TAL TECH

DOCTORAL THESIS

Exploration of Host-Agent-Environment Interactions Using Tools of Metagenomic Sequencing and Next Generation Phage Display

Mariliis Jaago

TALLINNA TEHNIKAÜLIKOOL
TALLINN UNIVERSITY OF TECHNOLOGY
TALLINN 2023

TALLINN UNIVERSITY OF TECHNOLOGY
DOCTORAL THESIS
56/2023

Exploration of Host-Agent-Environment Interactions Using Tools of Metagenomic Sequencing and Next Generation Phage Display

MARILIIS JAAGO

TAL TECH PRESS

TALLINN UNIVERSITY OF TECHNOLOGY
School of Science
Department of Chemistry and Biotechnology
The dissertation was accepted for the defence of the degree of Doctor of Philosophy in Chemistry and Biotechnology on 25/10/2023

Supervisor: Associate Professor Kaia Palm, PhD
Department of Chemistry and Biotechnology
Tallinn University of Technology
Principal Researcher
Protobios LLC
Tallinn, Estonia

Co-supervisor: Professor Tõnis Timmusk, PhD
Department of Chemistry and Biotechnology
Tallinn University of Technology
Tallinn, Estonia

Opponents: Dr Leopold Parts, PhD
Wellcome Sanger Institute
Hinxton, United Kingdom

Prof Reet Mändar, PhD
University of Tartu
Tartu, Estonia

Defence of the thesis: 24/11/2023, Tallinn

Declaration:
Hereby I declare that this doctoral thesis, my original investigation and achievement, submitted for the doctoral degree at Tallinn University of Technology has not been submitted for doctoral or equivalent academic degree.

Mariliis Jaago

---
signature

European Union
European Regional
Development Fund

Investing
in your future

ISSN 2585-6898 (publication)
ISBN 978-9916-80-069-0 (publication)
ISSN 2585-6901 (PDF)
ISBN 978-9916-80-070-6 (PDF)
Printed by Koopia Niini & Rauam

TALLINNA TEHNIKAÜLIKOOL
DOKTORITÖÖ
56/2023

Metagenoomi sekveneerimise ja järgmise põlvkonna faagidisplei kasutamine inimese eksposoomi kirjeldamisel

MARILIIS JAAGO

TAL TECH
KIRJASTUS

Contents

List of publications

The research publications underlying the doctoral thesis are the following:

I. Jaago, M., Timmusk, U. S., Timmusk, T., Palm, K. 2021. Drastic Effects on the Microbiome of a Young Rower Engaged in High-Endurance Exercise After a Month Usage of a Dietary Fiber Supplement. *Frontiers in Nutrition*, 8:654008. doi: 10.3389/fnut.2021.654008.

II. Jaago, M., Pupina, N., Rähni, A., Pihlak, A., Sadam, H., Vrana, N. E., Sinisalo, J., Pussinen, P., Palm, K. 2022. Antibody response to oral biofilm is a biomarker for acute coronary syndrome in periodontal disease. *Communications Biology*, 5 (1): 205, ARTN 205. doi: 10.1038/s42003-022-03122-4.

III. Sadam, H., Pihlak, A., Kivil, A., Pihelgas, S., **Jaago, M.**, Adler, P., Vilo, J., Vapalahti, O., Neuman, T., Lindholm, D., Partinen, M., Vaheri, A., Palm, K. 2018. Prostaglandin D2 Receptor DP1 Antibodies Predict Vaccine-induced and Spontaneous Narcolepsy Type 1: Large-scale Study of Antibody Profiling. *EBioMedicine*, 29, 47–59. doi: 10.1016/j.ebiom.2018.01.043.

Author's contribution to the publications

I. Author contributed to the experimental design, performed the data analysis and interpretation of results, and contributed to the writing of the manuscript.

II. Author contributed to the experimental design, contributed to the experiments, performed the data analysis and interpretation of results, and contributed to the writing of the manuscript.

III. Author contributed to the data analysis, to the interpretation of results and to the writing of the manuscript.

Other associated publications

Tamberg, L., **Jaago, M.**, Saalik, K., Sirp, A., Tuvikene, J., Shubina, A., Kiir, C. S., Nurm, K., Sepp, M., Timmusk, T., Palgi, M. 2020. Daughterless, the Drosophila orthologue of TCF4, is required for associative learning and maintenance of the synaptic proteome. *Disease Models & Mechanisms*, 13 (7): dmm042747. doi: 10.1242/dmm.042747.

Sadam, H., Pihlak, A., **Jaago, M.**, Pupina, N., Rähni, A., Toots, M., Vaheri, A., Nieminen, J. K., Siuko, M., Tienari, P. J., Palm, K. 2021. Identification of two highly antigenic epitope markers predicting multiple sclerosis in optic neuritis patients. *EBioMedicine*, 64, 103211. doi: 10.1016/j.ebiom.2021.103211.

Pupina, N., Avarlaid, A., Sadam, H., Pihlak, A., **Jaago, M.**, Tuvikene, J., Rähni, A., Planken, A., Planken, M., Kalso, E., Tienari, P.J., Nieminen, J.K., Seppänen, M.R.J., Vaheri, A., Lindholm, D., Sinisalo, J., Pussinen, P., Timmusk, T., Palm, K. 2022. Immune response to a conserved enteroviral epitope of the major capsid VP1 protein is associated with lower risk of cardiovascular disease. *EBioMedicine*, 76, 103835. doi: 10.1016/j.ebiom.2022.103835.

Rähni, A., **Jaago, M.**, Sadam, H., Pupina, N., Pihlak, A., Tuvikene, J., Annuk, M., Mägi, A., Timmusk, T., Ghaemmaghami, A.M., Palm, K. 2022. Melanoma-specific antigen-associated antitumor antibody reactivity as an immune-related biomarker for targeted immunotherapies. *Commun Med 2*, 48 (2022). doi: 10.1038/s43856-022-00114-7.

Jaago, M., Rähni, A., Pupina, N., Pihlak, A., Sadam, H., Tuvikene, J., Avarlaid, A., Planken, A., Planken, M., Haring, L., Vasar, E., Baćević, M., Lambert, F., Kalso, E., Pussinen, P., Tienari, P. J., Vaheri, A., Lindholm, D., Timmusk, T., Ghaemmaghami, A. M., Palm, K. 2022. Differential patterns of cross-reactive antibody response against SARS-CoV-2 spike protein detected for chronically ill and healthy COVID-19 naïve individuals. *Scientific Reports*, 12 (1): 16817. doi: 10.1038/s41598-022-20849-6.

Introduction

The relationship between the human host, various biological agents and the environment has been studied within the epidemiological triad approach to characterize infectious disease development and spread. The exposome refers to the totality of all biological and environmental exposures a person encounters throughout their life. Lately, exposome research has highlighted the importance of analysing exposures as a comprehensive whole to unravel normal human functioning and disease development, as well as improve disease prevention and targeted health interventions. Here, metagenomic sequencing and next generation phage display technologies were used to characterize parts of the human exposome in various health and disease backgrounds.

The human microbiome and virome play a vital role in shaping human health and well-being. The microbiome interacts with the body from early life through adulthood, contributing to digestion, nutrient absorption, synthesis of essential molecules, development and maintenance of the immune system, mental health, metabolism, and the modulation of drug efficacy. Diet and exercise have a crucial role in shaping the composition and diversity of the gut microbiome. In turn, the gut microbiome affects the physical fitness of the body. On the other hand, the microbiome has been linked to the development of various diseases. Smoking, for example, has been associated with alterations in the oral microbiome and an increased risk of periodontitis, a chronic gum inflammation. Periodontitis, in turn, has been linked to cardiovascular disease, potentially through the release of inflammatory molecules and the systemic dissemination of oral bacteria. Although there are still some challenges to exposome research which warrant consideration, it enables to gain insight into these intricate connections and highlights the potential impact of the microbiome on systemic health beyond the local environment.

In these studies, we evaluated the effect of dietary fibre supplement intake on the composition of an athlete's gut microbiome, providing valuable insights into the potential benefits of dietary interventions for athletes' microbiota. We discovered antigen epitopes that are specifically targeted in patients with narcolepsy and cardiovascular disease (CVD), shedding light on potential disease mechanisms. We characterized and compared the data analysis challenges and approaches in various study designs, providing insights into the strengths and limitations of each approach for exposome research. The results of the studies contribute to the overall characterization of the human exposome and provide useful technological and computational tools for further exposome research.

Abbreviations

ACS	acute coronary syndrome
AD	Alzheimer's disease
AGE	advanced glycation end product, product of nonenzymatic reactions between reducing sugars and proteins, lipids, or nucleic acids
ATAC-seq	assay for transposase-accessible chromatin using sequencing
BCL6	B-cell lymphoma 6 protein
BCR	B-cell receptor
cDNA	complementary DNA; copy DNA
CMV	Human cytomegalovirus
CNS	central nervous system
CRT	cyclic reversible termination
CVD	cardiovascular disease
dot ELISA	dot enzyme-linked immunosorbent assay in solid phase
DP1	prostaglandin D2 receptor 1
EBNA-1	Epstein Barr nuclear antigen 1
EBNA-2	Epstein Barr nuclear antigen 2
EBNA-6	Epstein Barr nuclear antigen 6
EBV	Epstein Barr virus
EC	entry complex
ELISA	enzyme-linked immunosorbent assay
FMT	fecal microbiota transplantation
FUT2	fucosyltransferase 2
F/B ratio	ratio of *Firmicutes* phylum bacteria to *Bacteroidetes*
G	gingivitis diagnosis/patients with gingivitis
gp42	glycoprotein 42
H	healthy controls with no gum inflammation diagnosis
HA	hemagglutinin
HC	healthy control
HHV-6	human herpesvirus 6
HHV-7	human herpesvirus 7
HLA	human leukocyte antigen
HMP	Human Microbiome Project
HSV-1	herpes simplex virus 1
HSV-2	herpes simplex virus 2
H1N1	Influenza A virus subtype
H1N1-HC	healthy controls with an influenza A(H1N1) infection exposure
IBD	Inflammatory bowel disease
IgG	immunoglobulin G
LPS	lipopolysaccharide, outer membrane component of Gram-negative bacteria
MHC-1	major histocompatibility complex class 1
miRNA	microRNA
MMP	matrix metalloprotease
MS	multiple sclerosis

MVA	Mimotope Variation Analysis
NGS	next-generation sequencing
NK	natural killer cells
no-CAD	healthy controls with no coronary artery disease diagnosis
NT1	narcolepsy type 1
P	periodontitis diagnosis/patients with periodontitis
PAMP	pathogen-associated molecular pattern
Pdmx-HC	Pandemrix-vaccine induced narcolepsy type 1
PGD2	prostaglandin D2
PRR	pattern recognition receptor
RNA-Seq	RNA sequencing
ROC	receiver operating characteristic
SBS	sequencing by synthesis
scRNA-seq	single cell RNA sequencing
SCFA	short-chain fatty acid
SLE	spatial lifecourse epidemiology
SLS	sequencing by ligation
SNA	single nucleotide addition
sNT1	sporadic narcolepsy type 1
s-CAD	stable coronary artery disease
Top2500	dataset of top 2500 peptides with highest abundance
TCR	T-cell receptor
T1D	type 1 diabetes
T2D	type 2 diabetes
VP26	Small capsomere-interacting protein; SCP; BFRF3
VZV	varicella-zoster virus
WES	whole-exome sequencing
WGS	whole genome sequencing

1 Relevance of the microbiome in the epidemiological triad and exposome approach

Current and emerging research approaches in biology are more-and-more focusing on studying a problem in a comprehensive manner, considering evidence from multiple different aspects. This chapter will give a brief overview about what is the epidemiological triad, why is it useful, and what are some of the directions this approach is taking. Furthermore, this chapter will introduce the term "exposome" and give examples of the importance of exposome research. Finally, the chapter exemplifies the relevance of incorporating the human microbiome studies in various comprehensive research.

The epidemiological triad is a classic model that provides a useful framework for understanding the complex interplay between the agent, host, and environment in the development and spread of infectious diseases (**Figure 1**). This model has informed the development of effective prevention and control strategies, and has been widely used in public health research and practice (reviewed in Budden et al. 2017).

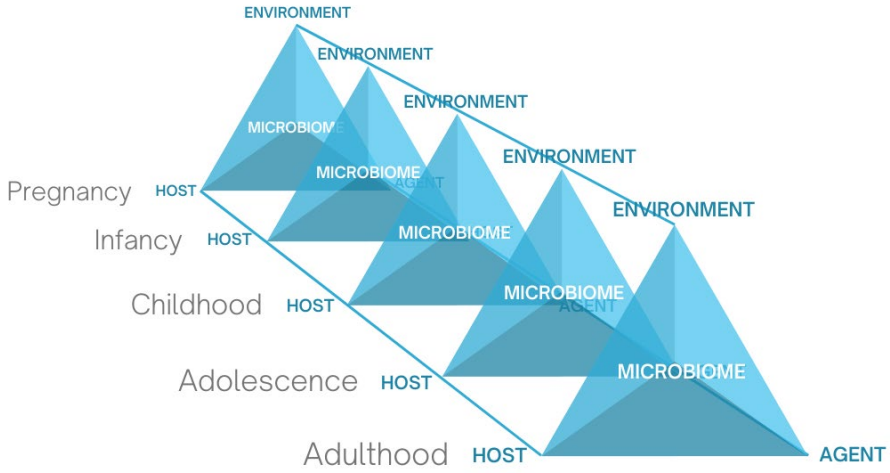


Figure 1. The epidemiological triad across a lifespan. *The epidemiological triad describes the interplay between the potentially affected human host, the disease-causing agent, the changing environment in which all components reside in. The microbiome can be thought of as an optional vector in the middle, affecting the host and transmitting any indirect effects from agents. Spatial lifecourse epidemiology emphasizes considering the history of the host's interactions with the environment and disease-causing agents, since early development to adulthood. Figure customized from (Jia et al. 2020).*

Changes in temperature, precipitation, and other environmental variables can affect the distribution and abundance of vector-borne diseases, such as malaria and dengue fever, as well as the survival and transmission of other infectious agents (Mora et al. 2022). For example in Europe, increased temperature and precipitation has been associated with higher incidence of West Nile Virus infection (Marcantonio et al. 2015; Moirano et al. 2018). During the COVID-19 pandemic, researchers discovered that fine-particle air pollution increased the population death rate from COVID-19 (X. Wu et al. 2020; Conticini, Frediani, and Caro 2020). Fine-particular air pollution has been determined as a risk factor for cardiopulmonary disease and lung cancer, which in turn can exacerbate serious infections (reviewed in Pope et al. 2020; Domingo and Rovira

2020). Advances in genomic sequencing technology have enabled researchers to identify and characterize infectious agents, as well as to better understand their evolutionary history and potential for transmission more accurately (reviewed in Loman and Pallen 2015). This has led to new insights into the epidemiology of infectious diseases, including the emergence of new pathogens and the evolution of drug resistance.

