. In routine care, human rhinovirus (Ct value: 30) was detected in a nasopharyngeal sample by PCR. CSF analysis was normal except for cytoalbuminologic dissociation. The patient fully recovered without specific treatment and was discharged after 9 days.

ADEM case criteria, including focal neurological deficits, were met and confirmed by the VACC-Tool. The stool sample collected on hospital day 2 tested positive for saffold virus by viral mNGS as confirmed by nested RT-PCR.

*Case Vignette: Adenovirus in a 13-year-old with ADEM*

A 13-year-old male presented with headache, vomiting, anosmia, decreased peripheral sensitivity and paresthesia, and transient sudden loss of consciousness witnessed by his parents. A somatosensory evoked potential test showed focal interruption of the median nerve. Blood testing revealed moderate leukocytosis. The patient fully recovered without specific treatment and was discharged after 4 days.

ADEM case criteria, including focal neurological deficits, were met as confirmed by the VACC-Tool. The stool sample collected on hospital day 1 tested positive for human adenovirus C by viral mNGS and confirmatory PCR.

## Discussion

This study introduced a nested precision medicine approach combining automated encephalitis and ADEM case classification in realtime using mobile health technology with viral mNGS and confirmatory PCR testing to diminish bias in the clinical *and* laboratory context. This study provided insight into three years of standardized digital infectious disease surveillance with an established total case number.

Following the precision screening and precision diagnosis steps based on consensus case criteria incorporated into the VACC-Tool<sup>4</sup>, we investigated whether agnostic mNGS might help in closing the 'diagnostic gap' in children with encephalitis and/or ADEM, with particular focus on viruses being the most frequent cause of those disease entities<sup>17</sup>.

In line with previous studies, few encephalitis and ADEM cases (13%) held a specific infectious disease routine care diagnosis within our cohort. Review of routine care medical records in our cohort showed non-uniform proceedings in the diagnostic work-up, including one diagnosis of "enterovirus infection" based on clinical suspicion only, exemplarily portraying the potential pitfalls of chart review and ICD-coding for surveillance purposes.

If the diagnostic work-up is prompted by a treating physician, it is usually confined to bacterial cultures and certain viruses, e.g. herpes or enteroviruses. Multiplex panels help to test for a certain range of pathogens, expanding the diagnostic search field by usually 10-20 pathogens<sup>7</sup>.

However, mNGS is an agnostic detection method and is being leveraged increasingly, e.g. as second-line diagnostics for "hard-to-diagnose cases"<sup>18</sup>. In contrast to PCR, mNGS laboratory protocols are not yet standardized, and sensitivities may vary. Our study underlines that the sensitivity and specificity of mNGS may not yet approximate that of established PCR assays, depending on methods used to enrich viruses, to generate DNA libraries, and the depth of sequencing<sup>1,18</sup>. Standardized protocols for viral mNGS, downstream data processing, and validation for CSF and respiratory tract specimens in addition to stool specimens will be prerequisites for the meaningful linkage of surveillance systems and databases<sup>19</sup>.

Viral mNGS and confirmatory PCR testing in this pediatric cohort identified novel or rare potential linkages to specific clinical cases that were ascertained using the VACC-Tool at the bedside: We described cases of ADEM linked to a) saffold virus, b) rotavirus, and c) adenovirus infection<sup>13</sup>, all of which are requiring further surveillance.

At this time, mNGS could support clinical decision-making by prompting isolation measures or avoiding inappropriate medication, e.g. antibiotics<sup>1</sup>. However, interpretation of mNGS results will become challenging, e.g. when multiple viruses are detected in one sample. Within our cohort, we describe a case of encephalitis linked to sapovirus infection who simultaneously tested positive for enteroviruses. Differentiation between cause or bystander effect needs to be clarified on a case-by-case basis. Evidently, rapid turnaround times and reasonable costs of mNGS will also be critical for mNGS results to become actionable<sup>6</sup>.

Apart from the detection method used, sample type and timing are critical. For (para-/post-)infectious ADEM and encephalitis, non-invasive stool samples may be the preferred specimen type, as they will potentially provide an extended diagnostic window beyond the time of viral clearance from CSF<sup>1,20</sup>. In our study, we increased diagnostic yield for enteroviruses using stool samples. However, comparative studies of mNGS from other body compartments, including blood, urine, saliva, or CSF are warranted<sup>21</sup>. Pooling of samples in our study was more cost-effective but likely impaired the sensitivity of viral mNGS. Also, future use of mNGS to 'close the diagnostic gap' in cases of encephalitis and ADEM will need to include parasites, bacteria, and fungi. Another limitation of the hospital-based surveillance was the lack of a control group, e.g. involving healthy individuals to investigate whether mNGS hits might have been a bystander effect. On a positive note, selection bias was avoided with this approach.

Future studies will incorporate more infectious/inflammatory CNS disease entities to provide distinct case classification results by delineating different disease entities from one another. As described previously, the VACC-Tool 2.0 beta-version allows for automated case classification of 14 different infectious/inflammatory CNS diseases, including facial nerve palsy, progressive multifocal leukoencephalopathy, or Guillain-Barré syndrome in full compliance with data standards across sites and data privacy<sup>22</sup>. Not only does use of the VACC-Tool help to ascertain diagnoses with minimal bias but it has the potential to flag clinically suspicious cases, prompting further diagnostic work-up, e.g. lumbar puncture or imaging studies to allow for highest diagnostic certainty in the clinical setting<sup>4</sup>.

Our results suggest that the combination of digital case classification and agnostic viral mNGS provides an innovative precision medicine approach to diminish bias and increase diagnostic yield in pediatric encephalitis and ADEM. Use of mobile health technology aiding healthcare experts to collect all pertinent

clinical information while the patient is readily accessible at the bedside will not only help to arrive at comprehensive but also unbiased diagnoses based on consensus case criteria in realtime and ultimately, to harmonize interpretation of potential linkages between disease entities and certain pathogens detected by agnostic detection methodologies, e.g. mNGS. Application of viral mNGS as first-line molecular detection method will require ongoing technical improvement, further evaluation, and confirmation using target-specific confirmatory testing.

CONFIDENTIAL

**Acknowledgements**

The authors thank the patients for participation.

**Contributor Statement**

Conceptualization and supervision of the work: BR. Development of methodology: BR. Conducting formal analysis: PO, XM, AH, BR. Provision of study materials, reagents, materials, patients, laboratory samples, computing resources, and other analysis tools: BR, AH. Data curation: PO, BR. Writing of the original draft: PO. Review and editing of the manuscript draft: all authors. Project administration: BR. Funding (in-kind): BR.

CONFIDENTIAL

## References

- 1 Ramachandran, P. S. & Wilson, M. R. Metagenomics for neurological infections - expanding our imagination. *Nat Rev Neurol* **16**, 547-556, doi:10.1038/s41582-020-0374-y (2020).
- 2 Khetsuriani, N., Holman, R. C. & Anderson, L. J. Burden of encephalitis-associated hospitalizations in the United States, 1988-1997. *Clin Infect Dis* **35**, 175-182, doi:10.1086/341301 (2002).
- 3 Collaborators, G. B. D. N. Global, regional, and national burden of neurological disorders, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol* **18**, 459-480, doi:10.1016/S1474-4422(18)30499-X (2019).
- 4 Obermeier, P. et al. Enabling Precision Medicine With Digital Case Classification at the Point-of-Care. *EBioMedicine* **4**, 191-196, doi:10.1016/j.ebiom.2016.01.008 (2016).
- 5 Sejvar, J. J. et al. Encephalitis, myelitis, and acute disseminated encephalomyelitis (ADEM): case definitions and guidelines for collection, analysis, and presentation of immunization safety data. *Vaccine* **25**, 5771-5792, doi:10.1016/j.vaccine.2007.04.060 (2007).
- 6 Jia, X. et al. A streamlined clinical metagenomic sequencing protocol for rapid pathogen identification. *Sci Rep* **11**, 4405, doi:10.1038/s41598-021-83812-x (2021).
- 7 Chiu, C. Y. & Miller, S. A. Clinical metagenomics. *Nat Rev Genet* **20**, 341-355, doi:10.1038/s41576-019-0113-7 (2019).
- 8 Karsch, K. et al. Human Parechovirus Infections Associated with Seizures and Rash in Infants and Toddlers. *Pediatr Infect Dis J* **34**, 1049-1055, doi:10.1097/INF.0000000000000802 (2015).
- 9 Allander, T., Emerson, S. U., Engle, R. E., Purcell, R. H. & Bukh, J. A virus discovery method incorporating DNase treatment and its application to the identification of two bovine parvovirus species. *Proc Natl Acad Sci U S A* **98**, 11609-11614, doi:10.1073/pnas.211424698 (2001).
- 10 Victoria, J. G. et al. Metagenomic analyses of viruses in stool samples from children with acute flaccid paralysis. *J Virol* **83**, 4642-4651, doi:10.1128/JVI.02301-08 (2009).
- 11 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).
- 12 Deng, X. et al. An ensemble strategy that significantly improves de novo assembly of microbial genomes from metagenomic next-generation sequencing data. *Nucleic Acids Res* **43**, e46, doi:10.1093/nar/gkv002 (2015).
- 13 Dhingra, A. et al. Molecular Evolution of Human Adenovirus (HAdV) Species C. *Sci Rep* **9**, 1039, doi:10.1038/s41598-018-37249-4 (2019).
- 14 Oka, T. et al. Detection of human sapovirus by real-time reverse transcription-polymerase chain reaction. *J Med Virol* **78**, 1347-1353, doi:10.1002/jmv.20699 (2006).
- 15 Adlhoch, C. et al. Highly sensitive detection of the group A Rotavirus using Apolipoprotein H-coated ELISA plates compared to quantitative real-time PCR. *Virol J* **8**, 63, doi:10.1186/1743-422X-8-63 (2011).
- 16 Ma, X. et al. Can we distinguish respiratory viral infections based on clinical features? A prospective pediatric cohort compared to systematic literature review. *Rev Med Virol* **28**, e1997, doi:10.1002/rmv.1997 (2018).
- 17 Granerod, J. et al. Causes of encephalitis and differences in their clinical presentations in England: a multicentre, population-based prospective study. *Lancet Infect Dis* **10**, 835-844, doi:10.1016/S1473-3099(10)70222-X (2010).
- 18 Kufner, V. et al. Two Years of Viral Metagenomics in a Tertiary Diagnostics Unit: Evaluation of the First 105 Cases. *Genes (Basel)* **10**, doi:10.3390/genes10090661 (2019).
- 19 Perlejewski, K. et al. Next-generation sequencing in the diagnosis of viral encephalitis: sensitivity and clinical limitations. *Sci Rep* **10**, 16173, doi:10.1038/s41598-020-73156-3 (2020).
- 20 Obermeier, P. E. et al. Acute Disseminated Encephalomyelitis After Human Parechovirus Infection. *Pediatr Infect Dis J* **35**, 35-38, doi:10.1097/INF.0000000000000928 (2016).

- 21 Edridge, A. W. D. *et al.* Viral Metagenomics on Cerebrospinal Fluid. *Genes (Basel)* **10**, doi:10.3390/genes10050332 (2019).
- 22 Hoppe, C. *et al.* Innovative Digital Tools and Surveillance Systems for the Timely Detection of Adverse Events at the Point of Care: A Proof-of-Concept Study. *Drug Saf* **39**, 977-988, doi:10.1007/s40264-016-0437-6 (2016).

CONFIDENTIAL



## 2.1.4 Poster A

Obermeier PE, Delwart E, Rath BA. **Acute disseminated encephalomyelitis linked to rotavirus infection - Novel insight from a prospective surveillance program.** Poster Presentation, European Clinical Case Reports Congress, Vienna, Austria, April 21-22, 2018.

### Acute disseminated encephalomyelitis (ADEM) linked to rotavirus infection - Novel insight from a prospective surveillance program



Patrick E Obermeier<sup>1,2,3</sup>, Eric Delwart<sup>4</sup>, Barbara A Rath<sup>1,2,3,5</sup>

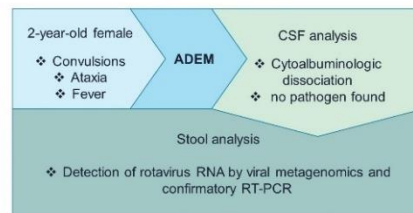
- 1 Vienna Vaccine Safety Initiative, Berlin, Germany.  
 2 Université de Bourgogne Franche-Comté, Besançon, France  
 3 Charité University Medical Center, Berlin, Germany  
 4 Blood Systems Research Institute, San Francisco, USA.  
 5 University of Nottingham, Nottingham, UK



#### BACKGROUND

Acute disseminated encephalomyelitis (ADEM) is a demyelinating disease of the central nervous system (CNS) accounting for considerable disease burden in infants and children. In most instances, a viral infection will precede neurological symptoms that include focal neurological deficits and/or signs of encephalopathy. Once a case of ADEM has been ascertained, it may be challenging to identify a viral cause or trigger of the disease. To date, we may know only a small fraction of viral pathogens that may be associated with para-/postinfectious ADEM.

#### CASE PRESENTATION



#### OBJECTIVE

To explore potential links between viral infections and ADEM or other neuroinflammatory conditions, we established a prospective syndromic surveillance program at the Charité Department of Pediatrics assessing all patients hospitalized with signs and symptoms of CNS infection/inflammation.

#### CONCLUSION

To our knowledge, this is the first report of confirmed ADEM associated with rotavirus infection. Standardized case ascertainment and unbiased pathogen detection may facilitate translational research exploring complex disease presentations and associated pathogens with minimal selection bias. However, large-scale prospective surveillance programs with paired cerebrospinal fluid (CSF) and stool sampling are warranted to further investigate the incidence of ADEM in the context of rotavirus infection.

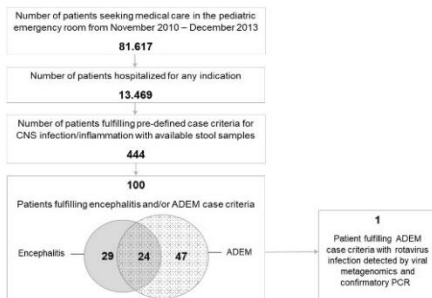
#### METHODS

The present study was conducted in the context of a Quality Management program for enhanced virologic and syndromic surveillance at the Charité Department of Pediatrics in collaboration with the Blood Systems Research Institute, San Francisco, California, USA.

Between November 2010 and December 2013, a total of 100 stool samples were collected from pediatric patients with confirmed encephalitis and/or ADEM (mean age 11 years, range 0.15-17.85 years, 49% male).

A mobile application, the VACC-Tool™, was used to ascertain clinical presentations at the point of care and to classify cases according to pre-defined clinical consensus case definitions.

Stool samples were tested using a viral metagenomics sequencing approach and subsequent confirmatory PCR testing.



The VACC-Tool™ is a mobile application allowing doctors to assess the patient's status at the point of care – thereby comparing the disease presentation to multiple case definitions at the same time. The data are in full compliance with international standards.

#### REFERENCES AND LINKS

- Obermeier P, Muehlhans S, Hoppe C, Karsch K, Tief F, Seeber L, Chen X, Conrad T, Boettcher S, Diedrich S, Rath B. Enabling Precision Medicine With Digital Case Classification at the Point-of-Care. *EBioMedicine*. 2016;4:191-196.
- Obermeier PE, Karsch K, Hoppe C, Seeber L, Schneider J, Muehlhans S, Chen X, Tief F, Kaindl AM, Weschke B, Böttcher S, Diedrich S, Rath B. Acute Disseminated Encephalomyelitis After Human Parechovirus Infection. *Pediatr Infect Dis J*. 2016;35:35-38.
- Hoppe C, Obermeier P, Muehlhans S, Alchikh M, Seeber L, Tief F, Karsch K, Chen X, Boettcher S, Diedrich S, Conrad T, Kistler B, Rath B. Innovative Digital Tools and Surveillance Systems for the Timely Detection of Adverse Events at the Point of Care: A Proof-of-Concept Study. *Drug Saf*. 2016;39(10):977-88. doi:10.1007/s40264-016-0437-6. PMID:27350063.
- Rath B. Vienna Vaccine Safety Initiative. *Hum Vaccin Immunother* 2017: 1-4.

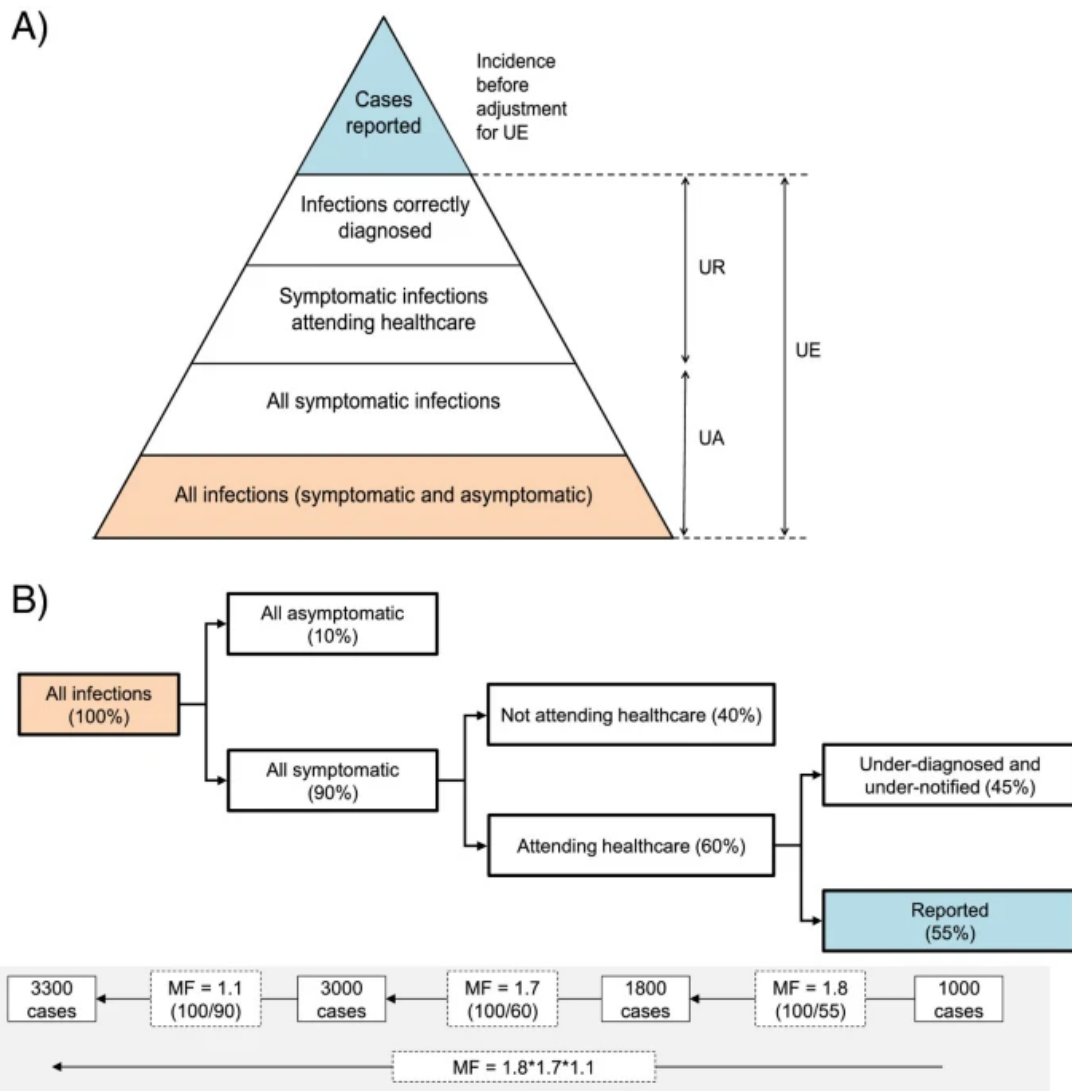
[info@vi-vi.org](mailto:info@vi-vi.org) | [www.vi-vi.org](http://www.vi-vi.org)



## 2.2 Acute Respiratory Infections

Influenza and other high-threat pathogens causing respiratory disease, such as (severe) acute respiratory infections ((S)ARI) and ILI, were identified among the top ten threats to global health in 2019 by WHO<sup>107</sup>. In 2020, their rating proved true as the COVID-19 pandemic caused by SARS-CoV-2 became a global menace<sup>108</sup>. The degree to which the world has suffered damage by this Public Health Emergency of International Concern might indicate that there exist unmet needs for the purpose of respiratory infectious disease surveillance and infection control.

Surveillance systems are in place for influenza, (S)ARI, and ILI aim to investigate the spread, disease severity and burden as well as molecular pathogen characteristics to provide guidance on outbreak management and patient care. For influenza and ILI, several surveillance systems have been compiling valuable information, e.g. the U.S. Influenza Surveillance System operated by stakeholders such as the Centers for Disease Control and Prevention (CDC) which are, at the same time, WHO collaborating centers contributing to GISRS on the global scale. Those well-established surveillance networks are now being leveraged increasingly for COVID-19 surveillance, too<sup>109,110</sup>. To this end, FluView and COVIDView by the CDC are two examples of surveillance based on multiple data streams, mainly relying on voluntary reports from vigilant health care providers located at public health, commercial, and clinical laboratories<sup>111,112</sup>. However, those voluntary reports often tend to be delayed and/or incomplete and biased, lacking standardization and thus bearing the risk of misperceptions and underestimation of disease burden<sup>113-115</sup>. **Figure 9**<sup>115</sup> displays an estimation of the extent to which infectious diseases in general might be under-ascertained and under-reported based on a study by Gibbons et al.<sup>115</sup>. Bias inherent with reports from healthcare professionals may arise from different levels of knowledge, lack of time in busy clinical care, or lack of access to and availability of reporting resources (e.g. computers or paper forms), for example<sup>53,116,117</sup>.



**Figure 9.** Conceptual pyramid generically displaying A) underestimation (UE), under-ascertainment (UA), and under-reporting (UR) of infectious diseases and B) an estimation of missed infectious disease cases among a fictional set of 1000 reports of diseased individuals based on multiplication factors (MF) derived from the literature<sup>115</sup>.

#### *Case Definitions and Shared Terminology in Surveillance Systems*

One means to account for the risk of bias and thus misinterpretation/ underestimation of (respiratory) infectious disease is that surveillance networks will arrive at common ground in terms of shared terminology, e.g. in that they make use of the same case definitions for ILI and (S)ARI (**Figure 9**<sup>118</sup>). In the multitude of surveillance systems and case definitions however, minor or major differences will impede comparability of data<sup>119</sup>. In the U.S. Outpatient Influenza-like Illness Surveillance Network (ILINet), leveraged for CDC's COVIDView, ILI is defined as a body "temperature of 100°F [37.8°C] or greater and a cough and/or a sore throat without a known cause other than influenza"<sup>110</sup> (see also **Figure 9**<sup>118</sup>).

Definition	Type	Sudden onset	General symptoms	Respiratory symptoms
ECDC	ILI	Yes	At least one among: fever, feverishness, headache, malaise, myalgia	At least one among: cough, sore throat, shortness of breath
WHO	ILI	No	Fever $\geq 38^{\circ}\text{C}$ with onset within the last 10 days	Cough
CDC	ILI	Yes	Fever $\geq 100^{\circ}\text{F}$ ( $37.8^{\circ}\text{C}$ ) Absence of a known cause other than influenza	At least one among: cough, sore throat
GROG	ARI	Yes	At least one among: fever $\geq 38^{\circ}\text{C}$ , headache, weakness, myalgia, chills	At least one among: cough, coryza, bronchitis, pharyngitis, shortness of breath, expectoration

**Figure 10.** Overview of case definitions for ILI and ARI<sup>118</sup>.

Of note, the 2020 ILINet/ CDC definition includes “sore throat” as ILI criterion while WHO has omitted it in the 2018 WHO revised definition<sup>120</sup>. A study by Casalegno et al. based on the French influenza surveillance network Groupes Régionaux d’Observation de la Grippe (GROG) showed that the revised WHO ILI case definition was the most accurate compared to the CDC and also the European Centre for Disease Prevention and Control (ECDC) ILI case definitions<sup>118</sup>. On another layer of standardization within case definitions themselves, terminology needs to be harmonized so that any data point (or variable) can be understood unambiguously, shared, and analyzed across sites. For example, standardized glossaries will have to clarify what the term “confirmed influenza case” actually signifies [“A patient who tests positive for influenza virus (flu) infection by an approved laboratory test”]<sup>121</sup>. For this purpose, regulatory authorities advocate the use of harmonized terminology and controlled vocabulary, e.g. by the Clinical Data Interchange Standards Consortium (CDISC)<sup>122-124</sup>. Also, new terms such as “COVID-19-like illness (CLI)”<sup>111</sup> will arise, requiring ongoing and sharp delineation from congeneric terms, i.e. ILI, to guarantee straightforward communication among healthcare experts.

A comprehensive systematic review of influenza-associated disease burden in children below 5 years of age conducted by Wang et al. and the Respiratory Virus Global Epidemiology Network published in 2020 exemplifies that a lack of standardization hinders meta-analyzability of data, thus requiring demand-oriented re-processing of source data and, in addition, *post hoc* alignment. Despite considerable efforts undertaken in this study, the authors acknowledge limitations resulting in “wide uncertainty ranges”<sup>125</sup> in

their analyses as a natural consequence of concurrent terminology and bias if standardized case criteria are not applied uniformly<sup>125</sup>.

The value of implementing case criteria as means of standardization is thus indisputable. Against this backdrop, practical considerations for preparing GISRS for the COVID-19 pandemic issued by WHO in May 2020 need to be considered with caution: According to GISRS interim guidance, “alternative surveillance systems, such as [...] coding-based surveillance for COVID-19”<sup>126</sup> shall be considered. Towards this end, a recent preliminary study investigating emergency pre-hospital visits with suspected COVID-19 patients in North Carolina, USA has specifically demonstrated that ICD-codes have poor accuracy<sup>127</sup> and might be subject to “hot topic bias”<sup>113,114</sup>.

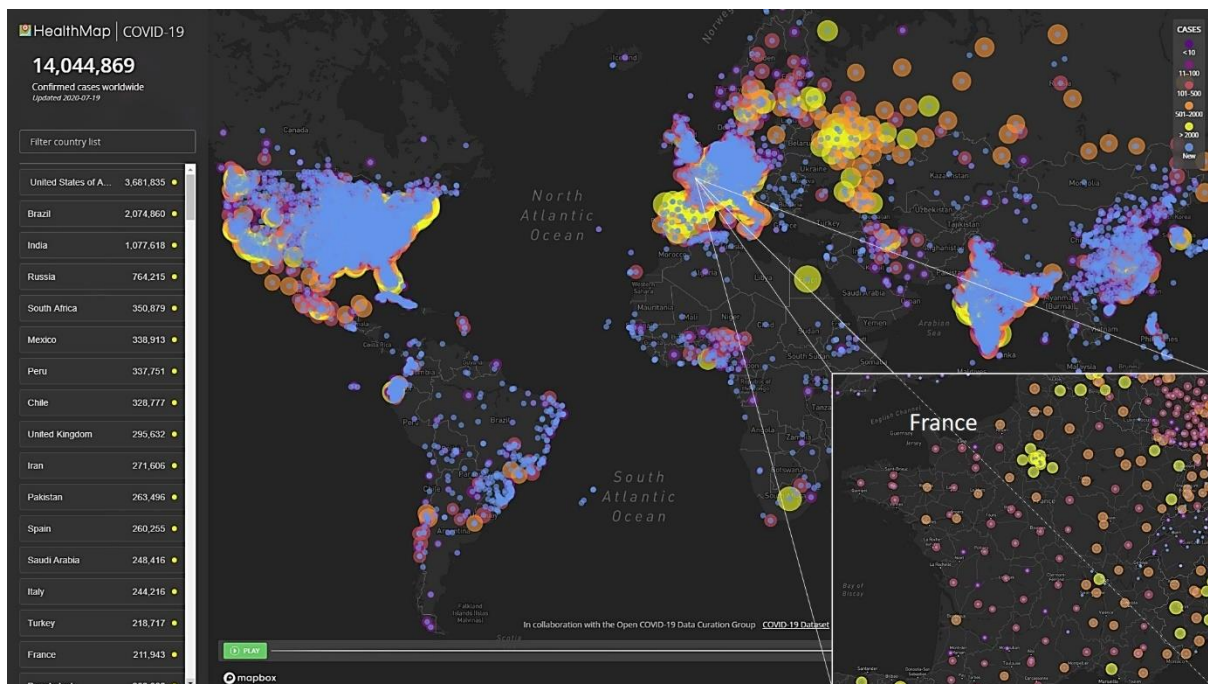
#### *Investigating Respiratory Infectious Disease*

“Alternative surveillance systems” imply, among others, digital surveillance, i.e. any surveillance that involves electronic health data<sup>128</sup>, including user-generated data as auxiliary means to supplement traditional surveillance investigating increasing infectious disease activity<sup>129-131</sup>. User-generated data driven infoveillance, i.e. trend analyses based on (longitudinal) metrics generated in digital format<sup>128,132-134</sup>, have already been leveraged in an attempt to predict a forthcoming pandemic, mainly focusing on influenza<sup>130,132,133,135</sup>. A notorious example of this approach is Google Flu Trends, operated by Google LLC (Limited Liability Company) from 2008-2015 aiming to nowcast and predict influenza activity in (near) real-time based on search queries<sup>136</sup>. Some studies stated that data obtained by web-based Google Flu Trends allowed to identify increased influenza activity approximately one week prior to CDC surveillance systems, especially before and during the 2009 influenza outbreak<sup>137-142</sup>. During the 2012/2013 influenza season however, it was concluded that Google Flu Trends results tended to be overestimating potential disease signals<sup>129,143,144</sup>. This and other reasons, including data privacy concerns and lack of methodological transparency have led to decommissioning of the service<sup>145</sup>.

A service still in place since 1994 is the ProMED-mail informal electronic outbreak reporting system aiming to monitor emerging infectious diseases globally with a rapid turnaround time of usually <24h, on that account bypassing traditional peer-review. Reports are usually screened by expert moderators and then rapidly shared with a community of currently >80.000 subscribers in >180 countries via e-mail, website

postings or alerts on Twitter and Facebook<sup>146 147</sup>. As for COVID-19, ProMED-mail was one of the first reporting an "urgent notice on the treatment of pneumonia of unknown cause" in Wuhan, China on 30<sup>th</sup> of December 2019, 11:59 pm<sup>148</sup>.

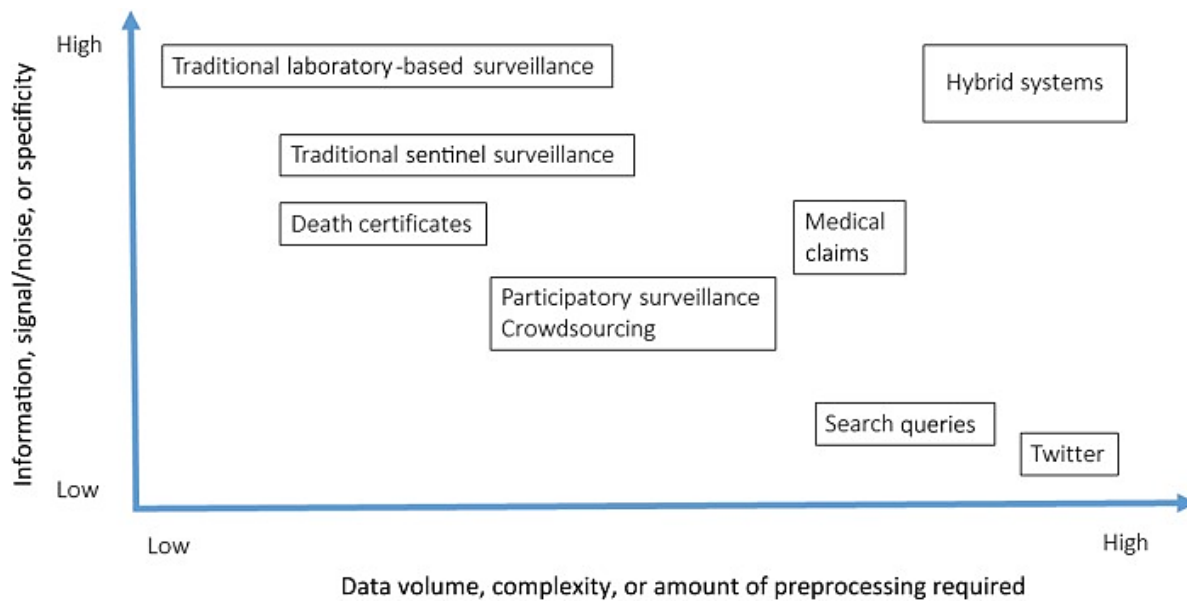
ProMED-mail alerts are also integrated into an informal non-profit partnering system for "querying, filtering, integrating and visualizing unstructured reports on disease outbreaks"<sup>149</sup> based on multiple data streams, including news media, called HealthMap<sup>146,149,150</sup>. During the COVID-19 pandemic, HealthMap has implemented an artificial intelligence approach towards early detection and visualization of confirmed COVID-19 cases<sup>150</sup> (**Figure 11**<sup>151</sup>), thereby aiming to forecast spread and provide guidance for health authorities<sup>149,150</sup>. Declared goal of the Initiative is also to mine those large amounts of data being gathered now during the SARS-CoV-2 outbreak in order to prepare for a potential next epi-/pandemic<sup>152</sup>.