Emerging technological tools offer promising opportunities within epidemiological research. Spatial lifecourse epidemiology (SLE) has been proposed as a comprehensive approach based on advanced spatial, location-aware, and artificial intelligence (AI) technologies to investigate long-term effects of measurable biological, environmental, behavioural, and psychosocial factors on individual risk for chronic diseases (Jia et al. 2020). In addition to conventional data sources, people's location data could be gathered using GPS in cell phones, history of medical test results and diseases from electronic health records, and socioeconomic factors and self-reported health issues from online surveys (Jia et al. 2020). AI-powered models could be built that measure the environmental, personal biologic characteristics, and interpersonal interactions to predict individual risk of chronic disease (Jia et al. 2020). Overall, these proposed approaches and other recent discoveries related to the epidemiological triad are helping to advance our understanding of the complex interactions between agents, hosts, and the environment in the development and spread of diseases, and can inform the development of targeted prevention and control strategies.

The totality of all environmental exposures throughout the human life is referred to as the exposome (Wild 2005a). Through exposome research environmental factors have been associated with disease, such as smoking as a risk factor for type 2 diabetes (T2D) or alcohol for high blood pressure (Hall et al. 2014; McGinnis, Brownstein, and Patel 2016). Today, environmental exposures have been linked to CVDs, pulmonary diseases, kidney disease, cancer, and neurological and mental disorders. The molecular mechanisms by which the environment impacts gene expression are numerous, including DNA methylation, histone modification, non-coding RNA function, extracellular vesicles, epitranscriptome, and the mitochondrial genome (reviewed in H. Wu, Eckhardt, and Baccarelli 2023). A certain type of chemical or biological agent can also affect the body through various mechanisms. For example, exposure to fine particulate matter (<2.5µm) ($PM_{2.5}$), a component in air pollution, results in gene expression changes and DNA hypomethylation in bronchial epithelial cells and the biological effect is greater when exposure is repetitive (S. K. Huang et al. 2021). On the other hand, urban dust induces the expression of inflammatory interleukin-6 and cyclooxygenase-II through downregulation of micro-RNA activity (Tsai et al. 2020). Ning and colleagues highlight the need for investigating multiple biological aspects in response to environmental agents such as $PM_{2.5}$ (Ning et al. 2022). Even though people can be exposed to the same environmental agents, individual variability in genetic background plays a role in the resulting biological response. For example, people with a genetic variant rs3825807 respond to fish oil supplementation with a decrease in serum triglycerides, a beneficial effect for the maintenance of healthy blood lipid levels, whereas people without the variant do not (Francis et al. 2021). Understanding the mechanisms by which the environment affects the human body allows to develop personalized health recommendations and better public health policies. Eckhardt and colleagues demonstrated that individual DNA methylation patterns could be used as a proxy for smoking-induced lung cancer and revealed patients with a higher risk of smoking-related health problems (Eckhardt et al. 2022).

The human microbiome, which refers to the collection of microorganisms that live on and within the human body, has emerged as an important factor in exposome research. There are around 100-500 bacterial species living inside our body at a certain time and the composition of the community is rather stable over time (Qin et al. 2010; Faith et al. 2013; Martínez, Muller, and Walter 2013). The main phyla of bacteria colonizing the human body are *Firmicutes*, *Bacteroidetes*, *Actinobacteria* and *Proteobacteria*, and the archeae found in humans are usually methanogens (Eckburg et al. 2005; Gill et al. 2006). Although there are only a few phyla associated with humans, the variation in the specific species of bacteria across individuals is very high to the extent that a unique "microbiome fingerprint" has been discussed (Vanhoutte et al. 2004; Eckburg et al. 2005; Rothschild et al. 2018). Although the microbial composition is more similar among relatives (reviewed in Dethlefsen et al. 2006), the largest effects on composition are environmental and dietary (Rothschild et al. 2018).

There are multiple reasons why focusing on the microbiome in exposome research is important for the betterment of population health. First is disease transmission – microbes can be transmitted from person to person and the composition of an individual's microbiome can influence their susceptibility to infections. Studies have shown this to be the case during sexual transmission and in the development of bacterial vaginosis (Mehta et al. 2020). Person-to-person transmission also occurs from mother to child during early development (Maqsood et al. 2019). The microbiome has also been linked to the development of numerous non-infectious diseases, including obesity, diabetes, and inflammatory bowel disease. For example, a study by Qin and colleagues found that the gut microbiome of individuals with T2D was significantly different from that of healthy individuals by having a decrease in beneficial butyrate-producing species and an increase in opportunistic pathogens such as *Escherichia coli* (*E. coli*) and *Clostridium* species (Qin et al. 2012). The third reasoning behind studying the microbiome through an epidemiological lens is biomarker discovery. Biomarkers derived from the microbiome can provide insights into disease diagnosis, prognosis, and treatment response. For example, a study by Zeller and colleagues found that a microbiome-based biomarker panel was able to accurately distinguish patients with colorectal cancer from healthy individuals (Zeller et al. 2014). Fourth, manipulating the microbiome has been proposed as a potential therapeutic approach for certain diseases. Clinical trials have demonstrated fecal microbiota transplantation (FMT) as an effective treatment for recurrent *Clostridium difficile* infections (Baunwall et al. 2020). Furthermore, FMT has been shown to improve the efficacy of immune checkpoint inhibitor therapy in melanoma treatment, whereas FMT efficacy is also being assessed in colorectal cancer treatment (reviewed in Kaźmierczak-Siedlecka et al. 2020; Biazzo and Deidda 2022).

Overall, the microbiome is an important area of research in the scope of exposome research, as it can provide insights into the transmission and development of both infectious and non-infectious diseases, deliver diagnostic biomarkers as well as potential therapeutic interventions.

2 Impact of microbiome and virome on human health

The human microbiome is a complex system containing different bacterial species living in the human body and the interplay between the human host and the microbiome influences many aspects of human health. Similarly, the human virome develops together with human development and affects the normal functioning of the body. This chapter gives an overview of the initial development of the human microbiome and virome as the human body develops from infancy. Next, this chapter discusses the various ways the microbiome and virome have a necessary role in the healthy functioning of the body. Additionally, a brief overview is given about the topic of vaccination as a necessary public health tool, however developing narcolepsy is given as an example of rare serious side effects of vaccination. Then the chapter gives an overview of how microbiome dysbiosis develops and how the gut and oral microbiome can contribute to adverse human health conditions, including chronic disease. Finally, the role of virome is discussed in the context of promoting or aggravating chronic disease. To read more about the role of the microbiome or virome in the scope of other diseases such as autoimmune diseases or cancer, see reviews by (Mesri, Feitelson, and Munger 2014; Ringelhan et al. 2015; Smatti et al. 2019).

2.1 Role of microbiome and virome in human health

The broad features of the bacterial microbiome develop during the first years of life (Palmer et al. 2007). A large part of the bacterial microflora is passed on from the mother to the infant (Maqsood et al. 2019), and environmental factors play an important role in the development of the microbiome (Palmer et al. 2007). When the infant starts to consume solid food, their gut microbiome changes drastically and starts to resemble an adult's microbiome (Palmer et al. 2007; Maqsood et al. 2019).

Human interaction with bacteria has an important role in the normal functioning of the individual (**Figure 2**). The bacteria living inside us aid in food digestion and energy metabolism, improve our resilience to pathogens and toxins, and help the maturation of our developing immune system. The gut microbiome breaks down plant-based polysaccharides into simpler compounds, e.g. short-chain fatty acids (SCFAs) such as butyric acid, that our colon epithelium cells can use for energy (reviewed in Flint 2004). The microbial fermentation process from polysaccharides to butyric acid can involve multiple steps and bacterial species. For example, *Bifidobacterium* spp secretes lactate which is then fermented by *Eubacterium hallii* into butyric acid which the host can use (Duncan, Louis, and Flint 2004; Belenguer et al. 2006). Our gut bacteria are under the constant monitoring by the immune system which allows the bacteria to operate in a specific niche but immediately responds to bacterial invasion to other tissues (reviewed in Tilman 2004). Non-pathogenic bacteria colonizing our gut or skin can also help protect us from pathogenic bacteria. Firstly, non-pathogenic bacteria take up the physical space of the pathogenic bacteria and compete for nutrients (reviewed in Tilman 2004). The second mechanism has a more direct effect as some species of non-pathogenic bacteria secrete compounds that inhibit the growth of pathogenic bacteria or are toxic to them (reviewed in Fons, Gomez, and Karjalainen 2000; Tilman 2004). Thirdly, some non-pathogenic bacteria can neutralize toxins secreted by pathogenic bacteria (reviewed in Pool-Zobel, Veeriah, and Böhmer 2005). For example, *B. fragilis* has been shown to affect our T-cell and cytokine response and has a defensive effect against colitis, colorectal cancer and viral encephalitis (Mazmanian et al. 2005; Shen et al. 2012; Y. K. Lee

et al. 2018; Ramakrishna et al. 2019). Our exposure to bacteria in childhood is important for the maturation of the developing immune system. For example, segmented filamentous bacteria are important in the development of Th17 cells, a subset of T helper cells, and thus protect us from developing inflammatory diseases later in life (reviewed in Schnupf, Gaboriau-Routhiau, and Cerf-Bensussan 2013).

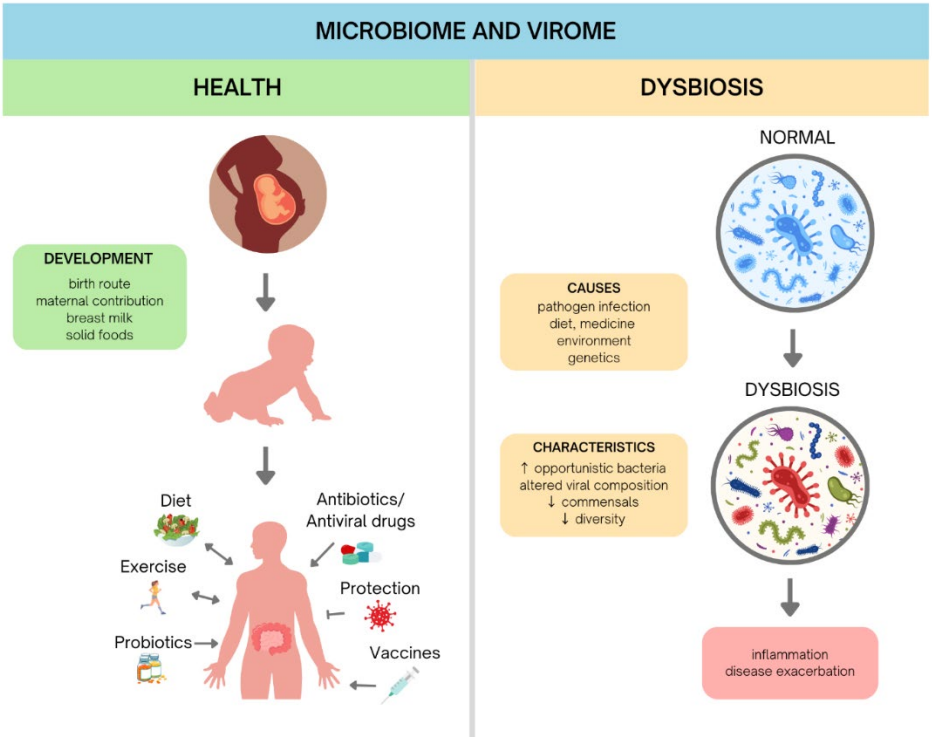


Figure 2. Development and interplay of the human host and the microbiome and virome in health and development of dysbiosis. *The development of the composition and functioning of the microbiome and virome start during development in the mother's womb. The birth route (vaginal versus caesarean section birth), contribution from the mother's microbiome, whether the developing child is given breast milk or milk formula or a combination of these, and what solid foods are introduced to the developing child (and in which order and proportion) all affect the microbiome and virome of the person in early years. Later in life, the microbiome and virome are influenced by diet, exercise, and use of probiotics, antibiotics, antiviral drugs and vaccines. Vice versa, the microbiome and virome influence physical performance and which foods are better ingestible. The microbiome and virome help protect the person from harmful biological agents. Some of the causes leading to a dysbiosis in the microbiome or virome are infection by certain pathogens, diet and medicine, the environment, and genetics. Dysbiosis is characterized by an increase in opportunistic bacteria, altered viral composition, decrease of commensal species and a decrease in species diversity. Dysbiosis can lead to inflammation or exacerbation of disease.*

Unfortunately, in the developed world, the Western diet is gaining popularity with its high fat content, abundance of refined sugars and protein and a decreased proportion of fruits and vegetables. The Western diet habits are associated with metabolic disorders such as obesity, T2D, and liver disease (reviewed in Liping Zhao 2013; J. L. Sonnenburg and Bäckhed 2016). The Western diet causes deterioration of the gut mucus layer with decreased thickness and increased permeability, whereas certain beneficial bacteria or

fibre intake rescues the harmful phenotype (Schroeder et al. 2018). High fructose diets are associated with chronic inflammation by promoting the survival of non-beneficial bacteria (Montrose et al. 2021). Dietary emulsifiers which are added to processed foods have been shown to alter the gut microbiome composition and induce bowel inflammation (Swidsinski et al. 2009; Chassaing, Van de Wiele, et al. 2017). Artificial sweeteners alter the microbiome composition and exacerbate glucose intolerance as shown by experiments done in mice (Suez et al. 2014). Low-fibre or high-fat diet also decreases microbiome diversity (E. D. Sonnenburg et al. 2016; Denou et al. 2016). The food industry often treats animals with small doses of antibiotics, which affect the microbiome composition and may have negative effects on metabolic processes as shown in mice (Cho et al. 2012).

The gut microbiome modulates the metabolism of dietary components, and studies have examined the correlation between measures of fitness and the gut microbiota. Butyrate-producing bacteria, *Bacteroides*, and the *Firmicutes/Bacteroidetes* (F/B) ratio have all been shown to correlate with aerobic fitness levels (VO2 max) (Estaki et al. 2016; Allen et al. 2018). However, other studies have shown contrasting results as in elderly people, higher exercise loads have correlated with a lower F/B ratio (Y. Yu et al. 2018). Few studies have directly investigated the effect of the gut microbiome on athletic performance by use of germ-free mice. These mice, colonized with specific bacterial species, e.g. *Eubacterium rectale* and *Bacteroides fragilis*, have shown improvements in exercise capacity compared to their germ-free counterparts (Hsu et al. 2015; W.-C. Huang et al. 2019). The gut microbiome may modulate adipose tissue thermogenic pathways and skeletal muscle anabolism and function, and a more diverse microbiome may be more beneficial for performance (reviewed in Moreno-Navarrete and Fernandez-Real 2019; Hawley 2020).