**Figure 11.** HealthMap for COVID-19 as of 19<sup>th</sup> of July 2020<sup>151</sup> illustrating an estimate of SARS-CoV-2 spread worldwide based on open source data.

Other approaches aim to incorporate open-source user data mainly from social media platforms, e.g. Twitter and Facebook<sup>130-132,153</sup>.

However, such trivial data sources usually entail tremendous background noise and lack any sort of standardization in the first instance. See **Figure 12**<sup>153</sup> for a graphical plot of information content as opposed to data volume of selected surveillance systems<sup>153</sup>.



**Figure 12.** Graphical plot of information content vs. data volume of selected surveillance systems<sup>153</sup>.

#### *Digital Surveillance within the Framework of Legislation*

As a side note, digital surveillance, especially where user-generated data are leveraged, will not only have to overcome bias but medicolegal aspects. According to federal, national, or international guidance, regulations, or laws, health data are usually considered ‘sensitive’, thus underlying protection<sup>154</sup>. In the US, for example, such data need to be handled under the Health Insurance Portability and Accountability Act of (HIPAA) Privacy Rule that will guarantee data be non-identifiable and any activity on a HIPAA cleared device or software will be logged and is traceable to the individual data entryist etc<sup>124</sup>. In Europe, health data need to be handled in compliance with the General Data Protection Regulation (GDPR), currently leaving part of the legal basis of current surveillance and infection control measures *in limbo*:

Regulation 2016/679 of the European Parliament and of the Council of 27<sup>th</sup> of April 2016 on the protection of natural persons regarding the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) say

- ♦ (52) “Derogating from the prohibition on processing special categories of personal data should also be allowed [...] where it is in the public interest to do so, in particular [...] for health security, monitoring and alert purposes, the prevention or control of communicable diseases and other serious threats to health.”<sup>155</sup> (Text with EEA relevance)

Revision of the current law will need to clarify what “the public interest”<sup>155</sup> might be, against the backdrop of an ongoing COVID-19 pandemic<sup>156</sup>. An increasing number of European Union member states are implementing digital surveillance tools including wearables and contact tracing apps to enhance COVID-19

infection control measures<sup>157,158</sup>. A review by Pepin et al. has shown that in several European countries, including France, Germany, Spain, Sweden, the Netherlands, and Italy, use of tracing software on a wearable allows to control whether individuals abide by home quarantine policy, for example<sup>158</sup>.

A Taiwanese research group developed a wearable device for the early detection of SARS-CoV-2 infection through the continuous measurement of body surface temperatures and heart rates<sup>159</sup>. In Israel, the Israel Security Agency (ISA) Shin Bet (aka "The unseen shield") is in control of COVID-19 epidemiological investigation by use of personal tracing of citizens<sup>160</sup>, bringing human rights defenders into the arena.

Within the European Union, Juan Fernando López Aguilar, chair of Parliament's Committee on Civil Liberties, Justice and Home Affairs pledged "we'll keep a close eye that EU law principles and rules are respected throughout the fight against Covid-19. That includes apps and technologies to control the spread patterns of the pandemics."<sup>161</sup>.

### **2.2.1 Agnostic Identification of (Re-)Emerging Viruses**

Besides efforts undertaken to diminish bias in (digital) syndromic surveillance by using standardized terminologies and case definitions in compliance with official regulations and laws, unbiased pathogen detection methods are warranted. In patients presenting with signs and symptoms of (S)ARI and ILI, the diagnostic tests ordered by physicians in routine care will usually include bacterial swabs and virological testing for influenza viruses, particularly during flu season. Depending on age, immune status of the patient, or outbreak situations, spontaneous physician testing requests might also include respiratory syncytial virus (RSV), human rhinovirus (HRV), human adenovirus (HAdV), parainfluenza virus (HPIV) and/or human coronaviruses (HCoV), specifically SARS-CoV-2 during the COVID-19 pandemic.

#### *Identifying Specific Virus Infection based on Clinical Presentations*

Differentiating viral pathogens based on clinical presentation may be challenging as has previously been shown by the Vienna Vaccine Safety Initiative in a systematic literature review and meta-analysis of 47 published clinical studies (PROSPERO registration number: CRD42017059557)<sup>36</sup>. During the COVID-19 pandemic, Shah et al. and the University of California, San Francisco (UCSF) COVID-19 Hospital Translational and Clinical Epidemiology Working Group investigated distinctive clinical features of COVID-19 patients in



comparison to the clinical presentation of patients with other viral respiratory infections. For this purpose, the authors performed retrospective chart review of adult (S)ARI and ILI patients at an academic medical center. In this study, COVID-19 patients turned out to have “a longer duration of symptoms, [they] were more likely to be admitted to the hospital, [they] had longer hospitalizations and were more likely to develop ARDS [acute respiratory distress syndrome]”<sup>162</sup>.

Of note, medical patient records used for retrospective chart review have repeatedly been shown to be inconclusive in many instances due to lack of pertinent information<sup>22,30,36,37,45</sup>, thus limiting the reliability of results and conclusions. Also, hospitalization rates of COVID-19 patients, length of hospital stay, and admission to the intensive care unit due to ARDS, for example, may be confounded by observer or ascertainment bias, as well as the availability of hospital beds in different phases of a “wave” or outbreak. The authors of the aforementioned study acknowledged that decision-making on hospitalization, in particular, may involve other aspects than disease severity only, i.e. “concern for late decompensation”<sup>163</sup> or “difficulties with hospital discharge due to requirements for isolation and infection control”<sup>70,163</sup>.

Difficulty inherent with retrospective chart review became evident with another retrospective study assessing a diagnostic model for COVID-19 based on radiological and clinical features across multiple sites in China, where authors left open the question of how they handled missing information in patient records. By all means, they concluded that clinical characteristics identified in their multicenter study did not differ between COVID-19 patients and non-COVID-19 (S)ARI/ ILI patients<sup>164</sup>.

As a side note, the distinction of the terms ILI and CLI (“COVID-19-like illness”)<sup>111</sup> as used by CDC appears particularly challenging against the backdrop of current scientific evidence.

The 2020/2021 ‘hot topic’ example of COVID-19 underlines that the identification of specific viruses based on clinical symptoms alone is, for now, hardly possible. Thorough diagnostic work-up is a key piece in the puzzle to identify the most probable cause of an infectious disease<sup>36</sup>. Unless laboratory-confirmed diagnoses take priority over subjective ‘clinical diagnoses’, diagnostic accuracy will have to be considered low, thus impairing clinical care (e.g. prescription of antibiotics) and blurring infectious disease surveillance. A comparative study by Alchikh et al. showed that a large proportion of viral (S)ARI and ILI cases remained unresolved in routine care while a virus was detected by independent systematic PCR testing in the very

same cases in the context of a conjunct prospective digital surveillance program conducted by the Vienna Vaccine Safety Initiative. The study also reinforced the importance of applying case criteria for case classification to enhance diagnostic accuracy. Especially revised WHO ILI criteria turned out to provide a balanced level of both, high sensitivity and specificity<sup>37</sup>.

#### *Accounting for the Unpredictable and Unknown*

Aiding physicians in the quest for the cause of a (S)ARI/ ILI case by broadening the panel of pathogens tested at once in a cost-efficient manner, multiplex assays have been developed. According to the manufacturer, the assay will yield a result for a total of 23 respiratory pathogens, covering both viruses and bacteria, within 45 minutes<sup>165</sup>. Overall, several validation studies attested good accuracy and results being actionable in some instances with this specific multiplex panel<sup>166-168</sup>, also for SARS-CoV-2<sup>169</sup>. However, limitations include the fact that variant strains or novel viruses, e.g. a notional ‘pathogen X’, may not be captured by pathogen-specific testing methodologies.

Accounting for this major limitation, agnostic viral mNGS has been leveraged not only for CNS infectious/ inflammatory diseases but also for (S)ARI and ILI in recent years. In those studies, authors conclude that mNGS has the potential to revolutionize infectious disease diagnostics due to its agnostic, i.e. sequence-independent nature<sup>170-190</sup>.

The real-world usefulness of mNGS in unresolved cases of (S)ARI and ILI was revealed in January 2020, when some of the first cases of unresolved pneumonia in China yielded positive results for SARS-CoV-2 by mNGS<sup>191</sup>.

The technique of mNGS again showed to be useful in the case of a 56-year-old female traveler from Wuhan, China, who presented with fever and fatigue to a hospital in Shanghai, the most populated metropolis worldwide. Using a multiplex panel, this patient tested all negative but by exploratory mNGS and orthogonal PCR, SARS-CoV-2 was detected rapidly<sup>192</sup>. The authors thus deemed mNGS to hold “great promises as a unique diagnostic, especially for patients with acute or severe infections”<sup>192</sup>.

Another example of mNGS real-world application during the COVID-19 pandemic is reported in a preprint from Cambodia where the COVID-19 index case was analyzed by mNGS. Despite Cambodia being resource-limited, the Cambodian authors endorse ongoing application of mNGS and advocate for expansion of mNGS

infrastructure in Cambodia and beyond<sup>193</sup>, serving as a shield for the early containment of COVID-19 and a notional 'pathogen X'.

Despite this great track record in developing as well as developed countries and several years of experience with mNGS application, this novel diagnostic approach has rarely been used for surveillance purposes. In June 2020, Le et al. on behalf of the Oxford University Clinical Research Unit (OUCRU) COVID-19 research group sent a letter to the editor of *Journal of Infection*, admonishing the fact that so far, only 3 publications "reporting the application of metagenomics to detect SARS-CoV-2 and co-infections, with a combined sample size of nine patients"<sup>194</sup> exist. This fact seemed highly deplorable given the opportunities inherent with agnostic mNGS for early and unbiased detection of (re-)emerging pathogens<sup>194</sup>.

Ultimately, evidence is growing that use of standardized criteria and shared terminology may be integrated in eHealth tools taken to the patient's bedside for the real-time capturing and output of critical data. eHealth in combination with agnostic pathogen testing (such as mNGS) will facilitate the unbiased detection of (re-)emerging infectious diseases.

In **Study 2 (Manuscript B)** of the PhD thesis, we provided a proof-of-principle investigating whether mNGS would help to increase diagnostic yield if applied in addition to PCR testing in pediatric ILI cases with clinical complications. This question was addressed within a quality improvement and digital surveillance program allowing for the standardized, combined assessment of clinical presentations and severity measurements.

### **2.2.2 Manuscript B**

Obermeier PE, Alchikh M, Ma X, Reiche J, Schweiger B, Rath B. **Leveraging digital surveillance to study emerging pathogens in influenza-like illness in children.** J Clin Virol Plus (submitted April 2021).

Leveraging digital surveillance to study emerging pathogens in influenza-like illness in children

Patrick E Obermeier<sup>1,2,3</sup>, Maren Alchikh<sup>1,2,3</sup>, Xiaolin Ma<sup>4</sup>, Janine Reiche<sup>4</sup>,

Brunhilde Schweiger<sup>4</sup>, Barbara A Rath<sup>1,3\*</sup>

1 Vienna Vaccine Safety Initiative, Berlin, Germany.

2 Charité University Medical Center, Berlin, Germany

3 UMR Chrono-Environnement, Université Bourgogne Franche-Comté, Besançon, France

4 National Reference Center for Influenza, Robert Koch-Institut, Berlin, Germany

\*Corresponding author: Barbara Rath, MD PhD, Vienna Vaccine Safety Initiative, 10437 Berlin, Germany, F: +49-30-69088616, Barbara.rath@vi-vi.org

Total word count: 1244/1250 words

Highlights (4/5)

- Health authorities recommend digital surveillance systems for infectious diseases (84/85)
- This study combines digital syndromic surveillance with shotgun virus detection (82/85)
- Mobile health tools enable standardized clinical assessments and severity scoring (83/85)
- Linkage of 2 agnostic methodologies increases diagnostic yield and diminishes bias (85/85)

**Abstract (250/250 words)**

**Background:** Public health authorities recommend the use of digital health technology to augment the monitoring of the impact of respiratory infectious diseases.

**Objectives:** To investigate a nested study of complicated influenza-like illness (ILI) combining digital technology for point-of-care clinical severity scoring with specific PCR testing and additional agnostic virus detection using metagenomics.

**Study Design:** In a hospital-based digital surveillance program, 6,073 pediatric patients fulfilling pre-defined ILI criteria and/or holding a physician ILI diagnosis (mean age 3.1 years, range 0-18.8 years, 56% male) were assessed using the ViVI Disease Severity ScoreApp from 12/2009-04/2015. Simultaneous nasopharyngeal samples were collected for PCR testing for influenza virus A/B, rhino-, metapneumo-, boca-, adeno-, parainfluenza, betacoronavirus, and respiratory syncytial virus. For the nested study, we selected a subset of 100/6,073 ILI patients presenting with a body temperature  $<39^{\circ}\text{C}$  and seizure (mean age 3.95 years, range 0-17.84 years, 59% male). Sample aliquots underwent viral metagenomics in addition to PCR testing.

**Results:** Disease severity was higher in the subset of ILI patients presenting with afebrile/low-febrile seizures compared to the total cohort average (mean ViVI Disease Severity Score 17.1 vs. 14.5,  $p<0.0001$ ). By PCR, eight different virus species were detected in the subset of 100 ILI patients with neurological complication. Metagenomics identified five additional virus species: herpesvirus-4 and -6, picobirnavirus, enterovirus-A and -D.

**Conclusions:** Standardized disease severity scoring allows for the unbiased comparison of detailed clinical presentations across cohorts with the potential to increase diagnostic yield and identify pathogens and clinical risk. Metagenomics may be used to complement PCR.

**Keywords:** Surveillance, influenza-like illness, disease severity, metagenomics, PCR, coronavirus, precision medicine

**Non-standard Abbreviations**

ViVI - Vienna Vaccine Safety Initiative

mNGS - metagenomic next-generation sequencing

CONFIDENTIAL

**Background**

Viral respiratory infections impose a substantial health threat. Children are among the main transmitters (1-4). To better understand the impact and spread of respiratory viral infections, WHO recommends the standardized “monitoring of the prevalence, severity and management of respiratory diseases” (1) and to “improve early diagnosis [...] through [...] the development of new tools” (1).

**Objectives**

This study uses a nested approach to explore the timely identification and characterization of a rare but significant specific disease presentation: afebrile/low-febrile seizures in patients with influenza-like illness (ILI). Within a well-described prospective surveillance cohort for pediatric ILI patients, we combined the use of innovative mobile health (m-Health) technology for clinical disease severity assessments with PCR and viral metagenomic next-generation sequencing (mNGS).



## Methods

### Study design

The analysis was conducted in the context of a previously described prospective surveillance and quality improvement program for pediatric ILI patients at the Charité Department of Pediatrics in collaboration with the National Reference Center for Influenza at the Robert Koch-Institute in Berlin, Germany. The program was approved by the Charité IRB (EA4/008/10). Informed consent procedures were waived for the purpose of infection control (5-11).

The cohort comprises 6,073 pediatric patients fulfilling pre-defined ILI criteria and/or holding a physician ILI diagnosis from 12/2009-04/2015 (mean age 3.1 years, range 0-18.8 years, 56% male) (5-8). All patients underwent standardized clinical assessments. Patients were assessed using the ViVI (Vienna Vaccine Safety Initiative) ScoreApp for real-time computation of the ViVI Disease Severity Score (ViVI ScoreApp; <https://score.vi-vi.org>), a published and validated 22-item weighed clinical score ranging from 0-48, with increasing values reflecting increasing disease severity (7).

### Sample collection and PCR testing

From all patients, respiratory specimens were collected for PCR testing for influenza virus A/B, human rhinovirus (HRV), metapneumovirus (HMPV), bocavirus (HBoV-1), adenovirus (HAdV), parainfluenza virus (HPIV1-4), betacoronaviruses (HCoV-NL63, HCoV-229E, HCoV-HKU1, HCoV-OC43), and respiratory syncytial virus (RSV) at the Robert Koch-Institute (8).

### Viral mNGS in 100 selected patient samples

We identified a subset of 100 ILI patients experiencing a seizure in the presence of no/low-grade fever, i.e. body temperature  $<39^{\circ}\text{C}$ . Nucleic acids were extracted from these patient samples using the Qiagen Viral RNA mini kit following manufacturer's instructions. Viral cDNA synthesis was performed separately on each sample. Resulting cDNAs from groups of ten samples were pooled, processed using the Nextera XT DNA sample preparation kit, and barcoded using distinct pairs of index barcodes per

pool. Sequencing was performed on the Illumina HiSeq4000 instrument (150 paired-end sequencing). Reads were debarcoded and bacterial/human reads subtracted. Cleaned reads were assembled using Ensemble Assembler.

#### Statistical analysis

Descriptive statistics were performed with SPSS 27.0. Statistical significance was assessed using two-sided t-test,  $p < 0.05$  was considered statistically significant.

CONFIDENTIAL

## Results

Among all patients (n=6,073), 632 (10.4%) experienced a seizure (mean age 2.69 years, range 0-17.84 years, 56% male). Of those, we identified a subset of 100 patients (15.8%) with afebrile/low-febrile seizure (mean age 3.95 years, range 0-17.84 years, 59% male), see Figure 1.

The average ViVI Score in the overall cohort (n=6,073) was significantly lower (14.52; range 0-34, 95%CI [14.34;14.64]) compared to the group of 632 ILI patients with a seizure (mean ViVI Score 17.51, range 5-34, 95%CI [17.12;17.9], >67th percentile of the cohort, p<0.0001) and the subset of 100 ILI patients with *afebrile/low-febrile* seizure (mean ViVI Score 17.11, range 5-30, 95%CI [16.18;18.04], >67<sup>th</sup> percentile of the cohort, p<0.0001).

Among the subset of 100 ILI patients with afebrile/low-febrile seizure, PCR yielded positive results in 65/100 cases for eight viral species, the majority of which were HBoV and HRV (24%, respectively), followed by influenza virus A/B (10%), HAdV (8%), RSV (7%), HCoV (6%), HPIV (5%), and HMPV (4%). Dual/multiple infections were detected in 23/100 cases, most frequently involving HBoV and HRV. Viral mNGS would have missed at least 47/100 cases where the PCR assays used herein had yielded a positive virus detection result, involving both high and low viral titer samples. However, viral mNGS extended diagnostic yield by five additional viral species: human herpesvirus-4 and -6, picobirnavirus, enterovirus-A and -D (Table 1).

## Discussion

We report a proof-of-concept study using a nested approach combining the use of an m-Health tool enabling standardized clinical assessments of disease severity with agnostic virus detection in addition to virus-specific PCR. The study was embedded in a digital surveillance program comprising a cohort of 6,073 pediatric ILI patients.

Specifically, we investigated ILI patients experiencing neurological complication, i.e. seizure, because disease burden is usually high in this patient group (12, 13) as demonstrated by significantly higher VIVI Disease Severity Scores in ILI patients experiencing a seizure compared to the overall cohort average in our study. Particularly, we chose to focus on ILI patients with *afebrile/low-febrile* seizure since in these cases the cause of the convulsion is less clear and the quest for a specific cause, such as viral infection, is warranted. In this subgroup of patients, the direct role of infection might outweigh the role of increased neuron excitability due to mere elevation/increase of the body temperature (14).

In patients with ILI, extended virus diagnostics are rarely part of routine care leaving the cause unresolved unless further investigations are performed, i.e. virus-specific PCR, ideally complemented by an agnostic diagnostic method.

In this study, PCR testing for eight viruses failed to establish a link between symptoms and viral infection in 35% of cases. Agnostic mNGS has the potential to increase diagnostic yield. In our study however, mNGS alone did not achieve the level of sensitivity of specific PCR testing with at least 47% NGS-negative samples compared with PCR results. On the other hand, additional viruses were identified, some of which had been associated with neurological complications elsewhere (15, 16). In this study, HHV-4 and HHV-6B were detected by mNGS, both of which might not be tested regularly in ILI patients (13, 17). Expanding the view for potential viruses might improve isolation measures and prevent erroneous use of antibiotics (13).

Another potential application for mNGS is outbreak investigation, as mNGS is bypassing selection bias. In this study, enterovirus-D68 was detected in the summer/fall of 2014, matching circulation of enterovirus-D68 elsewhere in Germany (18). At that time, enterovirus-D68 outbreaks were reported in North America (19) and a first case of ILI with severe neurological complication had been recognized in Europe (15). For evaluation of potentially meaningful infectious disease signals, unbiased early detection combined with in-depth clinical information is essential, as seen also during the onset of the COVID-19 pandemic (2).

Causality needs to be evaluated critically, on a case-by-case basis. The present study provides a proof-of-principle for standardized patient assessments using a validated m-Health tool allowing for correlations between disease severity and viral infection. Large-scale implementation of this approach is underway with the PEDSIDEA (Partnering for Enhanced Digital Surveillance of Influenza-like Disease and the Effect of Antivirals & Vaccines) project (20).

This study is limited by the small number of patients, pooling of samples, and restriction to respiratory samples when focusing on neurological complications. For cost reasons, mNGS testing could not be performed for reference groups, e.g. high-fever ILI patients experiencing a seizure.

Technical validity of mNGS results is limited and influenced by preanalytical sample handling, nucleic acid enrichment, and bioinformatics algorithms (2, 21). The diagnostic work-up of unresolved cases may also require bacteriological and other investigations.

**Conclusions**

Digital health technologies, such as the ViVI ScoreApp, provide immediate, standardized disease severity assessments at the point-of-care. Combined use with agnostic virus detection methodologies can help to increase diagnostic yield, diminish bias, and detect unexpected or novel viruses. Currently, mNGS may be used in conjunction with virus-specific PCR testing.

CONFIDENTIAL

**Authors' contributions**

Conceptualization and supervision of the work: BR, BS. Development of methodology: BR, BS. Conducting formal analysis: PO, MA, XM, JR, BR. Provision of study materials, reagents, materials, patients, laboratory samples, computing resources, and other analysis tools: BR, BS, JR. Data curation: PO, BR, BS. Writing of the original draft: PO. Review and editing of the manuscript draft: all authors. Project administration: BR. Funding (in-kind): BR, BS.

**Funding**

Funding was provided in-kind.

**Declaration of Competing Interest**

The authors report no declarations of interest.

**Acknowledgements**

The authors thank the patients for participation.

Table 1

Table 1. Patient baseline characteristics, viral metagenomic next-generation sequencing (mNGS) and PCR test results

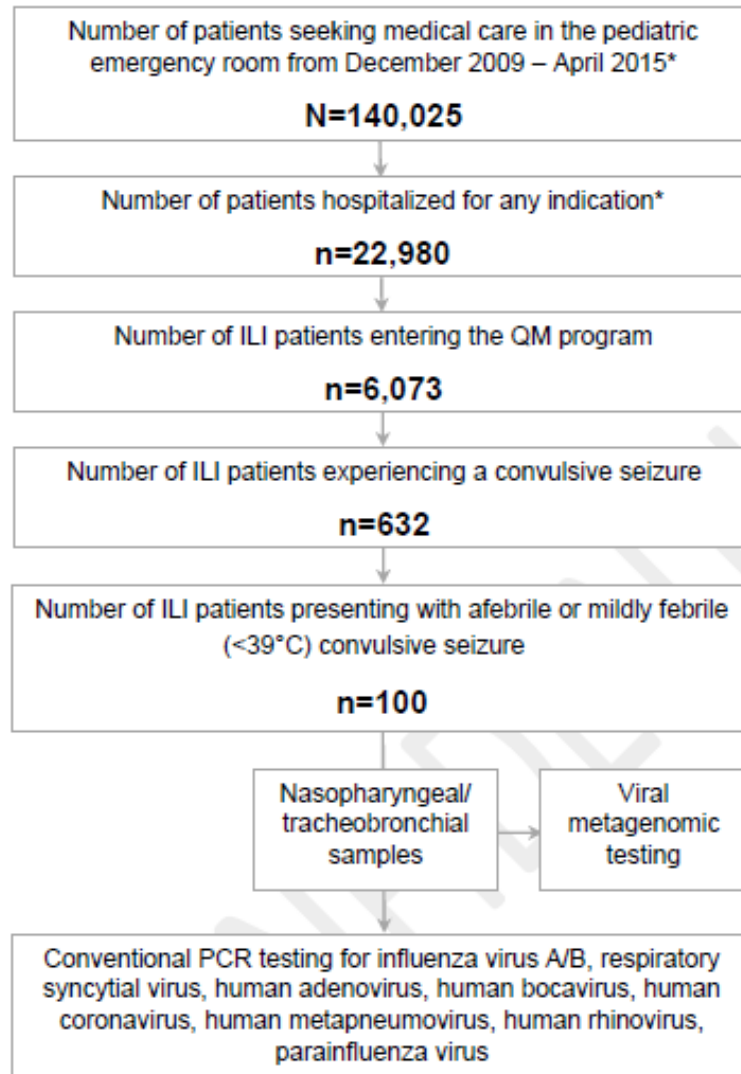
Pool	Sample	Collection date	Age in years	Gender	Disease Severity Score	PCR-Result (CT value)	mNGS Result
1	a	Dec/2009	2	male	13	HAdV (16), HBoV (37)	HHV-6B,
	b	Jan/2010	9	female	13	-	HAdV-A,
	c	Mar/2010	3	male	11	-	
	d	Jan/2011	1	male	11	-	HAdV-B
	e	Feb/2011	2	female	25	RSV (28)	
	f	Mar/2011	8	male	9	HBoV (38)	
	g	Apr/2011	1	male	22	HBoV (27)	
	h	Apr/2011	2	female	15	HRV (35), HBoV (35)	
	i	Jun/2011	0	female	7	HBoV (36)	
	j	Jul/2011	5	male	16	-	
2	a	Nov/2011	1	male	15	HRV (34)	HBoV-1,
	b	Nov/2011	1	female	21	HRV (35), HBoV (17)	EV-A,
	c	Nov/2011	1	female	15	HBoV (38), PIV (35)	HRV-C,
	d	Feb/2012	2	female	21	RSV (24)	RSV
	e	Mar/2012	2	male	23	HBoV (38)	
	f	Apr/2012	3	female	16	-	
	g	May/2012	1	female	18	HRV (38)	
	h	May/2012	1	male	26	-	
	i	Oct/2012	0	male	19	HBoV (37)	
	j	Oct/2012	2	male	15	HRV (34)	
3	a	Nov/2012	7	male	16	-	HAdV-C,
	b	Dec/2012	12	female	16	HRV (33)	
	c	Dec/2012	1	male	16	HAdV (39), HPIV (28)	PBV,
	d	Jan/2013	1	female	20	HBoV (19)	RSV
	e	Jan/2013	18	female	18	-	
	f	Jan/2013	4	male	14	RSV (29), HBoV (37)	
	g	Jan/2013	1	female	17	HAdV (22), HBoV (35)	
	h	Jan/2013	13	female	26	RSV (23)	
	i	Feb/2013	3	female	20	Flu A H1N1 (30)	
	j	Feb/2013	1	male	20	RSV (27)	
4	a	Feb/2013	2	female	17	Flu A H3N2 (37), HBoV (35)	Influenza virus A
	b	Feb/2013	2	female	16	HRV (33)	
	c	Feb/2013	1	male	18	Flu A H3N2 (26)	
	d	Feb/2013	15	female	22	HCoV-229E (34)	
	e	Mar/2013	3	female	13	HAdV (32)	
	f	Mar/2013	16	male	12	-	
	g	Mar/2013	4	female	14	-	
	h	Mar/2013	8	male	15	Flu A H3N2 (23)	



	i	Mar/2013	10	male	15	HMPV (30), HBoV (36)	
	j	Mar/2013	1	male	19	-	
5	a	Mar/2013	14	male	22	HBoV (37)	HAdV-C, EV-A, Influenza virus B, HMPV
	b	Apr/2013	5	female	18	Flu B Vic (30), HAdV (33), HBoV (37)	
	c	Apr/2013	3	male	27	HBoV (37)	
	d	Apr/2013	4	male	13	Flu B Yam (27)	
	e	Apr/2013	5	male	14	HRV (30), HMPV (30)	
	f	Apr/2013	5	male	23	HRV (32)	
	g	May/2013	2	female	21	HMPV (27)	
	h	May/2013	1	male	17	HAdV (24)	
	i	Jun/2013	3	male	12	-	
	j	Jul/2013	10	male	30	-	
6	a	Aug/2013	5	female	14	HRV (32)	HBoV-1
	b	Sep/2013	0	male	14	HAdV (34), HRV (27), HBoV (37), HCoV-HKU1 (31)	
	c	Sep/2013	4	female	15	PIV (37)	
	d	Oct/2013	4	male	20	-	
	e	Oct/2013	14	male	13	-	
	f	Nov/2013	1	male	20	HRV (22), HBoV (37)	
	g	Nov/2013	3	male	13	HRV (38), HBoV (38)	
	h	Nov/2013	15	female	21	HRV (26)	
	i	Nov/2013	3	male	16	HRV (37)	
	j	Nov/2013	1	male	20	HRV (29), HBoV (16)	
7	a	Nov/2013	3	female	28	HRV (33), HBoV (38)	HHV-4, Influenza virus B, RSV, HCoV-NL63
	b	Dec/2013	4	male	25	HRV (30)	
	c	Jan/2014	2	female	16	HBoV (35)	
	d	Jan/2014	1	female	20	-	
	e	Jan/2014	3	female	15	-	
	f	Feb/2014	1	male	22	RSV (30)	
	q	Feb/2014	13	male	18	-	
	h	Feb/2014	8	male	5	-	
	i	Feb/2014	1	male	21	RSV (26), HBoV (36), HCoV-NL63 (23)	
	j	Mar/2014	1	male	14	-	
8	a	Mar/2014	2	female	11	HRV (32), HMPV (25)	-
	b	Mar/2014	1	female	24	RSV (36)	
	c	Mar/2014	8	female	15	-	
	d	Apr/2014	2	male	10	HCoV-OC43 (34)	
	e	Apr/2014	2	male	16	-	
	f	Apr/2014	16	female	24	-	
	g	Apr/2014	5	female	21	-	
	h	Apr/2014	6	female	17	-	

	i	May/2014	1	male	16	-	
	j	Jun/2014	4	male	17	-	
9	a	Jun/2014	1	female	19	HRV (35)	EV-A, EV-D68, HCoV-OC43
	b	Jul/2014	1	female	14	HAdV (33)	
	c	Aug/2014	0	female	13	-	
	d	Aug/2014	1	male	21	HAdV (29), HRV (30)	
	e	Oct/2014	3	female	13	HPIV (25), HCoV-OC43 (24)	
	f	Oct/2014	0	male	13	-	
	g	Nov/2014	0	male	17	HRV (40), HBoV (37), HPIV (20)	
	h	Nov/2014	8	male	17	HRV (29)	
	i	Dec/2014	9	male	16	Flu A H3N2 (27), HPIV (26)	
	j	Dec/2014	5	female	16	-	
10	a	Dec/2014	5	male	23	-	Influenza virus A, HPIV-3, HCoV-NL63
	b	Jan/2015	3	male	16	Flu A H3N2 (20)	
	c	Feb/2015	4	male	19	-	
	d	Feb/2015	0	male	7	-	
	e	Feb/2015	0	female	11	-	
	f	Feb/2015	0	male	16	HBoV (37)	
	g	Mar/2015	13	female	14	Flu B Yam (36)	
	h	Mar/2015	1	female	22	Flu A H3N2 (30), HRV (31)	
	i	Mar/2015	1	male	12	HCoV-NL63 (29)	
	j	Mar/2015	1	male	19	-	