Other studies have investigated the effect of probiotic supplementation on athletic performance. Common probiotic bacteria used were strains of *Lactobacillus* or *Bifidobacterium*. Most studies have investigated the effect of probiotic supplementation on aerobic exercise performance measures (Lamprecht et al. 2012; Jäger et al. 2016; Y.-M. Chen et al. 2016; Scheiman et al. 2019; Soares et al. 2019), while some have investigated strength and anaerobic outcomes (Jäger et al. 2016; Y.-M. Chen et al. 2016; Townsend et al. 2018). The effects on performance variables have been highly mixed between studies, with some finding beneficial effects on performance parameters such as time-to-fatigue (Y.-M. Chen et al. 2016; W.-C. Huang et al. 2018; 2019; Scheiman et al. 2019; Soares et al. 2019), while others finding no effects (Lamprecht et al. 2012; Townsend et al. 2018). The use of antibiotics in mouse models has been studied to determine the potential effects of a lack of gut microbes and their metabolites on exercise capacity and muscle function. Recent studies have demonstrated that antibiotic treatment decreased the exercise capacity of mice and reduced muscle glycogen, and that this phenotype could be rescued by either natural reseeding or acetate infusion (Nay et al. 2019; Okamoto et al. 2019). The relative abundance of *Firmicutes* was increased in antibiotic-treated mice while *Bacteroidetes*, alpha diversity, and fecal bacterial DNA concentration was reduced (Nay et al. 2019; Okamoto et al. 2019). A low microbiota-available carbohydrate diet (LMC) *versus* a high microbiota-available carbohydrate (HMC) diet was also tested by Okamoto and colleagues. The reduced exercise capacity phenotype in LMC mice was rescued when mice were given a fecal microbiota transplant from HMC mice and a dose of inulin before exercise (Okamoto et al. 2019). Prebiotics are non-digestible food ingredients, usually fibres, which support the growth of gut microbiota such as *Bifidobacterium* and *Lactobacillus* spp. by acting as an

additional nutrient source (reviewed in Zmora, Suez, and Elinav 2019). Although there are a few studies and the results are individual, prebiotic supplementation can increase SCFA production, an important energy source, which may suggest a beneficial effect to physical performance. On the other hand, some synbiotics, which are a combination of pre- and probiotics, have had beneficial effects for active people by improving gut oxidative status and mucosal immunity (reviewed in Calero, Rincón, and Marqueta 2020).

In addition to the many bacterial species residing in the human body since infancy and into the old age, many viruses also inhabit the body and affect the human health. It is estimated that the human body holds a comparable amount of viral particles to its own cells and bacterial cells (reviewed in Shkoporov and Hill 2019). Three components can be distinguished in the healthy virome: 1) viruses that are not replicating in the body, 2) viruses that infect the microbiome of the body, and 3) viruses that infect human cells and persist in the body (Koonin, Dolja, and Krupovic 2021). Phages comprise the largest part of the human virome (~90%), either infecting and lysing bacterial hosts or incorporating the viral genome into the bacterial genetic material and becoming dormant. Major phages in the gut are *Myoviridae* and *Podoviridae*. Most common eukaryotic viruses include *Anelloaviridae* (ssDNA virus), *Adenoviridae* and *Herpesviridae* (dsDNA viruses), and *Reoviridae* and *Picornaviridae* (RNA viruses) which cause gut infection. Around 8% of the human genome consists of retroviral insertions which can be expressed and thus affect the host (reviewed in Tamayo-Trujillo et al. 2023).

Similarly to the bacterial composition, an individual's virome is fairly stable across time yet varies greatly when compared to other individuals (Reyes et al. 2010; Abeles et al. 2014; Shkoporov et al. 2019; reviewed in Aggarwala, Liang, and Bushman 2017; Shkoporov and Hill 2019). There are many factors that influence the virome of an adult, first of which are geography and ethnicity (Holtz et al. 2014; Zuo et al. 2020). Somewhat associated with geography, diet also influences the gut bacterial community and thus the gut phage community (Minot et al. 2011). Age is another factor as overall virome diversity increases by adulthood and then decreases with age (Gregory et al. 2020). Human genetics has an effect on the virome as certain genetic traits confer impaired resistance to viral infection, such as mutations that render RNA polymerase III defective decrease interferon-associated antiviral immunity (Ramanathan et al. 2020) or genetic T-cell CD28 deficiency which supports skin papillomavirus persistence (Béziat et al. 2021). Other genetic variants are protective against viral infection, such as having *fucosyltransferase 2* (*FUT2*) deficiency protects against norovirus due to *FUT2* gene product being necessary for viral capsid binding (Lindesmith et al. 2003). The human virome starts to develop early in life. Although the infant's gut is (nearly) sterile from viruses at birth, virus particles can be detected already at week 1 (Breitbart et al. 2008). Most of these viruses are phages that influence the survival of their bacterial hosts, thereby indirectly affecting the developing bacterial gut microbiome (Breitbart et al. 2008; Maqsood et al. 2019). The early virome composition depends on the birth route as infants born out of vaginal birth showed a higher virome diversity compared to their caesarean-birth counterparts (McCann et al. 2018). In another study, the birth route was shown to affect the phage-bacterial host interactions (Breitbart et al. 2008). Feeding infants with breast milk (at least partially) compared to an all-formula diet affected the infant gut virome and was connected with better health (Liang, Zhao, et al. 2020; reviewed in Turin and Ochoa 2014). Interestingly, rota- picorna-, and polyomaviruses, which cause viral gastroenteritis, can also be found in the viromes of healthy infants (Lim et al. 2015; Oude Munnink and van der Hoek 2016; Liang, Zhao, et al. 2020).

There are numerous benign viruses co-existing with the human host. *Anelloviridae*, for example, is a family of viruses commonly found in the majority of people (reviewed in Bendinelli et al. 2001) and that are drastically increased in patients with impaired immune response (Norman et al. 2015; Young et al. 2015). Anelloviruses indicate immunocompetence in transplant recipients and human immunodeficiency virus patients (Cebriá-Mendoza et al. 2021). Furthermore, colonization of the gut by viruses is needed for the correct functioning of intestinal lymphocytes (Liu et al. 2019) or for protection against gut inflammation through toll-like receptor activation (J.-Y. Yang et al. 2016). The "hygiene hypothesis" has been expanded to include the importance of viruses and targeted introduction of beneficial lactobacilli at early childhood reduces the risk of eczema and hay fever (Wickens et al. 2018). However, the composition of the virome can become unstable in the event of disease, as for example human cytomegalovirus (HCMV) can cause severe clinical symptoms and Epstein-Barr virus (EBV) can cause cancer in immunocompromised patients (reviewed in Meyding-Lamadé and Strank 2012).

2.1.1 Examples of importance and adverse effects of vaccines

In addition to the normal interaction between the developing human virome and the human body, the human health is also affected by certain medical interventions, such as vaccines. Since the 1950s, immunization has been used as an important public health tool and extensive vaccination plans have been put into place all around the world. This has significantly improved child mortality rates and even lead to some potentially lethal diseases like smallpox being eradicated (reviewed in Pollard and Bijker 2021). Due to strict regulatory oversight and well-documented clinical trials, vaccines are a safe and effective intervention to prevent diseases. There are some common mild side effects (also well documented) which include injection site pain, fever, malaise, headache, and mild viraemia. Serious side effects from vaccines are very rare, with anaphylaxis being the most common if the recipient has an allergy to any of the vaccine components or to traces of materials used during vaccine production.

During the COVID-19 pandemic of 2020-2023, fast development and distribution of different SARS-CoV-2 vaccines helped protect populations against serious complications and death from SARS-CoV-2 infection (Henry et al. 2021). On the other hand, vaccination against SARS-CoV-2 was associated with certain rare but serious side effects such as thrombosis, Guillain-Barré syndrome, acute transverse myelitis (neurological disease), myocarditis and pericarditis, and glomerular disease (Mushtaq et al. 2022). At least 10% of patients with SARS-CoV-2 infection developed long COVID which can be characterized by a wide variety of symptoms and comorbidities including cardiovascular, thrombortic and cerebrovascular disease, T2D, and chronic fatigue. One of the potential causes of long COVID is immune dysregulation and possible reactivation of underlying pathogens such as human herpesviruses (HHVs) (reviewed in Davis et al. 2023). We have also demonstrated the variable potential immune response to antigens of SARS-CoV-2 among the general population as well as potential antigenic mimicry with HHVs (Jaago, Rähni, et al. 2022).

Due to their sparse numbers, very rare side effects are usually only recognized through post-marketing surveillance. For example, after being administered the Pandemrix (Pdmx) H1N1 influenza A vaccine, people with a genetic susceptibility (carrying the *HLA-DQB1*06:02*) were more likely to develop narcolepsy (reviewed in Pollard and Bijker 2021). Pdmx is an AS03-adjuvanted monovalent vaccine which was used broadly in many European countries during the 2009-2010 H1N1 Influenza A outbreak, with Finland,

Sweden, Norway and Ireland reaching the highest coverage rates. The majority of the Pdmx-associated cases were of people developing narcolepsy type 1 (NT1) which is a form of narcolepsy where orexin-producing neurons in the hypothalamus are depleted due to an autoimmune attack (reviewed in Hallberg et al. 2019). The patients with NT1 usually suffer from increased daytime sleepiness, cataplexy (transient muscle weakness), wake-time hallucinations, sleep paralysis and disturbed sleep. Other biomarkers helping to diagnose narcolepsy are low levels of orexin in the cerebrospinal fluid and early onset of rapid eye movement phase sleep soon after falling asleep (reviewed in Bassetti et al. 2019).

2.2 Microbiome and virome as causative agents of health problems

Dysbiosis refers to a state in which the composition of the microbiome deviates from the normal and results in harm to the human host due to the rise of opportunistic pathogens. Opportunistic pathogens are usually found in low levels, but in the right conditions can cause serious disease. The members of *Enterobacteriaceae* family, which includes *Escherichia coli*, are usually commensal but can establish a dominant infection and contribute to gut inflammation and inflammatory bowel disease (IBD) (Ubeda et al. 2010; Winter et al. 2010; Buffie et al. 2012). The second feature common to microbiome dysbiosis is the loss of commensal constituents. Patients with ulcerative colitis or Chron's disease (both types of IBD) have decreased levels of *Bacteroidetes* and *Lachnospiraceae* (Frank et al. 2007), bacteria which are beneficial to gut health (reviewed in Bäckhed et al. 2005). Fortunately, in some cases re-introduction of these beneficial species has been shown to reverse pathological processes and alleviate disease symptoms. The third characteristic of dysbiosis is the decrease in microbiome alpha diversity, which represents the mean diversity of species at a given site and is often represented in Shannon's or Simpson's index value (reviewed in Mosca, Leclerc, and Hugot 2016; Thukral 2017). There are several causes to microbiome dysbiosis. Firstly, infection by a pathogen and subsequent inflammatory processes in the site can alter the micro-environment or directly or indirectly affect the survival of existing species (discussed in Hrncir 2022). Secondly, diet, medicine or environmental factors (e.g. pollution) can trigger dysbiosis. In fact, the largest effect the human host can have over its gut bacteria comes through diet (Zhang et al. 2010; 2012). The specific ways in which diet affects the microbiome have been discussed above. Thirdly, an individual's microbiome can be affected by their genetics. For example, patients with a genetic disposition for lactose intolerance had a higher abundance of *Bifidobacterium* in the gut microbiome (Bonder et al. 2016). Upregulation of vitamin D receptor, whose ligands include vitamin D but also microbial metabolites and fatty acids, has been linked to IBD and a decrease in *Parabacteroides* abundance in the gut (J. Wang et al. 2016). Fourth, as discussed above, maternal contribution early in life and the household environment affect the microbiome greatly (Palmer et al. 2007; Koenig et al. 2011; Lax et al. 2014; Maqsood et al. 2019).

An aberrant host-microbe interaction with gut bacteria has been implicated in the pathophysiology of Chron's disease and ulcerative colitis (Libertucci et al. 2018; W. Zhong et al. 2019; Britton et al. 2019) and patients with these diseases have a less diverse gut microbiome community (Ott et al. 2004; Manichanh et al. 2006; Harry Sokol et al. 2008; H. Sokol et al. 2009). Insulin-resistance-induced high blood glucose levels are associated with a decreased microbiota-epithelium distance which is a measure of the thickness of the healthy gut mucus layer (Chassaing, Raja, et al. 2017). A high-fat diet has been shown

to increase blood lipopolysaccharide (LPS) levels which in turn alters the gut microbial community and promotes obesity and insulin-resistance (Cani et al. 2007; 2008). Gut inflammation can boost *E. coli*, a bacterium with direct genotoxic effects, and thereby increase the risk of developing colorectal cancer (Cuevas-Ramos et al. 2010; Arthur et al. 2012; Pleguezuelos-Manzano et al. 2020). *Fusobacterium nucleatum* (*F. nucleatum*) has been isolated from colorectal cancer tissue (Castellarin et al. 2012). *Helicobacter pylori* infection and the host's response to it has been implicated in exacerbating symptoms in Alzheimer's disease (AD) patients (Kountouras et al. 2007; Roubaud-Baudron et al. 2012).

Although the gut microbiome is the largest and most studied portion of the whole human microbiome, bacterial communities residing in the oral cavity are also important for normal functioning of the body. However, the oral microbiome can also become dysbiotic and illicit negative effects on health. The mouth is a complex ecosystem with several distinct habitats for microbial colonization with over 750 prokaryotic taxa being identified in the oral cavity, of which around 50% are officially named (Human Oral Microbiome Database). The oral microbiome is maintained by both host- and microbe-derived factors. The salivary components, including secretory immunoglobulin A, lactoferrin, lactoperoxidase, lysozyme, statherin, and histatins, regulate the microbiome and keep it in balance (reviewed in Philip D. Marsh et al. 2016). Proteins from saliva and gingival crevicular fluid form the acquired pellicle, which regulates bacterial attachment to dental and epithelial surfaces and protects against acid effects (reviewed in Siqueira, Custodio, and McDonald 2012). Although the oral cavity is heavily colonized by bacteria, the monitoring and constant regulation by the host's immune system prevents most acute infections (reviewed in Zaura et al. 2014).