Figure 1. Flowchart



\*numbers extrapolated based on a 3-year average

## References

1. Forum of International Respiratory Societies. 2017. The Global Impact of Respiratory Disease - Second Edition. Sheffield, European Respiratory Society.
2. Zhang N, Wang L, Deng X, Liang R, Su M, He C, Hu L, Su Y, Ren J, Yu F, Du L, Jiang S. 2020. Recent advances in the detection of respiratory virus infection in humans. *J Med Virol* 92:408-417.
3. Lee BY, Shah M. 2012. Prevention of influenza in healthy children. *Expert Rev Anti Infect Ther* 10:1139-52.
4. Sun D, Zhu F, Wang C, Wu J, Liu J, Chen X, Liu Z, Wu Z, Lu X, Ma J, Peng H, Xiao H. 2020. Children Infected With SARS-CoV-2 From Family Clusters. *Front Pediatr* 8:386.
5. Tief F, Hoppe C, Seeber L, Obermeier P, Chen X, Karsch K, Muehlhans S, Adamou E, Conrad T, Beresniak A, Schweiger B, Adam T, Rath B. 2016. An inception cohort study assessing the role of pneumococcal and other bacterial pathogens in children with influenza and ILI and a clinical decision model for stringent antibiotic use. *Antivir Ther* 21:413-24.
6. Chen X, Pouran Yousef K, Duwe S, Karsch K, Grover S, Wahlisch S, Obermeier P, Tief F, Muehlhans S, Seeber L, von Kleist M, Schweiger B, Rath B. 2014. Quantitative influenza follow-up testing (QIFT)—a novel biomarker for the monitoring of disease activity at the point-of-care. *PLoS One* 9:e92500.
7. Rath B, Conrad T, Myles P, Alchikh M, Ma X, Hoppe C, Tief F, Chen X, Obermeier P, Kisler B, Schweiger B. 2017. Influenza and other respiratory viruses: standardizing disease severity in surveillance and clinical trials. *Expert Rev Anti Infect Ther* 15:545-568.
8. Ma X, Conrad T, Alchikh M, Reiche J, Schweiger B, Rath B. 2018. Can we distinguish respiratory viral infections based on clinical features? A prospective pediatric cohort compared to systematic literature review. *Rev Med Virol* 28:e1997.
9. Rath B, Chen X, Spies V, Muehlhans S, Obermeier P, Tief F, Seeber L, Karsch K, Milde J, Skopnik H, Schweiger B, Duwe SC. 2017. Prospective surveillance of antiviral resistance in hospitalized infants less than 12 months of age with A(H3N2) influenza infection and treated with oseltamivir. *Antivir Ther* 22:515-522.
10. Rath B, Tief F, Obermeier P, Tuerk E, Karsch K, Muehlhans S, Adamou E, Duwe S, Schweiger B. 2012. Early detection of influenza A and B infection in infants and children using conventional and fluorescence-based rapid testing. *J Clin Virol* 55:329-33.
11. Tuttle R, Weick A, Schwarz WS, Chen X, Obermeier P, Seeber L, Tief F, Muehlhans S, Karsch K, Peiser C, Duwe S, Schweiger B, Rath B. 2015. Evaluation of novel second-generation RSV and influenza rapid tests at the point of care. *Diagn Microbiol Infect Dis* 81:171-6.
12. Jones T, Jacobsen SJ. 2007. Childhood febrile seizures: overview and implications. *Int J Med Sci* 4:110-4.
13. Carman KB, Calik M, Karal Y, Isikay S, Kocak O, Ozcelik A, Yazar AS, Nuhoglu C, Sag C, Kilic O, Dinleyici M, Lacinel Gurlevik S, Yimenicioglu S, Ekici A, Perk P, Tosun A, Isik I, Yazar C, Arslantas D, Dinleyici EC. 2019. Viral etiological causes of febrile seizures for respiratory pathogens (EFES Study). *Hum Vaccin Immunother* 15:496-502.
14. Mosili P, Maikoo S, Mabandla MV, Qulu L. 2020. The Pathogenesis of Fever-Induced Febrile Seizures and Its Current State. *Neurosci Insights* 15:2633105520956973.
15. Lang M, Mirand A, Savy N, Henquell C, Maridet S, Perignon R, Labbe A, Peigue-Lafeuille H. 2014. Acute flaccid paralysis following enterovirus D68 associated pneumonia, France, 2014. *Euro Surveill* 19.
16. Mohammadpour Touserani F, Gainza-Lein M, Jafarpour S, Brinegar K, Kapur K, Loddenkemper T. 2017. HHV-6 and seizure: A systematic review and meta-analysis. *J Med Virol* 89:161-169.
17. Bartolini L, Piras E, Sullivan K, Gillen S, Bumbut A, Lin CM, Leibovitch EC, Graves JS, Waubant EL, Chamberlain JM, Gaillard WD, Jacobson S. 2018. Detection of HHV-6 and EBV and

Cytokine Levels in Saliva From Children With Seizures: Results of a Multi-Center Cross-Sectional Study. *Front Neurol* 9:834.

18. Reiche J, Bottcher S, Diedrich S, Buchholz U, Buda S, Haas W, Schweiger B, Wolff T. 2015. Low-level Circulation of Enterovirus D68-Associated Acute Respiratory Infections, Germany, 2014. *Emerg Infect Dis* 21:837-41.
19. Messacar K, Abzug MJ, Dominguez SR. 2016. 2014 outbreak of enterovirus D68 in North America. *J Med Virol* 88:739-45.
20. Rath B. 2018. Vienna Vaccine Safety Initiative. *Hum Vaccin Immunother* 14:1038-1041.
21. Deng X, Naccache SN, Ng T, Federman S, Li L, Chiu CY, Delwart EL. 2015. An ensemble strategy that significantly improves de novo assembly of microbial genomes from metagenomic next-generation sequencing data. *Nucleic Acids Res* 43:e46.

CONFIDENTIAL

### 3 MINING THE BIG PICTURE

---

#### 3.1 Whole-Genome Analysis – Leveraging Big Data in Virology

Big data in virology have become eminently important for diagnostics, research, epidemiology, and public health with the advent of (m)NGS. While first-generation sequencing, mainly Sanger sequencing, of longer genomic regions was cumbersome, (m)NGS has the potential to swiftly provide gigabases of genomic information in unprecedented depth in relatively short time and at reasonable costs<sup>195,196</sup>.

During the COVID-19 pandemic for example, NGS and whole-genome sequence analysis have been leveraged for SARS-CoV-2 origin tracing, tracing of transmission chains, investigation of SARS-CoV-2-susceptible organs, analysis of changes in the host's microbiome following COVID-19, vaccine development or the investigation of potential therapeutics<sup>196</sup>. The door is open to a plethora of other opportunities of mining full viral genomes, including in-depth studies of virus quasispecies or drug resistance genes<sup>195-197</sup>.

One advantage of agnostic (m)NGS is its ability to amplify and detect any genomic information without necessarily knowing it *a priori*. This basic property has proven key for the discovery of novel pathogens, including one of the most prevalent group of viruses in acute respiratory infections across different age groups: human adenoviruses. As of September 2021, a total of 104 different HAdV genotypes belonging to one of 7 species (A-G) have been assigned by the Human Adenovirus Working Group, recently based on the “genomics-based paradigm and algorithm”<sup>198</sup> particularly focusing on whole genome, penton base, hexon loop 1, hexon constant, and fiber knob sequences<sup>198</sup>. Under this current scientific consensus, whole-genome sequencing has almost become a *conditio sine qua non* for sound HAdV typing<sup>199</sup>. Indeed, some of the most recent novel HAdV genotypes have been discovered by mNGS analysis of clinical samples, including the most recent genotype HAdV-C104 which turned out to be a recombinant strain harboring penton base and hexon gene sequences clustering with HAdV-1 and a fiber gene sequence clustering with HAdV-2 but segregating itself in phylogenetic analysis of the whole genome sequence<sup>200</sup>. Recombination has been well-known to be an important factor contributing to diversity of adenovirus species D but for species C this has been poorly studied so far<sup>201</sup>.

For this reason, we investigated molecular evolution of 51 HAdV species C full-genome sequences obtained by NGS in **Study 3 (Manuscript C)** of the PhD thesis, thereby accomplishing re-assignment of two strains formerly labeled HAdV-C2 to novel genotype HAdV-C89<sup>201</sup>.

### 3.2 Manuscript C

Dhingra A, Hage E, Ganzenmueller T, Böttcher S, Hofmann J, Hamprecht K, Obermeier P, Rath B, Hausmann F, Dobner T, Heim A.. **Molecular Evolution of Human Adenovirus (HAdV) Species C**. Sci Rep 2019; 9:1039.



# SCIENTIFIC REPORTS

OPEN

## Molecular Evolution of Human Adenovirus (HAdV) Species C

Akshay Dhingra<sup>1,6</sup>, Elias Hage<sup>1,6</sup>, Tina Ganzenmueller<sup>1,4,6</sup>, Sindy Böttcher<sup>2</sup>, Jörg Hofmann<sup>3</sup>, Klaus Hamprecht<sup>4</sup>, Patrick Obermeier<sup>2,7,8</sup>, Barbara Rath<sup>7,8</sup>, Fabian Hausmann<sup>5</sup>, Thomas Dobner<sup>5</sup> & Albert Heim<sup>1,6</sup>

Received: 24 September 2018

Accepted: 23 November 2018

Published online: 31 January 2019

Currently, 88 different Human Adenovirus (HAdV) types are grouped into seven HAdV species A to G. Most types (57) belong to species HAdV-D. Recombination between capsid genes (hexon, penton and fiber) is the main factor contributing to the diversity in species HAdV-D. Noteworthy, species HAdV-C contains so far only five types, although species HAdV-C is highly prevalent and clinically significant in immunosuppressed patients. Therefore, the evolution of species HAdV-C was studied by generating 51 complete genome sequences from circulating strains. Clustering of the whole genome HAdV-C sequences confirmed classical typing results (fifteen HAdV-C1, thirty HAdV-C2, four HAdV-C5, two HAdV-C6). However, two HAdV-C2 strains had a novel penton base sequence and thus were re-labeled as the novel type HAdV-C89. Fiber and early gene region 3 (E3) sequences clustered always with the corresponding prototype sequence but clustering of the E4 region indicated recombination events in 26 out of the 51 sequenced specimens. Recombination of the E1 gene region was detected in 16 circulating strains. As early gene region sequences are not considered in the type definition of HAdVs, evolution of HAdV-C remains on the subtype level. Nonetheless, recombination of the E1 and E4 gene regions may influence the virulence of HAdV-C strains.

Human adenoviruses (HAdVs) are double-stranded, non-enveloped, linear DNA viruses of 34–36 kbp length<sup>1,2</sup>. Currently, there are 88 different HAdV types known, which have been classified into seven species A to G and new adenovirus types continue to emerge<sup>1,3</sup>. Types were exclusively defined as serotypes (in cross-neutralization) up to type 51, for newer types a genotype definition was mostly used which requires either novel sequences or recombinant phylogeny in genes coding for major capsid proteins<sup>4</sup>.

The majority of HAdV types belong to species HAdV-D (57 types) followed by species HAdV-B (16 types) (<http://hadvwg.gmu.edu/>). Homologous recombination among capsid genes (hexon, penton and fibre) is imperative in contributing to the high diversity of species HAdV-D types. In contrast, rapid selection of novel capsid gene sequences is the major factor for diversity in species HAdV-B<sup>1,3,5,6</sup>, although a few recombinant HAdV types of species B have also been described<sup>7</sup>. Recombination of capsid protein genes may diversify the tissue tropism of novel HAdV types and thus enhance the pathogenicity and virulence of the new viruses<sup>3,8,9</sup>.

Although infections with species HAdV-C types are highly prevalent, evolution of species HAdV-C did not result in multiple novel types, neither by recombination between the main capsid genes nor by rapid selection of e.g. immune escape mutations in the highly variable immunogenic loops of hexon. Despite species HAdV-C comprises only 5 types so far (1, 2, 5, 6 and 57), these are clinically more significant than species HAdV-B and -D in causing severe manifestations in immunocompromised patients, in particular in allogeneic hematopoietic stem cell transplant (HSCT) recipients<sup>10–12</sup>. More than half of adenovirus infections in immunocompromised hosts are associated with species C type 1 and 2<sup>11–14</sup>. HAdV also accounts for 15% of upper respiratory tract infections and 5% of lower respiratory tract infections in adult and pediatric immunocompetent patients<sup>2,15</sup>. Species HAdV-C has been detected in the majority of respiratory tract infections followed by species B<sup>16,17</sup>. After primary infection, HAdV-C DNA can persist in a latent state in lymphoid cells and asymptomatic, intermittent shedding

<sup>1</sup>Hannover Medical School, Institute of Virology, Hannover, Germany. <sup>2</sup>Robert Koch Institut, FG 15, Nationales Referenzzentrum für Poliomyelitis und Enteroviren, Berlin, Germany. <sup>3</sup>Institute of Medical Virology, Helmut-Ruska-Haus, Charité Medical School, Berlin, Germany. <sup>4</sup>Institut für Medizinische Virologie und Epidemiologie der Viruserkrankungen, Universitätsklinikum Tübingen, Tübingen, Germany. <sup>5</sup>Heinrich Pette Institute, Leibniz Institute for Experimental Virology, Hamburg, Germany. <sup>6</sup>German Centre for Infection Research (DZIF), partner site Hannover-, Braunschweig, Germany. <sup>7</sup>Vienna Vaccine Safety Initiative, Berlin, Germany. <sup>8</sup>Laboratoire Chrono-environnement, Université Bourgogne Franche-Comté, Besançon, France. Correspondence and requests for materials should be addressed to A.H. (email: heim.albert@mh-hannover.de)

of infectious virus in feces can be observed for many years<sup>10,18,19</sup>. Immunosuppression frequently leads to HAdV reactivation and severe clinical manifestations, such as disseminated adenoviral disease after HSCT<sup>2,8,19–21</sup>.

The high prevalence of infections with HAdV-C types in combination with the long-term latency of HAdV-C DNA should increase the probability of superinfection with another HAdV-C type and thus promote the evolution of novel HAdV-C types by intertypic recombination. Still, only 5 types of HAdV-C have been described so far and little is known about the factors affecting pathogenicity and evolution of HAdV-C types and circulating HAdV-C strains. A recent study applying whole genomic sequencing to adenovirus specimens from immunocompromised pediatric patients confirmed the predominance of species HAdV-C strains in this population and detected an infection chain (transmission of HAdV-A31 between the included patients)<sup>22</sup>.

In the study presented here, 51 whole genome HAdV-C sequences were generated by next-generation sequencing of clinical isolates and diagnostic specimens and analyzed for recombinant phylogeny of genomes and diversity of immunogenic capsid proteins. Surprisingly, multiple recombination of early gene regions was found to be predominant in the evolution of these clinically relevant HAdV-C strains, whereas recombination of genes for major capsid proteins was almost absent and therefore only one novel type of species HAdV-C was found.

## Results

**Molecular typing by imputed serology.** All specimens were initially typed by sequencing of loops 1 and 2 of the main neutralization determinant  $\epsilon$ . This molecular typing procedure gives the same results as the classical serological typing technique, neutralization testing, and thus is called “imputed serology”. 15 (out of the 51) specimens identified as type 1, 30 specimens as type 2, 4 specimens as type 5, 2 specimens as type 6 but none as type 57 (Table 1).

Nucleic acid sequences coding for the neutralization determinant of types HAdV-C1, -C2 and -C6 were highly conserved (identity 97.85–100% in loop1 and 97.89–100% in loop2, respectively), whereas these sequences were found to be more diverse in HAdV-C5 strains. In the latter, the loop 1 sequence identity was 96.20–99.12% and the loop 2 identity was 97.16–100% at the nucleotide level, at the amino acid level 97.38%–98.44% and 97.22% to 98.76%, respectively (Fig. 1).

**Molecular phylogeny of complete genomic sequences and major capsid proteins.** Clustering of complete genomic sequences confirmed initial typing by imputed serology (Fig. 2).

In general, clustering of hexon and fiber gene sequences constantly followed clustering of complete genomic sequences, excluding intertypic recombination affecting hexon and fiber gene in the phylogeny of HAdV-C (Fig. 3).

Two strains (29C2 and 47C2) with a genomic backbone of type HAdV-C2 had a unique penton base sequence and presented as a novel cluster in the phylogenetic tree with a bootstrap value of 99% (Fig. 4). Compared to all other HAdV-C prototype sequences strains 29C2 and 47C2 had a unique substitution (A363E) and a unique deletion of P364 in the hypervariable RGD loop of the penton base. This loop binds with its RGD amino acid motif to cellular integrins. Additionally, strains 29C2 and 47C2 had a L154Q substitution in the second hypervariable loop of the penton base compared to the species HAdV-C consensus sequence. Due to their unique penton base sequence, strains 29C2 and 47C2 were re-labelled as a novel HAdV-C type 89, with strain 29C2 as the prototype, designated as Adenovirus C human/DEU/Hannover/2015/89[P89H2F2]. Several amino acid substitutions were observed in the hypervariable region 1 (HVR1) and the hypervariable RGD loop which binds to the secondary cellular receptor in comparison to the other prototype sequences of HAdV-C (Fig. 5, for nucleotide sequences please see Supplementary Fig. S1).

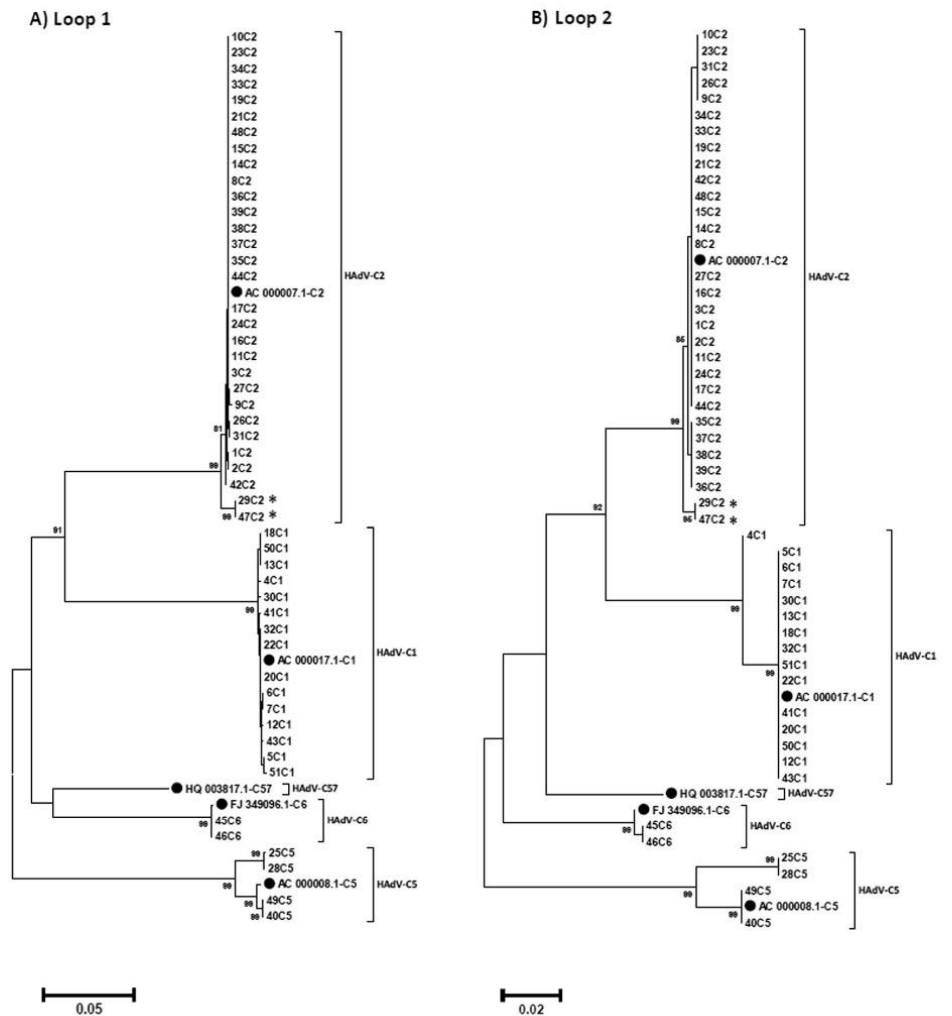
Clustering of penton gene sequences suggested recombination of the penton base gene region in 9 genomes generated from other strains (9C2, 10C2, 23C2, 25C5, 26C2, 28C5, 31C2, 40C5, 49C5). Actually none of the analyzed 51 HAdV-C genomes had a penton sequence clustering with prototype 5. The only four type 5 strains (25C5, 28C5, 40C5, 49C5) found in this study were more closely related to type 6 regarding their penton sequence although this was not confirmed by high bootstrap values. Five HAdV type 2 strains (9C2, 10C2, 23C2, 26C2, 31C2) had penton sequences clustering with type 1, but also without a significant bootstrap value. As a recombinant phylogeny of the penton gene could not be confirmed by bootscanning, these 9 strains were not labelled as new genotypes. However, deduced amino acid sequences of the strains 9C2, 10C2, 23C2 and 26C2 were identical to type 1 and the deduced amino sequence of strain 31C2 had only a single amino acid substitution (Q377E) compared to the HAdV-C1 prototype sequence.

**Molecular phylogeny of early gene regions.** Multiple recombination events and novel sequence stretches were observed in the E1, E2 and E4 gene regions, whereas (with the single exception of strain 43C1) clustering of the E3 sequences was consistent with typing results in imputed serology and clustering of complete genomic sequences. The E3 sequence derived from strain 43C1 clustered with type 57 in the phylogenetic tree of the E3 region (Fig. S2). Although this clustering was not supported by a significant bootstrap value, a recombinant origin of strain 43C1 was indicated by bootscanning for the 5' part of the E3 region (about nt. position 27,500–29,000) (Fig. S3, 43C1). Bootscanning for the remaining part of the E3 gene region was not conclusive due to the closely related type 2 and type 57 prototype sequences, but multiple alignments of the deduced amino acid sequences of E3 gene products confirmed that strain 43C1 was closely related to these (or a common ancestor) but not to type HAdV-C1.

Surprisingly, five circulating strains (12C1, 20C1, 43C1, 45C6 and 46C6) clustered with type 57 in the E4 gene region (supported by a 100% bootstrap value and confirmed by bootscanning), although circulation of HAdV-C57 was not observed in this study. E4 sequences of multiple type 1 and type 2 strains (4C1, 5C1, 9C2, 10C2, 13C1, 18C1, 23C2, 26C2, 31C2, 32C1, 41C1, 50C1 and 51C1) clustered with type 5 (Fig. 6). This finding was not supported by a significant bootstrap value, but bootscanning of these 13 sequences confirmed a

Specimen (strain #/ species/type)	Year	Imputed serology (partial hexon sequencing)	Source	Derived from immunosuppressed patient (+/-)	Clustering of complete genomic sequence with prototype	Relabelled as novel type	Accession Number
1C2	2000	2	Feces	+	2	—	MH121072
2C2	2000	2	Blood	+	2	—	MH121070
3C2	2002	2	Blood	+	2	—	MH121071
4C1	2002	1	Urine	+	1	—	MH121073
5C1	2002	1	Blood	+	1	—	MH121074
6C1	2008	1	Blood	+	1	—	MH121075
7C1	2008	1	Feces	+	1	—	MH121076
8C2	2009	2	URT	—	2	—	MH121077
9C2	2009	2	URT	—	2	—	MH121078
10C2	2010	2	URT	—	2	—	MH121079
11C2	2012	2	LRT	—	2	—	MH121080
12C1	2012	1	URT	—	1	—	MH121081
13C1	2012	1	Feces	+	1	—	MH121082
14C2	2012	2	Feces	+	2	—	MH121083
15C2	2012	2	Blood	+	2	—	MH121084
16C2	2012	2	Feces	—	2	—	MH121085
17C2	2013	2	LRT	—	2	—	MH121086
18C1	2013	1	Feces	+	1	—	MH121087
19C2	2013	2	Blood	+	2	—	MH121088
20C1	2013	1	Blood	+	1	—	MH121089
21C2	2014	2	Blood	+	2	—	MH121090
22C1	2014	1	URT	—	1	—	MH121091
23C2	2014	2	Blood	+	2	—	MH121092
24C2	2014	2	Blood	+	2	—	MH121093
25C5	2014	5	Feces	+	5	—	MH121094
26C2	2014	2	Feces	+	2	—	MH121095
27C2	2015	2	URT	—	2	—	MH121096
28C5	2015	5	Feces	+	5	—	MF681662
29C2	2015	2	Feces	+	2	89	MH121097
30C1	2015	1	Feces	+	1	—	MH121098
31C2	2015	2	URT	—	2	—	MH121099
32C1	2015	1	URT	—	1	—	MH121100
33C2	2015	2	Feces	+	2	—	MH121101
34C2	2015	2	Feces	+	2	—	MH121102
35C2	2015	2	Blood	+	2	—	MH121103
36C2	2015	2	Feces	+	2	—	MH121104
37C2	2015	2	Feces	+	2	—	MH121105
38C2	2016	2	Feces	+	2	—	MH121106
39C2	2016	2	Blood	+	2	—	MH121107
40C5	2015	5	Feces	+	5	—	MH121118
41C1	2016	1	Feces	+	1	—	MH121108
42C2	2017	2	Feces	+	2	—	MH121109
43C1	2017	1	Feces	+	1	—	MH121110
44C2	2017	2	Feces	+	2	—	MH121111
45C6	2017	6	Feces	+	6	—	MH121112
46C6	2017	6	Feces	+	6	—	MH121113
47C2	2017	2	Feces	+	2	89	MH121114
48C2	2017	2	Feces	+	2	—	MH121115
49C5	2017	5	Feces	+	5	—	MH121119
50C1	2017	1	Feces	+	1	—	MH121116
51C1	2017	1	URT	—	1	—	MH121117

**Table 1.** Origin and characteristics of species HAdV-C strains analyzed in this study (URT: upper respiratory tract, LRT: lower respiratory tract).

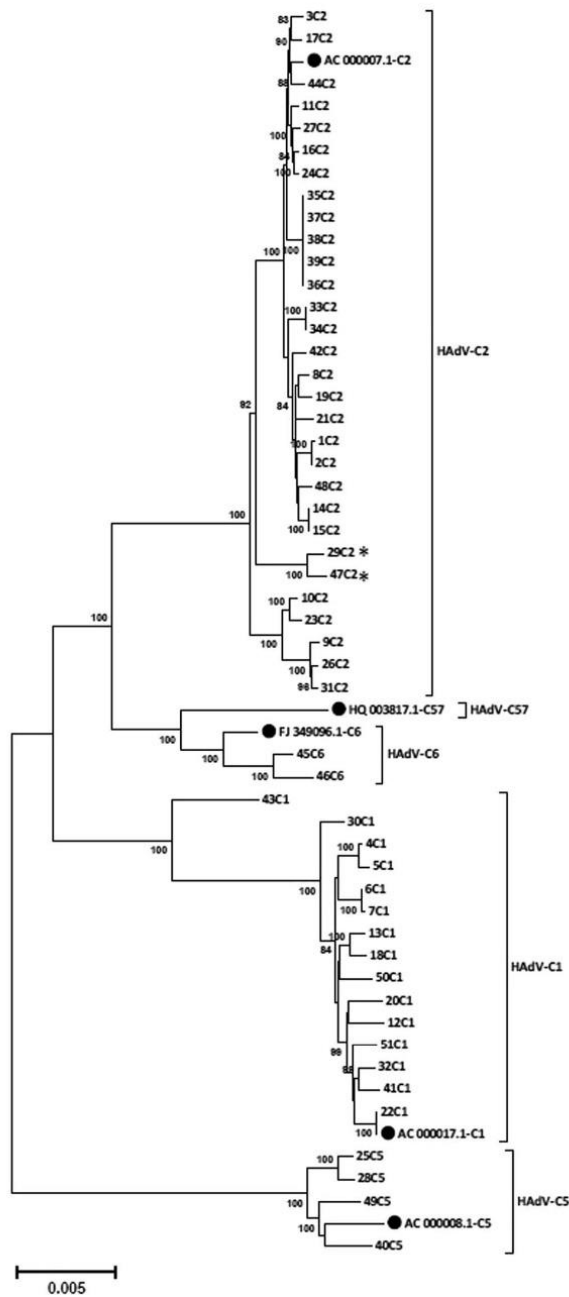


**Figure 1.** Phylogenetic analysis of the nucleic acid sequences of the neutralization  $\epsilon$  determinant loop 1 (A) and loop 2 (B) of the hexon. Genbank sequences of the five HAdV-C prototypes are highlighted by a black dot (labelling indicates accession number-species and type). The neighbor-joining tree was generated based on the Kimura two-parameter model with MEGA7. Bootstrap values  $< 80\%$  are not robust and therefore not depicted. \*Strains 29C2 and 47C2 were renamed as the novel type HAdV-C89.

recombinant origin of the E4 sequence. Moreover, these isolates had deduced amino acid sequences identical to the sequence of prototype HAdV-C5 in E4ORF3, E4ORF4 and E4ORF6/7. Five other type 2 strains (3C2, 29C2, 33C2, 34C2, 47C2) were more distantly related to the HAdV-C5 prototype sequence, but the recombinant origin of these E4 sequences was also confirmed by bootscanning only. E4ORF3 and E4ORF4 deduced amino sequences of 3C2, 29C2, 33C2, 34C2, 47C2 were also identical to the HAdV-C5 prototype sequence but E4ORF6/7 was not. Vice versa, three type 5 strains (25C5, 28C5 and 49C5) clustered with type 1 in the E4 region. Although the bootstrap value of this cluster was not significant, probably due to the very low sequence diversity between prototype 1 and prototype 2 E4 sequences, recombination was confirmed by bootscanning. Finally, three type 1 strains (6C1, 7C1 and 30C1) clustered with type 2 in the E4 region with a 92% bootstrap probability but recombination of the E4 region was not supported by bootscanning.