The balance between different species in the oral microbiome is important for maintaining oral health, and dysbiosis occurs when there is a disturbance in this balance and is often caused by modifiable factors such as poor oral hygiene, smoking, and changes in saliva flow and composition (Philip D. Marsh, Head, and Devine 2014; J. Wu et al. 2016a). Disease does usually not result from the presence of traditionally pathogenic bacteria but from a shift in the natural balance of the microbiota, allowing normally harmless bacteria to grow to higher proportions (reviewed in Phil D. Marsh, Head, and Devine 2015). The current understanding of dysbiosis development is that certain low-abundance microbial pathogens can cause inflammatory disease by affecting the host's immune response and composition of the microbiota. The presence of biofilm alone is not enough to enable progression into periodontitis – severe gum inflammation – but instead interactions between immune response mediators and the biofilm are required (reviewed in Meyle and Chapple 2015). Dysbiosis leads to the destruction of periodontal tissue by aberrant host inflammatory response (reviewed in George Hajishengallis 2014). The tissue breakdown provides resident bacteria with nutrients. The oral bacteria have evolved to use inflammation-derived nutrients, which supports further dysbiosis and tissue damage and results in a continuous pathogenic cycle (reviewed in G. Hajishengallis 2014). Some of the most prominent periodontitis-associated bacteria are *Porphyromonas gingivalis* and *Aggregatibacter actinomycetemcomitans* which thrive in a dysbiotic oral environment (reviewed in George Hajishengallis 2014). Smoking is a risk factor for many disease conditions, and it also promotes dysbiosis within the oral bacterial community. For example, a study performed in the USA observed that smokers had a significantly lower proportion of Proteobacteria and certain other taxa, whereas *Streptococcus* abundance was enriched. Functional analysis suggested that the depleted bacterial taxa were important actors in metabolising carbohydrates (J. Wu et al. 2016b).

A study with African American people have shown, however, that the compositional dysbiosis could be reversed by stopping the cigarette smoking habit (Y. Yang et al. 2019).

Dysbiosis in periodontal disease can trigger bacteraemia, leading to the systemic dissemination of oral bacteria, which can cause a number of systemic diseases including cardiovascular disease, rheumatoid arthritis, adverse pregnancy outcomes, stroke, inflammatory bowel disease and colorectal cancer, respiratory tract infection, meningitis or brain abscesses, lung, liver or splenic abscesses, appendicitis, pneumonia, and diabetes (reviewed in Y. W. Han and Wang 2013; de Pablo et al. 2009; Chapple, Genco, and Working group 2 of joint EFP/AAP workshop 2013; Dewhirst et al. 2010). Interestingly, while physical activity has beneficial effects on health, professional athletes have on average a higher incidence of certain oral inflammatory conditions, such as dental caries, periodontitis, or dental erosion. The proposed causes for the oral inflammation in athletes include nutritional differences (high amount of carbohydrates and sports drinks), eating disorders, repeating mouth dryness during training, and insufficient knowledge of best health behaviours. If left unchecked, local oral inflammation could develop into systemic inflammation (reviewed in Needleman et al. 2015).

IBD is a chronic inflammatory condition of the gastrointestinal tract, where dysregulated host-oral microbiota interactions play a crucial role in its pathogenesis (reviewed in Ni et al. 2017). Although the acidic conditions of the stomach and the existing intestinal microbiota usually prevent transmission to the lower gastrointestinal tract, some oral bacteria can still colonize the upper intestine (Bolei Li et al. 2019; reviewed in Lawley and Walker 2013). Several bacterial species have been associated with gut inflammation or IBD, including *Campylobacter concisus*, *F. nucleatum*, *Veillonella* spp., *Prevotella* spp., *Streptococcus* spp., and *Klebsiella* spp. (Man et al. 2010; Strauss et al. 2011; Vieira-Silva et al. 2019; Atarashi et al. 2017; Pascal et al. 2017). The development of IBD occurs over multiple progressive steps and can be exacerbated by virulence factors produced by these bacteria, but also by the use of antibiotics (reviewed in Read, Curtis, and Neves 2021). Inflammatory state of the gut microflora has also been observed in patients with coronary artery disease (CAD) CAD compared to healthy people (Jie et al. 2017). Furthermore, there is evidence that CAD patients also have alterations in their oral microbiome composition (reviewed in Acharya et al. 2017). The interaction between the host and the microbiome in the periodontium can trigger or exacerbate processes leading to atherosclerosis via innate immune system activation, bacteraemia, or directly by cytokines and inflammatory proteins (CARDIoGRAMplusC4D Consortium et al. 2013; reviewed in Grundtman and Wick 2011).

In addition to the bacterial microbiome having a negative effect on our health, the human virome can also promote or aggravate chronic conditions. Differences in virome composition have been observed in paediatric and adult IBD (Fernandes et al. 2019; Zuo et al. 2019; Liang, Conrad, et al. 2020), hypertension (M. Han et al. 2018), and colorectal cancer (Nakatsu et al. 2018). Specific virome composition has been characterized preceding type 1 diabetes (T1D) onset (G. Zhao et al. 2017; Wook Kim et al. 2019) and children with repeated exposure to enteroviruses at age 1-2 years have an increased risk for developing coeliac disease. HHVs are a group of double-stranded DNA viruses that have coevolved with humans throughout history. There are nine known HHVs routinely infecting humans: herpes simplex virus type 1 and 2 (HSV-1 and HSV-2), varicella-zoster virus (VZV), EBV, HCMV, and human herpesvirus 6 and 7 (HHV-6 and HHV-7) (reviewed in L. Zhong et al. 2023). HHVs are ubiquitous in human populations

and their latent or lytic infection can cause a range of clinical manifestations, from asymptomatic infections to severe and life-threatening diseases. During HHV entry, specific glycoproteins are required for membrane fusion. Even though HHVs infect different cell types, these enter the cell in a similar process. Most importantly, entry complex (EC) which is based on glycoproteins H and L, induces conformational changes in fusogen glycoprotein B (gB) which then prompts membrane fusion. Vaccines and neutralizing antibodies against EC and gB are tested in clinical trials to prevent HHV infection (reviewed in L. Zhong et al. 2023). The clinical manifestations of HHV infections can vary depending on the virus involved and the immune status of the host (reviewed in Griffiths and Reeves 2021).

HHVs have been linked to several chronic diseases, such as Alzheimer's disease. In a Taiwanese study, HSV infection increased the risk of developing dementia > 2 fold, whereas in another study, HHV-6A (type A) and HHV-7 were found in post-mortem brain tissue of patients with AD (Readhead et al. 2018; Tzeng et al. 2018). The trigger of neurodegenerative disease onset has been proposed to be the activation of latent HHV infection. Studies have shown that some types of HHVs, such as HSV-1 and HHV-6, can infect and replicate in brain cells, leading to chronic inflammation and neuronal damage. Furthermore, the reactivation of latent HHVs in the CNS can trigger an immune response and the release of pro-inflammatory cytokines, which are linked to neurodegeneration and cognitive decline (reviewed in Harris and Harris 2015). There is growing evidence that HHVs are associated with CVD. In particular, several studies have implicated HSV-1 and HCMV in the development of atherosclerosis and CVD (Zhu et al. 2000; G. C. Wang et al. 2010). HSV-1 and HCMV cause chronic low-grade inflammation and immune activation, which can contribute to the development and progression of atherosclerosis (Al-Ghamdi 2012; Y. peng Wu et al. 2016). Epidemiological studies have suggested that individuals with a history of HSV-1 infection as evident by serum antibodies have an increased risk of developing CVD (Sorlie et al. 1994; Zhu et al. 2000). For long, HSV-1, HCMV, and EBV have been shown to be present in atherosclerotic plaques and herpesviral DNA is found in human carotid and coronary artery specimens (Shi and Tokunaga 2002).

2.2.1 Pathogens and viruses contributing to molecular mimicry-associated disease

There have been multiple potential mechanisms in discussion about the HHV contribution to autoimmune disease, which include molecular mimicry, bystander activation, and epitope spreading. **Molecular mimicry** occurs when infectious pathogens express antigenic epitopes that structurally resemble epitopes of self-antigens, leading to the direct triggering of auto-reactive T-cells. For example, epitopes of EBV have been shown to resemble self-proteins and facilitate antibody binding (Lang et al. 2002). Delayed molecular mimicry, proposed by 't Hart and colleagues, is a variation of this model in which latent chronic infections create a repertoire of virus-specific memory T-cells, which can be reactivated when they encounter molecular mimicry motifs present in self-antigens ('t Hart, Hintzen, and Laman 2009). **Bystander activation** involves the accidental activation of autoreactive T-cells by cytokines produced by virus-specific immune cells or host cell destruction by viral infection leading to the release of cryptic epitopes, including self-antigens (Halenius and Hengel 2014). **A multistep process** has also been proposed where 1) epitopes on viruses interact with autoantibodies in susceptible individuals, and then 2) an unrelated immune system activation leads to autoimmunity (McCoy, Tsunoda, and Fujinami 2006). **Epitope spreading** occurs when antigens released from the primary

lesions in the target tissue prime a range of potentially autoreactive T-cells, leading to chronic autoimmunity. Fujinami et al suggested the “fertile field” hypothesis, which is also in line with ‘t Hart’s and McCoy’s view. The fertile field hypothesis proposes that exposure to a potential immunogen is usually harmless, but in specific circumstances, such as during a viral infection, the immune environment changes lead to a dysregulated immune response. Therefore, a viral infection can create a fertile field where immune responses to antigens can develop, as was explored by our group in the context of COVID-19 disease (Jaago, Rähni, et al. 2022). Additionally, primed autoreactive T-cells by viral infections also create a fertile field because later events might trigger the expansion and activation of these cells, leading to autoimmune disease (Fujinami et al. 2006). In summary, HHVs are a diverse group of viruses that can cause a wide range of clinical manifestations and are associated with various acute and chronic diseases. The interactions between HHVs and the human host are complex (**Figure 3**) and involve both viral and host factors, making the development of effective prevention and treatment strategies challenging.

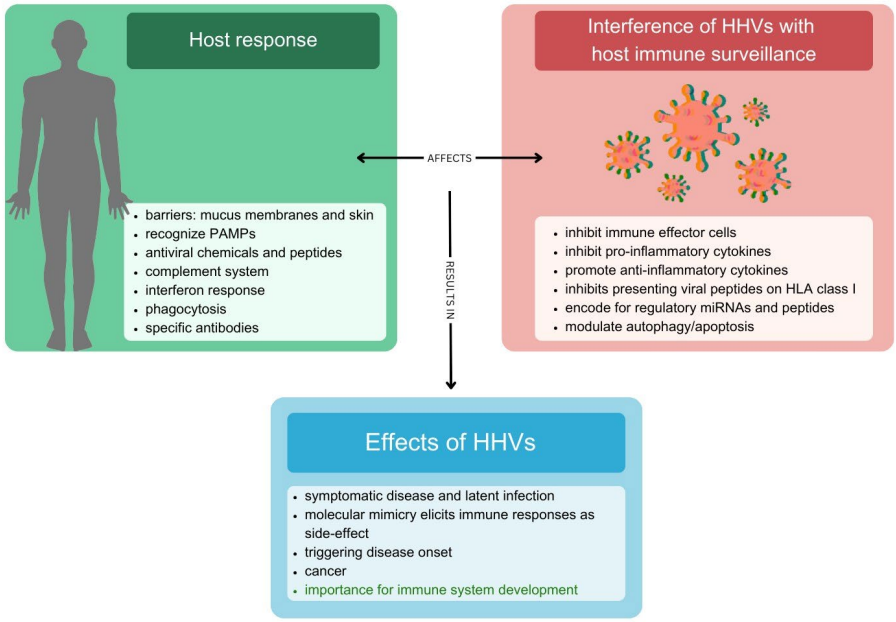


Figure 3. Interaction between HHVs and human immune system response and the (long-term) effects of HHVs. *The host has multiple protective strategies to protect itself from HHV infection: physical barriers such as mucus membranes and skin, recognition of pathogen-associated molecular patterns (PAMPs), deployment of antiviral chemicals and peptides, activation of the complement system and the interferon response, presence and functioning of phagocytic cells, and the development of specific antibodies. HHVs interfere with the host’s immune system by inhibiting immune effector cells, pro-inflammatory cytokines, presentation of viral peptides on HLA class I molecules, promote anti-inflammatory cytokines, encode for regulatory micro-RNAs (miRNAs) and peptides and modulate autophagy and apoptoic processes in host cells. Both the host’s response and HHV immune evasion strategies affect one another. Some long term effects that can arise from successful infection by HHVs include sympomatic disease and latent infection, molecular mimicry eliciting immune response as a side-effect, triggering disease onset, and cancer. However, infection by HHVs has a formative impact on the development of the immune system. Bullet points inspired by review articles (Jasinski-Bergner, Mandelboim, and Seliger 2020; Ike et al. 2020; Verzosa et al. 2021).*

3 Exploring the exposome

Exposome research is a comprehensive approach to studying a process in greater detail and from multiple aspects. Therefore, exposome studies could benefit from high-throughput and high-resolution methods that are robust and easy to use. This chapter will give an overview of some of the suitable tools such as metagenomics and Mimotope Variation Analysis (MVA). Furthermore, this chapter will introduce some innovative approaches that can be undertaken such as integrative omics and machine learning (ML) applications. Finally, the chapter will discuss the advantages of exposome research as well as some current challenges that warrant solving with a focus on data analysis.

3.1 Innovative tools (MVA and metagenomic sequencing) for data gathering

Studying the human microbiome has come a long way since its start. Initially, researchers could only study those species that were cultivable, for example when studying the microbiota of dental plaque (Gibbons et al. 1964). Late 1990s saw the first use of microarray technology to detect pathogens in microbiome samples (reviewed in Miller and Tang 2009). A polymerase chain reaction (PCR) step before the microarray increased the sensitivity of the assay (reviewed in Miller and Tang 2009). In 2004, a panel for detecting 40 of the most common human intestinal bacteria was introduced which used the 16S rRNA gene as a differentiating target (R.-F. Wang et al. 2004). The 16S rRNA genes encode for the RNA portion of the small subunit of the prokaryotic ribosome and the conserved nature of the genes allow to determine phylogenetic relationships between found bacteria (Woese and Fox 1977). Different technologies have emerged that allow fast detection and describing of pathogens (both bacterial and viral) based on nucleic acid sequences, including TaqMan PCR and Luminex xTAG bead-based systems (Navidad et al. 2013; Chapela, Garrido-Maestu, and Cabado 2015; Jiang et al. 2017). The Axiom Microbiome Array can detect 12’000 species of human- and animal-related bacteria, viruses, archaea, fungi and protozoa from a wide array of sampling materials, enabling its use both in a clinical and veterinary setting (Thissen et al. 2019). The Human Microbiome Project (HMP) was launched in 2008 that uses next generation sequencing (NGS) methods and today includes data on 32’000 samples which takes up almost 50 terabytes of storage (HMP homepage).