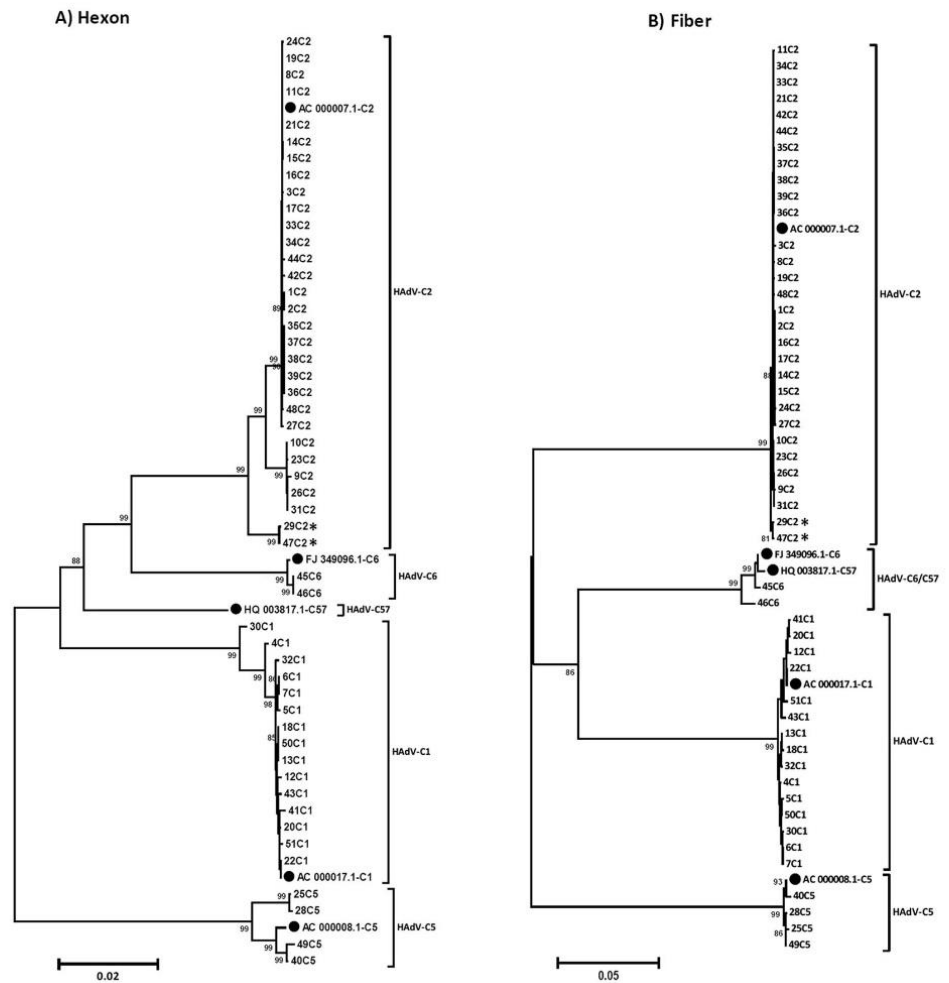
As gene products of E4 and E1 interact with each other, similar clustering of strains in the E1 phylogenetic tree as observed in the E4 phylogenetic tree was anticipated. Surprisingly, phylogenetic analyses of the E1 region showed a completely different clustering (Fig. 6).

Only a single sequence (strain 40C5) clustered with prototype HAdV-C5 in the E4 region and in E1 region. Three type 2 strains (9C2, 26C2, 31C2) clustered with prototype HAdV-C1 in the E1 region (supported by a 100% bootstrap value and confirmed by bootscanning) indicating an E1 sequence of recombinant origin. There were 13 other type 2 strains (1C2, 2C2, 8C2, 14C2, 15C2, 19C2, 21C2, 35C2, 36C2, 37C2, 38C2, 39C2 and 48C2)



**Figure 2.** Phylogenetic analysis of whole-genome sequences. Clustering of nucleic acid sequences of circulating strains with Genbank sequences of the five species HAdV-C prototypes (highlighted by a black dot, labelling indicates accession number-species and type). The neighbor-joining tree was generated based on the Kimura two-parameter model with MEGA7. Bootstrap values < 80% are not robust and therefore not depicted. \*Strains 29C2 and 47C2 were renamed as the novel type HAdV-C89.

and one type 5 strain (49C5) that also clustered with type HAdV-C1 prototype sequence in the E1 gene region. Recombination was confirmed by bootscanning, but clustering in the tree was non-significant (low bootstrap value). Surprisingly, one strain (40C5) had an intra-E1 recombination event in its phylogeny indicated by bootscan analysis with the E1A region closely related to prototype HAdV-C1 and the E1B region (and most of its genomic backbone) closely related to HAdV-C5 (Fig. S3).



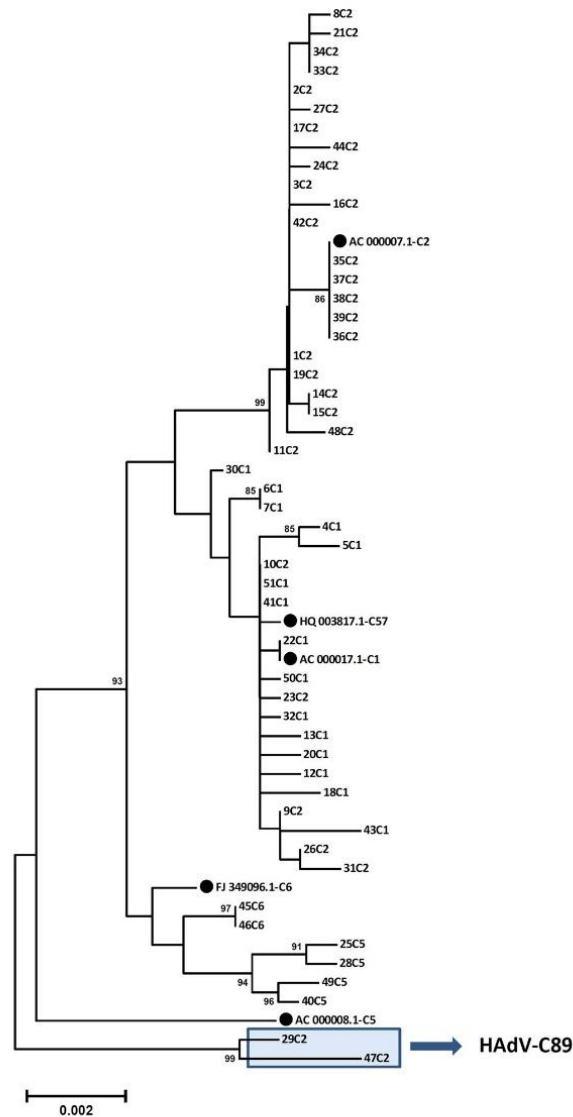
**Figure 3.** Phylogenetic analysis of the major capsid genes hexon (A) and fiber (B). Clustering of circulating strains and Genbank sequences of HAdV-C prototypes (highlighted by a black dot, labelling indicates accession number-species and type). The neighbor-joining tree was generated based on the Kimura two-parameter model with MEGA7. Bootstrap values < 80% are not robust and therefore not depicted. \*Strains 29C2 and 47C2 were renamed as the novel type HAdV-C89.

In the E2A region five type 2 isolates (9C2, 10C2, 23C2, 26C2, 31C2) and one type 6 strain (45C6) were found to have a recombinant phylogeny and clustering with the prototype HAdV-C1 (supported by significant bootstrap values) (Fig. S4). Moreover, the two isolates representing the novel prototype HAdV-C89 (29C2 and 47C2) had a unique E2A sequence compared to all other isolates and prototypes.

Genes of the E2B region (encoding e.g. the DNA polymerase) are highly conserved between all HAdV-C prototypes and circulating strains analyzed in this study with the exception of strains 43C1 and 50C1. These clustered on their own and had unique amino acid substitutions in comparison with all other isolates in the deduced amino acid sequence (Fig. S4).

**Molecular phylogeny of minor capsid proteins.** Sequence diversity of all six minor capsid proteins (IIIa, V, VI, VII, VIII, and IX) was very low and a recombinant phylogeny could not be detected. Interestingly, two circulating type 2 strains (33C2 and 34C2, derived from the same patient) had a deletion of three amino acids (position 447 to 449) as compared to the prototype HAdV-C2 sequence in the IIIa protein (data not shown). In the minor capsid protein V a deletion of two amino acids (serine and threonine at 228–229 aa position) was detected in two other circulating strains (1C2 and 2C2).

Minor capsid proteins VII and VIII were highly conserved at the nucleotide and amino acid level. In the amino acid alignment of minor capsid protein IX, 11 out of the 30 type 2 isolates (1C2, 2C2, 8C2, 9C2, 14C2,



**Figure 4.** Phylogenetic analysis of the major capsid gene penton base. Clustering of circulating strains and prototype sequences (highlighted by a black dot, labelling indicates accession number-species and type). Strains 29C2 and 47C2 were highlighted with a blue box and renamed as novel type HAdV-C89. The neighbor-joining tree was generated based on the Kimura two-parameter model with MEGA7. Bootstrap values < 80% are not robust and therefore not depicted.

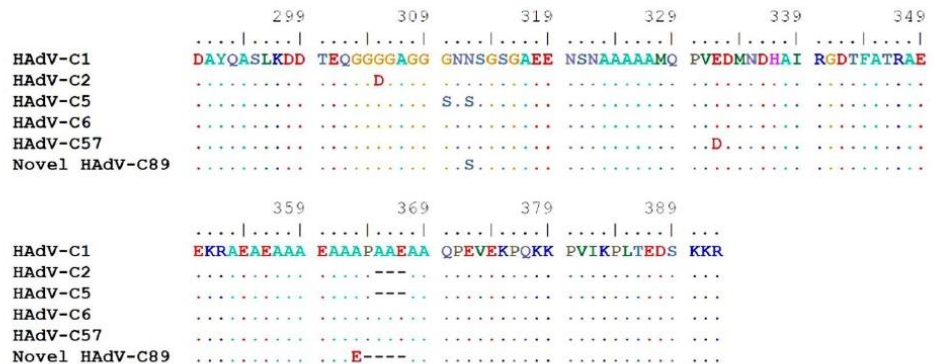
15C2, 19C2, 21C2, 26C2, 31C2, 42C2) were found to be more similar to prototype HAdV-C1 as they contain two amino acid substitutions (A3T and D123E) as compared to prototype HAdV-C2 protein IX amino acid sequence.

### Discussion

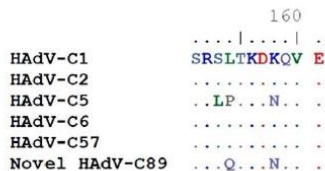
In the present study, analysis of the complete genomes of 51 circulating species HAdV-C strains revealed multiple recombination events of early gene regions (e.g. E1 and E4) that may contribute to the evolution of species HAdV-C. To our knowledge this has never been described before.

The E4 region encodes for six proteins (ORF1, ORF2, ORF3, ORF4, ORF6 34 K and ORF6/7), which are known to have one or more function required for lytic growth of virus, transcription of viral genes and replication<sup>23–25</sup>. E4 sequences of the HAdV-C1 and -C2 prototypes and many of the analyzed type 1 and 2 circulating strains were highly conserved; therefore conclusions on recombination between these two types were not feasible. In spite of this, 21 out of 45 analyzed type 1 and 2 strains displayed a recombinant E4 region. Five of these 21 strains clustered with HAdV-C57 in the E4 region and 14 of these strains had HAdV-C5 like E4

## A) RGD loop



## B) HVR1



**Figure 5.** Multiple amino acid alignments of the hypervariable loops of the penton base. (A) RGD loop sequences and (B) hypervariable region 1 sequences of all HAdV-C prototypes including the novel HAdV-C89.

sequences, although this was confirmed by bootscanning only. Therefore, HAdV-C5 like and HAdV-C57 like E4 sequences were highly “overrepresented” in the analyzed circulating virus strains compared to their typing results in imputed serology and their genomic backbone. This indicated a positive selection for these recombinant type 1 and 2 strains that have an E4 region sequence of type 5 or 57. This positive selection was observed both in the 39 analyzed strains originating from immunosuppressed patients, where these can cause severe disseminated infections with high virus loads in peripheral blood (as e.g. strain 40C5<sup>26</sup>) and in the 12 other analyzed strains originating from usually mild respiratory infections of immunocompetent patients.

Although the sequence diversity between type 5 and types 1 and 2 is low in the E4 gene region, type 5 has substantial amino acid substitutions in its E4 ORFs compared to types 1 and 2. Prototypes of HAdV-C1, -C2, -C6 and many of the analyzed circulating strains shared an identical amino acid E4ORF4 sequence but prototype HAdV-C5 had a distinctive substitution of serine at position 67 that was also present in seventeen circulating strains clustering with prototype 5 in the E4 gene region. The novel prototype HAdV-C89 strains (29C2 and 47C2) also cluster with the HAdV-C5 sequence, but have three other amino acid substitutions which may affect the binding of E4ORF4 to PP2A, a key function for this E4 protein<sup>27–30</sup>.

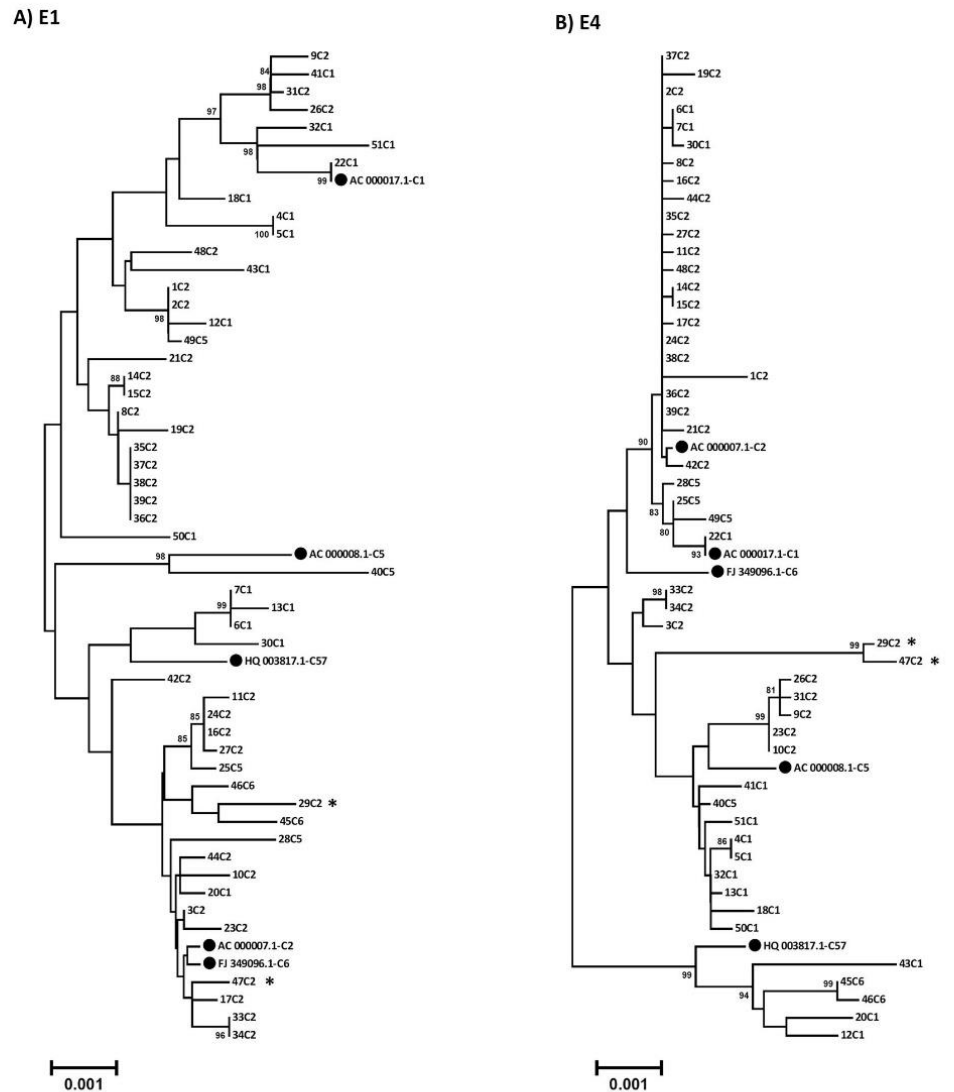
The deduced amino acid sequences of the E4ORF6/7 region of prototypes HAdV-C2, -C6, -C57 and many circulating HAdV-C strains were identical. However, prototype HAdV-C5 and 13 circulating strains (4C1, 5C1, 9C2, 10C2, 13C1, 18C1, 23C2, 26C2, 31C2, 32C1, 41C1, 50C1 and 51C1) have a distinctive K at position 68 which might serve as a SUMO2/3 acceptor site in the E4orf6/7 protein, thus affecting promoter activation by E4orf6/7 or its stability (Melling, personal communication).

Moreover, the deduced amino acid sequence of the E4ORF3<sup>31,32</sup> is identical in prototypes HAdV-C1, -C2, -C6 and -C57, but the prototype of HAdV-C5 has one substitution L33R that was conserved in all circulating recombinant HAdV-C strains clustering with this prototype's E4 region sequence.

The E4ORF1 gene product binds via its C-terminal PDZ binding motif (S/T-X-Y) to proteins containing PDZ domains and as expected this motif was highly conserved in all our E4ORF1 sequences<sup>33,34</sup>. With exception of the C-terminal region, multiple amino acid substitutions and deletions were observed in circulating strains like 41C1 and 1C2, but will probably not affect the function of this motif. Interestingly, these mutations were only observed in the circulating virus strains whereas in all species HAdV-C prototypes the complete E4ORF1 amino acid sequence is conserved.

The overall genetic diversity of the E1 region was higher than in the E4 region and also many (17 out of 51) circulating HAdV-C strains were found to be recombinant in the E1 region (confirmed by bootscan analysis although only three of these strains clustered with significant bootstrap value). As E1 and E4 proteins interact with each other, a similar clustering of strains in the E1 region as in the E4 region could be expected, but surprisingly almost all circulating strains except 49C5 clustered differently in these phylogenetic trees. The E1 region encodes only three functional proteins (E1A, E1B-19K and E1B-55K). E1A has a conserved region 4 (CR4) at amino acid position 202–210 which mediates interactions with host cell proteins such as the corepressor CtBP<sup>35</sup>.





**Figure 6.** Phylogenetic analysis of the early gene regions E1 (A) and E4 (B). Clustering of circulating strains and prototype sequences (highlighted by a black dot, labelling indicates accession number-species and type). The neighbor-joining tree was generated based on the Kimura two-parameter model with MEGA7. Bootstrap values < 80% are not robust and therefore not depicted. \*Strains 29C2 and 47C2 were renamed as the novel type HAdV-C.89.

Three circulating strains (43C1, 45C6, 46C6) had few substitutions in the CR4 region that might affect the interaction with host cell proteins<sup>36</sup>. Four strains (9C2, 26C2, 31C2 and 41C1) had a P140S substitution in the E1B-19K sequence which may affect the Bcl-2- homology region 2 (BH2) known to inhibit apoptosis by the E1B protein<sup>37</sup>. Apart from this single substitution the E1B-19K amino acid sequence was observed to be highly conserved. The N-terminal region of large E1B-55k protein represents an intrinsically disordered protein domain<sup>38</sup> and correspondingly the sequence of this region was less conserved than the C-terminal region in the circulating strains.

The E3 region contains seven genes affecting the host immunity after virus infection but not essential for the replication of HAdV-C<sup>39</sup>. In our study, the E3 gene region sequences of circulating strains were highly conserved and clustered with their respective prototype sequence, with the exception of strain 43C1. This recombination is clearly evident because the E3 sequences of different HAdV-C prototypes are highly divergent. The type 1 strain 43C1 clustered with the prototype HAdV-C57 E3 region sequence (Fig. S2) and thus is the only candidate to study further the effects of a recombinant E3 region in species HAdV-C. With this single exception, clustering suggested coevolution of the E3 region with the major capsid proteins hexon and fiber in species HAdV-C, whereas in species HAdV-D, homologous recombination in the E3 region (specifically CR1 genes) has been confirmed previously<sup>40</sup>.

Recombination events between the genes encoding major capsid proteins (hexon, fiber and penton) are an important mechanism in the evolution of species HAdV-D types, but almost absent in the evolution of species HAdV-C strains as confirmed by our sequence data<sup>1,5,40</sup>. Conserved sequence segments (called “universal conserved segments”) are present in HAdV-D genomes, e.g. in the penton base gene, the hexon gene and multiple ones in the E3, fiber and E4 genes, that drive the recombination of these regions<sup>41,42</sup>. Obviously, this mechanism is absent in species HAdV-C. Therefore, recombination of genes encoding major capsid proteins seems to be almost impossible in species HAdV-C, as also observed in the present study. The single exception to this rule of non-recombinant evolution of major capsid proteins in species HAdV-C is type HAdV-C57. Its penton base sequence is closely related to type HAdV-C1, its fiber to type HAdV-C6 and it has a unique hexon sequence and thus a novel neutralization determinant sequence<sup>14</sup>. Surprisingly, all circulating strains had neutralization epitope sequences identical (or almost identical) to the prototype sequences (Fig. 1) although a rapid evolution of these loop 1 and 2 sequences could be driven by an immune escape mechanism. This was previously described in species HAdV-B<sup>1,3,5,6</sup> but was obviously absent in species HAdV-C. Perhaps, E3 gene products interfering with the immune-response against HAdV-C type infected cells<sup>39,43,44</sup> could be so effective that there was almost no positive selection of immune escape mutations.

In this study, another novel prototype of species HAdV-C has been described (type 89). It was typed initially as a HAdV-C2 strain by conventional methods, because both hexon and fiber sequences, as well as most of its genetic backbone are closely related to the HAdV-C2 prototype sequence. However, HAdV-C89 was found to have a novel penton base sequence. In general, penton base sequences of species HAdV-C types are far less diverse as in species HAdV-D, but interestingly the novel type HAdV-C89 penton base sequence is most diverse compared to the old prototypes in the functional RGD loop. Although amino acid residues adjacent to the RGD motif were conserved (Fig. 5), the conformation of the RGD site may be different and thus, binding properties of type 89 to the secondary cellular receptor compared to other HAdV-C types<sup>45,46</sup>. Moreover, HAdV-C89 has a recombinant E4 region (Fig. 6) closely related to HAdV-C5 suggesting a recombinant phylogeny of this region, which may also influence its virulence. On the other hand, prevalence of type 89 in the present study was low, representing only two out of 51 strains. These were isolated from different patients, strain 29C2 in 2015 and strain 47C2 in 2017, respectively. Prevalence of HAdV-C89 strains in this study was about equal to HAdV-C6 but, for comparison, not a single strain of HAdV-C57 was detected in the present study.

In conclusion, simple typing procedures, as merely a partial sequencing of the hexon gene of a diagnostic specimen, seem to be sufficient to describe the “core region” of the HAdV-C genome. This includes the hexon and the fibre gene and the E3 region, which was considered to be important for immune escape and probably long term persistence of HAdV-C strains after primary infection. On the other hand, these routine typing strategies will not detect novel penton base sequences and thus would have missed the novel type 89. Moreover, future studies employing complete genomic sequencing on circulating HAdV-C strains are needed to study the clinical significance of recombination in the E1 and E4 regions as e.g. the predominance of type 5 derived E4 regions in the present study.

## Methods

**Specimens and strains.** A total of 51 species HAdV-C strains were obtained between the years 2000 and 2017 from clinical specimens originating from the upper respiratory tract (URT,  $n = 9$ ), lower respiratory tract (LRT,  $n = 2$ ), feces ( $n = 27$ ), blood ( $n = 12$ ) and urine ( $n = 1$ ). For 11 of these specimens adenoviruses were isolated on A549 cell cultures prior to sequencing, direct sequencing was applied to the other specimens (Table 1). Specimens originated from 40 patients; strains 6C1, 7C1 and 30C1 originated from a single patient as did strains 35C2, 36 C2, 37C2, 38 C2 and 39C2 from another patient. Strains 1C2 and 2C2 also originated from a single patient, as did strains 4C1 and 5C1, 14C2 and 15 C2, 33 C2 and 34 C2, 45C6 and 46C6. Specimens originated from different regions of Germany. Table 1 shows an overview of specimens sequenced, year of isolation and Genbank Accession Number.

**Nucleic acid extraction and viral load quantitation.** Extraction of DNA from 200  $\mu$ l clinical material or cell culture supernatant was performed as described by QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany). HAdV DNA was quantified by a generic, quantitative real-time PCR as described previously<sup>47</sup>.

**Imputed serology by partial sequencing.** Initial typing was performed by Sanger sequencing of parts of the neutralization epitope (loop 1 and loop 2) of the hexon gene (imputed serology) and the fibre knob gene as previously described<sup>48</sup>.

**Enrichment of HAdV DNA and library preparation.** Library preparation of 42 specimens (including all cell culture isolates) with sufficient HAdV DNA load (a ratio of HAdV/human genomic DNA  $> 1,000$ ) was performed using Nextera XT DNA library preparation kit (Illumina, San Diego, CA, USA) with 1 ng of DNA as input following manufacturers’ guideline. Purified DNA of the other nine specimens (6C1, 19C2, 20C1, 23C2, 24C2, 35C2, 47C2, 48C2 and 50C1) was sheared by sonication, and sequencing libraries were prepared directly (KAPA library preparation kit, KAPA Biosystems, Wilmington, MA, USA) as previously described<sup>49</sup>, with a few modifications. DNA fragments were end-repaired, A-tailed and ligated to NEBnext Illumina adapter (New England Biolabs, Ipswich, MA, USA). After PCR pre-amplification (6–14 cycles) with adapter specific primers, up to 750 ng of DNA was target enriched for HAdV fragments with RNA baits. These 120-mer RNA baits were designed to span the length of the positive strand of 24 Genbank HAdV-C1, -C2 and -C5 reference genomes (SureSelectXT kit, Agilent Technologies, CA, USA).

**Next generation sequencing and bioinformatics.** After quality control, libraries were sequenced on an Illumina MiSeq platform (2\*300 bp paired-ends run, v3 sequencing chemistry). The obtained sequence data

was quality controlled and de novo assembled using CLC Genomics Workbench v9 (Qiagen, <https://www.qiagen-bioinformatics.com/products/clc-genomics-workbench/>) and SPAdes v3.6<sup>50</sup> with minimum average coverage of 22 and maximum average coverage 78111 per nucleotide. Genome annotations were transferred from the most closely related type and cross-checked by using an ORF finder followed by BLAST analysis using Geneious v10 (<https://www.geneious.com/>).

**Phylogenetic and recombination analysis.** A multiple alignment of whole genomic sequences was constructed using fast Fourier transforms (MAFFT)<sup>51</sup>. This software (<https://www.ebi.ac.uk/Tools/msa/mafft/>) was applied using the default gap parameters. Pairwise alignment, comparisons and visualization of genomes were performed on BioEdit version 7.2.0 (<http://www.mbio.ncsu.edu/BioEdit/page2.html>). Bootstrapped, neighbor-joining (Kimura 2-parameter model) phylogenetic trees with 1000 replicates were constructed using MEGA v7 software. Similarity plots and recombination detection (bootscan approach) was performed using Simplot software (version 3.5.1) with a window size of 1000 bp and step size of 200bp<sup>52</sup>. The following complete genomic nucleotide sequences representing all prototypes of the HAdV-C species were used in the analysis (Genbank accession numbers in parentheses): HAdV-C1 (AC000017), HAdV-C2 (AC000007), HAdV-C5 (AC000008), HAdV-C6 (FJ349096) and HAdV-C57 (HQ003817). Furthermore, amino acid alignments for each gene were constructed with ClustalW program.

**Ethical statement.** All methods were carried out in accordance with relevant guidelines and regulations. For scientific use of routine anonymized data, ethical approval is not required in Germany (confirmed by a letter of the local ethical committee "Ethikkommission der Medizinischen Hochschule Hannover", 9th of April, 2018). This study analyzed molecular phylogeny of the species human adenovirus C and did not include experimentation with human tissue, therefore informed consent of patients is not required for this type of study.

## References

1. Robinson, C. M. *et al.* Molecular evolution of human adenoviruses. *Sci Rep* **3**, 1812, <https://doi.org/10.1038/srep01812> (2013).
2. Lion, T. Adenovirus infections in immunocompetent and immunocompromised patients. *Clinical microbiology reviews* **27**, 441–462, <https://doi.org/10.1128/CMR.00116-13> (2014).
3. Hashimoto, S. *et al.* Recombinant type Human mastadenovirus D85 associated with epidemic keratoconjunctivitis since 2015 in Japan. *Journal of medical virology* **90**, 881–889, <https://doi.org/10.1002/jmv.25041> (2018).
4. Seto, D., Chodosh, J., Brister, J. R. & Jones, M. S. & Members of the Adenovirus Research, C. Using the whole-genome sequence to characterize and name human adenoviruses. *Journal of virology* **85**, 5701–5702, <https://doi.org/10.1128/JVI.00354-11> (2011).
5. Robinson, C. M., Seto, D., Jones, M. S., Dyer, D. W. & Chodosh, J. Molecular evolution of human species D adenoviruses. *Infect Genet Evol* **11**, 1208–1217, <https://doi.org/10.1016/j.meegid.2011.04.031> (2011).
6. Ismail, A. M. *et al.* Genomic analysis of a large set of currently and historically important human adenovirus pathogens. *Emerg Microbes Infect* **7**, 10, <https://doi.org/10.1038/s41426-017-0004-y> (2018).
7. Yoshitomi, I., Sera, N., Gonzalez, G., Ianaoka, N. & Fujimoto, T. First isolation of a new type of human adenovirus (genotype 79), species Human mastadenovirus B (B2) from sewage water in Japan. *Journal of medical virology* **89**, 1192–1200, <https://doi.org/10.1002/jmv.24749> (2017).
8. Walsh, M. P. *et al.* Evidence of molecular evolution driven by recombination events influencing tropism in a novel human adenovirus that causes epidemic keratoconjunctivitis. *PLoS One* **4**, e5635, <https://doi.org/10.1371/journal.pone.0005635> (2009).
9. Dehghan, S. *et al.* Computational analysis of four human adenovirus type 4 genomes reveals molecular evolution through two interspecies recombination events. *Virology* **443**, 197–207, <https://doi.org/10.1016/j.virol.2013.05.014> (2013).
10. Garnett, C. T., Erdman, D., Xu, W. & Gooding, L. R. Prevalence and quantitation of species C adenovirus DNA in human mucosal lymphocytes. *Journal of virology* **76**, 10608–10616 (2002).
11. Lion, T. *et al.* Monitoring of adenovirus load in stool by real-time PCR permits early detection of impending invasive infection in patients after allogeneic stem cell transplantation. *Leukemia* **24**, 706–714, <https://doi.org/10.1038/leu.2010.4> (2010).
12. Al Qurashi, Y. M., Guiver, M. & Cooper, R. J. Sequence typing of adenovirus from samples from hematological stem cell transplant recipients. *Journal of medical virology* **83**, 1951–1958, <https://doi.org/10.1002/jmv.22204> (2011).
13. Leen, A. M. & Rooney, C. M. Adenovirus as an emerging pathogen in immunocompromised patients. *British journal of haematology* **128**, 135–144, <https://doi.org/10.1111/j.1365-2141.2004.05218.x> (2005).
14. Walsh, M. P. *et al.* Computational analysis of two species C human adenoviruses provides evidence of a novel virus. *J Clin Microbiol* **49**, 3482–3490, <https://doi.org/10.1128/JCM.00156-11> (2011).
15. Ison, M. G. Adenovirus infections in transplant recipients. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America* **43**, 331–339 (2006).
16. Hong, J. Y. *et al.* Lower respiratory tract infections due to adenovirus in hospitalized Korean children: epidemiology, clinical features, and prognosis. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America* **32**, 1423–1429, <https://doi.org/10.1086/320146> (2001).
17. Rocholl, C., Gerber, K., Daly, J., Pavia, A. T. & Byington, C. L. Adenoviral infections in children: the impact of rapid diagnosis. *Pediatrics* **113**, e51–56 (2004).
18. Garnett, C. T. *et al.* Latent species C adenoviruses in human tonsil tissues. *Journal of virology* **83**, 2417–2428, <https://doi.org/10.1128/JVI.02392-08> (2009).
19. Kosulin, K. *et al.* Persistence and reactivation of human adenoviruses in the gastrointestinal tract. *Clin Microbiol Infect* **22**, 381 e381–381 e388, <https://doi.org/10.1016/j.cmi.2015.12.013> (2016).
20. Veltrop-Duits, L. A. *et al.* High titers of pre-existing adenovirus serotype-specific neutralizing antibodies in the host predict viral reactivation after allogeneic stem cell transplantation in children. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America* **52**, 1405–1413, <https://doi.org/10.1093/cid/cir231> (2011).
21. Ganzenmueller, T. & Heim, A. Adenoviral load diagnostics by quantitative polymerase chain reaction: techniques and application. *Rev Med Virol* **22**, 194–208, <https://doi.org/10.1002/rmv.724> (2012).
22. Houldcroft, C. J. *et al.* Use of Whole-genome Sequencing of Adenovirus in Immunocompromised Paediatric Patients to Identify Nosocomial Transmission and Mixed-genotype Infection. *J Infect Dis*, <https://doi.org/10.1093/infdis/jiy323> (2018).
23. Tauber, B. & Dobner, T. Adenovirus early E4 genes in viral oncogenesis. *Oncogene* **20**, 7847–7854, <https://doi.org/10.1038/sj.onc.1204914> (2001).
24. Tauber, B. & Dobner, T. Molecular regulation and biological function of adenovirus early genes: the E4 ORFs. *Gene* **278**, 1–23 (2001).
25. Leppard, K. N. E4 gene function in adenovirus, adenovirus vector and adeno-associated virus infections. *The Journal of general virology* **78**(Pt 9), 2131–2138, <https://doi.org/10.1099/0022-1317-78-9-2131> (1997).

26. Voigt, S., Hofmann, J., Edelmann, A., Sauerbrei, A. & Kuhl, J. S. Brincidofovir clearance of acyclovir-resistant herpes simplex virus-1 and adenovirus infection after stem cell transplantation. *Transpl Infect Dis* **18**, 791–794, <https://doi.org/10.1111/tid.12582> (2016).
27. Shtrichman, R., Sharf, R., Barr, H., Dobner, T. & Kleinberger, T. Induction of apoptosis by adenovirus E4orf4 protein is specific to transformed cells and requires an interaction with protein phosphatase 2A. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 10080–10085 (1999).
28. Miron, M. J. *et al.* Localization and importance of the adenovirus E4orf4 protein during lytic infection. *Journal of virology* **83**, 1689–1699, <https://doi.org/10.1128/JVI.01703-08> (2009).
29. Mui, M. Z. *et al.* Identification of the adenovirus E4orf4 protein binding site on the B55alpha and Cdc55 regulatory subunits of PP2A: Implications for PP2A function, tumor cell killing and viral replication. *PLoS pathogens* **9**, e1003742, <https://doi.org/10.1371/journal.ppat.1003742> (2013).
30. Kleinberger, T. & Shenk, T. Adenovirus E4orf4 protein binds to protein phosphatase 2A, and the complex down regulates E1A-enhanced junB transcription. *Journal of virology* **67**, 7556–7560 (1993).
31. Forrester, N. A. *et al.* Adenovirus E4orf3 targets transcriptional intermediary factor 1gamma for proteasome-dependent degradation during infection. *Journal of virology* **86**, 3167–3179, <https://doi.org/10.1128/JVI.06583-11> (2012).
32. Vink, E. I. *et al.* Impact of Adenovirus E4-ORF3 Oligomerization and Protein Localization on Cellular Gene Expression. *Viruses* **7**, 2428–2449, <https://doi.org/10.3390/v7052428> (2015).
33. James, C. D. & Roberts, S. Viral Interactions with PDZ Domain-Containing Proteins- An Oncogenic Trait? *Pathogens* **5**, <https://doi.org/10.3390/pathogens5010008> (2016).
34. Chung, S. H., Frese, K. K., Weiss, R. S., Prasad, B. V. & Javier, R. T. A new crucial protein interaction element that targets the adenovirus E4-ORF1 oncoprotein to membrane vesicles. *Journal of virology* **81**, 4787–4797, <https://doi.org/10.1128/JVI.02855-06> (2007).
35. Zhang, Q., Yao, H., Vo, N. & Goodman, R. H. Acetylation of adenovirus E1A regulates binding of the transcriptional corepressor CtBP. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 14323–14328, <https://doi.org/10.1073/pnas.011283598> (2000).
36. Cohen, M. J. *et al.* Dissection of the C-terminal region of E1A redefines the roles of CtBP and other cellular targets in oncogenic transformation. *Journal of virology* **87**, 10348–10355, <https://doi.org/10.1128/JVI.00786-13> (2013).
37. Han, J., Sabbatini, P. & White, E. Induction of apoptosis by human Nbk/Bik, a BH3-containing protein that interacts with E1B 19K. *Molecular and cellular biology* **16**, 5857–5864 (1996).
38. Sieber, T. *et al.* Intrinsic disorder in the common N-terminus of human adenovirus 5 E1B-55K and its related E1BN proteins indicated by studies on E1B-93R. *Virology* **418**, 133–143, <https://doi.org/10.1016/j.virol.2011.07.012> (2011).
39. Lichtenstein, D. L., Toth, K., Doronin, K., Tollefson, A. E. & Wold, W. S. Functions and mechanisms of action of the adenovirus E3 proteins. *International reviews of immunology* **23**, 75–111 (2004).
40. Singh, G. *et al.* Homologous recombination in E3 genes of human adenovirus species D. *Journal of virology* **87**, 12481–12488, <https://doi.org/10.1128/JVI.01927-13> (2013).
41. Gonzalez, G. *et al.* Intertypic modular exchanges of genomic segments by homologous recombination at universally conserved segments in human adenovirus species D. *Gene* **547**, 10–17, <https://doi.org/10.1016/j.gene.2014.04.018> (2014).
42. Gonzalez, G., Koyanagi, K. O., Aoki, K. & Watanabe, H. Interregional Coevolution Analysis Revealing Functional and Structural Interrelatedness between Different Genomic Regions in Human Mastadenovirus D. *Journal of virology* **89**, 6209–6217, <https://doi.org/10.1128/JVI.00515-15> (2015).
43. Burgert, H. G. *et al.* Subversion of host defense mechanisms by adenoviruses. *Curr Top Microbiol Immunol* **269**, 273–318 (2002).
44. Windheim, M. *et al.* A unique secreted adenovirus E3 protein binds to the leukocyte common antigen CD45 and modulates leukocyte functions. *Proceedings of the National Academy of Sciences of the United States of America* **110**, E4884–E4893, <https://doi.org/10.1073/pnas.1312420110> (2013).
45. Ruoslahti, E. RGD and other recognition sequences for integrins. *Annu Rev Cell Dev Biol* **12**, 697–715, <https://doi.org/10.1146/annurev.cellbio.12.1.697> (1996).
46. Madisch, I. *et al.* Phylogenetic analysis and structural predictions of human adenovirus penton proteins as a basis for tissue-specific adenovirus vector design. *Journal of virology* **81**, 8270–8281, <https://doi.org/10.1128/JVI.00048-07> (2007).
47. Heim, A., Ebnet, C., Harste, G. & Pring-Akerblom, P. Rapid and quantitative detection of human adenovirus DNA by real-time PCR. *Journal of medical virology* **70**, 228–239, <https://doi.org/10.1002/jmv.10382> (2003).
48. Madisch, I., Harste, G., Pommer, H. & Heim, A. Phylogenetic analysis of the main neutralization and hemagglutination determinants of all human adenovirus prototypes as a basis for molecular classification and taxonomy. *Journal of virology* **79**, 15265–15276, <https://doi.org/10.1128/JVI.79.24.15265-15276.2005> (2005).
49. Wilkie, G. S. *et al.* First fatality associated with elephant endotheliotropic herpesvirus 5 in an Asian elephant: pathological findings and complete viral genome sequence. *Sci Rep* **4**, 6299, <https://doi.org/10.1038/srep06299> (2014).
50. Nurk, S. *et al.* Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *Journal of computational biology: a journal of computational molecular cell biology* **20**, 714–737, <https://doi.org/10.1089/cmb.2013.0084> (2013).
51. Katoh, K. & Toh, H. Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform* **9**, 286–298, <https://doi.org/10.1093/bib/bbn013> (2008).
52. Lole, K. S. *et al.* Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *Journal of virology* **73**, 152–160 (1999).

## Acknowledgements

The authors thank Gabriele Harste, Jenny Witthuhn and Sandra Flucht for excellent technical assistance and Eric Delwart for providing virological analysis in-kind. This work was supported by the SFB 900 and A. Dhingra was supported by the graduate program “Infection Biology” of the Hannover Biomedical Research School (HBRS).

## Author Contributions

Akshay Dhingra carried out sequencing and phylogenetic analysis, discussed the results and contributed to writing of the manuscript. Elias Hage designed sequence analysis procedures and has given technical suggestions on methods. Tina Ganzenmueller contributed to the data analysis and writing of the manuscript. Sindy Böttcher, Jörg Hofmann, Klaus Hamprecht, Barbara Rath, and Patrick Obermeier have provided strains for sequencing, performed virus isolation and sequencing (Jörg Hofmann). Thomas Dobner and Fabian Hausmann analyzed E4 sequence data and contributed to writing of the manuscript. Albert Heim has supervised the project and took lead in writing the manuscript. All authors provided critical feedback and helped shape the research, analysis and manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-37249-4>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019

### 3.3 Machine Learning Analysis for Pattern Recognition of Adenovirus Infection

Once the scene has been set for standardized and comprehensive clinical data capturing aided by mobile applications at the bedside to meet the level of highly detailed laboratory data obtained by systematic and agnostic mNGS, a plethora of new possibilities open up to boost our understanding of viral infections.

In the context of the two digital surveillance programs mentioned above, with several hundred pediatric patients assessed using mHealth technology developed by the Vienna Vaccine Safety Initiative and with detailed virological test results, including NGS, we undertook the unique challenge to investigate differences in real-world disease severity with clinical presentations of certain viral infections. For this purpose, we deployed our previously published and validated ViVI Disease Severity Score (DSS) implemented in the ViVI ScoreApp. This app was specifically designed to improve infectious disease management and patient care. It allows for efficient work flow in busy clinical settings, such as emergency rooms, with its user-friendly format according to the principles of Design Thinking<sup>202</sup>. Following data entry, the app computed a score reflecting disease severity with a numerical value between 0-48 (increasing values reflecting increasing disease severity)<sup>30,39,203</sup>. In addition to traditional epidemiological outcome measures, such as hospitalization or mortality, the ViVI Disease Severity Score provides a detailed, reproducible, CDISC standardized, and meta-analyzable measure with considerably lower risk of bias<sup>30</sup>. Following the WHO guide on Pandemic Influenza Severity Assessment (PISA), a severity measure should “take into account the presence or absence of underlying chronic diseases or conditions”<sup>204</sup> which was readily realized with the ViVI Risk Factor Score, incorporated in the ViVI ScoreApp, as well. Not only did the ViVI Disease Severity and Risk Factor Scores mirror fine nuances in seriousness of illness on the individual patient level and over time but the comprehensive set of parameters assessed (see **Figure 13**<sup>30</sup>) could be analyzed further using a machine learning methodology to portray characteristic clinical patterns of certain viral infections and in certain patient groups, e.g. in different pediatric age brackets. Following a systematic literature review by Ma et al. performed in 2018<sup>36</sup> we understood that based on clinical symptoms alone, viral infections cannot be differentiated from one another reliably. However, if we combined all the available data captured with the ViVI Score App, including pertinent negative data (i.e. the *absence* of a symptom), a disease severity measure, and underlying conditions, we assumed we could gain novel insight.

### The ViVI Disease Severity Score (ViVI Score)

$$\begin{aligned} &= \\ &\text{Disease Severity with Signs and Symptoms of Uncomplicated disease (DSU; weighed 1}\times\text{)} \\ &\text{PLUS} \\ &\text{Disease Severity with Signs and Symptoms of Complicated disease (DSC; weighed 3}\times\text{)} \end{aligned}$$

---

#### DSU 1: Fever

- Evidence of fever (defined as any measurement in current disease episode  $\geq 38^{\circ}\text{C}$ )

#### DSU 2: Cough

- Evidence of cough

#### DSU 3: Pharyngitis

- Evidence of sore throat or inflamed throat on examination

#### DSU 4: Rhinitis

- Evidence of coryza/rhinitis on examination

#### DSU 5: Headache

- Evidence of headache or pain in head/neck area (using age-appropriate measures)

#### DSU 6: Myalgia

- Evidence of muscle pain on exam (using age appropriate techniques in young children)

#### DSU 7: Malaise

- Level of reduction in general well-being  $\geq 5$  on a scale from 0-10

#### DSU 8: Diarrhea

- Evidence of diarrhea, i.e.  $\geq 3$  runny bowel movements (or  $\geq 3$  more/day or baseline)

#### DSU 9: Vomiting

- Evidence of vomiting (at least once)
- 

#### DSC 1: High and prolonged fever

- Body temperature  $>40^{\circ}\text{C}$  for  $\geq 3$  days

#### DSC 2: Dyspnea

One or more of the following:

- Evidence of shortness of breath (dyspnea, labored breathing, respiratory distress)
- Evidence of difficulty breathing
- Evidence of tachypnea (using age-appropriate standards)
- Need for mechanical ventilation or extracorporeal membrane oxygenation (ECMO)

#### DSC 3: Hypoxia

One or more of the following:

- Evidence of cyanosis (incl. turning blue during seizures)
- Evidence of hypoxia ( $\text{O}_2$  sat  $<93\%$ )
- Evidence of  $\text{O}_2$  requirement (incl. blow-by oxygen)
- Evidence of respiratory failure and/or need for mechanical ventilation or ECMO

#### DSC 4: Hemoptysis

- Evidence of bloody/colored sputum

#### DSC 5: Altered/ loss of consciousness

One or more of the following:

- Evidence of CNS involvement (e.g. encephalopathy, encephalitis)
- Evidence of altered mental status
- Evidence of Glasgow Coma Scale (GCS) or Infant Face Scale  $<15$  and/or personality change
- Evidence of unconsciousness (other than postictal) or/and
- Evidence of drowsiness or difficulty to arouse (incl. lethargy and/or decreased levels of activity)
- Evidence of dizziness
- Evidence of confusion
- Evidence of severe weakness (incl. floppiness in infants)
- Evidence of paralysis

#### DSC 6: Seizure

- Evidence of seizures

#### DSC 7: Dehydration

One or more of the following:

- Evidence of severe dehydration (documented dehydration, need for IV-therapy, or Base Excess  $<-7$  on blood gas analysis)
- Evidence of decreased urine output and/or need for hemofiltration/dialysis

#### DSC 8: Exacerbation of chronic disease

- Exacerbation of chronic disease (incl. asthma, chronic hepatic, cardiovascular, or renal disease, diabetes mellitus, or metabolic disease)

#### DSC 9: Septic shock or multi-organ failure

One or more of the following:

- Evidence of septic shock
- Evidence of secondary complications (renal/multi-organ failure, rhabdomyolysis, myocarditis)
- Evidence of hypotension and/or need for vasopressor support

#### DSC 10: Need for hospitalization

- Assessor's judgment that the patient should be admitted to an inpatient ward (regardless of cost, availability of hospital beds, and other external factors)

#### DSC 11: Lower respiratory tract infection/superinfection

One or more of the following:

- Evidence of lower respiratory tract disease (pneumonia, bronchitis, pulmonary rales, wheezing/obstruction, need for mechanical ventilation/ECMO incl. clinical, laboratory, or radiological evidence of bacterial superinfection in the lower respiratory tract)

#### DSC 12: Upper respiratory tract infection/superinfection

One or more of the following:

- Evidence of upper respiratory tract disease (cough, coryza, red/sore throat, ear ache)
- Evidence of upper respiratory tract bacterial superinfection (incl. laboratory, radiological, or clinical findings, such as purulent drainage, bulging tympanic membrane, positive Strep A rapid test or microbiology result)

#### DSC 13: Need for admission to the intensive care unit (ICU)

One or more of the following:

- Assessor's judgment that patient would benefit from admission to the ICU (including intermediate care)
- Assessor's judgment that patient would benefit from assisted respiration (incl. BiPAP, CPAP)
- Assessor's judgment that patient would benefit from mechanical ventilation or ECMO

**Figure 13.** The ViVI Disease Severity Score (ViVI Score) components and calculation formula<sup>30</sup>.

In **Study 4 (Manuscript D)** of this PhD thesis, we investigated whether the application of a dual precision medicine approach linking A) standardized digital surveillance facilitating disease severity scoring, machine learning analysis and pattern recognition to B) systematic PCR testing and comprehensive NGS would enable the identification of key clinical characteristics of respiratory adenovirus infection in different pediatric age brackets. This proof-of-concept study was exemplarily conducted for adenoviruses, in place of any viral infection of interest. Of note, the synthesis of unbiased syndromic and virological surveillance allowed the discovery of a novel emerging virus associated with fatal disseminated neonatal adenovirus infection: HAdV-D80.



### **3.4 Manuscript D**

Obermeier PE, Heim A, Biere B, Hage E, Alchikh M, Conrad T, Schweiger B, Rath, B. **Linking digital surveillance and in-depth virology to study clinical patterns of viral respiratory infections in vulnerable patient populations.** iScience (submitted October 2021).

**Linking digital surveillance and in-depth virology to study clinical patterns of viral respiratory infections in vulnerable patient populations**

Patrick E Obermeier<sup>1,2,3</sup>, Albert Heim<sup>4</sup>, Barbara Biere<sup>5</sup>, Elias Hage<sup>4</sup>, Maren Alchikh<sup>1,2,3</sup>,  
Tim Conrad<sup>6</sup>, Brunhilde Schweiger<sup>5</sup>, Barbara A Rath<sup>1,2,3\*\*</sup>

1 Vienna Vaccine Safety Initiative, Pediatric Infectious Diseases, Berlin, Germany

2 Charité University Medical Centre, Department of Pediatrics, Berlin, Germany

3 UMR Chrono-environnement, Université Bourgogne Franche-Comté, Besançon, France

4 National Reference Laboratory for Adenoviruses, Hannover Medical School, Hannover, Germany

5 National Reference Centre for Influenza, Robert Koch-Institute, Berlin, Germany

6 Department of Mathematics and Computer Science, Freie Universität Berlin, Berlin, Germany

**\*\*Corresponding Author/ Lead Contact:** Barbara A. Rath, MD PhD, Vienna Vaccine Safety Initiative, Berlin, Germany, [barbara.rath@vi-vi.org](mailto:barbara.rath@vi-vi.org), @Vi\_VIorg

**Summary (150/150):** To improve the identification and management of viral respiratory infections we established a clinical and virologic surveillance program for pediatric patients fulfilling pre-defined case criteria of influenza-like illness and viral respiratory infections. The program resulted in a cohort comprising 6,073 patients (56% male, median age 1.6 years, range 0-18.8 years), where every patient was assessed with a validated disease severity score at the point-of-care using the VIVI ScoreApp. We used machine learning and agnostic feature selection to identify characteristic clinical patterns. We tested all patients for human adenoviruses, 571 (9%) were positive.

Adenovirus infections were particularly common and mild in children  $\geq 1$  month of age but rare and potentially severe in neonates: with lower airway involvement, disseminated disease and a 50% mortality rate (n=2/4). In one fatal case, we discovered a novel virus: HAdV-80. Standardized surveillance leveraging digital technology helps to identify characteristic clinical patterns, risk factors, and emerging pathogens.

**Keywords:** m-Health, digital surveillance, disease severity, adenovirus, emerging pathogens, feature selection, influenza-like illness, infectious diseases, pattern recognition, precision medicine

CONFIDENTIAL

**The Bigger Picture:** (991/1,000 characters)

Viral respiratory infections pose a substantial health threat prompting public health authorities and healthcare experts to find new avenues for the timely identification of emerging pathogens and risk groups presenting with specific clinical patterns. During a 6-year digital surveillance program using a mobile health app, the ViVI ScoreApp, we identified clinical severity and symptom patterns in 6,073 pediatric patients with signs of influenza-like illness and viral respiratory infection. This was combined with PCR and next-generation sequencing. We deployed this precision medicine approach to investigate whether the identification and management of viral respiratory infections could be improved. Our findings underline opportunities of linking the use of mobile health technology at the bedside with machine learning and pattern recognition analysis for the early and unbiased signal detection of laboratory confirmed virus infections in clinical care, surveillance, and research.

**Highlights:**

- We used mobile health technology to enable clinical pattern recognition (73/85)
- The ViVI ScoreApp provided precision data for cross-cohort meta-analysis (75/85)
- Neonates with adenovirus infection are at risk of severe or fatal disease outcomes (84/85)
- In one neonate with disseminated disease, we found a new adenovirus: HAdV-D80 (79/85)

**eTOC Blurb:** (80/80 words)

To improve the identification and management of viral respiratory infections we introduce a dual precision medicine approach combining the utilization of a mobile health application for the identification of patient risk groups, characteristic clinical patterns, and emerging viral infections with minimal bias combined with systematic virological testing. Precision medicine leveraging mobile health apps offers the opportunity to aid healthcare professionals in their daily clinical practice and in research aimed at recognizing emerging pathogens and individual risk in vulnerable patient populations.

**Maturity Level Categorization:** DSML 2: Proof-of-Concept: Data science output has been formulated, implemented, and tested for one domain/problem

CONFIDENTIAL

## Introduction

Precision medicine aims to “classify individuals into subpopulations that differ in their susceptibility to a particular disease”<sup>1</sup>. The identification of patterns in disease presentation is particularly compelling for the early identification of new and emerging viral pathogens, e.g. in virus epi-/pandemics. Thus, health authorities advocate the implementation of precision methodology and use of artificial intelligence, including machine learning in both, clinical care and research of viral respiratory infectious diseases. The goal is to improve outbreak detection, in-depth epidemiological analysis, and the quality of patient care<sup>2,3</sup>.

For the identification of patient groups with a high risk of a particularly severe clinical course, a well-established outcome measure is needed, where the risk of interrater variability and bias is at a minimum. To this end, the Vienna Vaccine Safety Initiative (VIVI), an international non-profit research institute focusing on innovation in infectious diseases & vaccines, has developed and validated a standardized VIVI Disease Severity Score (VIVI Score, <https://score.vi-vi.org>), for instant clinical assessments at the point-of-care. The VIVI Score is available via mobile application capturing the patient’s complex clinical presentation in real-time. Based on a validated set of clinical parameters in line with World Health Organization criteria of uncomplicated and complicated influenza-like illness (ILI), a composite score is generated within 3 minutes, ranging from 0-48, and reflecting increasing disease severity with increasing values. The VIVI Score has since been validated in a 6,000 children cohort in Germany<sup>4</sup>, a multi-center project in Europe<sup>5</sup>, and in community clinics in the United States comprising both adults and children<sup>6</sup>. The studies have shown that the ability to obtain an instant, well-standardized measure of clinical severity allows for rapid cross-cohort comparison and the identification of outliers in real-time<sup>4</sup>. This will enable the timely stratification of “individuals into subpopulations”<sup>1</sup>, as would be the requirement for precision medicine.

Children may be affected by a number of viral respiratory viral infections with varying susceptibility and disease severity in different subgroups<sup>7</sup>. For example, respiratory syncytial virus (RSV) infection poses a major health threat to neonates and infants<sup>8</sup> while other respiratory viruses such as SARS-CoV-

2 might usually be associated with milder disease in young children<sup>7,9</sup>. However, such assumptions have traditionally been made based on retrospective chart review, which tends to be prone to observer bias and/or incomplete of data<sup>10</sup> or “anecdotal evidence” lacking standardization.

These limitations become apparent also with regards to one of the most common respiratory viral pathogens in young children, human adenoviruses (HAdV). HAdV infection of the respiratory tract is usually considered mild and self-limiting. In certain high risk groups, such as immunocompromised individuals, HAdV infection can rapidly evolve into disseminated and fatal disease. Anecdotal reports suggest that HAdV infection may also be serious in the very young, i.e. neonates. It has been associated with significant morbidity and mortality<sup>11-28</sup>, but systematic prospective studies of HAdV incidence and severity in healthy children are lacking.

With this precision medicine study, we aim to investigate characteristic clinical patterns and disease severity of HAdV infection in children presenting with ILI based on a real-world dataset obtained during a 6-year digital/virologic surveillance program combining standardized patient assessments and disease severity scoring using mobile health (m-Health) technology for machine learning and pattern recognition analysis with in-depth virological testing in all patients.

The current study provides a proof-of-concept for improving quality of care and disease management through precision data and artificial intelligence. In addition, we exemplify the potential of point-of-care clinical severity assessments for the early identification of novel/emerging respiratory viral pathogens.

## Results

A total of 6,073 ILI patients participated in the digital syndromic and virologic surveillance program, 571 (9%) of which tested HAdV-positive (**Figure 1**), with the majority (65%) being <5 years of age. In detail, 4 neonates (0-28 days of age), 156 infants (29 days – 12 months of age), 214 toddlers (13-24 months of age), 150 pre-school children (3-5 years of age), and 47 schoolchildren or adolescents (6-18 years of age) tested HAdV-positive (see **Figure 2** for age distributions). In 44% of all HAdV-positive patients (n=249/571) no other viral pathogen was detected. **Table 1** gives an overview of baseline patient characteristics.

### *Above-Average Disease Severity in HAdV-Positive Neonates*

Among all 6,073 ILI patients, the mean baseline ViVI Score value reflecting disease severity was 14.5, representing the 50<sup>th</sup> percentile of the Cohort whereas the mean baseline ViVI Score value of all 571 HAdV-positive patients was 13.8 (≈46<sup>th</sup> percentile), thus indicating below-average disease severity in HAdV-positive patients in general.

Among all HAdV-positive patients (n = 571), mean ViVI Scores were lower in HAdV mono-infections (n=249, mean baseline ViVI Scores of 12.8, ≈40<sup>th</sup> percentile) compared to HAdV-positive patients with at least one viral co-infection (n = 344, mean baseline ViVI Scores of 14.6, ≈50<sup>th</sup> percentile). Patients with HAdV mono-infection also had the lowest rates of need for hospitalization, critical care, and oxygen supplementation compared to the overall cohort average and patients with co-infection(s) (**Table 1**).

Comparing disease severity among different age groups of HAdV-positive patients however, neonates were the only group presenting with an above-average baseline ViVI Score of 14.8 (≈57<sup>th</sup> percentile) and HAdV-positive neonates were the only HAdV-infected patient group where case fatalities occurred (see clinical vignettes below, **Neonate #1** and **#2**). The case fatality rate of HAdV-positive neonates was the highest of the entire cohort (0.3% vs. 0.1% overall case fatality rate).



*Clinical Vignettes of Fatal HAdV Infection*

**Neonate #1:** A male, previously healthy full-term neonate presented on day-of-life 13 with cough/rhinitis and  $T_{max}$  39.4°C, rapidly progressing to generalized compartment syndrome, acute respiratory distress syndrome (ARDS), and encephalitis with fatal outcome in week 5 of life, despite intensive care, extracorporeal membrane oxygenation (ECMO), and ventricular-peritoneal shunt. ViVI Scores peaked at 33 (99<sup>th</sup> percentile). Intravenous immunoglobulin and cidofovir (1 mg/kg body weight, combined with probenecid) were administered on hospital days 6 and 14. Tracheobronchial lavage samples on day 4 tested positive for HAdV species D (Cq value 24). No alternative pathogens were detected. HAdV species D was also detected in blood, stool, cerebrospinal fluid, and urine. **Figure 3A** illustrates disease progression with continuously increasing ViVI Score values reflecting progressive disease, as well as intensifying pulmonary opacification and abdominal distension on babygram images. **Figure 3B** shows virus load dynamics in different body compartments with a plateau in blood samples and decreasing virus loads in urine after cidofovir administration.

**Neonate #2:** A female pre-term neonate with persistent ductus arteriosus and left ventricular dysfunction developed ARDS requiring ECMO following delivery. The ViVI Score was 22 (90<sup>th</sup> percentile). Nasopharyngeal swabs tested positive for HAdV (Cq value 33) and human bocavirus (Cq value 35); an external ear swab grew *Listeria monocytogenes* serotype 4b. The patient expired on day 2 of life despite maximum life support, high-dose intravenous corticosteroids, and catecholamines.

*Severe Lower Airway Infection is Characteristic of HAdV-Positive Neonates*

Feature selection analysis was performed to identify clinical characteristics associated with HAdV infection in different age groups. Feature selection confirmed that HAdV-positive neonates have a distinctive clinical presentation compared to older children. Severe lower respiratory tract infections, often requiring critical care, are hallmarks of adenovirus infection in this vulnerable age group (relative importance levels  $\geq 80$  and  $\geq 60$ , respectively). Each of the neonatal cases in the digital/virologic surveillance program showed significant lower airway disease leading to respiratory distress and

hospitalization. Children >2 years of age were more likely to experience mild upper respiratory or systemic symptoms such as fever and headache, rarely requiring hospitalization. **Figure 4** depicts relative variable importance levels for top-5 clinical features of HAdV-positive patients in different age strata.

*Molecular Characterization of HAdV-D80 – A Novel Adenovirus*

A tracheobronchial lavage sample from **Neonate #1** underwent whole-genome sequencing at the German Reference Laboratory for Adenoviruses, leading to the discovery of a novel adenovirus, HAdV-D80 (GenBank accession number: KY618679).

Whole-genome phylogenetic analysis revealed HAdV-D22 as the closest relative of the novel HAdV-D80 genotype (**Figure 5A**). See **Figures 5-6** for phylogenetic clustering of the novel HAdV-D80 gene regions with other HAdV-D types.

**Discussion**

We report a precision medicine study combining systematic virological testing with real-time digital clinical assessments using a previously validated disease severity score via mobile application. This nested approach allowed us to eliminate reporting bias while performing comprehensive machine learning and pattern recognition analysis, and to determine the real-world impact of specific respiratory viral infections in a pediatric ILI cohort.

*Arriving at Precision Diagnoses Using m-Health Technology*

We established a standard operating procedure for the prospective precision screening<sup>29</sup> of all consecutive pediatric patients presenting to one of the busiest pediatric hospitals in Europe, during an observation period of six years<sup>4,30</sup>. The rigorous use of precision medicine methodology and pre-defined entry criteria<sup>31</sup> helped to diminish reporting bias, with a known denominator of 6,073 ILI patients. Enabling precision analysis, each patient was assessed using the VIVI Score mobile application at the point-of-care. The app allows the user to assess the most critical items depicting individual risk and disease severity in patients with respiratory viral infections. The severity score has been described elsewhere<sup>4</sup>, and is based on systematic literature review and prospective validation in more than 10,000 patients to date<sup>4,6</sup>.

Key parameters reflecting the patient's clinical presentation were collected, requiring the user to determine the presence and absence of specific signs and symptoms at the time of assessment. This allows capturing pertinent negative data with precision, e.g. the *absence* of a specific sign or symptom as well as its presence. This provides a significant advantage over data mining from routine care medical records, which do not always allow the distinction between the absence of a finding versus non-assessment. By forcing the user to make this discernment at the bedside, the VIVI Score Mobile App eventually eliminated indeterminate data<sup>4,10,32</sup>. In addition, the app verified whether a patient fulfilled a simplified ILI case definition (body temperature  $\geq 38^{\circ}\text{C}$  and  $\geq 1$  respiratory symptom/s)<sup>4,30,32</sup>. Verification of case criteria at the point-of-care has been shown to be particularly useful in prospective

surveillance, providing highly standardized datasets while avoiding selection and ascertainment bias<sup>33</sup>. Of note, the resulting precision dataset facilitates downstream applications, such as machine learning and pattern recognition analysis, as well as the testing of cases against alternative case definitions by the US Centers for Disease Control and Prevention (CDC) or the European CDC, thus facilitating meta-analyzability of data<sup>32</sup>.

#### *Honing in on Emerging Pathogens - Adenovirus Infection in ILI Patients*

Using pre-defined precision algorithms, this study is focused on ILI patients who tested positive for HAdV. Specifically, we aimed to investigate disease severity and characteristic clinical patterns associated with HAdV infection and to delineate potential age strata at risk of severe clinical manifestation.

Human adenoviruses are emerging pathogens which can cause mild to severe or even fatal disease. In healthy individuals and especially in children, HAdV infection is usually deemed common and mild<sup>34</sup>. Therefore, testing might be performed rarely in routine care<sup>32</sup>. With rigorous PCR testing for HAdV of any ILI patient regardless of the doctor's clinical suspicion within our inception cohort, we specified that respiratory HAdV infection is most common in children aged 1-5 years, but rare in neonates.

Our understanding of disease severity associated with HAdV infection in children presenting with ILI symptoms usually comes from anecdotal or epidemiologic reports and retrospective chart review<sup>34-37</sup>. To measure disease severity, case fatality rates or surrogate parameters such as 'need for hospitalization/ critical care', or 'oxygen requirement' have traditionally been used. These however, tend to be coarse (fatal vs. non-fatal) or biased<sup>4</sup>. To enable precision disease severity scoring, we used the ViVI Disease Severity Score at the point-of-care. We found out that HAdV infection in hospitalized children in general was associated with below-average disease severity. In neonates though, HAdV infection led to above-average disease severity (57<sup>th</sup> percentile) with a mortality rate of 50%. Our study is the first to provide a precision measure of disease severity associated with HAdV infection in pediatric ILI patients, delineating neonates at increased risk of severe HAdV infection.