For microbiome studies, two main approaches are used: sequencing based on amplicon genes (e.g. 16S rRNA) or shotgun metagenomics. Gene amplicon sequencing is a widely used technology for determining microbiome composition. 16S rRNA is often used as the sequencing target due to its conserved nature and short length (~1500 bp) and it performs well for determining bacterial species (reviewed in Janda and Abbott 2007; Woo et al. 2008). Fungal diversity has been often assessed using internal transcribed spacers as amplicons, although certain other targets have proven more reliable (De Filippis et al. 2017). Metagenomics, on the other hand, allows for full genome sequencing and does not rely on the presence of certain target genes. This allows to sequence all microbes, including the unknowns. Because viruses have such diverse genomes, metagenomics approach enables to sequence the human virome (Kristensen et al. 2010; reviewed in Quince et al. 2017). Furthermore, metagenomics provides substantially more information than gene amplicon sequencing. On the other hand, shotgun metagenomics is still more expensive and it is hindered by the lack of available reference genomes and annotations in public databases (reviewed in Quince et al. 2017).

All the major NGS platforms are employed in microbiome studies: 454 pyrosequencing, Illumina platforms iSeq, MiniSeq, MiSeq, HiSeq, NextSeq, Ion Torrent; as well as third-generation sequencing approaches SOLiD and Nanopore (Margulies et al. 2005; "Illumina | Next-Generation Sequencing (NGS)" 2023; "Ion Torrent | Thermo Fisher" 2023; Clarke et al. 2009; one of first applications of SOLiD in Valouev et al. 2008). As the cost of sequencing has decreased, more and more sequencing data can be generated and analyzed. This has led to the rise in importance of data analysis pipelines and tools. Some of the most popular pipelines are EBI, MGRAST, QIIME/QIIME 2, and MOTHUR, all with slightly differing implementations, from which the researcher needs to select an appropriate tool based on their expertise (reviewed in Malla et al. 2019). Today, targeted genetic sequencing has helped identify antibiotic resistance in *Bacillus anthracis*, *Yersinia pestis*, and *Francisella tularensis*, all highly pathogenic bacteria (Stefan, Koehler, and Minogue 2016). Clinical metagenomics has been used in diagnosing infections in a routine clinical setting, decreasing test turnaround, as well as assaying the close to real-time transmission patterns of SARS-CoV-2 and tracking it in waste water (Geoghegan et al. 2020; Larsen and Wigginton 2020; Tao et al. 2022).

As environmental exposures affect the human immune system response, comprehensive profiling and mapping of exposome-immunome characteristics can benefit (chronic) immune-related disease research (Ronsmans et al. 2022). Omics technologies have brought innovative tools to immunology and immune system profiling, supporting large-scale research projects on genome-exposome interactions: the ImmVar study, the Human Immunology Project Consortium, the Milieu Intérieur study and the Human Functional Genomics Project (Brusic et al. 2014; De Jager et al. 2015; Thomas et al. 2015; Ter Horst et al. 2016; Y. Li et al. 2016). Clinical studies of the mechanisms of immune system-related diseases commonly utilize bulk DNA sequencing (DNA-seq), RNA sequencing (RNA-seq) and single-cell RNA sequencing (scRNA-seq), and assay for transposase-accessible chromatin using sequencing (ATAC-seq) technologies (Regev et al. 2017; Momozawa et al. 2018; Schultze and Rosenstiel 2018; Plichta et al. 2019). B- and T-cells are effector cells of the adaptive immune system. B- and T-cell receptor (BCR and TCR, respectively) sequencing focuses on the highly variable binding sites of the receptors and thus allows to characterize the host's response to pathogens (Schultheiß et al. 2020). Computational tools for TCR and BCR repertoire profiling have improved to an extent that information about TCRs and BCRs can be extracted from regular transcriptomics datasets. For example, tools such as MiXCR (Bolotin et al. 2017), TRUST (Bo Li et al. 2016), and V'DJer (Mose et al. 2016) can been used on RNA-seq data generated from a bulk tumour tissue. Furthermore, TraCeR (Stubbington et al. 2016), TRAPeS (Afik et al. 2017), BASIC (Canzar et al. 2017), BraCeR (Lindeman et al. 2018), and VDJPuzzle (Eltahla et al. 2016; 2016) are only some of the tools developed for extracting TCR or BCR profiles from single cell RNA-seq data.

MVA is a high-throughput phage-display method which allows to profile the immunoglobulin G (IgG) antibody immune response to random 12-mer peptides (Sadam et al. 2018) (**Figure 4**). First, patient sera containing IgG antibodies is incubated with the library of phages displaying 12-mer peptides, allowing for antibodies to bind specifically to peptides as antigens. As the IgG antibodies are bound to magnetic beads, the successful antibody-peptide complexes can be retained in the well as unbound antibodies and phage particles are washed away. Next, the phages are lysed and the DNA encoding for the displayed peptide is amplified with PCR. During PCR, custom barcode sequences are added to each sample, which allow pooling the samples together and

analyzing them in multiplex fashion. The displayed peptide encoding sequences are then sequenced using Illumina platform and de-multiplexed using barcodes, yielding around 3 million ~50 basepair reads per sample. DNA reads are then translated *in silico* into 12-mer peptide sequences and information about their relative abundance is calculated. Peptides together with their 1) sequence and 2) detected abundance values (reads) comprise the individual's immunoprofile. These peptides have mimicked the true epitopes of the antibodies in sample and are thus termed "mimotopes". Immunoprofiles of subjects can then be analyzed for individual-, infection-, or disease-specific features or used in longitudinal monitoring of the patient's health (Sadam et al. 2018; 2021; Jaago, Pupina, et al. 2022; Pupina et al. 2022; Rähni et al. 2022; Jaago, Rähni, et al. 2022). The advantages of MVA for the application of exposome research are discussed in more detail in the Results chapter.

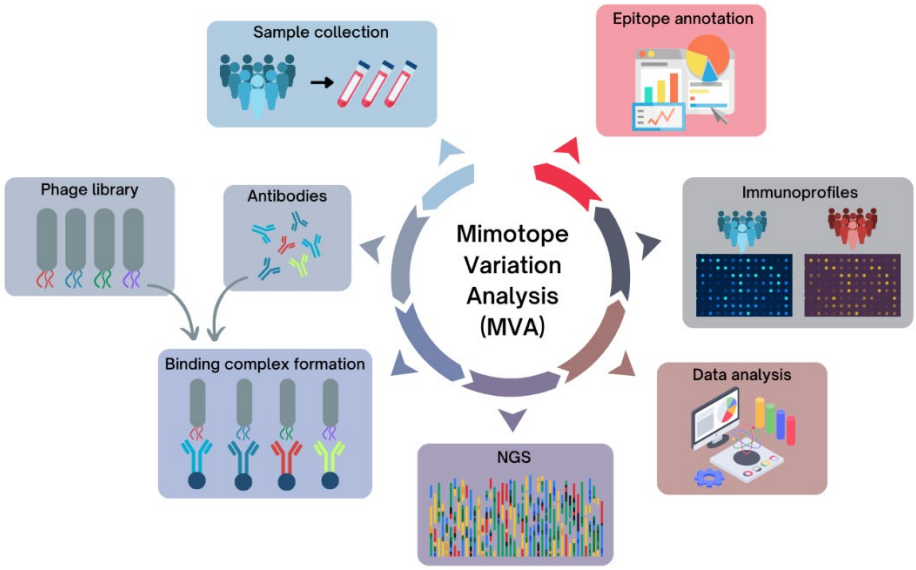


Figure 4. Overview of Mimotope Variation Analysis (MVA) process. *Serum or plasma samples are collected from patients. The IgG antibodies in the samples are incubated with phage library-displayed 12-mer peptides and magnetic beads to allow for triplex (IgG antibody + phage displaying 12mer peptides + magnetic bead) formation. These triplexes are barcoded per sample and purified and phage genomic insertions encoding for displayed peptides are sequenced using NGS. Then data is processed and analyzed to obtain individual immunoprofiles, which then can be mined for group-differential characteristics of antibody immune response. Epitopes of interest are annotated using online databases and scientific literature.*

In addition to new technologies for data acquisition, innovative approaches are being developed for the integration and processing of existing data, such as integrative omics and ML techniques. Integrative omics as a scientific approach is gaining adoption and application rapidly. Integrative omics refers to the analysis of multiple omics data, such as genomic, transcriptomic, proteomic, metagenomic etc, in parallel, to gain a more systematic insight about the studied samples. Different omics data have been combined to generate a multi-faceted view of the most important pathogenic bacteria associated with sepsis as they transition into the bloodstream, providing insights for personalized treatment and future drug discovery (Mu et al. 2023). Others have developed "ageing

clocks" to measure the patient's cellular, tissue, or whole organism biological age as a first step to combating age-related diseases at the root level (reviewed in Rutledge, Oh, and Wyss-Coray 2022). Novel emerging technologies are aiming to integrate information from genomics, proteomics, and transcriptomics to translate molecular immunology into clinics. Below is a subjective selection of a subset of attractive trending technologies. **Single-cell multiomics**: scRNA-seq combined with cell-surface protein expression data in a droplet microfluidics setup, cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq) and RNA expression and protein sequencing assay (REAP-seq) (Stoeckius et al. 2017; Peterson et al. 2017; "Single-Cell Multiomics Guide") methods allow to characterize cell types in greater molecular detail, generating large datasets with aims to reveal new insights into health and disease. A combination of barcode enabled antigen mapping (BEAM) with single-cell sequencing was used to profile the state of the cellular immune system after recovery from COVID-19, identifying important host-antigen interactions (Finnegan et al. 2022). **Spatial transcriptomics** adds 2D positional data to the RNA-seq dataset of a given tissue, by adding barcoded reverse transcription primers directly on a tissue slice and performing the copy or complementary DNA (cDNA) synthesis on the tissue surface, and then performing RNA-seq (Ståhl et al. 2016). Spatial transcriptomics approach has been used, for example, in studying location-associated characteristics of cardiomyocytes in the human heart during neonatal development (Sylvén et al. 2023). In neurobiology, spatial transcriptomics has been used for characterizing the development of the human hippocampus during development or for delineating disease progression, such as identifying T-cells intruding into the central nervous system (CNS) in multiple sclerosis (MS) (S. Zhong et al. 2020; Kaufmann et al. 2021). **Single-molecule protein sequencing:** Reed and colleagues have demonstrated dye-assisted sequencing of a single protein molecule with the ability to also distinguish single amino acid substitutions and posttranslational modifications (Reed et al. 2022).

Although there are some technical and data analytics challenges to be solved, the future of integrative omics is bright. As biosensors become compact and comfortable wearables, the continuous information they collect give valuable data on a person. An electrochemical regenerating wearable biosensor could detect different amino acids in trace amounts from a person's sweat in real time, enabling to potentially identify impaired metabolic function early on (M. Wang et al. 2022). Intelligent contact lenses are produced that allow performing an electroretinogram to diagnose ocular diseases or measuring inner eye pressure during glaucoma management in real time (T. Y. Kim et al. 2021; K. Kim et al. 2021). Another smart contact lens enables to assess glucose levels in the tear liquid and respond to aberrant levels with releasing drugs from a wearable reservoir to manage diabetic retinopathy (Keum et al. 2020). Biosensors can be either attached to items humans carry (watches, phones, clothes, woven into textile), wearable on the skin, wearable as skin patches, or implantable, and these measure physical (skin conductivity, electromagnetic field, proximity to cellular towers, gyroscope), biochemical (ions, small metabolites), or biological (microbe, virus particle abundance) factors (reviewed in Smith, Li, and Tse 2023). Some of the potential advancements for future biosensors include independent power generation, larger array of biomarkers measured, more microfluidics systems, and more transparent data collection and preservation policies (reviewed in Smith, Li, and Tse 2023).

As the multi-omics datasets grow and are combined with external environment measurements, socioeconomic and spatial data, the necessity ML algorithms in data

analysis pipelines increases. To integrate variable data types (e.g. genomic data with transcriptomic), there are ML and AI tools available which operate on the basis of similarity or correlation, or utilize network analysis, Bayesian, multivariate or fusion approach (reviewed in Subramanian et al. 2020). Different tools can be used for different use cases including the identification of subgroups in the data (disease subtyping), selecting for group-wide disease biomarkers or extracting details about the disease mechanism (reviewed in Subramanian et al. 2020). High dimensionality of exposome data can also be combated using the Gatekeepers approach, which first robustly selects for Gatekeeper metabolite biomarkers – those associated with both external factors as well as other metabolites – and follows with a refined search for other biomarkers (M. Yu et al. 2021). The following is a selection of some of the real-life use cases of ML in exposome research and application. An artificial neural network (ANN)-based approach predicts the metabolic profile and classifies IBS patients based on their gut microbiome composition, whereas another tool analyzes the microbiome composition in a time series to predict patient's diet, drug use or food allergies (Le et al. 2020; X. Chen et al. 2021). SkinBug integrates metabolic reaction data to predict the metabolising of molecules exposed to the skin (e.g. cosmetics, pollution, medicine) by the skin microbiome (Jaiswal et al. 2021). As a result of studying the exposure of breast cancer treatment, an ML model can predict individual therapy response based on the tumour's whole exome and transcriptomics markers in combination with pathological staining images (Sammut et al. 2022). A supervised ML tool helps to screen scientific literature and extract exposure biomarkers into a curated database named Exposome-Explorer, decreasing the need for manual resources and increasing the speed of knowledge sharing across researchers (Lamurias et al. 2021). In drug development, ML based tools have been used to predict the interaction and affinity of potential drug molecules to their targets, thus decreasing the cost of drug development due to shorter preclinical study phase. Zeng and colleagues were able to identify novel targets to existing therapeutics (Zeng et al. 2020). By repurposing drugs with known safety characteristics and developed supply chains they could also decrease the overall cost of drug development (Zeng et al. 2020).

As datasets grow, data storage and computing power must be readily available to perform analyses in real time. Cloud computing and cloud storage services offer needed performance and flexibility. Although cloud computing has advanced significantly over the years, there are still challenges of using cloud computing for computational biology. As discussed by Luber and colleagues, the existing off-the-shelf tools built for cloud computing can be inaccessible, cumbersome, and costly. Second, batch job-oriented cloud computing systems like Amazon Web Services Batch, Google preemptible Virtual Machines, Apache Spark, and MapReduce implementations can be closed source, restrictively licensed, or locked into their own ecosystems (Luber et al. 2018). Therefore, Luber and colleagues propose a tool called Aether (Luber et al. 2018). Aether is a tool that leverages linear programming to minimize cloud computing costs while being constrained by user needs and cloud capacity. Aether is better than current solutions as it allows for efficient and cost-effective analysis of large metagenomic data sets by bidding on underutilized computing nodes in the cloud, resulting in significant cost and time savings (Luber et al. 2018). However, when dealing with sensitive medical data, stringent protection measures need to be considered beforehand, as suggested by a checklist for assessing quality of cloud computing services (Kobayashi et al. 2022). Aware of the cybersecurity risks, Japan is building a national cloud computing infrastructure for genome analyses (reviewed in Ogasawara 2022).

Taken together, the landscape of possible technologies and approaches for exposome data acquisition is great and the use of ML tools in analyzing the multi-omics data has elevated the field of study and moved it further along on the way of clinical application.