Use of the VIVI Score App offered the opportunity of portraying a vivid picture of what HAdV infection would typically look like in different age strata. Using a machine learning methodology adjusted for small sample sizes in our study, we showed that severe lower respiratory tract infection was characteristic of HAdV-positive neonates. Fever was a distinctive feature among older HAdV-positive children but rarely present in neonates. Given our findings, adenovirus diagnostics should be performed in afebrile neonates with lower respiratory tract infection or acute respiratory distress, especially if no other pathogen is found.

#### *Discovery of a Novel Virus - A Note on HAdV-D80*

In-depth virological analysis of HAdV strains has previously proven to be of clinical relevance in that certain species and genotypes may be associated with distinct clinical pictures and/or disease severity, e.g. HAdV-B7 accounting for severe respiratory tract infection<sup>36,38,39</sup>. In our study, whole-genome sequencing of the clinical isolate of one deceased HAdV-positive neonate led to the discovery of a novel HAdV genotype, according to the Human Adenovirus Working Group: HAdV-D80.

HAdV-D80 belongs to species (D) and shows recombination events “hijacking” neutralization determinants, which may contribute to immune evasion and increased virulence<sup>40</sup>. Disease severity and virus dissemination as seen in the neonatal HAdV-D80 index case presented herein may also be driven by immature immune responses<sup>41</sup>, lack of maternal antibodies, and the timing of virus transmission<sup>7</sup>. Our findings affirm previous studies that the host’s age may have considerable impact on infectious disease outcomes, thus corroborating the investigation of vulnerable patient groups, e.g. neonates with HAdV infection.

Our study illustrates the need for licensed adenovirus treatment options. By now, no antivirals are licensed for HAdV treatment in the neonatal age bracket. Currently, cidofovir is the only licensed adenovirus treatment, but (off-label) use in children remains limited due to its nephrotoxicity. Brincidofovir, a cidofovir ester with favorable bioavailability and safety profiles may become an alternative<sup>42</sup> but has only been tested in subjects aged  $\geq 2$  months of age<sup>43</sup>.

### *Handling Real-World Patient Data*

Within this digital surveillance study, we successfully used m-Health technology to gather real-world patient data. Real-world data have been leveraged increasingly, especially during the coronavirus infectious disease 2019 (COVID-19) pandemic. Use of real-world evidence has become common practice for those aiming to improve pandemic preparedness and response<sup>44,45</sup>. Regulatory agencies and stakeholders from both, academia and industry, endorse compliance with framework conditions for real-world data to become actionable<sup>44-48</sup>. In a contemporary opinion article, Pillai outlined key premises and avenues towards meaningful real-world data and systems in the context of pandemic preparedness<sup>45</sup>. The VIVI Score Mobile App was designed to facilitate timely and accurate diagnoses in real-time and to enable comprehensive clinical assessments and decision support, including a standardized disease severity measure. Data harmonization and compliance with international data standards were among the key concerns in the development process. Not only is the VIVI Score comparable and meta-analyzable across different sites but the app is constructed in full compliance with Clinical Data Interchange Standards Consortium (CDISC) standards<sup>49</sup>, allowing for swift data interoperability and instant readouts to regulatory agencies issuing the same standards<sup>4</sup>, e.g. the United States Food and Drug Administration (FDA)<sup>50</sup>.

### *Limitations*

The project had several strengths and limitations. First, the number of HAdV-positive neonates was very low. Albeit this being a finding itself, it impedes statistical analysis. For this reason, we refrained from performing classical significance testing but applied biostatistics analysis and machine learning methodology adjusted for very small case numbers. Second, only the HAdV-D80 prototype isolate underwent deep sequencing allowing for exact phylogenetic analysis. However, this isolate was the only sample characterized as species (D) by Fluorescence Melting Curve Analysis (FMCA), thereby attracting special attention in a set of samples from patients with respiratory disease. Third, this was

a single-center study, therefore potentially restricting the variance of findings, especially regarding the circulation of HAdV genotypes. Yet, the number of patients seen during six years of surveillance was relatively high in the setting of one of Europe's largest, supra-regional pediatric tertiary care centers.

### *Outlook*

Use of the methodologies described herein at multiple sites in both children and adults and including patient-generated data will be continued within the Partnering for Enhanced Digital Surveillance of Influenza-like Disease and the Effect of Antivirals and vaccines (PEDSIDEA) project<sup>5</sup>. This will allow additional investigation of factors influencing disease severity associated with viral respiratory infections, e.g. biomarkers such as single nucleotide polymorphisms of the host<sup>51,52</sup>. Also, further clarification of the role of emerging pathogens such as HAdV-D80 is warranted, including investigation of its role at the opposite extreme of age, i.e. the elderly. Future studies will focus on the importance of potential virus-virus interactions in case one individual has been infected by multiple viruses<sup>53</sup>. In this regard, use of novel virus detection techniques, e.g. agnostic metagenomic next-generation sequencing will soon find its way into the diagnostic workup, thus allowing to diminish bias in the laboratory context and enable detection of pathogens that have so far been unknown as has been the case with MERS-CoV or SARS-CoV-1 and -2.

### *Conclusion*

We expect the precision medicine approach combining the use of m-Health technology at the bedside with machine learning and pattern recognition analysis and in-depth virology to be useful for virus outbreak situations by facilitating timely signal detection and clinical decision support.

By the example of respiratory HAdV infection, we identified neonates being at risk of life-threatening disease based on a real-world patient data set. In this vulnerable patient group, rapid-turnaround diagnostics and effective antivirals are needed. Whole-genome sequencing lead to the discovery of a novel virus: HAdV-D80.

**Experimental Procedures***Resource Availability**Lead Contact*

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact Barbara Rath (barbara.rath@vi-vi.org, @Vi\_VIorg).

*Materials Availability*

The VIVI Disease Severity Score Mobile App can be downloaded via <https://score.vi-vi.org/> or from the App Store or Google Play store.

*Data Availability*

The datasets supporting the current study have not been deposited in a public repository because of patient data privacy but are available from the Lead Contact on request by qualified researchers.

*Method Details**Overall Study Design*

From 12/2009-04/2015, 6,073 pediatric in- and outpatients (median age: 1.6 years, range 0-18.8 years, 56% male) with ILI (fever  $\geq 38^{\circ}\text{C}$  and  $\geq 1$  respiratory symptom/s and/or physician diagnosis of ILI) participated in a prospective surveillance and quality management (QM) program at the Charité Department of Pediatrics in Berlin, Germany (institutional review board number: EA4/008/10) in collaboration with the National Reference Centre for Influenza at the Robert Koch-Institute<sup>4,30</sup>.

The study was performed in compliance with the ethical standards of the Declaration of Helsinki and according to German laws. Written informed consent was waived for the purpose of quality improvement and infection control given the observational nature of the study. Verbal informed consent was obtained from all patients and/or parents or caretakers of underage patients.



As described previously<sup>4,30</sup>, an independent QM team used a mobile application that was designed based on the user-centered principles of design-thinking to assess patients in a standardized manner, computing the ViVi (Vienna Vaccine Safety Initiative) Disease Severity Score in real-time<sup>4,30,54,55</sup>. The ViVi Score is a published and validated 22-item weighed composite clinical score ranging from 0-48, with increasing values reflecting increasing disease severity<sup>4,56</sup>. ViVi Score parameters are weighed in accordance with WHO criteria for uncomplicated, complicated, and progressive disease<sup>31</sup>.

In severely ill patients, follow-up clinical assessments and PCR-testing were performed every other day until resolution of symptoms and/or hospital discharge. Nasopharyngeal and/or tracheobronchial specimens were delivered to the Robert Koch-Institute for PCR-testing for human adenoviruses as well as influenza virus A/B, respiratory syncytial virus (RSV), human rhinovirus, metapneumovirus, bocavirus, parainfluenza, and betacoronaviruses in all patients<sup>4,30,56</sup>. When disseminated disease was suspected clinically, blood, cerebrospinal fluid, bronchial lavage, and urine samples were also tested, if available.

Age groups were defined as follows:

- Neonates: 0-28 days
- Infants: 29 days-12 months
- Toddlers: 13-24 months
- Pre-school children: 3-5 years
- Schoolchildren/adolescents: 6-18 years.

The Cohort was developed in compliance with the STROBE (Strengthening the Reporting of Observational studies in Epidemiology) statement<sup>57</sup>. **Figure 1** displays the precision screening process.

#### *Adenovirus Molecular Testing*

In all patients, real-time PCR-testing for HAdV was performed with a set of oligonucleotides capable of detecting species A-F<sup>58</sup>, with modified reaction conditions at the Robert Koch-Institute. Quantitation cycle (Cq) values during PCR-testing were used to estimate virus load. Cq values are reversely proportional to the logarithm of the virus load, i.e. low Cq values indicate high virus load<sup>59</sup>. In patients with disseminated disease, the PCR amplicon was genotyped by Fluorescence Melting Curve Analysis (FMCA)<sup>58</sup>. Virus culture was performed on A549 cells in Minimum Essential Medium in a closed system. Cells were cultured in 25 cm<sup>2</sup> flasks and infected with 50-200 µl of sterile filtered sample material. Following cytopathic effect, cells were freeze-thawed and harvested for further analysis.

#### *Whole-Genome Sequencing and Phylogenetic Analysis*

For whole-genome sequencing, HAdV clinical isolates were transferred to the Reference Laboratory for Adenoviruses in Hannover, Germany. A genomic library was prepared using the Nextera XT DNA sample preparation kit (Illumina, San Diego, CA)<sup>60</sup>. A total DNA input of 1 ng extracted from cell culture supernatant was simultaneously fragmented and ligated to adaptors using a transposon-based approach. The library was subjected to quality control prior to sequencing on an Illumina MiSeq, generating 300 bp paired-end reads. Fastq-file quality was tested before de novo assembly using CLC Genomics Workbench (Qiagen, version 8, Aarhus, Denmark). Phylogenetic analysis was performed in Mega7 (Kumar, Stecher, and Tamura 2015). Neighbor-joining trees were constructed applying a Kimura 2-parameter model approach and bootstrapping was performed with 500 replicates.

#### *Statistics and Bioinformatic Analysis*

Descriptive analyses were performed using SPSS Statistics 22 (IBM Corp. Armonk, NY). Clinical features were analyzed using the Correlation-based Feature Selection Measure to explore distinctive disease patterns associated with HAdV infection<sup>61</sup>. Variable importance was computed following the permutation principle of the mean decrease in accuracy importance. We used the Permutation Feature Importance Measure to compare significance levels across clinical variables. This method has

advantages over other measures in the case of small patient groups since it works directly on the data and does not have any underlying model assumptions<sup>62</sup>. The method measures the importance of a feature by calculating the increase of the model's prediction error after the feature values have been permuted. A feature's importance increases proportionally with the increase of the model error (or decrease of the accuracy) after the permutation, since in this case the model is sensitive to this feature for the prediction and it is thus important. The relative importance of the most important feature is set to 100 (non-dimensional); values of other features are scaled accordingly.

CONFIDENTIAL

**Table 1.** Baseline patient characteristics.

	<b>Total (n=6,073)</b>	<b>HAdV-positive (n=571, 9%)</b>			<b>HAdV-negative (n=5,502, 91%)</b>
<b>Median Age (range)</b>	1.6 years (0-18.8 years)	1.5 years (0-17.1 years)			1.6 years (0-18.8 years)
		Overall (n=571)	HAdV mono-infection <sup>†</sup> (n=249)	Viral co-infection detected <sup>†</sup> (n=322)	
<b>Gender</b>					
<b>Male</b>	3,399 (56%)	314 (55%)	139 (56%)	175 (54%)	3,085 (56%)
<b>Female</b>	2,674 (44%)	257 (45%)	110 (44%)	147 (46%)	2,417 (44%)
<b>Outpatient</b>	2,033 (33%)	239 (42%)	116 (47%)	123 (38%)	1,794 (33%)
<b>Neonates<sup>a</sup></b>	9 (0%)	0 (0%)	0 (0%)	0 (0%)	9 (1%)
<b>Infants<sup>b</sup></b>	496 (24%)	55 (23%)	24 (10%)	31 (10%)	441 (25%)
<b>Toddlers<sup>c</sup></b>	474 (23%)	82 (34%)	36 (15%)	46 (14%)	392 (22%)
<b>Pre-school children<sup>d</sup></b>	592 (29%)	80 (33%)	45 (18%)	35 (11%)	512 (29%)
<b>Schoolchildren/   adolescents<sup>e</sup></b>	462 (23%)	22 (9%)	11 (4%)	11 (3%)	440 (25%)
<b>Inpatient</b>	4,040 (67%)	332 (58%)	133 (53%)	199 (62%)	3,708 (67%)
<b>Neonates<sup>a</sup></b>	127 (3%)	4 (1%)	1 (0%)	3 (1%)	123 (3%)
<b>Infants<sup>b</sup></b>	1,408 (35%)	101 (30%)	39 (16%)	62 (19%)	1,307 (35%)
<b>Toddlers<sup>c</sup></b>	969 (24%)	132 (40%)	42 (17%)	90 (28%)	837 (23%)
<b>Pre-school children<sup>d</sup></b>	861 (21%)	70 (21%)	36 (15%)	34 (14%)	791 (21%)
<b>Schoolchildren/   adolescents<sup>e</sup></b>	675 (17%)	25 (8%)	15 (6%)	10 (3%)	650 (18%)
<b>O<sub>2</sub> therapy</b>	1,513 (25%)	98 (17%)	29 (12%)	69 (21%)	1,415 (26%)
<b>Neonates<sup>a</sup></b>	41 (3%)	2 (2%)	1 (0%)	1 (0%)	39 (3%)
<b>Infants<sup>b</sup></b>	529 (35%)	28 (29%)	8 (3%)	20 (6%)	501 (35%)
<b>Toddlers<sup>c</sup></b>	405 (27%)	40 (41%)	10 (4%)	30 (9%)	365 (26%)
<b>Pre-school children<sup>d</sup></b>	365 (24%)	23 (23%)	9 (4%)	14 (4%)	342 (24%)
<b>Schoolchildren/   adolescents<sup>e</sup></b>	173 (11%)	5 (5%)	1 (0%)	4 (1%)	168 (12%)
<b>Need for ICU<sup>†</sup> admission</b>	1,303 (21%)	71 (12%)	25 (10%)	46 (14%)	1,232 (22%)
<b>Neonates<sup>a</sup></b>	75 (6%)	2 (3%)	1 (0%)	1 (0%)	73 (6%)
<b>Infants<sup>b</sup></b>	531 (41%)	19 (27%)	3 (1%)	16 (5%)	512 (42%)
<b>Toddlers<sup>c</sup></b>	240 (18%)	26 (37%)	10 (4%)	16 (5%)	214 (17%)
<b>Pre-school children<sup>d</sup></b>	238 (18%)	16 (23%)	6 (2%)	10 (3%)	222 (18%)
<b>Schoolchildren/   adolescents<sup>e</sup></b>	219 (17%)	8 (11%)	5 (2%)	3 (1%)	211 (17%)

<b>Fatal outcome</b>	13 (0.1%)	2 (0.3%)	1 (0%)	1 (0%)	11 (0.2%)
<b>Neonates<sup>a</sup></b>	2 (15%)	2 (100%)	1 (0%)	1 (0%)	0 (0%)
<b>Infants<sup>b</sup></b>	5 (38%)	0 (0%)	0 (0%)	0 (0%)	5 (45%)
<b>Toddlers<sup>c</sup></b>	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
<b>Pre-school children<sup>d</sup></b>	2 (15%)	0 (0%)	0 (0%)	0 (0%)	2 (18%)
<b>Schoolchildren/   adolescents<sup>e</sup></b>	4 (31%)	0 (0%)	0 (0%)	0 (0%)	4 (36%)
<b>Mean baseline ViVI Score (range)</b>	14.5 (0-34)	13.8 (0-34)	12.8 (2-34)	14.6 (0-29)	14.6 (0-33)
<b>Neonates<sup>a</sup></b>	14.0 (2-25)	14.8 (10- 22)	10 (10)	16.3 (12-22)	13.9 (2-25)
<b>Infants<sup>b</sup></b>	14.6 (0-33)	13.7 (2-27)	12.3 (4-27)	14.6 (2-27)	14.7 (0-33)
<b>Toddlers<sup>c</sup></b>	15.3 (0-32)	14.3 (0-31)	13.5 (5-31)	14.8 (0-29)	15.4 (0-32)
<b>Pre-school children<sup>d</sup></b>	14.6 (0-32)	13.6 (0-28)	13.0 (2-28)	14.3 (0-28)	14.7 (0-32)
<b>Schoolchildren/   adolescents<sup>e</sup></b>	13.4 (0-34)	12.7 (3-34)	11.3 (3-34)	14.3 (6-25)	13.4 (0-33)

<sup>a</sup>Neonates: 0-28 days

<sup>b</sup>Infants: 29 days-12 months

<sup>c</sup>Toddlers: 13-24 months

<sup>d</sup>Pre-school children: 3-5 years

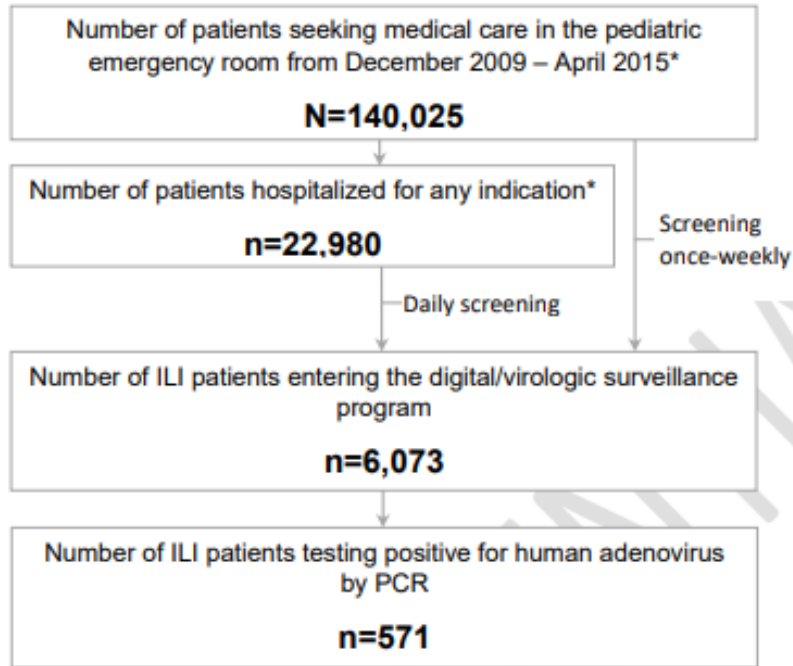
<sup>e</sup>Schoolchildren/adolescents: 6-18 years

\*Intensive care unit

<sup>†</sup>PCR testing performed for influenza A/B, respiratory syncytial virus, rhinovirus, metapneumovirus, bocavirus, parainfluenza, and coronavirus in addition to HAdV

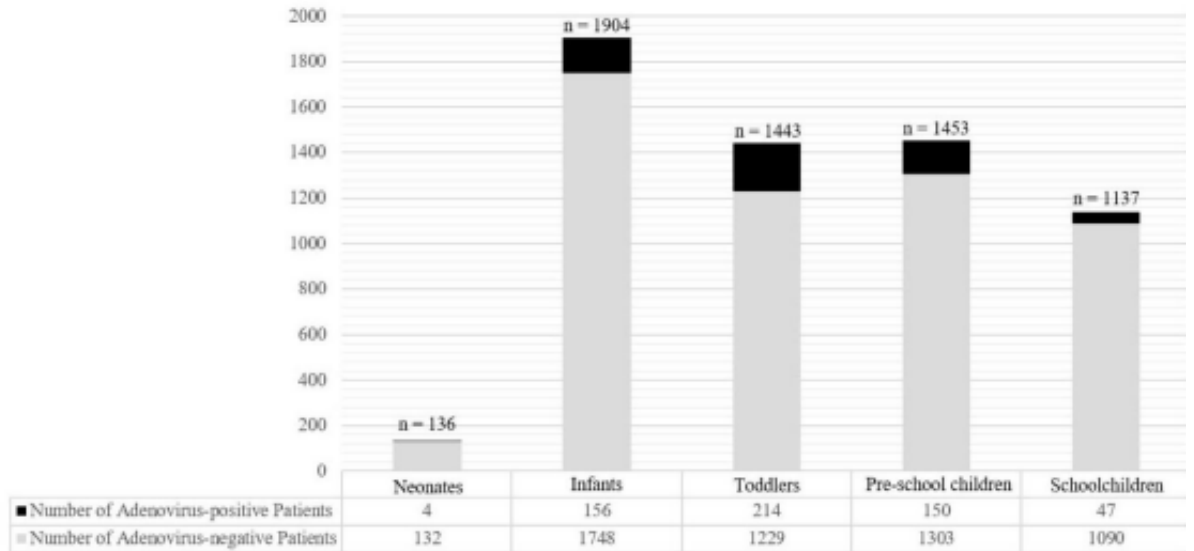
## Figures

Figure 1.



\*numbers extrapolated based on a 3-year average

Precision screening flow chart

**Figure 2.****Age Groups:**

Neonates: 0-28 days

Infants: 29 days-12 months

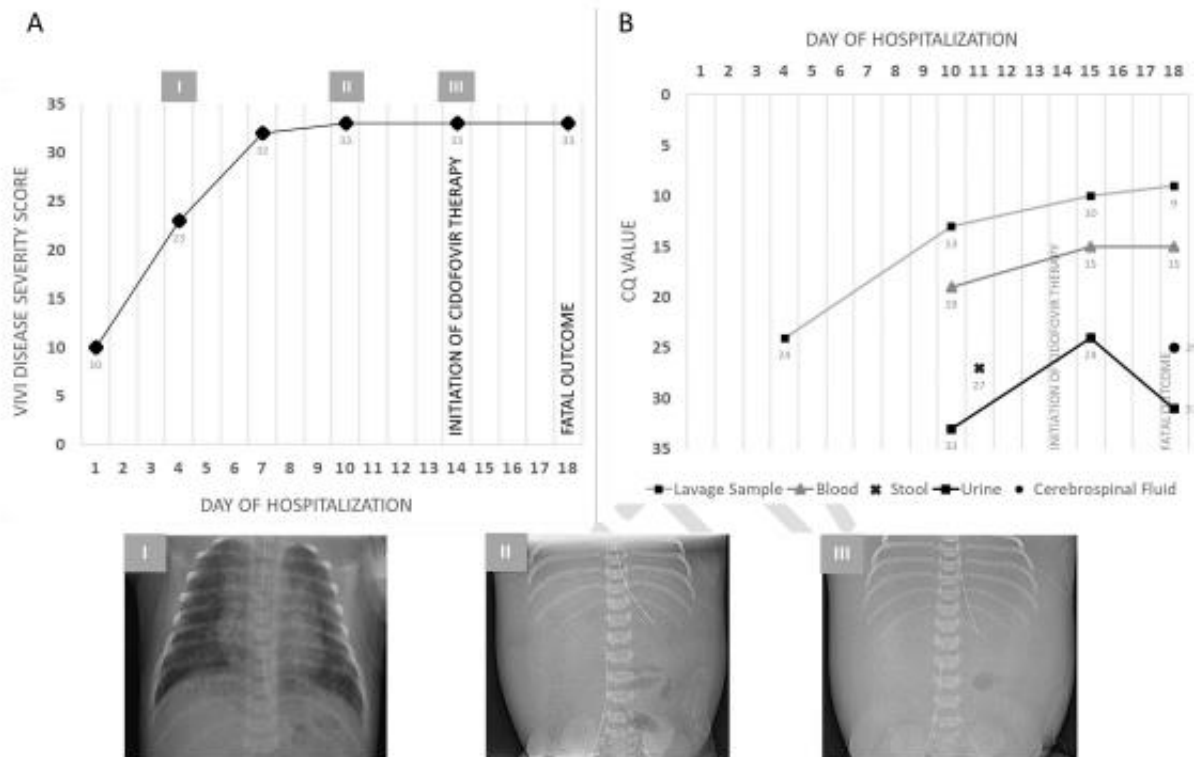
Toddlers: 13-24 months

Pre-school children: 3-5 years

Schoolchildren/adolescents: 6-18 years

Age distribution of HAdV-positive (dark gray) and HAdV-negative (light gray) patients within the digital/virologic surveillance cohort. Age groups: Neonates: 0-28 days, infants: 29 days-12 months, toddlers: 13-24 months, pre-school children: 3-5 years, schoolchildren/adolescents: 6-18 years.

Figure 3.



VIVI Score dynamics illustrating disease progression in the HAdV-D80 index case. Initial X-ray on hospital day 4 (I) showed diffuse patchy lung infiltrates, corresponding to respiratory distress syndrome grade I. Follow-up imaging on hospital day 10 (II) showed homogenic lung opacity due to diffuse bilateral atelectasis ('white lungs'), corresponding to respiratory distress syndrome grade IV. Follow-up imaging on hospital day 14 (III) revealed generalized compartment syndrome with ongoing abdominal swelling, pleural and pericardial effusion. B) Adenovirus load kinetics in different body compartments over time: Cq values (inverted vertical axis) during PCR-testing were used to estimate virus load. Abbreviations: VIVI – Vienna Vaccine Safety Initiative, Cq – Quantitation Cycle.

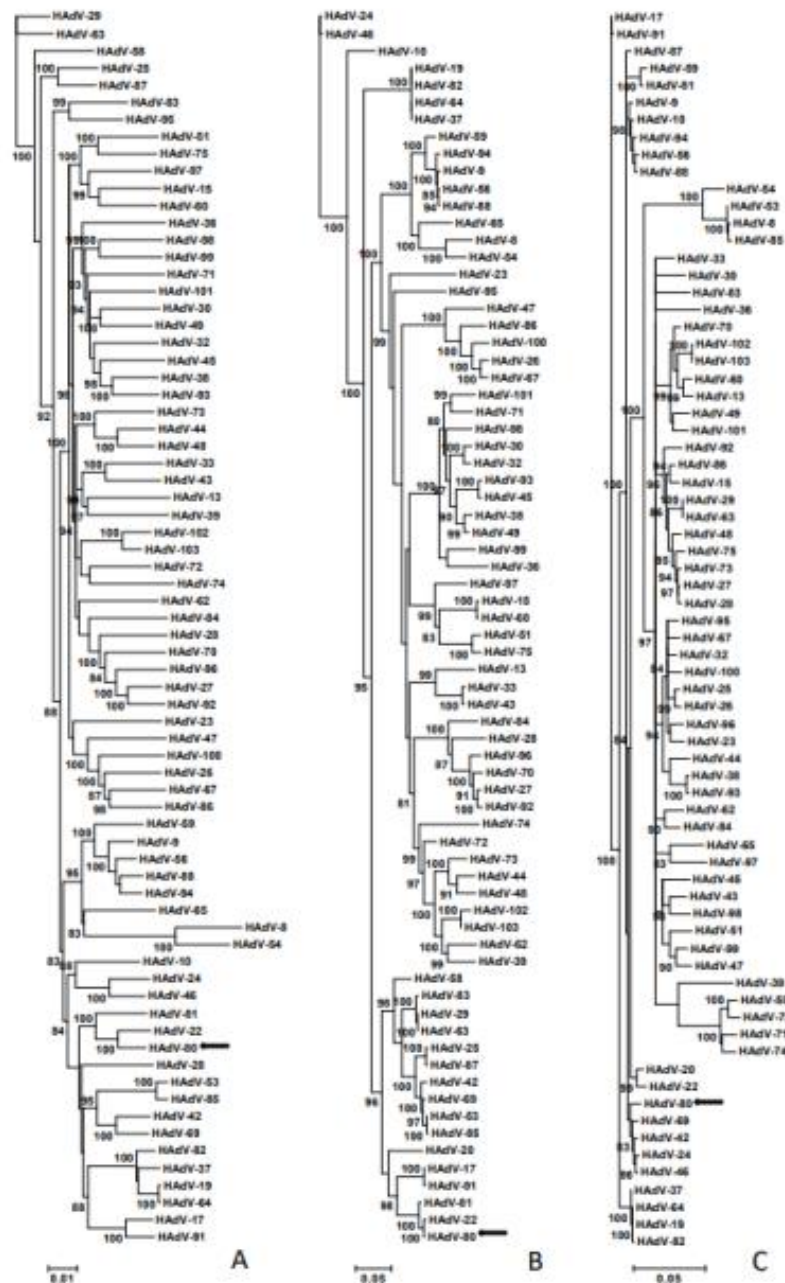


**Figure 4.**

	Neonates	Infants	Toddlers	Pre-school children	Schoolchildren/ adolescents
Hospitalization				Black	
ICU admission		Gray			
Oropharyngeal inflammation	Black		Black		
Cough		Gray			Black
Rhinitis			Black		Gray
Clinical/radiological signs of upper respiratory tract	Black				
Clinical/radiological signs of lower respiratory tract	Gray	Gray			
Hypoxia (O <sub>2</sub> -sat <93%)			Black		
Dyspnea					Black
High fever (>40°C) beyond 3 days					Gray
Malaise (reduced general condition)				Black	
Diarrhea			Black	Black	
Headache					Gray

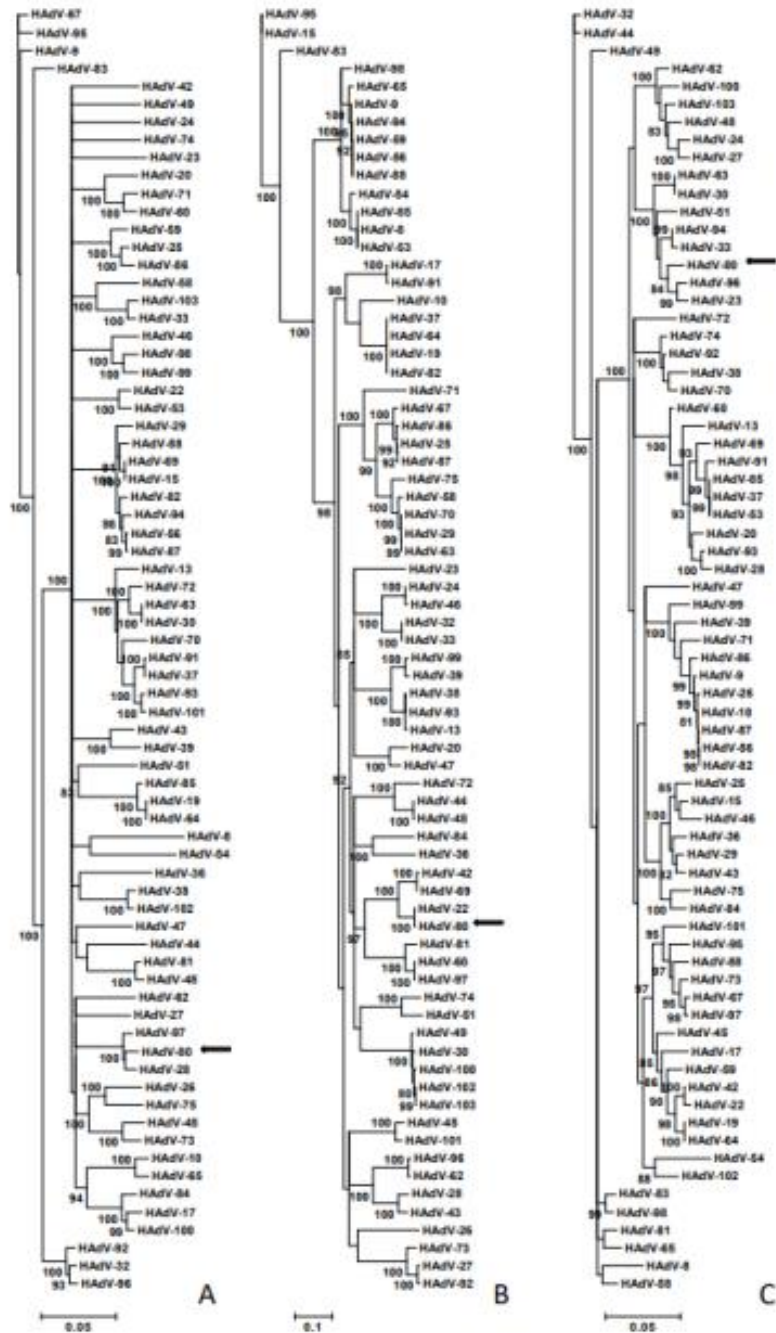
Relative variable importance analysis for top-5 clinical features of HAdV-positive patients (n=571) using the Permutation Feature Importance Measure. The relative feature importance of the most important feature is set to 100. The values of the other features are scaled accordingly relative to the most important feature. Level of importance of top-5 features is reflected as follows: white indicates importance <33, gray indicates importance ≥33 and <66, black indicates importance ≥66.

Figure 5.



Phylogenetic clustering of the HAdV-D80 A) whole genome, B) early gene region 4 (E4), and C) early gene region 3 (E3) sequence with all other HAdV-D prototype sequences. Bootstrap values >80% were considered robust. Arrow indicates the clustering position of HAdV-D80.

Figure 6.



Phylogenetic clustering of the HAAdV-D80 A) hexon gene, B) fiber gene, C) penton base gene sequence with all other HAAdV-D prototype sequences. Bootstrap values >80% were considered robust. Arrow indicates the clustering position of HAAdV-D80.

**Acknowledgements**

For data management and sample collection, we thank Christian Hoppe, Susann Muehlhans, Franziska Tief, Katharina Karsch, Xi Chen, and Lea Seeber. The authors are also grateful to Eleni Adamou and the laboratory teams at the National Reference Centre for Influenza at the Robert Koch-Institute in Berlin and the National Reference Laboratory for Adenoviruses at Hannover Medical School.

**Author Contributions**

Conceptualization and supervision of the work: BR, BS. Development of methodology: BR, BS. Conducting formal analysis: BR, BS, AH, BB, TC, MA, EH, PO. Provision of study materials, reagents, materials, patients, laboratory samples, computing resources, and other analysis tools: BR, BS, AH, TC. Data curation: BR, BS, AH, PO. Writing of the original draft: PO. Review and editing of the manuscript draft: all authors. Project administration: BR. Funding (in-kind): BR, BS.

**Declaration of Interests**

The authors declare no competing interests.

## References

1. Committee on a Framework for Developing a New Taxonomy of Disease; National Research Council. Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease. Washington DC, The National Academies Press 2011.
2. Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med* 2015; **372**(9): 793-5.
3. Gwinn M, MacCannell D. Infectious Diseases: Precision Medicine for Public Health. 2018. <https://blogs.cdc.gov/genomics/2015/09/24/infectious-diseases/> (accessed March 06, 2021).
4. Rath B, Conrad T, Myles P, et al. Influenza and other respiratory viruses: standardizing disease severity in surveillance and clinical trials. *Expert Rev Anti Infect Ther* 2017; **15**(6): 545-68.
5. Rath B, Maltezou HC, Papaevangelou V, et al. Partnering for enhanced digital surveillance of influenza-like disease and the effect of antivirals and vaccines (PEDSIDEA). *Influenza Other Respir Viruses* 2019; **13**(4): 309-18.
6. Rath B, Seng S. Partnering for Enhanced Digital Surveillance of Influenza-like Disease and the Effect of Antivirals and Vaccines (PEDSIDEA) - First Results in the United States. 2020. <https://doi.org/10.1542/peds.146.1.MeetingAbstract.29> (accessed 27th of September 2021).
7. Glynn JR, Moss PAH. Systematic analysis of infectious disease outcomes by age shows lowest severity in school-age children. *Sci Data* 2020; **7**(1): 329.
8. Li X, Willem L, Antillon M, Bilcke J, Jit M, Beutels P. Health and economic burden of respiratory syncytial virus (RSV) disease and the cost-effectiveness of potential interventions against RSV among children under 5 years in 72 Gavi-eligible countries. *BMC Med* 2020; **18**(1): 82.
9. Panetta L, Proulx C, Drouin O, et al. Clinical Characteristics and Disease Severity Among Infants With SARS-CoV-2 Infection in Montreal, Quebec, Canada. *JAMA Netw Open* 2020; **3**(12): e2030470.
10. Obermeier P, Muehlhans S, Hoppe C, et al. Enabling Precision Medicine With Digital Case Classification at the Point-of-Care. *EBioMedicine* 2016; **4**: 191-6.
11. Abzug MJ, Levin MJ. Neonatal adenovirus infection: four patients and review of the literature. *Pediatrics* 1991; **87**(6): 890-6.
12. Ronchi A, Doern C, Brock E, Pugni L, Sanchez PJ. Neonatal adenoviral infection: a seventeen year experience and review of the literature. *J Pediatr* 2014; **164**(3): 529-35 e1-4.
13. Elnifro EM, Cooper RJ, Dady I, Hany S, Mughal ZM, Klapper PE. Three nonfatal cases of neonatal adenovirus infection. *J Clin Microbiol* 2005; **43**(11): 5814-5.
14. Henquell C, Boeuf B, Mirand A, et al. Fatal adenovirus infection in a neonate and transmission to health-care workers. *J Clin Virol* 2009; **45**(4): 345-8.
15. Debast SB, van Heyst AF, Bergman KA, Galama JM. A case of severe adenovirus pneumonia in a neonate. *Clin Microbiol Infect* 1996; **1**(4): 278-9.
16. Chiou CC, Soong WJ, Hwang B, Wu KG, Lee BH, Wang HC. Congenital adenoviral infection. *Pediatr Infect Dis J* 1994; **13**(7): 664-5.
17. Baserga M, Chan B. Hematochezia and thrombocytopenia in a 3-day-old infant: Congenital adenoviral infection. *J Neonatal Perinatal Med* 2018; **11**(3): 335-8.
18. Censoplano N, Gorga S, Waldeck K, Stillwell T, Rabah-Hammad R, Flori H. Neonatal Adenovirus Infection Complicated by Hemophagocytic Lymphohistiocytosis Syndrome. *Pediatrics* 2018; **141**(Suppl 5): S475-S480.
19. Kelley CJ. A fatal case of neonatal adenovirus infection. *Neonatal Netw* 2010; **29**(5): 297-305.
20. Montone KT, Furth EE, Pietra GG, Gupta PK. Neonatal adenovirus infection: a case report with in situ hybridization confirmation of ascending intrauterine infection. *Diagn Cytopathol* 1995; **12**(4): 341-4.
21. Rieger-Fackeldey E, Aumeier S, Genzel-Boroviczeny O. Disseminated adenovirus infection in two premature infants. *Infection* 2000; **28**(4): 237-9.
22. Auletta JJ, Sanchez PJ, Meyer EK, et al. Adjuvant haploidentical virus-specific T lymphocytes for treatment of disseminated adenovirus infection in a premature infant. *J Allergy Clin Immunol* 2019; **144**(2): 594-7 e4.

23. Kim JS, Han HS, Park SH, Chun YK, Lee HJ, Chi JG. Neonatal adenoviral pneumonia--report of three autopsy cases. *J Korean Med Sci* 1997; **12**(2): 146-50.
24. Bajonowski T, Wiegand P, Cecchi R, et al. Detection and significance of adenoviruses in cases of sudden infant death. *Virchows Arch* 1996; **428**(2): 113-8.
25. Castelli JB, Siciliano RF, Vieira RD, Aiello VD, Strabelli TM. Fatal adenoviral necrotizing bronchiolitis case in a post-cardiac surgery intensive care unit. *Braz J Infect Dis* 2011; **15**(3): 285-7.
26. Kinney JS, Hierholzer JC, Thibeault DW. Neonatal pulmonary insufficiency caused by adenovirus infection successfully treated with extracorporeal membrane oxygenation. *J Pediatr* 1994; **125**(1): 110-2.
27. Angella JJ, Connor JD. Neonatal infection caused by adenovirus type 7. *J Pediatr* 1968; **72**(4): 474-8.
28. Moallem M, Song E, Jaggi P, Conces MR, Kajon AE, Sanchez PJ. Adenovirus and "Culture-Negative Sepsis" in a Preterm Neonate. *AJP Rep* 2016; **6**(4): e417-e20.
29. Wang M, Zhou Y, Zong Z, et al. A precision medicine approach to managing 2019 novel coronavirus pneumonia. *Precis Clin Med* 2020; **3**(1): 14-21.
30. Tief F, Hoppe C, Seeber L, et al. An inception cohort study assessing the role of pneumococcal and other bacterial pathogens in children with influenza and ILI and a clinical decision model for stringent antibiotic use. *Antivir Ther* 2016; **21**(5): 413-24.
31. WHO. Clinical management of human infection with pandemic (H1N1) 2009: revised guidance 2009. [http://www.who.int/csr/resources/publications/swineflu/clinical\\_management\\_h1n1.pdf](http://www.who.int/csr/resources/publications/swineflu/clinical_management_h1n1.pdf) (accessed 10th of January 2017).
32. Alchikh M, Conrad T, Hoppe C, et al. Are we missing respiratory viral infections in infants and children? Comparison of a hospital-based quality management system with standard of care. *Clin Microbiol Infect* 2019; **25**(3): 380 e9- e16.
33. Cameron AR, Meyer A, Faverjon C, Mackenzie C. Quantification of the sensitivity of early detection surveillance. *Transbound Emerg Dis* 2020; **67**(6): 2532-43.
34. Gaunt ER, Harvala H, McIntyre C, Templeton KE, Simmonds P. Disease burden of the most commonly detected respiratory viruses in hospitalized patients calculated using the disability adjusted life year (DALY) model. *J Clin Virol* 2011; **52**(3): 215-21.
35. Wu PQ, Zeng SQ, Yin GQ, et al. Clinical manifestations and risk factors of adenovirus respiratory infection in hospitalized children in Guangzhou, China during the 2011-2014 period. *Medicine (Baltimore)* 2020; **99**(4): e18584.
36. Cai R, Mao N, Dai J, et al. Genetic variability of human adenovirus type 7 circulating in mainland China. *PLoS One* 2020; **15**(4): e0232092.
37. Goikhman Y, Drori Y, Friedman N, et al. Adenovirus load correlates with respiratory disease severity among hospitalized pediatric patients. *Int J Infect Dis* 2020; **97**: 145-50.
38. Gray GC, McCarthy T, Lebeck MG, et al. Genotype prevalence and risk factors for severe clinical adenovirus infection, United States 2004-2006. *Clin Infect Dis* 2007; **45**(9): 1120-31.
39. Lee J, Choi EH, Lee HJ. Clinical severity of respiratory adenoviral infection by serotypes in Korean children over 17 consecutive years (1991-2007). *J Clin Virol* 2010; **49**(2): 115-20.
40. Walsh MP, Chintakuntlawar A, Robinson CM, et al. Evidence of molecular evolution driven by recombination events influencing tropism in a novel human adenovirus that causes epidemic keratoconjunctivitis. *PLoS One* 2009; **4**(6): e5635.
41. Adkins B, Leclerc C, Marshall-Clarke S. Neonatal adaptive immunity comes of age. *Nat Rev Immunol* 2004; **4**(7): 553-64.
42. Ison MG, Hayden RT. Adenovirus. *Microbiol Spectr* 2016; **4**(4).
43. ClinicalTrials.gov [Internet]. Bethesda (MD): National Library of Medicine (US). 6 Studies found for: brincidofovir | Adenovirus | Child. [https://clinicaltrials.gov/ct2/results?term=brincidofovir&cond=Adenovirus&Search=Apply&age\\_v=&age=0&gndr=&type=&rslt=](https://clinicaltrials.gov/ct2/results?term=brincidofovir&cond=Adenovirus&Search=Apply&age_v=&age=0&gndr=&type=&rslt=) (accessed 15th of May 2021).
44. Drury RE, O'Connor D. Symptom study app provides real-world data on COVID-19 vaccines. *Lancet Infect Dis* 2021.

45. Pillai P. How Do Data Bolster Pandemic Preparedness and Response? How Do We Improve Data and Systems to Be Better Prepared? *Patterns (N Y)* 2021; **2**(1): 100190.
46. Berger ML, Sox H, Willke RJ, et al. Good practices for real-world data studies of treatment and/or comparative effectiveness: Recommendations from the joint ISPOR-ISPE Special Task Force on real-world evidence in health care decision making. *Pharmacoepidemiol Drug Saf* 2017; **26**(9): 1033-9.
47. Dolgin E. Core Concept: The pandemic is prompting widespread use-and misuse-of real-world data. *Proc Natl Acad Sci U S A* 2020; **117**(45): 27754-8.
48. Jarow JP, LaVange L, Woodcock J. Multidimensional Evidence Generation and FDA Regulatory Decision Making: Defining and Using "Real-World" Data. *JAMA* 2017; **318**(8): 703-4.
49. Souza T, Kush R, Evans JP. Global clinical data interchange standards are here! *Drug Discov Today* 2007; **12**(3-4): 174-81.
50. U.S. Food & Drug Administration. STUDY DATA STANDARDS: WHAT YOU NEED TO KNOW. 2017. <https://www.fda.gov/files/drugs/published/Study-Data-Standards--What-You-Need-to-Know.pdf> (accessed 15th of May 2021).
51. Ghafouri-Fard S, Noroozi R, Vafaei R, et al. Effects of host genetic variations on response to, susceptibility and severity of respiratory infections. *Biomed Pharmacother* 2020; **128**: 110296.
52. Forbester JL, Humphreys IR. Genetic influences on viral-induced cytokine responses in the lung. *Mucosal Immunol* 2021; **14**(1): 14-25.
53. Nickbakhsh S, Mair C, Matthews L, et al. Virus-virus interactions impact the population dynamics of influenza and the common cold. *Proc Natl Acad Sci U S A* 2019.
54. Rath B. Editorial: Communicating vaccine safety - the Vienna Vaccine Safety Initiative. *Curr Drug Saf* 2015; **10**(1): 3-4.
55. Rath B. Vienna Vaccine Safety Initiative. *Hum Vaccin Immunother* 2017: 1-4.
56. Ma X, Conrad T, Alchikh M, Reiche J, Schweiger B, Rath B. Can we distinguish respiratory viral infections based on clinical features? A prospective pediatric cohort compared to systematic literature review. *Rev Med Virol* 2018; **28**(5): e1997.
57. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* 2007; **370**(9596): 1453-7.
58. Chmielewicz B, Nitsche A, Schweiger B, Ellerbrok H. Development of a PCR-based assay for detection, quantification, and genotyping of human adenoviruses. *Clin Chem* 2005; **51**(8): 1365-73.
59. Chen X, Pouran Yousef K, Duwe S, et al. Quantitative influenza follow-up testing (QIFT)--a novel biomarker for the monitoring of disease activity at the point-of-care. *PLoS One* 2014; **9**(3): e92500.
60. Hage E, Gerd Liebert U, Bergs S, Ganzenmueller T, Heim A. Human mastadenovirus type 70: a novel, multiple recombinant species D mastadenovirus isolated from diarrhoeal faeces of a haematopoietic stem cell transplantation recipient. *J Gen Virol* 2015; **96**(9): 2734-42.
61. Hall MA. Correlation-Based Feature Subset Selection for Machine Learning. Hamilton, New Zealand: University of Waikato; 1998.
62. Strobl C, Boulesteix AL, Zeileis A, Hothorn T. Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC Bioinformatics* 2007; **8**: 25.

## 4 CONCLUSIONS AND PERSPECTIVES

---

The four studies reported in this PhD thesis have led to the following conclusions for enhanced infectious disease surveillance and management

- 1) The combination of digital case classification based on standardized case criteria and agnostic viral mNGS can diminish ascertainment bias and increase diagnostic yield in pediatric encephalitis and ADEM. Standardized clinical data for precision diagnoses based on consensus case criteria will enable harmonized interpretation of potential linkages between disease entities and certain pathogens. For now, application of viral mNGS in encephalitis and ADEM patients may require confirmation using target-specific confirmatory testing for technical limitations.
- 2) Use of the ViVI ScoreApp provides standardized disease severity assessments at the point-of-care. Combined use with an agnostic virus detection method helps to increase diagnostic yield, diminish bias, and detect unexpected viruses. In ILI patients, mNGS should be used in conjunction with virus-specific PCR testing.
- 3) For mere detection and cursory identification, partial sequencing of viral genomes might suffice, as shown with HAdV infection within the German national surveillance system. As opposed to whole-genome sequencing however, partial sequencing can impede detection of recombinant and novel strains, leading to erroneous conclusions. Linking whole-genome data to detailed and standardized clinical information might help to better understand the clinical impact of a virus strain.
- 4) A dual precision medicine approach combining digital surveillance and in-depth virology can be useful in the context of virus outbreak situations by facilitating rapid signal detection, decision support models based on pattern recognition analysis, and infection control measures. In an exemplary proof-of-concept study, we showed that HAdV infection is rare in neonates but can rapidly progress to life-threatening disease with age-specific clinical characteristics. Whole-genome sequencing lead to the discovery of a novel virus: HAdV-D80.



Beyond the above-mentioned conclusions, the PhD project at issue has provided food for thought on open research questions, follow-up studies, and further real-world application to better prepare for virus epidemics and pandemics and to ultimately curb disease burden.

In Studies 1 and 2, we successfully used viral mNGS in small subsets of patients, serving as proof-of-concept. However, ongoing large-scale application is warranted to investigate the clinical relevance of findings and to enable the identification of very rare and/or novel associations. For this purpose, standard operating procedures for mNGS, including sample preparation, target enrichment, and bioinformatics work steps are a prerequisite<sup>195,205-208</sup>. To facilitate the implementation of agnostic methodologies in diagnostic microbiological laboratories, national and international guidelines need to account for the ever-growing evidence of mNGS serving as useful auxiliary means to resolve complex cases. To this end, health authorities such as FDA-ARGOS (dAtabase for Regulatory-Grade micrObial Sequences) are committed to help shape basic framework conditions for quality-controlled and standardized transition of innovative mNGS into routine diagnostic workflows<sup>209</sup>. Future studies will shed light on other particular fields of (m)NGS application besides recombination analysis as discussed in Study 3, e.g. the practical value of ultra-deep sequencing to investigate quasispecies or single nucleotide polymorphisms with potential implications on pathogenicity, drug resistance and/or vaccine escape<sup>27,195,197,208,210</sup>.

As indicated in Study 4, host factors including but not limited to age and pre-existing conditions will influence disease severity and susceptibility. On the molecular level, -omics analyses characterizing the individual immune response are of high interest. Originating from the same two digital surveillance programs, we continue to link highly granular clinical data including a meta-analyzable disease severity measure to potential biomarker molecules, e.g. alarmins as regulatory proteins of immune response that could help to identify patient risk groups ahead of time in the proper sense of precision medicine<sup>211,212</sup>.

Study 4 has also demonstrated the immense value of standardized real-world data enabling subsequent machine learning analysis to portray the clinical manifestations of viral diseases and to eventually arrive at decision support models for virus prediction. Nailing down a potential pathogen based on clinical

symptoms, seasonality, and pre-disposing factors will help to allocate resources, especially in low-resource settings where diagnostic tests might be sparse and early virus identification is key to prevent infectious disease spread or to distribute medication.

If we continue to collect standardized data deploying digital health technology over multiple seasons, bioinformatics analysis will perpetuate steady adjustment of diagnostic algorithms, thus accounting for potential changes in clinical pictures associated with certain pathogens. On this behalf, further application of the ViVI Score App is already underway under the umbrella of the multicenter PEDSIDEA (Partnering for Enhanced Digital Surveillance of Influenza-like Disease and the Effect of Antivirals and vaccines) project<sup>39</sup>. Within this quality improvement program, special focus lies on the implementation of patient-reported outcomes captured by a user-friendly application, the ViVI ScoreMe App that will help patients keep track of their symptoms based on scientific criteria (<https://www.vi-vi.org/pedsidea---at-home.html>). For the comprehensive monitoring of infectious diseases involving patient-generated health data, future directions will also include leveraging smart tech such as wearables to collect respiratory and heart rates, body temperature, locomotion patterns or even to perform invasive measurements of specific metabolites<sup>213</sup>.

In any case, future infectious disease surveillance will generate a ‘high volume’ of data, at ‘high velocity’, with a ‘high variety’ of data forms and ultimately, ‘high veracity’ of data will have to be preserved to guarantee big data be of ‘high value’ in real-world application<sup>214</sup>. In the new era of emerging and re-emerging infectious diseases, not only medicolegal aspects and ethical questions on how to generate and curate big health data will have to be considered but standardization in any aspect will be key to meet the challenges coming along with *“one of the few genuine adventures left in the world”*<sup>1</sup>.

## RÉFÉRENCES

1. Zinsser H. Rats, Lice, and History. Boston: Little, Brown & Co; 1935. Chapter I, para 4.
2. Valleron AJ, Meurisse S, Boelle PY. Historical Analysis of the 1889–1890 Pandemic in Europe. *International Journal of Infectious Diseases* 2008; **12**(E95).
3. Valleron AJ, Cori A, Valtat S, Meurisse S, Carrat F, Boelle PY. Transmissibility and geographic spread of the 1889 influenza pandemic. *Proc Natl Acad Sci U S A* 2010; **107**(19): 8778-81.
4. Baginsky A. Die Influenza-Epidemie 1889/90. In: Guttmann S, Leyden E, editors.: Verlag von J.F. Bergmann; 1892.
5. Mulder J, Masurel N. Pre-epidemic antibody against 1957 strain of Asiatic influenza in serum of older people living in the Netherlands. *Lancet* 1958; **1**(7025): 810-4.
6. Taubenberger JK, Morens DM, Fauci AS. The next influenza pandemic: can it be predicted? *JAMA* 2007; **297**(18): 2025-7.
7. Nguyen-Van-Tam JS, Hampson AW. The epidemiology and clinical impact of pandemic influenza. *Vaccine* 2003; **21**(16): 1762-8.
8. Honigsbaum M. "An inexpressible dread": psychoses of influenza at fin-de-siecle. *Lancet* 2013; **381**(9871): 988-9.
9. Kousoulis AA, Tsoucalas G. Infection, contagion and causality in Colonial Britain: the 1889-90 influenza pandemic and the British Medical Journal. *Infez Med* 2017; **25**(3): 285-91.
10. Taubenberger JK, Hultin JV, Morens DM. Discovery and characterization of the 1918 pandemic influenza virus in historical context. *Antivir Ther* 2007; **12**(4 Pt B): 581-91.
11. Simon CE. THE FILTERABLE VIRUSES. *Physiological Reviews* 1923; **3**(4).
12. Gelderblom H, Krüger D. Helmut Ruska (1908–1973): His Role in the Evolution of Electron Microscopy in the Life Sciences, and Especially Virology. *Advances in Imaging and Electron Physics* 2014; **182**: 1-94.
13. Ackermann HW. The first phage electron micrographs. *Bacteriophage* 2011; **1**(4): 225-7.
14. Zhang Y, Hung T, Song J, He J. Electron microscopy: essentials for viral structure, morphogenesis and rapid diagnosis. *Sci China Life Sci* 2013; **56**(5): 421-30.
15. Woolhouse M, Scott F, Hudson Z, Howey R, Chase-Topping M. Human viruses: discovery and emergence. *Philos Trans R Soc Lond B Biol Sci* 2012; **367**(1604): 2864-71.
16. Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* 1998; **5**(10): R245-9.
17. Xie G, Yu J, Duan Z. New strategy for virus discovery: viruses identified in human feces in the last decade. *Sci China Life Sci* 2013; **56**(8): 688-96.
18. Handelsman J. Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev* 2004; **68**(4): 669-85.
19. Declich S, Carter AO. Public health surveillance: historical origins, methods and evaluation. *Bull World Health Organ* 1994; **72**(2): 285-304.
20. Ziegler T, Mamahit A, Cox NJ. 65 years of influenza surveillance by a World Health Organization-coordinated global network. *Influenza Other Respir Viruses* 2018; **12**(5): 558-65.
21. World Health Organization. WHO Recommended Surveillance Standards. Second edition, 1999.
22. Obermeier P, Muehlhans S, Hoppe C, et al. Enabling Precision Medicine With Digital Case Classification at the Point-of-Care. *EBioMedicine* 2016; **4**: 191-6.

23. Forum of International Respiratory Societies. The Global Impact of Respiratory Disease – Second Edition. Sheffield, European Respiratory Society, 2017.
24. Office of Public Health Scientific Services. Centers for Disease Control and Prevention. PUBLIC HEALTH SURVEILLANCE PREPARING FOR THE FUTURE. 2018.
25. Tief F, Hoppe C, Seeber L, et al. An inception cohort study assessing the role of pneumococcal and other bacterial pathogens in children with influenza and ILI and a clinical decision model for stringent antibiotic use. *Antivir Ther* 2016; **21**(5): 413-24.
26. Rath B, Tief F, Obermeier P, et al. Early detection of influenza A and B infection in infants and children using conventional and fluorescence-based rapid testing. *J Clin Virol* 2012; **55**(4): 329-33.
27. Rath B, von Kleist M, Tief F, et al. Virus load kinetics and resistance development during oseltamivir treatment in infants and children infected with Influenza A(H1N1) 2009 and Influenza B viruses. *Pediatr Infect Dis J* 2012; **31**(9): 899-905.
28. Rath B, Tief F, Karsch K, et al. Towards a personalised approach to managing influenza infections in infants and children - food for thought and a note on oseltamivir. *Infect Disord Drug Targets* 2013; **13**(1): 25-33.
29. Tuttle R, Weick A, Schwarz WS, et al. Evaluation of novel second-generation RSV and influenza rapid tests at the point of care. *Diagn Microbiol Infect Dis* 2015; **81**(3): 171-6.
30. Rath B, Conrad T, Myles P, et al. Influenza and other respiratory viruses: standardizing disease severity in surveillance and clinical trials. *Expert Rev Anti Infect Ther* 2017; **15**(6): 545-68.
31. Rath B, Chen X, Spies V, et al. Prospective surveillance of antiviral resistance in hospitalized infants less than 12 months of age with A(H3N2) influenza infection and treated with oseltamivir. *Antivir Ther* 2017; **22**(6): 515-22.
32. Chen X, Pouran Yousef K, Duwe S, et al. Quantitative influenza follow-up testing (QIFT)--a novel biomarker for the monitoring of disease activity at the point-of-care. *PLoS One* 2014; **9**(3): e92500.
33. Karsch K, Obermeier P, Seeber L, et al. Human Parechovirus Infections Associated with Seizures and Rash in Infants and Toddlers. *Pediatr Infect Dis J* 2015; **34**(10): 1049-55.
34. Hoppe C, Obermeier P, Muehlhans S, et al. Innovative Digital Tools and Surveillance Systems for the Timely Detection of Adverse Events at the Point of Care: A Proof-of-Concept Study. *Drug Saf* 2016; **39**(10): 977-88.
35. Obermeier PE, Karsch K, Hoppe C, et al. Acute Disseminated Encephalomyelitis After Human Parechovirus Infection. *Pediatr Infect Dis J* 2016; **35**(1): 35-8.
36. Ma X, Conrad T, Alchikh M, Reiche J, Schweiger B, Rath B. Can we distinguish respiratory viral infections based on clinical features? A prospective pediatric cohort compared to systematic literature review. *Rev Med Virol* 2018; **28**(5): e1997.
37. Alchikh M, Conrad T, Hoppe C, et al. Are we missing respiratory viral infections in infants and children? Comparison of a hospital-based quality management system with standard of care. *Clin Microbiol Infect* 2019; **25**(3): 380 e9- e16.
38. Obermeier P. Automated case classification of neurologic adverse events at the point-ofcare and enhanced virologic surveillance – new avenues towards quality improvement and precision medicine Charité University Medical Center; 2018.
39. Rath B, Maltezou HC, Papaevangelou V, et al. Partnering for enhanced digital surveillance of influenza-like disease and the effect of antivirals and vaccines (PEDSIDEA). *Influenza Other Respir Viruses* 2019; **13**(4): 309-18.
40. Sheleme T, Bekele F, Ayela T. Clinical Presentation of Patients Infected with Coronavirus Disease 19: A Systematic Review. *Infect Dis (Auckl)* 2020; **13**: 1178633720952076.

41. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA* 2020; **323**(13): 1239-42.
42. Collaborators GBDN. Global, regional, and national burden of neurological disorders, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol* 2019; **18**(5): 459-80.
43. World Health Organization. Neurological disorders: public health challenges, 2006.
44. Kakooza-Mwesige A, Mohammed AH, Kristensson K, Juliano SL, Lutwama JJ. Emerging Viral Infections in Sub-Saharan Africa and the Developing Nervous System: A Mini Review. *Front Neurol* 2018; **9**: 82.
45. Rath B, Magnus M, Heininger U. Evaluating the Brighton Collaboration case definitions, aseptic meningitis, encephalitis, myelitis, and acute disseminated encephalomyelitis, by systematic analysis of 255 clinical cases. *Vaccine* 2010; **28**(19): 3488-95.
46. Xing XW, Zhang JT, Ma YB, et al. Metagenomic Next-Generation Sequencing for Diagnosis of Infectious Encephalitis and Meningitis: A Large, Prospective Case Series of 213 Patients. *Front Cell Infect Microbiol* 2020; **10**: 88.
47. Wilson MR, Sample HA, Zorn KC, et al. Clinical Metagenomic Sequencing for Diagnosis of Meningitis and Encephalitis. *N Engl J Med* 2019; **380**(24): 2327-40.
48. McRae JE, Quinn HE, Saravanos GL, et al. Paediatric Active Enhanced Disease Surveillance (PAEDS) annual report 2016: Prospective hospital-based surveillance for serious paediatric conditions. *Commun Dis Intell (2018)* 2019; **43**.
49. Centers for Disease Control and Prevention. Case definitions for infectious conditions under public health surveillance. *MMWR Recomm Rep* 1997; **46**(RR-10): 1-55.
50. Smura T, Blomqvist S, Kolehmainen P, et al. Aseptic meningitis outbreak associated with echovirus 4 in Northern Europe in 2013-2014. *J Clin Virol* 2020; **129**: 104535.
51. Tapiainen T, Prevots R, Izurieta HS, et al. Aseptic meningitis: case definition and guidelines for collection, analysis and presentation of immunization safety data. *Vaccine* 2007; **25**(31): 5793-802.
52. Twilt M. Precision Medicine: The new era in medicine. *EBioMedicine* 2016; **4**: 24-5.
53. Verheij RA, Curcin V, Delaney BC, McGilchrist MM. Possible Sources of Bias in Primary Care Electronic Health Record Data Use and Reuse. *J Med Internet Res* 2018; **20**(5): e185.
54. Ramachandran PS, Wilson MR. Diagnostic Testing of Neurologic Infections. *Neurol Clin* 2018; **36**(4): 687-703.
55. Hanson KE. The First Fully Automated Molecular Diagnostic Panel for Meningitis and Encephalitis: How Well Does It Perform, and When Should It Be Used? *J Clin Microbiol* 2016; **54**(9): 2222-4.
56. World Health Organization. Laboratory biosafety manual. Third Edition. 2004.
57. Freed NE, Myers CA, Russell KL, et al. Diagnostic discrimination of live attenuated influenza vaccine strains and community-acquired pathogenic strains in clinical samples. *Mol Cell Probes* 2007; **21**(2): 103-10.
58. Huang HS, Tsai CL, Chang J, Hsu TC, Lin S, Lee CC. Multiplex PCR system for the rapid diagnosis of respiratory virus infection: systematic review and meta-analysis. *Clin Microbiol Infect* 2018; **24**(10): 1055-63.
59. Nelson PP, Rath BA, Fragkou PC, Antalis E, Tsiodras S, Skevaki C. Current and Future Point-of-Care Tests for Emerging and New Respiratory Viruses and Future Perspectives. *Front Cell Infect Microbiol* 2020; **10**: 181.
60. BioMÉRIEUX. BIOFIRE® FILMARRAY® ME Panel. 2020. <https://www.biomerieux-diagnostics.com/filmarray-meningitis-encephalitis-me-panel> (accessed 05th July 2020).