3.2 Advantages of exposome research and challenges with data analysis

The exposome approach presents a novel way to investigate the underlying causes of health disparities and the biological mechanisms that link environmental exposures and human health. Various definitions of exposome exist, but the latest definitions classify exposome into three categories, namely internal exposome (such as metabolism, gut microbiome, and inflammation), specific external exposome (including environmental pollutants, occupation, education, and diet), and general external exposome (such as socioeconomic status and climate) (**Figure 5**) (Wild 2005b; 2012). There are numerous advantages to studying the exposome as a whole and not only focusing on the individual parts. Firstly, monitoring the exposome in a longitudinal way can identify important time windows where environmental challenges affect the health the most. For example, fine air particle exposure during the prenatal period has been associated with susceptibility to asthma and allergic outcomes (A. Lee et al. 2018). The large-scale EXPOSOMICS project aims to better understand the exposome by integrating data from different cohorts representing different life stages, such as prenatal, childhood, adolescence, and adulthood (reviewed in Agache et al. 2019).

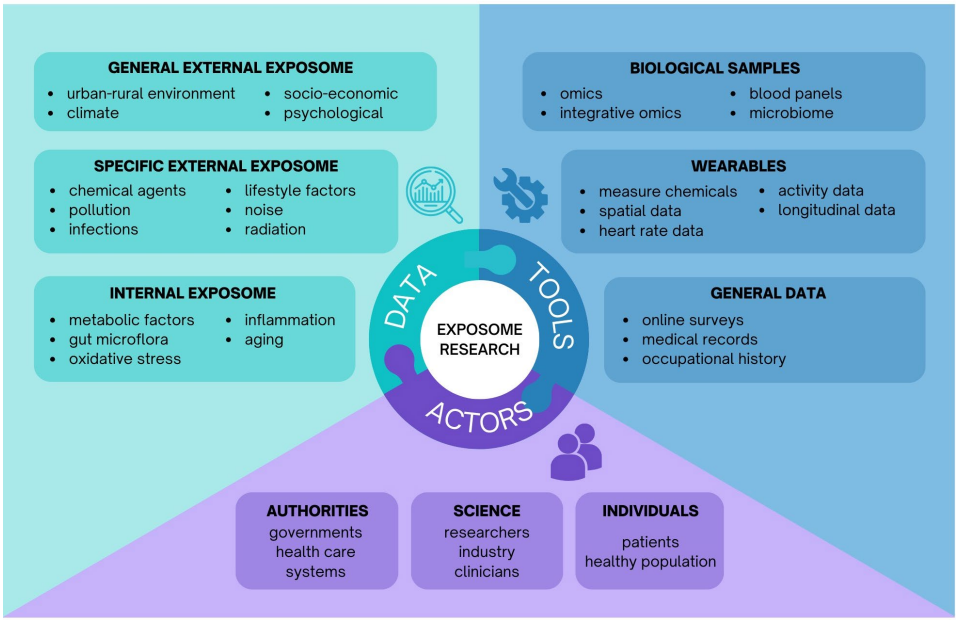


Figure 5. Overview of what data exposome research focuses on, the various tools used in exposome research, and the important actors involved in, regulating, or benefitting from exposome research. *Exposome research collects and analyses data concerning the general external exposome (e.g. overall urban-rural environment and climate, socio-economic status, and psychological/cultural overview), the specific external exposome (e.g. chemical agents and pollution, infections, lifestyle decisions, noise, and radiation), and the internal exposome (e.g. metabolic factors, gut microflora, oxidative stress, inflammation, and aging). Some of the ways data can be collected is taking biological samples (e.g. blood panels, microbiome samples) and performing (integrative) omics analyses, using wearables to gather continuous readings about environmental chemicals, spatial/locational data, heart rate and activity measures. Wearables allow for easier longitudinal datasets to be collected. In addition, general data can be gathered via online surveys, or by accessing medical records and occupational history. Results and insights from exposome research can be beneficial for authorities such as governments and health care systems in improving the overall health of the population. Exposome research can benefit from multi-disciplinary and multi-sector co-operation by researches, clinicians and the industry. Data can be collected both from patients as well as the healthy population. The data section is according to (Ferrante et al. 2022).*

Exposome research would promote using patient-centric molecular, environmental, and clinical data for precision medicine. The main challenges include integrating and analysing heterogeneous data sets (**Figure 5**), making personalized decisions using targeted but potentially noisier subsets of data, and ensuring data quality and reproducibility. The Biomedical Data Translator (BDT) program, sponsored by the National Institutes of Health, aims to address these challenges by developing a data integration and translation tool that can query knowledge sources describing human biology. Today, the BDT program or "Translator" includes 250 distinct knowledge sources, including environmental exposure data, socioeconomic data, electronic health record data, and open data sources on molecular biology, biological pathways, systems biology, and chemical structures and drug targets. Some use cases highlighted for "Translator" include 1) exploring the relationship between diseases and medication in a genetic or environmental background,

2) explaining the association between occurrences of different diseases, and 3) repurposing existing drugs for other diseases (Fecho et al. 2022; Unni et al. 2022). The BDT team has also developed a "Biolink Model" for presenting related information to the user, which has been used for aggregating and visualizing COVID-19 associated data in light of drug discovery, biomarker identification and other use cases (Reese et al. 2021; Unni et al. 2022). The Human Early Life Exposome (HELIX) project aims to characterize the impact of early life exposures to later stages of life by investigating more than 1300 pairs of mother and child (Maitre et al. 2022). Having measured >100 exposures, ranging from chemical to lifestyle-associated to social factors, the HELIX project has identified >1100 connections with a significant biological response. Exposures during pregnancy mostly lead to DNA methylation changes whereas distinct diet-, toxin-, or climate-related factors encountered during childhood resulted in a broad range of measurable changes. With access to dietary questionnaire data, animal product intake was identified as the probable route for toxic metal and other chemical element intake, and fruit and vegetable intake as a source of pesticide intake (Maitre et al. 2022). Exposome research in the field of age-related illnesses has identified specific pollution agents, physical, lifestyle, medication, as well as gut microbiota, inflammation, and socioeconomic factors that promote adverse health problems (reviwed in Ding et al. 2022). These studies exemplify how combining multiple data sources to generate a more comprehensive view of the impact of exposures can aid in disease research, drug discovery, and better public health policies.

There are many challenges today that need to be addressed to fully harness the power of exposome research.

- One of the major concerns is how to integrate data across multiple methodological approaches (conventional survey instruments or advanced sensors, geographic information systems, and biomonitoring and biomarker data), as different tools provide data in different formats, thus increasing the difficulty of comparison and integration (reviewed in Siroux, Agier, and Slama 2016; Turner et al. 2017).
- Wearable devices, such as sensors and monitors, are necessary to achieve this but there are compliance problems, where participants forget or refuse to wear the monitors, which in turn can lead to incomplete data or biased results. Behavioural changes are a concern, as people may change their behaviour in response to being monitored, which can affect the accuracy of the data collected (DeBord et al. 2016; Agache et al. 2019). Another challenge is the potential issue of radiofrequency radiation from wearable devices (Lagorio et al. 2021).
- Levels of environmental contaminants can vary throughout a person's life, and individual lifestyle changes can increase or decrease exposures (DeBord et al. 2016). Furthermore, certain periods of organ development have been identified as more susceptible to environmental exposures with regard to asthma and allergic outcomes (Agache et al. 2019). Therefore, it is crucial to measure the individual exposure across a long timeframe, which is cumbersome and results in large amounts of data.
- Massive data obtained from high-throughput platforms are sensitive to experimental conditions and can be affected by measurement errors (DeBord et al. 2016). Optimal sample size should be considered for the analysis to see

any significant common trends among a heterogenous sample pool (Schultze and Rosenstiel 2018). It is crucial to properly measure, document and quality control procedures to enable replication and meta-analyses of the results (reviewed in Siroux, Agier, and Slama 2016; Tran et al. 2020; Le Goallec et al. 2020).

As technological readiness improves and the challenges mentioned above are solved, exposome research will allow to profile the vast space of host-agent-environment interactions in a comprehensive manner that reflects the nature of the interactions in a more accurate way.

4 Aims of the study

Human health is affected by the environment through life, with the microbiome acting as both a target and a mediator of the environmental exposure. The diet and exercise routines of athletes are closely monitored and curated to increase athletic performance. However, less is known about how each individual responds to certain dietary regimes and how they impact the gut microbiome of the individual. Here we employed a gut microbiome profiling approach to gain more insight into the personalized response of a young athlete to dietary change. Host immunity, whether generated by natural infection or vaccination, is another variable that shapes the current environment for pathogens. In rare cases, H1N1 Influenza and the associated Pdmx® vaccination in 2009 were specific environmental triggers for narcolepsy onset. We were curious to examine the pathogen and vaccine induced immune response leading to disease. Another disease where the host's individual immune response has a high impact on pathogenesis, is periodontitis – chronic gum inflammation which is associated with dysbiosis in the composition of oral bacterial and viral community. Moreover, CVD is relatively common in patients with periodontal disease. Analysing the humoral immune response may provide information about the association of periodontitis and CAD and may provide valuable biomarkers to detect people with risk of disease earlier. Here we employ an unbiased comprehensive approach to analyze immune response profiles of patients with narcolepsy in a vaccine-background and, on the other hand, of patients with CAD in a periodontitis background. We also provide an analysis of the different data analysis workflows that can be used with different study designs and datasets.

The aims of the thesis were as following:

- Assess the heterogeneity of host-microbe interaction in the scope of gut microbiome profiling and antibody immune response profiling (**publication I, II, III**);
- characterize and compare data analysis challenges and approaches in different study designs (**publication I, II, III**);
- characterize the effect of dietary fibre supplement intake on an athlete's gut microbiome composition (**publication I**);
- characterize the effects of diet and exercise or poor lifestyle habits on host-microbiome interactions in the scope of gut microbiome and antibody immune response (**publication I, II**);
- find antigen epitopes that are specifically targeted in patients with narcolepsy or CVD (**publication II, III**);
- exemplify the potential of using MVA in exposome research and highlight considerations for data analysis (**publication II, III**).

5 Materials and Methods

The following methods were used to analyze antibody immune response characteristics and gut microbiome in the scope of this thesis:

- Analysis of gut microbiome sequencing data (publication I)
- Mimotope Variation Analysis (publications II and III)
- Data pre-processing and quality control steps (publications II and III)
- Feature selection (publication I, II and III)
- Identification of group classificators (publication I, II and III)
- Sequence alignment analysis (publications II and III)
- Disease classification analysis and biomarker prediction model building (publication II)
- Statistical tests (publication I, II, and III)
- Data visualization (publication I, II, and III)
- ELISA (publications II and III)
- Dot ELISA (publications II and III)
- Phage propagation (publication III) *
- Western blot (publication III) *

* not performed by author

6 Results and discussion

6.1 High-throughput profiling uncovers both intra- and interindividual heterogeneity in exposome profiles

The study of the human exposome is highly inter-disciplinary, deals with various sample materials and therefore requires many powerful experimenting methods. Given the broad nature of the exposome, ideal methods applied for studies are high-throughput and comprehensive, allowing to both reduce resource requirements (time and funds) but at the same time yield a large amount of data to give an accurate depiction of the underlying biology. However, having large amounts of data points to describe something may well result in a highly heterogenous picture. Here we have employed both microbiome profiling as well as MVA immunoprofiling to characterize the gut microbiome and humoral immune response, respectively, to exemplify the usefulness of both methods for exposome research, and to highlight the consideration of data analysis.

The gut microbiome has been shown to have a reciprocal relationship with athletic performance, as one influences the other and *vice-versa*. Furthermore, diet also has an impact on both the gut microbiome and athletic performance. Therefore, we aimed to profile the gut microbiome changes of a young athlete in response to dietary fibre supplementation in the period of active training and competition (**publication I**). The data was gathered from one person and included the following: training load data as minutes of activity per day, start and endpoint of regularly consuming a dietary fibre supplement, and three gut microbiome profile snapshots taken at the start of the study, before starting the fibre supplementation and after ending the 30-day fibre supplementation. The microbiome profiling data consisted of bacterial taxonomic groups with their abundance, consisting of groups at the phylum, genus, and species level. Although the gut microbiome profile consisted mostly of a few major bacterial phyla, diversity at the genus and species level was great. Across all three timepoints, 50 different genera were detected on average at a high enough abundance to accurately show in the profiling report. However, the abundance and proportions differed across timepoints, having certain genera emerge or disappear from the list and totalling 77 genera overall. On the other hand, there were 25 bacteria detected on species level on average across timepoints, totalling 32 altogether. One of the challenges with data analysis in this study was to identify suitable comparison data, as all the data points were taken from a single individual. Therefore, the data points were compared to the community average values provided together with the microbiome profiling report. However, one limitation of this approach is that the average community probably represents a less active person, and their microbiome composition could differ accordingly. Indeed, this could be observed from the higher average abundance of *Veillonella dispar* in the gut microbiome of the athlete compared to the control group. Another challenge with the data analysis was that not all taxa were represented in all timepoints, which made it difficult to assess the relative increase or decrease of the abundance. Therefore, we opted to visually represent the abundance values instead, to give a more accurate depiction of the underlying data and the heterogeneity of the microbiome profile. All-in-all, the gut microbiome profiling service with NGS was a beneficial tool to assess the microbiome diversity of an athlete during active fibre supplementation and training period, helping to piece together the complex exposome response of the human body.

The humoral antibody response in humans develops over time and is reflective of the surrounding environment, bacteria, viruses, and other organism exposures a person has accrued over the course of their everyday life (Lamont, Koo, and Hajishengallis 2018; Brodin and Davis 2017). Individuals differ from one-another based on overall immune cell and cytokine composition, and the specific composition of the antibody response is highly diverse across different people (reviewed in Brodin and Davis 2017). The humoral immune system has an important role in the defence a human body has against outside threats. However, antibody response can also facilitate a pathological autoimmune effect where the immune system incorrectly targets its own self proteins or other surfaces, causing a negative effect to health, as can be seen with T1D or MS for example (reviewed in Kawasaki 2014; Filippi et al. 2018). Given the many ways antibodies can influence a person's health status, either in a supportive and protective or pathogenic manner, it is beneficial to have means of analysing the antibody repertoire in its entirety to identify disease-associated biomarkers or monitor the overall health. MVA is a high throughput and unbiased method for profiling the comprehensive repertoire of a person's antibody targets. Using MVA we obtain data of short 12-mer sequences of peptides which have been recognized by antibodies in the analyzed sample. As the peptides have been similar enough to facilitate binding by an antibody, these peptide sequences mimic the true epitopes of the antibodies and are therefore termed mimotopes. Combining the sequences with data about how frequently the given sequences were captured by antibodies and observed in the dataset, we obtain the MVA antibody response profiles for each individual sample which can also be termed as MVA immunoprofiles.

We have employed MVA immunoprofiling technique on various clinical cohorts to analyze the links between antibody response characteristics and different clinical outcomes (Sadam et al. 2018; 2021; Rähni et al. 2022; Pupina et al. 2022; Jaago, Pupina, et al. 2022). Here we generated MVA antibody response profiles for patients with NT1, either caused by Pdmx vaccine (Pdmx-NT1, n=10) or sporadic (sNT1, n=6), H1N1-infected patients (H1N1-HC, n=16), Pdmx-vaccinated healthy controls (Pdmx-HC, n=16), and non-vaccinated healthy controls (Other HC, n=16) (**Table 1**). By comparing the mimotope sequences and abundance values in the immunoprofiles, we observed high heterogeneity across different patients (**publication III**). Among the Top2500 peptide dataset, which comprises the mimotope sequences of the 2500 most abundant peptide sequences per each sample, 86% of the mimotopes were unique to one sample (**publication III**). Although being highly heterogenous, the Top2500 peptide dataset did contain a shared portion, where the mimotope sequences were observed in multiple samples, indicating to a shared immune response feature (**publication III**). Most of the shared mimotopes among the Top2500 peptide dataset were observed in two samples (60.7%), 35.7% were seen in 3-10 samples, and 3.6% of the mimotopes were seen in more than 10 samples (**publication III**). Therefore, by using MVA to profile the humoral immune response in patients, we could see that there were common features shared by many people, however the majority of the response was highly diverse.

Table 1. Description of patients in clinical cohorts used in MVA immunoprofiling studies. *The group size (n), median age at sampling, gender proportions, ethnicity, Pdmx vaccination status, coronary artery disease, and periodontitis diagnoses are shown. n – group size; NT1 – narcolepsy type 1; sNT1 – sporadic narcolepsy type 1; Pmdx – Pandemrix, vaccine against influenza A(H1N1); HC – healthy controls, no NT1 diagnosis; Pdmx-HC – Pandemrix-vaccinated healthy controls; H1N1-HC – influenza A(H1N1)-infected controls; other HC – healthy Estonian controls; ACS – acute coronary syndrome; s-CAD – stable coronary artery disease; no-CAD – healthy controls with no coronary artery disease diagnosis; P – periodontitis diagnosis/patients with periodontitis; G – gingivitis diagnosis/patients with gingivitis; H – patients with no gum inflammation diagnosis; M – male; F – female; N/A – not available; Fin – Finnish ethnicity; Est – Estonian ethnicity.*

	n	**Median age at sampling**	**Gender (M/F)**	**Ethnicity**	**Pdmx-vacci-nation**	**Diagnosis**
Narcolepsy cohort						
NT1						
Pdmx-NT1	10	14	5/5	Fin	+	NT1
sNT1	6	22	1/5	Fin	+	NT1
HC						
Pdmx-HC	16	N/A	2/12	Fin	+	No NT1
H1N1-HC	16	21	0/16	Fin	-	No NT1
Other HC	16	34.5	6/10	Est	-	No NT1
Corogene cohort						
ACS						
P	16	63	14/2	Fin	N/A	ACS P
G	9	59	6/3	Fin	N/A	ACS G
H	7	61	4/3	Fin	N/A	ACS
s-CAD						
P	16	64	12/4	Fin	N/A	s-CAD P
G	9	69	7/2	Fin	N/A	s-CAD G
H	7	62	7/0	Fin	N/A	s-CAD
no-CAD						
P	16	64	10/6	Fin	N/A	P
G	9	61	4/5	Fin	N/A	G
H	7	55	3/4	Fin	N/A	No ACS No s-CAD No P No G

Next we generated MVA immunoprofiles for 96 individuals from the Finnish Corogene cohort (Vaara et al. 2012). The patients were diagnosed with acute coronary syndrome (ACS, n=32), stable coronary artery disease (s-CAD, n=32), or were controls (no-CAD, n=32) (**Table 1**). Furthermore, the same patients had also gone through oral exams to assess their gum health and were given a periodontitis diagnosis (P, n=48), gingivitis diagnosis (G, n=27), or deemed gum-healthy with no inflammation (H, n=21) (**Table 1**). Altogether we obtained 14.5 million peptide sequences across all 96 samples (**publication II**). The heterogeneity across immunoprofiles of patients was observed in the large amount of data – 14.5 million unique peptide sequences for the 96 samples analyzed. Furthermore, when extracting the most abundant and shared mimotope sequences, which had to be present in minimum 10% of the samples, the resulting dataset decreased down to hundreds of thousands of mimotopes (**publication II**). As MVA enables to capture many millions of mimotope leads per sample, our aim was also to verify that the high heterogeneity could not stem from experimental noise. Technical replicate data show a Pearson correlation of 0.87 at a single mimotope level between duplicate samples, suggesting high reproducibility of the results and no major effects from experimental noise (**publication II**). Taken together, these results suggest to a large heterogenous component within the individual MVA immunoprofiles and subsequently, within the complex repertoire of antibodies.

As discussed above in the Literature chapter, we are constantly exposed to common bacteria and viruses throughout our whole life and the human immune system co-develops together with the microbiome and the virome. Part of the heterogenous component of the humoral immune response is likely due to the individual history of the exact types of exposures, the time of exposures and the body's current health status at that time, as well as the order in which the exposures happened. Furthermore, the heterogenous component also encompasses the genetic variability of people, including the individual human leukocyte antigen (HLA) alleles and BCR and TCR repertoires. Thirdly, the environment in which children grew up in, together with its physio-chemical properties, as well as diet and exercise are a likely factor to affect the individual immune response profile. Nevertheless, it is highly important to assess the individuality in samples and to take it into account in further analysis and conclusions. One way to better evaluate and characterize the heterogenous component of MVA immunoprofiles is to increase the number of samples being analyzed while also focusing on a certain clinical diagnosis group and trying to decrease the number of background variables.

Although both included the data from MVA, the narcolepsy cohort and the CVD cohort exemplify different approaches that can be taken with data analysis (**Table 2**). Firstly, as the amount of initial data points is very high, feature creation and subsequent selection must be applied. Feature selection in MVA usually includes reducing the initially 12-mer peptide mimotopes into shorter consensus sequences, termed motifs. Motifs can be shorter in length and have only certain amino acid positions set whereas other positions can harbour variable amino acids. With the Corogene cohort, a non-discriminatory approach was chosen and patients were first divided into groups based on their clinical background (**publication II**). Then for each group, the top, most abundant and shared peptides were extracted from the dataset and subjected to a pattern detection algorithm to identify a smaller number of core motif sequences. This allowed to reduce the number of features to analyze further in the study without compromising much with the comprehensiveness of the dataset. However, this also meant that further analysis was required to identify features that would be group-differential. On the other hand,

with the narcolepsy study, a discriminatory approach was chosen for feature selection (**publication III**). Firstly, patients were again divided into groups based on their clinical background and the top, most abundant and shared peptides within each group were extracted. However, next the peptides from one group were compared against peptides from another group by scanning for any shorter core sequence motifs that were differentially detected across groups. This resulted in a smaller number of features to analyze further but which were also displaying group-differential characteristics. Although with the non-discriminatory approach no group-differential features were first selected, this would be achieved with the next step of determining features with most optimal sensitivity and specificity values for group classification. A receiver operating characteristic (ROC) curve was used in both studies to select for motifs that were either characteristic to the disease or control group (**publication II, III**). Finally, the group-stratifying motifs were annotated using different sequence databases. Due to the autoimmune background of narcolepsy, among other datasets, self protein sequence data was used to find regions similar to the identified motif mimotopes (**publication III**). On the other hand, the patients in the Corogene cohort had a chronic illness- and oral microbiome-related background, warranting the use of bacterial and common viral protein databases for annotation (**publication II**). After having selected the annotation databases, within the narcolepsy study, peptide sequences were mostly used for alignment analysis (**publication III**), whereas within the Corogene cohort, shorter core motif sequences were mostly used (**publication II**). Due to the greater number and diversity of the peptide sequences within the immunoprofile data, this allows for a more detailed alignment and could result in a clearer signal-to-noise ratio of the alignment, making it easier to detect true epitopes. On the other hand, using shorter core motif sequences is less resource-demanding and more robust, allowing to identify important epitopes that may not be entirely identical to the experimental data but similar enough to warrant an antibody-epitope interaction. All-in-all, there are different suitable approaches to take when analysing MVA immunoprofile data and a clear analysis plan should be designed based on the properties of the dataset.

Table 2. Similarities and differences of metagenomics and MVA immunoprofile data analysis approaches across studies.

	Athlete gut microbiome study	**Narcolepsy cohort**	**Corogene cohort**
Number of patients	1	48	96
Number of time points per sample	3	1	1
Feature selection	-	Discriminatory	Non-discriminatory
Identify group classificators	-	ROC	ROC
Alignment databases	-	Self proteins, Common human viruses	Oral bacteria, Common human viruses
Alignment input	-	Mostly peptides	Mostly core motifs

Taken together, both gut microbiome profiling with NGS and MVA are methods that can provide beneficial information about the interactions of humans with their environment, consisting of both non-living actors (diet) as well as living actors (microbiome and virome (with a caveat to term "living")). However, due to the large amount of data both methods generate, it is important to design an optimal data analysis plan that would suit best with the objectives of the study and the characteristics of the underlying data.

6.2 Characteristics of gut and oral microbiome composition due to lifestyle differences

6.2.1 Diet and exercise

Although there are multiple inheritable and non-heritable factors that influence the antibody response, the highest impact comes from the plethora of different viruses and microorganisms that we have exposure to. For example, the gut microbiota has been associated with athletic performance as certain bacterial species can promote cardiorespiratory fitness or increase in muscle mass (reviewed in Sohail et al. 2019; Ticinesi et al. 2019). Diet and exercise intensity have a high impact on the gut microbiota composition (reviewed in Das and Nair 2019), allowing athletes to affect the proportional balance towards bacterial species that offer performance benefits.

We analyzed the gut microbiome profile of a young athlete before and after their intense training period, as well as after having completed a month-long session of taking a dietary fibre supplement. To assess the overall diversity of the gut microbiome, we calculated the Shannon index value based on the microbiome profiling report. The Shannon index values were 2.11 and 2.08 before and after the training period, respectively, indicating no significant change in gut microbiome diversity (Hutcheson's modified t-test, p-value > 0.05, **publication I**). However, we observed a significant decrease in the microbiome community diversity after having completed a month-long dietary fibre course, as the Shannon index had decreased to 1.67 (Hutcheson's modified t-test, p-value < 0.05, **publication I**). This seems to be in contrast with most research which associate fibre intake with a higher microbial diversity (reviewed in Cronin et al. 2021). Interestingly, although the overall diversity had decreased after fibre intake, the proportion of *Firmicutes* bacteria had increased by 20% (**publication I**). Additionally, the dietary fibre intake had a drastic negative effect on the abundance of certain other major bacteria phyla: *Verrucomicrobia* (-94%) and *Cyanobacteria* (-97%) (**publication I**). By analysing the gut microbiome, we were able to classify the profile of the gut microbiome of the young athlete. Firstly, the ratio of *Firmicutes* phylum bacteria to *Bacteroidetes* (F/B ratio) was relatively stable during the observation period but significantly higher than in the control group in all time points (t-test of one mean, p-value < 0.05, **publication I**). An increase of *Firmicutes* bacteria has been shown to follow a training period and a higher F/B ratio has been associated with cardiorespiratory fitness (Estaki et al. 2016; Allen et al. 2018; Durk et al. 2019). Secondly, the microbiome composition indicated that the athlete had an enterotype II microbiome, where the abundance of *Prevotella* family bacteria is greater than abundances of *Bacteroides* and *Ruminococcus* families (**publication I**). After the intake of the dietary fibre supplement, the abundance of *Prevotella* increased by 41.7% (**publication I**), which is consistent with other research suggesting a link between fibre consumption and an abundance of *Prevotella* (G. D. Wu et al. 2011).

The association of a healthy dietary and exercise regime with the gut microbiome composition is a two-way influence (reviewed in Boisseau, Barnich, and Koechlin-Ramonatxo 2022). It has been shown that athletes have a higher diversity of gut bacteria compared to the average person (Mohr et al. 2020). The gut bacteria play an important role in nutrient uptake and energy metabolism, which are crucial processes for athletic performance. *Prevotella* has been associated with exercise duration, as subjects with a larger proportion of *Prevotella* among the gut microflora were reported to log more exercise time during a week (Petersen et al. 2017). Competing in an endurance sport such as rowing could then be benefitted by having an enterotype II microbiome with a larger proportion of *Prevotella*. A decreased diversity of microbes has been associated with immune-regulated poor health conditions (Valdes et al. 2018; Lozupone et al. 2012). However, prolonged heavy exercise instead affects microbial composition and increases intestinal permeability, which in turn can result in bacterial toxins, food antigens, commensal or pathogenic bacteria entering into systemic circulation and promoting systemic inflammation (Karl et al. 2017; reviewed in Vanuytsel, Tack, and Farre 2021). According to various studies, up to 90% of athletes complain of gastrointestinal problems, including vomiting, abdominal pain, and diarrhoea, and the prevalence of problems is higher in elite sports than in recreational activities (reviewed in de Oliveira, Burini, and Jeukendrup 2014).

6.2.2 Adverse health habits

Specific bacterial compositions are associated with negative effects on human health, for example periodontitis – a severe gum inflammation – is linked to the abundance of red- and orange-complex bacteria (reviewed in Sigmund S. Socransky and Haffajee 2002; Lamont, Koo, and Hajishengallis 2018). We analyzed the samples of 96 patients from the Finnish Corogene cohort where some of the patients had a diagnosis of gum inflammation (Vaara et al. 2012). The oral health of the patients was assessed, and patients were classified into gum-healthy (H, n=21), gingivitis (G, n=27), or periodontitis group (P, n=48). The proportion of currently active or ex-smokers was significantly higher among the patients who had a periodontitis diagnosis (Chi^2, p-value < 0.01, **publication II**). This is in agreement with other data, showing that smoking is a risk factor for developing periodontitis (reviewed in Nociti Jr, Casati, and Duarte 2015). Smoking introduces multiple harmful compounds to the oral ecosystem, which can influence the proper balance of different bacterial species (reviewed in Macgregor 1989). For example, a Danish study of saliva bacteria profiles found that compared to non-smokers, smokers have a larger abundance of *Streptococcus sobrinus* and *Eubacterium* [11][G-3] brachy (Belstrøm et al. 2014), bacterial taxa which are both associated with poor dental health (Loesche 1986; Loesche et al. 1975; Lafaurie et al. 2022). Cigarette smoking also has a positive effect on *Porphyromonas gingivalis*, a known periodontal pathogen, by allowing *P. gingivalis* to elicit a reduced inflammatory host response and thereby colonize the tissue more effectively (Bagaitkar et al. 2010).