61. Mostyn A, Lenihan M, O'Sullivan D, et al. Assessment of the FilmArray® multiplex PCR system and associated meningitis/encephalitis panel in the diagnostic service of a tertiary hospital. *Infection Prevention in Practice* June 2020; **2**(2).
62. Tansarli GS, Chapin KC. Diagnostic test accuracy of the BioFire(R) FilmArray(R) meningitis/encephalitis panel: a systematic review and meta-analysis. *Clin Microbiol Infect* 2020; **26**(3): 281-90.
63. Ginocchio CC, Zhang F, Malhotra A, et al. Development, technical performance, and clinical evaluation of a NucliSens basic kit application for detection of enterovirus RNA in cerebrospinal fluid. *J Clin Microbiol* 2005; **43**(6): 2616-23.
64. Debiasi RL, Tyler KL. Molecular methods for diagnosis of viral encephalitis. *Clin Microbiol Rev* 2004; **17**(4): 903-25, table of contents.
65. Lona-Reyes JC, Valdez-Nunez AL, Cordero-Zamora A, Rea-Rosas A, Ascencio-Esparza EP, Orozco-Alatorre LG. [Meningoencefalitis: etiología infecciosa en pacientes pediátricos de un hospital de referencia]. *Bol Med Hosp Infant Mex* 2018; **75**(4): 231-6.
66. Ambrose HE, Granerod J, Clewley JP, et al. Diagnostic strategy used to establish etiologies of encephalitis in a prospective cohort of patients in England. *J Clin Microbiol* 2011; **49**(10): 3576-83.
67. Granerod J, Ambrose HE, Davies NW, et al. Causes of encephalitis and differences in their clinical presentations in England: a multicentre, population-based prospective study. *Lancet Infect Dis* 2010; **10**(12): 835-44.
68. Glaser CA, Honarmand S, Anderson LJ, et al. Beyond viruses: clinical profiles and etiologies associated with encephalitis. *Clin Infect Dis* 2006; **43**(12): 1565-77.
69. Benninger F, Steiner I. CSF in acute and chronic infectious diseases. *Handb Clin Neurol* 2017; **146**: 187-206.
70. Wilson MR, O'Donovan BD, Gelfand JM, et al. Chronic Meningitis Investigated via Metagenomic Next-Generation Sequencing. *JAMA Neurol* 2018; **75**(8): 947-55.
71. Saha S, Ramesh A, Kalantar K, et al. Unbiased Metagenomic Sequencing for Pediatric Meningitis in Bangladesh Reveals Neuroinvasive Chikungunya Virus Outbreak and Other Unrealized Pathogens. *mBio* 2019; **10**(6).
72. Xing XW, Zhang JT, Ma YB, Zheng N, Yang F, Yu SY. Apparent performance of metagenomic next-generation sequencing in the diagnosis of cryptococcal meningitis: a descriptive study. *J Med Microbiol* 2019; **68**(8): 1204-10.
73. Kufner V, Plate A, Schmutz S, et al. Two Years of Viral Metagenomics in a Tertiary Diagnostics Unit: Evaluation of the First 105 Cases. *Genes (Basel)* 2019; **10**(9).
74. Haston JC, Rostad CA, Jerris RC, et al. Prospective Cohort Study of Next-Generation Sequencing as a Diagnostic Modality for Unexplained Encephalitis in Children. *J Pediatric Infect Dis Soc* 2019.
75. Edridge AWD, Deijs M, van Zeggeren IE, et al. Viral Metagenomics on Cerebrospinal Fluid. *Genes (Basel)* 2019; **10**(5).
76. Miller S, Naccache SN, Samayoa E, et al. Laboratory validation of a clinical metagenomic sequencing assay for pathogen detection in cerebrospinal fluid. *Genome Res* 2019; **29**(5): 831-42.
77. Rodriguez C, Gricourt G, Ndebi M, et al. Fatal Encephalitis Caused by Cristoli Virus, an Emerging Orthobunyavirus, France. *Emerg Infect Dis* 2020; **26**(6): 1287-90.
78. Hong NTT, Anh NT, Mai NTH, et al. Performance of Metagenomic Next-Generation Sequencing for the Diagnosis of Viral Meningoencephalitis in a Resource-Limited Setting. *Open Forum Infect Dis* 2020; **7**(3): ofaa046.

79. Goldberg B, Sichtig H, Geyer C, Ledebner N, Weinstock GM. Making the Leap from Research Laboratory to Clinic: Challenges and Opportunities for Next-Generation Sequencing in Infectious Disease Diagnostics. *mBio* 2015; **6**(6): e01888-15.
80. Meyding-Lamadé U. LEITLINIE Virale Meningoenzephalitis: Kommission Leitlinien der Deutschen Gesellschaft für Neurologie, 2018.
81. Stahl JP, Azouvi P, Bruneel F, et al. Guidelines on the management of infectious encephalitis in adults. *Med Mal Infect* 2017; **47**(3): 179-94.
82. Greenlee JE. Encefalite. MANUALE MSD Versione per i professionisti; 2019.
83. Ellul M, Solomon T. Acute encephalitis - diagnosis and management. *Clin Med (Lond)* 2018; **18**(2): 155-9.
84. Sigfrid L, Perfect C, Rojek A, et al. A systematic review of clinical guidelines on the management of acute, community-acquired CNS infections. *BMC Med* 2019; **17**(1): 170.
85. Mokili JL, Rohwer F, Dutilh BE. Metagenomics and future perspectives in virus discovery. *Curr Opin Virol* 2012; **2**(1): 63-77.
86. Lipkin WI. The changing face of pathogen discovery and surveillance. *Nat Rev Microbiol* 2013; **11**(2): 133-41.
87. Ruegamer T, Hoffmann R, Rohrhofer A, et al. Inhibition of HIV-1 infection by human pegivirus type 1-derived peptides is affected by human pegivirus type 1 genotype and HIV-1 coreceptor tropism. *AIDS* 2018; **32**(14): 1951-7.
88. Resch B, Puchas C, Resch E, Urlsberger B. Epidemiology of Respiratory Syncytial Virus-related Hospitalizations and the Influence of Viral Coinfections in Southern Austria in a 7-year Period. *Pediatr Infect Dis J* 2020; **39**(1): 12-6.
89. Hill AB. The Environment and Disease: Association or Causation? *Proc R Soc Med* 1965; **58**: 295-300.
90. Geneletti S, Gallo V, Porta M, Khoury MJ, Vineis P. Assessing causal relationships in genomics: From Bradford-Hill criteria to complex gene-environment interactions and directed acyclic graphs. *Emerg Themes Epidemiol* 2011; **8**(1): 5.
91. Iturriza-Gomara M, Auchterlonie IA, Zaw W, Molyneaux P, Desselberger U, Gray J. Rotavirus gastroenteritis and central nervous system (CNS) infection: characterization of the VP7 and VP4 genes of rotavirus strains isolated from paired fecal and cerebrospinal fluid samples from a child with CNS disease. *J Clin Microbiol* 2002; **40**(12): 4797-9.
92. Harvala H, Broberg E, Benschop K, et al. Recommendations for enterovirus diagnostics and characterisation within and beyond Europe. *J Clin Virol* 2018; **101**: 11-7.
93. Cordey S, Schibler M, L'Huillier AG, et al. Comparative analysis of viral shedding in pediatric and adult subjects with central nervous system-associated enterovirus infections from 2013 to 2015 in Switzerland. *J Clin Virol* 2017; **89**: 22-9.
94. Vu DL, Kaiser L. The concept of commensal viruses almost 20 years later: redefining borders in clinical virology. *Clin Microbiol Infect* 2017; **23**(10): 688-90.
95. Cadwell K, Patel KK, Maloney NS, et al. Virus-plus-susceptibility gene interaction determines Crohn's disease gene Atg16L1 phenotypes in intestine. *Cell* 2010; **141**(7): 1135-45.
96. Schatorje E, van der Flier M, Seppanen M, et al. Primary immunodeficiency associated with chromosomal aberration - an ESID survey. *Orphanet J Rare Dis* 2016; **11**(1): 110.
97. Chowdhury P, Khan SA. Significance of CCL2, CCL5 and CCR2 polymorphisms for adverse prognosis of Japanese encephalitis from an endemic population of India. *Sci Rep* 2017; **7**(1): 13716.
98. Nogales A, M LD. Host Single Nucleotide Polymorphisms Modulating Influenza A Virus Disease in Humans. *Pathogens* 2019; **8**(4).

99. Loeb M, Eskandarian S, Rupp M, et al. Genetic variants and susceptibility to neurological complications following West Nile virus infection. *J Infect Dis* 2011; **204**(7): 1031-7.
100. International Committee on Taxonomy of Viruses. Genus: Orthobunyavirus. 2019. [https://talk.ictvonline.org/ictv-reports/ictv\\_online\\_report/negative-sense-rna-viruses/bunyavirales/w/peribunyaviridae/1238/genus-orthobunyavirus](https://talk.ictvonline.org/ictv-reports/ictv_online_report/negative-sense-rna-viruses/bunyavirales/w/peribunyaviridae/1238/genus-orthobunyavirus) (accessed 12th of July 2020).
101. Edridge AWD, Deijs M, Namazzi R, et al. Novel Orthobunyavirus Identified in the Cerebrospinal Fluid of a Ugandan Child With Severe Encephalopathy. *Clin Infect Dis* 2019; **68**(1): 139-42.
102. Sejvar JJ, Kohl KS, Bilynsky R, et al. Encephalitis, myelitis, and acute disseminated encephalomyelitis (ADEM): case definitions and guidelines for collection, analysis, and presentation of immunization safety data. *Vaccine* 2007; **25**(31): 5771-92.
103. Loyse A, Burry J, Cohn J, et al. Leave no one behind: response to new evidence and guidelines for the management of cryptococcal meningitis in low-income and middle-income countries. *Lancet Infect Dis* 2019; **19**(4): e143-e7.
104. Fulton BD, Proudman DG, Sample HA, et al. Exploratory analysis of the potential for advanced diagnostic testing to reduce healthcare expenditures of patients hospitalized with meningitis or encephalitis. *PLoS One* 2020; **15**(1): e0226895.
105. Leon KE, Schubert RD, Casas-Alba D, et al. Genomic and serologic characterization of enterovirus A71 brainstem encephalitis. *Neurol Neuroimmunol Neuroinflamm* 2020; **7**(3).
106. Nilsson A, Edner N, Albert J, Ternhag A. Fatal encephalitis associated with coronavirus OC43 in an immunocompromised child. *Infect Dis (Lond)* 2020; **52**(6): 419-22.
107. World Health Organization. Ten threats to global health in 2019. 2019. <https://www.who.int/news-room/feature-stories/ten-threats-to-global-health-in-2019> (accessed 16th of July 2020).
108. Tadini E, Papamidimitriou-Olivgeris M, Opota O, et al. [SARS-CoV-2, a point in the storm]. *Rev Med Suisse* 2020; **16**(692): 917-23.
109. Lipsitch M, Swerdlow DL, Finelli L. Defining the Epidemiology of Covid-19 - Studies Needed. *N Engl J Med* 2020; **382**(13): 1194-6.
110. Centers for Disease Control and Prevention. Coronavirus Disease 2019 (COVID-19). Purpose and Methods. 15th of July 2020 2020. <https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/purpose-methods.html> (accessed 16th of July 2020).
111. Centers for Disease Control and Prevention. COVIDView Week 27, ending July 4, 2020. 09th of July 2020 2020. <https://www.cdc.gov/coronavirus/2019-ncov/covid-data/pdf/covidview-07-10-2020.pdf> (accessed 16th of July 2020).
112. Centers for Disease Control and Prevention. Weekly U.S. Influenza Surveillance Report. 10th of July 2020 2020. <https://www.cdc.gov/flu/weekly/index.htm> (accessed 16th of July 2020).
113. Choi BCK. Corrigendum to "The Past, Present, and Future of Public Health Surveillance". *Scientifica (Cairo)* 2018; **2018**: 6943062.
114. Choi BC. The past, present, and future of public health surveillance. *Scientifica (Cairo)* 2012; **2012**: 875253.
115. Gibbons CL, Mangen MJ, Plass D, et al. Measuring underreporting and underascertainment in infectious disease datasets: a comparison of methods. *BMC Public Health* 2014; **14**: 147.
116. Cain DM, Detsky AS. Everyone's a little bit biased (even physicians). *JAMA* 2008; **299**(24): 2893-5.



117. Althubaiti A. Information bias in health research: definition, pitfalls, and adjustment methods. *J Multidiscip Healthc* 2016; **9**: 211-7.
118. Casalegno JS, Eibach D, Valette M, et al. Performance of influenza case definitions for influenza community surveillance: based on the French influenza surveillance network GROG, 2009-2014. *Euro Surveill* 2017; **22**(14).
119. Aguilera JF, Paget WJ, Mosnier A, et al. Heterogeneous case definitions used for the surveillance of influenza in Europe. *Eur J Epidemiol* 2003; **18**(8): 751-4.
120. Fitzner J, Qasmieh S, Mounts AW, et al. Revision of clinical case definitions: influenza-like illness and severe acute respiratory infection. *Bull World Health Organ* 2018; **96**(2): 122-8.
121. Centers for Disease Control and Prevention. Glossary of Influenza (Flu) Terms. 2017. <https://www.cdc.gov/flu/about/glossary.htm> (accessed 16th of July 2020).
122. Hajduk GK, Jamieson NE, Baker BL, Olesen OF, Lang T. It is not enough that we require data to be shared; we have to make sharing easy, feasible and accessible too! *BMJ Glob Health* 2019; **4**(4): e001550.
123. Badawy R, Hameed F, Bataille L, et al. Metadata Concepts for Advancing the Use of Digital Health Technologies in Clinical Research. *Digit Biomark* 2019; **3**(3): 116-32.
124. Bernstein AB, Sweeney MH, Centers for Disease C, Prevention. Public health surveillance data: legal, policy, ethical, regulatory, and practical issues. *MMWR Suppl* 2012; **61**(3): 30-4.
125. Wang X, Li Y, O'Brien KL, et al. Global burden of respiratory infections associated with seasonal influenza in children under 5 years in 2018: a systematic review and modelling study. *Lancet Glob Health* 2020; **8**(4): e497-e510.
126. World Health Organization. Preparing GISRS for the upcoming influenza seasons during the COVID-19 pandemic – practical considerations. Interim Guidance as of 26 May 2020. 2020. [https://apps.who.int/iris/bitstream/handle/10665/332198/WHO-2019-nCoV-Preparing\\_GISRS-2020.1-eng.pdf?sequence=1&isAllowed=y](https://apps.who.int/iris/bitstream/handle/10665/332198/WHO-2019-nCoV-Preparing_GISRS-2020.1-eng.pdf?sequence=1&isAllowed=y).
127. Fernandez AR, Crowe RP, Bourn S, et al. COVID-19 Preliminary Case Series: Characteristics of EMS Encounters with Linked Hospital Diagnoses. *Prehosp Emerg Care* 2020: 1-23.
128. Choi J, Cho Y, Shim E, Woo H. Web-based infectious disease surveillance systems and public health perspectives: a systematic review. *BMC Public Health* 2016; **16**(1): 1238.
129. Lamos V, Miller AC, Crossan S, Stefansen C. Advances in nowcasting influenza-like illness rates using search query logs. *Sci Rep* 2015; **5**: 12760.
130. Gupta A, Katarya R. Social media based surveillance systems for healthcare using machine learning: A systematic review. *J Biomed Inform* 2020; **108**: 103500.
131. Sooknanan J, Comissiong DMG. Trending on Social Media: Integrating Social Media into Infectious Disease Dynamics. *Bull Math Biol* 2020; **82**(7): 86.
132. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. *J Med Internet Res* 2009; **11**(1): e11.
133. Eysenbach G. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. *AMIA Annu Symp Proc* 2006: 244-8.
134. Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection--harnessing the Web for public health surveillance. *N Engl J Med* 2009; **360**(21): 2153-5, 7.
135. Barclay E. Predicting the next pandemic. *Lancet* 2008; **372**(9643): 1025-6.
136. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature* 2009; **457**(7232): 1012-4.

137. Carneiro HA, Mylonakis E. Google trends: a web-based tool for real-time surveillance of disease outbreaks. *Clin Infect Dis* 2009; **49**(10): 1557-64.
138. Pervaiz F, Pervaiz M, Abdur Rehman N, Saif U. FluBreaks: early epidemic detection from Google flu trends. *J Med Internet Res* 2012; **14**(5): e125.
139. Martin LJ, Xu B, Yasui Y. Correction: Improving Google Flu Trends Estimates for the United States through Transformation. *PLoS One* 2015; **10**(4): e0122939.
140. Martin LJ, Xu B, Yasui Y. Improving Google Flu Trends estimates for the United States through transformation. *PLoS One* 2014; **9**(12): e109209.
141. Yang S, Santillana M, Kou SC. Accurate estimation of influenza epidemics using Google search data via ARGO. *Proc Natl Acad Sci U S A* 2015; **112**(47): 14473-8.
142. Cook S, Conrad C, Fowlkes AL, Mohebbi MH. Assessing Google flu trends performance in the United States during the 2009 influenza virus A (H1N1) pandemic. *PLoS One* 2011; **6**(8): e23610.
143. Butler D. When Google got flu wrong. *Nature* 2013; **494**(7436): 155-6.
144. Olson DR, Konty KJ, Paladini M, Viboud C, Simonsen L. Reassessing Google Flu Trends data for detection of seasonal and pandemic influenza: a comparative epidemiological study at three geographic scales. *PLoS Comput Biol* 2013; **9**(10): e1003256.
145. Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. *Science* 2014; **343**(6176): 1203-5.
146. Carrion M, Madoff LC. ProMED-mail: 22 years of digital surveillance of emerging infectious diseases. *Int Health* 2017; **9**(3): 177-83.
147. Rolland C, Lazarus C, Giese C, Monate B, Travert AS, Salomon J. Early Detection of Public Health Emergencies of International Concern through Undiagnosed Disease Reports in ProMED-Mail. *Emerg Infect Dis* 2020; **26**(2): 336-9.
148. ProMED-mail. PRO/AH/EDR> Undiagnosed pneumonia - China (HU): RFI. 2019. <https://promedmail.org/promed-post/?id=20191230.6864153> (accessed 19th of July 2020).
149. Freifeld CC, Mandl KD, Reis BY, Brownstein JS. HealthMap: global infectious disease monitoring through automated classification and visualization of Internet media reports. *J Am Med Inform Assoc* 2008; **15**(2): 150-7.
150. Xu B, Kraemer MUG, Open C-DCG. Open access epidemiological data from the COVID-19 outbreak. *Lancet Infect Dis* 2020; **20**(5): 534.
151. Open COVID-19 Data Working Group. Detailed Epidemiological Data from the COVID-19 Outbreak. 2020. <https://www.healthmap.org/covid-19/> (accessed 19th of July 2020).
152. Cho A. Artificial intelligence systems aim to sniff out signs of COVID-19 outbreaks. 2020. <https://www.sciencemag.org/news/2020/05/artificial-intelligence-systems-aim-sniff-out-signs-covid-19-outbreaks> (accessed 19th of July 2020).
153. Simonsen L, Gog JR, Olson D, Viboud C. Infectious Disease Surveillance in the Big Data Era: Towards Faster and Locally Relevant Systems. *J Infect Dis* 2016; **214**(suppl\_4): S380-S5.
154. Gilbert GL, Degeling C, Johnson J. Communicable Disease Surveillance Ethics in the Age of Big Data and New Technology. *Asian Bioeth Rev* 2019; **11**(2): 173-87.
155. EUR-Lex. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance). 2016. <https://eur-lex.europa.eu/eli/reg/2016/679/oj> (accessed 18th of July 2020).
156. Bengio Y, Ippolito D, Janda R, et al. Inherent privacy limitations of decentralized contact tracing apps. *J Am Med Inform Assoc* 2020.

157. Buckee CO, Balsari S, Chan J, et al. Aggregated mobility data could help fight COVID-19. *Science* 2020; **368**(6487): 145-6.
158. Pepin JL, Bruno RM, Yang RY, et al. Wearable Activity Trackers for Monitoring Adherence to Home Confinement During the COVID-19 Pandemic Worldwide: Data Aggregation and Analysis. *J Med Internet Res* 2020; **22**(6): e19787.
159. Chung YT, Yeh CY, Shu YC, et al. Continuous temperature monitoring by a wearable device for early detection of febrile events in the SARS-CoV-2 outbreak in Taiwan, 2020. *J Microbiol Immunol Infect* 2020; **53**(3): 503-4.
160. Amit M, Kimhi H, Bader T, Chen J, Glassberg E, Benov A. Mass-surveillance technologies to fight coronavirus spread: the case of Israel. *Nat Med* 2020.
161. European Parliament. Covid-19 tracing apps: ensuring privacy and data protection. 2020. <https://www.europarl.europa.eu/news/en/headlines/society/20200429STO78174/covid-19-tracing-apps-ensuring-privacy-and-data-protection> (accessed 19th of July 2020).
162. Shah SJ, Barish PN, Prasad PA, et al. Clinical features, diagnostics, and outcomes of patients presenting with acute respiratory illness: A retrospective cohort study of patients with and without COVID-19. *EClinicalMedicine* 2020; **27**: 100518.
163. Shah SJ, Barish PN, Prasad PA, et al. Clinical features, diagnostics, and outcomes of patients presenting with acute respiratory illness: a comparison of patients with and without COVID-19. *medRxiv* 2020.
164. Chen X, Tang Y, Mo Y, et al. A diagnostic model for coronavirus disease 2019 (COVID-19) based on radiological semantic and clinical features: a multi-center study. *Eur Radiol* 2020.
165. BioMÉRIEUX. BIOFIRE® Respiratory 2.1 plus Panel. 2020. <https://www.biomerieux-diagnostics.com/filmarrayr-respiratory-panel> (accessed 25th of July 2020).
166. Buchan BW, Windham S, Balada-Llasat JM, et al. Practical Comparison of the BioFire FilmArray Pneumonia Panel to Routine Diagnostic Methods and Potential Impact on Antimicrobial Stewardship in Adult Hospitalized Patients with Lower Respiratory Tract Infections. *J Clin Microbiol* 2020; **58**(7).
167. Reischl AT, Schreiner D, Poplawska K, et al. The clinical impact of PCR-based point-of-care diagnostic in respiratory tract infections in children. *J Clin Lab Anal* 2020; **34**(5): e23203.
168. Qavi AJ, McMullen A, Burnham CD, Anderson NW. Repeat Molecular Testing for Respiratory Pathogens: Diagnostic Gain or Diminishing Returns? *J Appl Lab Med* 2020.
169. Creager HM, Cabrera B, Schnaubelt A, et al. Clinical evaluation of the BioFire(R) Respiratory Panel 2.1 and detection of SARS-CoV-2. *J Clin Virol* 2020; **129**: 104538.
170. Wang J, Han Y, Feng J. Metagenomic next-generation sequencing for mixed pulmonary infection diagnosis. *BMC Pulm Med* 2019; **19**(1): 252.
171. Li Y, Sun B, Tang X, et al. Application of metagenomic next-generation sequencing for bronchoalveolar lavage diagnostics in critically ill patients. *Eur J Clin Microbiol Infect Dis* 2020; **39**(2): 369-74.
172. Yang L, Haidar G, Zia H, et al. Metagenomic identification of severe pneumonia pathogens in mechanically-ventilated patients: a feasibility and clinical validity study. *Respir Res* 2019; **20**(1): 265.
173. Lewandowski K, Xu Y, Pullan ST, et al. Metagenomic Nanopore Sequencing of Influenza Virus Direct from Clinical Respiratory Samples. *J Clin Microbiol* 2019; **58**(1).
174. van Rijn AL, van Boheemen S, Sidorov I, et al. The respiratory virome and exacerbations in patients with chronic obstructive pulmonary disease. *PLoS One* 2019; **14**(10): e0223952.

175. Perez-Sautu U, Wiley MR, Iglesias-Caballero M, et al. Target-independent high-throughput sequencing methods provide evidence that already known human viral pathogens play a main role in respiratory infections with unexplained etiology. *Emerg Microbes Infect* 2019; **8**(1): 1054-65.
176. Charalampous T, Kay GL, Richardson H, et al. Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. *Nat Biotechnol* 2019; **37**(7): 783-92.
177. Qi C, Hountras P, Pickens CO, et al. Corrigendum: Detection of respiratory pathogens in clinical samples using metagenomic shotgun sequencing. *J Med Microbiol* 2020; **69**(1): 146.
178. Qi C, Hountras P, Pickens CO, et al. Detection of respiratory pathogens in clinical samples using metagenomic shotgun sequencing. *J Med Microbiol* 2019; **68**(7): 996-1002.
179. Wang J, Li Y, He X, et al. Gemykibivirus Genome in Lower Respiratory Tract of Elderly Woman With Unexplained Acute Respiratory Distress Syndrome. *Clin Infect Dis* 2019; **69**(5): 861-4.
180. Langelier C, Kalantar KL, Moazed F, et al. Integrating host response and unbiased microbe detection for lower respiratory tract infection diagnosis in critically ill adults. *Proc Natl Acad Sci U S A* 2018; **115**(52): E12353-E62.
181. Zhang D, Lou X, Yan H, et al. Metagenomic analysis of viral nucleic acid extraction methods in respiratory clinical samples. *BMC Genomics* 2018; **19**(1): 773.
182. Houldcroft CJ, Beale MA, Sayeed MA, Qadri F, Dougan G, Mutreja A. Identification of novel adenovirus genotype 90 in children from Bangladesh. *Microb Genom* 2018; **4**(10).
183. Zinter MS, Dvorak CC, Mayday MY, et al. Pulmonary Metagenomic Sequencing Suggests Missed Infections in Immunocompromised Children. *Clin Infect Dis* 2019; **68**(11): 1847-55.
184. Yang Y, Walls SD, Gross SM, Schroth GP, Jarman RG, Hang J. Targeted Sequencing of Respiratory Viruses in Clinical Specimens for Pathogen Identification and Genome-Wide Analysis. *Methods Mol Biol* 2018; **1838**: 125-40.
185. Li H, Gao H, Meng H, et al. Detection of Pulmonary Infectious Pathogens From Lung Biopsy Tissues by Metagenomic Next-Generation Sequencing. *Front Cell Infect Microbiol* 2018; **8**: 205.
186. Li Y, Deng X, Hu F, et al. Metagenomic analysis identified co-infection with human rhinovirus C and bocavirus 1 in an adult suffering from severe pneumonia. *J Infect* 2018; **76**(3): 311-3.
187. Madi N, Al-Nakib W, Mustafa AS, Habibi N. Metagenomic analysis of viral diversity in respiratory samples from patients with respiratory tract infections in Kuwait. *J Med Virol* 2018; **90**(3): 412-20.
188. Langelier C, Zinter MS, Kalantar K, et al. Metagenomic Sequencing Detects Respiratory Pathogens in Hematopoietic Cellular Transplant Patients. *Am J Respir Crit Care Med* 2018; **197**(4): 524-8.
189. Emonet S, Lazarevic V, Pugin J, Schrenzel J, Ruppe E. Clinical Metagenomics for the Diagnosis of Hospital-acquired Infections: Promises and Hurdles. *Am J Respir Crit Care Med* 2017; **196**(12): 1617-8.
190. Lewandowska DW, Schreiber PW, Schuurmans MM, et al. Metagenomic sequencing complements routine diagnostics in identifying viral pathogens in lung transplant recipients with unknown etiology of respiratory infection. *PLoS One* 2017; **12**(5): e0177340.
191. Chen L, Liu W, Zhang Q, et al. RNA based mNGS approach identifies a novel human coronavirus from two individual pneumonia cases in 2019 Wuhan outbreak. *Emerg Microbes Infect* 2020; **9**(1): 313-9.

192. Ai JW, Zhang Y, Zhang HC, Xu T, Zhang WH. Era of molecular diagnosis for pathogen identification of unexplained pneumonia, lessons to be learned. *Emerg Microbes Infect* 2020; **9**(1): 597-600.
193. Manning JE, Bohl JA, Lay S, et al. Rapid metagenomic characterization of a case of imported COVID-19 in Cambodia. *bioRxiv* 2020.
194. Van Tan L, Thi Thu Hong N, My Ngoc N, et al. SARS-CoV-2 and co-infections detection in nasopharyngeal throat swabs of COVID-19 patients by metagenomics. *J Infect* 2020.
195. Capobianchi MR, Giombini E, Rozera G. Next-generation sequencing technology in clinical virology. *Clinical Microbiology and Infection* 2013; **19**(1): 15-22.
196. Chen X, Kang Y, Luo J, et al. Next-Generation Sequencing Reveals the Progression of COVID-19. *Front Cell Infect Microbiol* 2021; **11**:632490.
197. Duwe SC, Wedde M, Birkner P, Schweiger B. Genotypic and phenotypic resistance of pandemic A/H1N1 influenza viruses circulating in Germany. *Antiviral Res* 2011; **89**(1): 115-8.
198. Human Adenovirus Working Group. <http://hadvwg.gmu.edu/> (accessed 6th of September 2021).
199. Zou X, Fan Y, Zhang W, et al. Metagenomics facilitated complete genome sequencing of adenovirus untyped by real-time PCR. *J Infect* 2018; **77**(2): 158-64.
200. Ji T, Li L, Li W, et al. Emergence and characterization of a putative novel human adenovirus recombinant HAdV-C104 causing pneumonia in Southern China. *Virus Evolution* 2021; **7**(1).
201. Dhingra A, Hage E, Ganzenmueller T, et al. Molecular Evolution of Human Adenovirus (HAdV) Species C. *Sci Rep* 2019; **9**(1): 1039.
202. Seeber L, Conrad T, Hoppe C, et al. Educating parents about the vaccination status of their children: A user-centered mobile application. *Prev Med Rep* 2017; **5**: 241-50.
203. Rath B, Penttinen P. Incidence, Severity and Impact of Influenza: a joint meeting organised by the ISIRV Epidemiology Group and ECDC, Stockholm, 2019. *Euro Surveill* 2019; **24**(23).
204. World Health Organization. Pandemic Influenza Severity Assessment (PISA) (<https://apps.who.int/iris/bitstream/handle/10665/259392/WHO-WHE-IHM-GIP-20172-engpdf;jsessionid=90F6A894043AAAF51F2FB604C81A4F64?sequence=1>) 2017.
205. Deng X, Achari A, Federman S, et al. Author Correction: Metagenomic sequencing with spiked primer enrichment for viral diagnostics and genomic surveillance. *Nat Microbiol* 2020; **5**(3): 525.
206. Deng X, Achari A, Federman S, et al. Metagenomic sequencing with spiked primer enrichment for viral diagnostics and genomic surveillance. *Nat Microbiol* 2020; **5**(3): 443-54.
207. Ladner JT, Beitzel B, Chain PS, et al. Standards for sequencing viral genomes in the era of high-throughput sequencing. *mBio* 2014; **5**(3): e01360-14.
208. Lee ER, Parkin N, Jennings C, et al. Performance comparison of next generation sequencing analysis pipelines for HIV-1 drug resistance testing. *Sci Rep* 2020; **10**(1): 1634.
209. Sichtig H, Minogue T, Yan Y, et al. FDA-ARGOS is a database with public quality-controlled reference genomes for diagnostic use and regulatory science. *Nat Commun* 2019; **10**(1): 3313.
210. Lu IN, Muller CP, He FQ. Applying next-generation sequencing to unravel the mutational landscape in viral quasispecies. *Virus Res* 2020; **283**: 197963.
211. Di Salvo E, Di Gioacchino M, Tonacci A, Casciaro M, Gangemi S. Alarmins, COVID-19 and comorbidities. *Ann Med* 2021; **53**(1): 777-85.
212. Kapurniotu A, Gokce O, Bernhagen J. The Multitasking Potential of Alarmins and Atypical Chemokines. *Front Med (Lausanne)* 2019; **6**: 3.

213. Iqbal SMA, Mahgoub I, Du E, Leavitt MA, Asghar W. Advances in healthcare wearable devices. *npj Flex Electron* 2021; 5(9) (<https://doi.org/10.1038/s41528-021-00107-x>).
214. Jain A. The 5 V's of big data. 2016. <https://www.ibm.com/blogs/watson-health/the-5-vs-of-big-data/>.