T2D is also a disease that is associated with poor lifestyle decisions. Some research suggest that there is an association between T2D and the human body's bacterial ecosystem (Z. Chen et al. 2021). Furthermore, it has been shown that the risk of developing periodontitis is increased by 2 to 3 times in patients with diabetes, where most of the research is focused on T2D, although T1D has also been associated in teens and children (reviewed in Preshaw and Bissett 2019). T2D often correlates with chronic inflammation (mostly due to innate immunity) in the body, including the periodontium

where oral bacteria reside, and diabetics have a higher bacterial LPS-induced inflammation (reviewed in Nassar, Kantarci, and Van Dyke 2007; Mesia et al. 2016). On the other hand, periodontitis has been shown to exacerbate diabetic complications, possibly through LPS-induced cytokine production which in turn induces advanced glycation end product (AGE)-mediated cytokine response in diabetes (Grossi and Genco 1998). As a positive note, increased physical activity has been shown to reduce periodontitis risk, potentially because it also promotes better insulin sensitivity and glucose metabolism (Merchant et al. 2003). Taken all of the above, we were also interested in investigating the relationship between periodontitis and diabetes. Within the Finnish Corogene cohort where some patients had periodontitis, 19% of the patients were diagnosed with diabetes (insulin dependent and non-insulin dependent combined) (Vaara et al. 2012). However, we observed no significant proportional difference of patients with diabetes within patients within the gum-healthy or periodontitis groups (Chi^2, p-value > 0.05, **publication II**).

To investigate the interactions between the human host and bacterial microflora in more detail, individual MVA response profiles from 96 patients in the Corogene cohort were studied. As MVA provides information about the target epitopes of the antibody repertoire detected from the blood plasma, we compared the MVA-obtained mimotopes to the primary sequence information of known periodontal pathogens. We assessed the individual antibody response to 7 bacterial species, which have been specifically associated with periodontitis (*P. gingivalis*, *Tannerella forsythia*, *Prevotella intermedia*, *Fusobacterium nucleatum*, *Campylobacter rectus*, *Aggregatibacter actinomycetemcomitans*, and *Porphyromonas endodontalis*). For this we performed an alignment analysis, where we aligned MVA-obtained antibody response mimotope profiles to the primary sequences of proteins from the bacteria. As a result, we observed a more abundant predicted immune response to epitopes of *P. gingivalis* and *A. actinomycetemcomitans* compared to *F. nucleatum* (Mann-Whitney U test, **** p-value < 0.0001, **publication II**). We identified an abundant core mimotope pattern P..T.[P][R] which existed in 45% and 30% of *P. gingivalis* and *A. actinomycetemcomitans* immunodominant antigens, respectively, but only in 10% among *F. nucleatum* antigens (**publication II**). *P. gingivalis* is regarded as one of the keystone bacteria involved in the progression of periodontal disease and it's correlated with more severe periodontitis (Bostanci and Belibasakis 2012; Holt and Ebersole 2005; S. S. Socransky et al. 1998), whereas *A. actinomycetemcomitans* has been associated with localized aggressive periodontitis (J. Slots, Reynolds, and Genco 1980; Haubek et al. 2008). *F. nucleatum*, a commensal bacterium, has been shown to support the survival of *P. gingivalis* in an oral biofilm setting (Bradshaw et al. 1998). Research findings show that both the synergistic cooperation of the oral (opportunistic and pathogenic) bacteria, as well as the host's immune response play an important role in the pathogenesis of periodontitis (reviewed in Sedghi, Bacino, and Kapila 2021).

Our findings of potential antibody immune response targets on the oral bacteria associated with periodontal disease contribute to the existing knowledge and warrant future studies into the detailed interplay of the host-microbe interactions. Furthermore, it is interesting to contemplate whether we could improve our overall health and avoid overweight/obesity-related comorbidities by deliberately controlling our microbiome through physical exercise and proper diet. However, more insight is needed, and many challenges must be overcome in order to introduce this strategy into everyday personal life or public health plans.

6.3 Antibody profiles enable to identify epitopes that are differentially recognized in case *versus* control cohorts

We hypothesized that there were differences in the antibody immune response between healthy individuals and patients with a certain clinical condition. Within the narcolepsy study cohort, Pdmx-vaccinated patients with NT1 diagnosis had a lower antibody response to major antigens of H1N1 as compared to Pdmx-vaccinated healthy controls (**publication III**). As described in the literature overview, herpesviral and certain bacterial infections such as *Helicobacter pylori* are common among the human population and have been associated with many pathologies. Therefore, we used ELISA to measure the IgG immune response against HCMV, EBV capsid antigen (CA), and *H. pylori* among patients in the Corogene cohort. We found decreased IgG serology against EBV CA in gum-healthy patients (H) who had ACS diagnosis compared to the no-CAD group (Mann-Whitney U, * p-value < 0.05) (**Figure 1**). Similarly, in periodontitis patients we observed a decreased IgG response against HCMV in patients with s-CAD or ACS diagnosis compared to the no-CAD group (Mann-Whitney U, * p-value < 0.05, *** p-value < 0.001) (**Figure 1**). On the other hand, we observed that patients with periodontitis were significantly more likely to be seropositive for *H. pylori* compared to groups with better periodontal health (Chi^2, p-value < 0.05) (**Figure 1**). An association between *H. pylori* and periodontitis has been observed before, as *H. pylori* had been detected in dental plaques and its detection rate correlated with a more severe state of periodontal disease (Zheng and Zhou 2015). Taken together, we could characterize differences in the overall seropositivity to *H. pylori* among patients with different periodontal status, and IgG seroresponse to HCMV or EBV CA protein was differentiating between CAD groups (**Figure 1**).

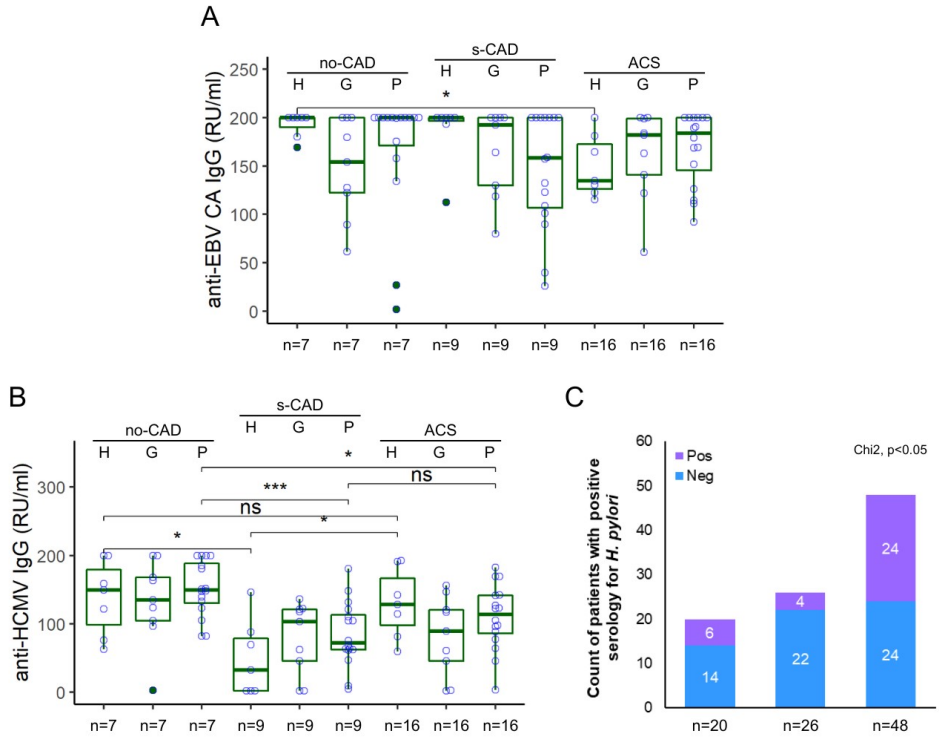


Figure 6. IgG seroreactivity to common human viruses and H. pylori in patients with NT1, CAD, or periodontitis. *IgG seroreactivity was measured using commercial ELISA test kits. Levels of the detected IgG response shown in relative units (RU)/ml. A. IgG immune response against influenza A(H1N1) in NT1 patients and healthy controls. Statistical test: ANOVA, ***, p<0.0001. NT1 – narcolepsy type 1 (n=16); Pdmx-HC – Pandemrix-vaccinated healthy controls (n=16); H1N1-HC – inluenza A(H1N1) infected healthy Finnish controls (n=16); HC – healthy Estonian controls (n=16). B. IgG immune response against EBV CA. C. IgG immune response against HCMV antigens. D. Proportion of patients IgG seropositive for anti-H. pylori antibodies across periodontal disease groups. Statistical test: Chi2, p<0.05. B-D. Statistical test: Mann-Whitney U, *** p<0.001, * p<0.05, ns, p>0.05. ACS – acute coronary syndrome; s-CAD – stable coronary artery disease; no-CAD – patients with no coronary artery disease diagnosis; H – periodontally healthy; G – gingivitis: P – periodontitis; HCMV – human cytomegalovirus; EBV CA – Epstein-Barr virus capsid antigen.*

When analysing the antibody immune response of patients within the Corogene cohort using MVA, we identified 5 major mimotopes which were differentially targeted across clinical diagnosis subgroups (**publication II**). Antibody response against a mimotope with a consensus sequence P..T.PR (mimotope A) (Proline-X-X-Threonine-X-Proline-Arginine where X denotes any amino acid) was observed to be stronger among patients with a periodontitis diagnosis who did not have coronary artery disease (no-CAD) (**publication II**). The consensus sequence P..T.PR (mimotope A) is similar to some of the consensus sequences (type 2, 4, and 5) delineated in the bacterial alignment analysis as immunogenic features with a more frequent response (**publication II**). Therefore, patients with periodontitis but no CAD diagnosis who showed a higher response against P..T.PR mimotope could have a different immune response, and thus effect the composition of the oral periodontal bacterial community as compared to those periodontitis patients

who did have a CAD diagnosis. Having periodontitis increases the likelihood of a person having atherosclerotic vascular disease, although the causality and the potential mechanism behind this link is not yet clear (Lockhart et al. 2012). Periodontitis and CAD both have an immunoinflammatory background and are known to share many of the same risk factors: unhealthy lifestyle, smoking, diabetes mellitus (Kott et al. 2020; reviewed in Tsuchida 2020; Kobiyama and Ley 2018). Within our studies, we also identified a characteristic immune response profile in patients with a smoking background or with a diabetes diagnosis background (**publication II**). Taken together, we have provided a way of identifying the phases of CAD by using immune response profile differences, some of which targeted specific epitopes of oral bacteria (**publication II**).

We have shown that the immune response profiles are different across clinical groups within the narcolepsy cohort (**publication III**). After identifying the specific mimotopes which were being differentially targeted by the immune system, our next aim was to classify the true epitopes against which the immune response had been created. Within the MVA immunoprofile data we determined a peptide mimotope with the sequence RVLAPALDSWGT which was enriched in the profiles of patients with NT1 as compared to other clinical groups (**publication III, Table 3**). Annotation analysis using sequence alignment approach identified the second extracellular loop region of the human prostaglandin D2 receptor 1 (DP1) (UniProt accession Q13258) as having high sequence homology to the peptide mimotope. Furthermore, patients from the Pdmx-NT1 and sNT1 groups also exhibited a higher predicted antibody immune response to protein BCL6 (**Table 3**), which regulates the germinal centre reaction where B-cells are selectively driven to produce higher-affinity antibodies (ref in Pei et al. 2017). These findings of specific immune response characteristics in patients with NT1 highlight the complex nature of the disease and suggest to an interplay of different molecular targets and pathways.

Table 3. Human and virus protein epitopes associated with a differential immune response in disease. *UniProt accession codes: human DP1 - Q13258; human BCL6 - P41182; Influenza A(H1N1) HA - C4RUW8; EBV VP26 – Q3KSU9; EBV ENBA6 - P03204. DP1 – prostaglandin D2 receptor DP1; BCL6 - B-cell lymphoma 6 protein; HA – hemagglutinin; VP26 - Small capsomere-interacting protein; EBNA6 – Epstein-Barr nuclear antigen 6; NT1 – narcolepsy type 1; ACS – acute coronary syndrome; no-CAD – patients with no coronary artery disease diagnosis.*

Antigen	Epitope sequence	Amino acid position	Diagnosis association	Other info
Human DP1	RVLAPALD	94-101	NT1	
Human BCL6	EGLKPAAPSA	279–288	Loosely NT1	(Zandian et al. 2017)
Influenza A(H1N1) HA	KLESTRIYQIL	521-531	Natural H1N1 infection	
EBV VP26	QPQDTAPRGARKKQ	153-176	ACS in G No-CAD in P	(Loebel et al. 2017)
EBV EBNA6	PAPQAPYQGYQEPPA-PQAPYQGYQEPPP-PQAPYQGYQEP	741-779	P with smoking	(Falk et al. 1995; Loebel et al. 2017)

When analysing the Corogene cohort of patients, some with CAD and/or periodontitis diagnosis, we observed a strong antibody immune response to a mimotope with the pattern P..T.PR. Our previous work has identified this mimotope as originating from an immunodominant region QPQDTAPRGARKKQ on Small capsomere-interacting protein (VP26) protein of EBV, encompassing amino acid positions 153-176 (Sadam et al. 2018; 2021). Alignment analysis performed using individual MVA immunoprofile data and the primary sequence of VP26 protein resulted in high predicted immunoreactivity to C-terminus of VP26 (**publication II**). Interestingly, the MVA-predicted reactivity to mimotope P..T.PR was significantly higher in patients with a periodontitis diagnosis compared to gum-healthy controls (Mann-Whitney U, * p-value < 0.05) (**publication II**) (**Table 3**). To validate MVA findings, we performed a dot ELISA analysis, where phage particles displaying the C-terminal VP26 epitope were fixed onto nitrocellulose membrane. Patient sera predicted as positive for anti-P..T.PR reactivity by MVA exhibited higher signals at the dot ELISA experiment (Mann-Whitney U, **** p-value < 0.0001, **publication II**). Taken together, we have observed a higher antibody response to C-terminal epitope of VP26 of EBV in patients with a periodontitis diagnosis. We also performed an alignment analysis to describe any proteins of the common periodontal bacteria or common HHVs with sequence similarity to the P..T.PR mimotope which could potentially facilitate cross-reactive immunity. We identified transmembrane protein signal peptidase I (Uniprot accession Q7MTG1) and transcription termination factor Rho (Q7MX79) of *P. gingivalis,* and isoleucine-tRNA ligase (C9R644) of *A. actinomycetemcomitans* as sharing sequence similarity to mimotope P..T.PR (**publication II**). In addition, we also predicted mimotope B to be associated with a protein of EBV (**publication II**). Alignment analysis showed that mimotope B with the sequence P[Y]..[Y]Q shares sequence similarity to EBV protein Epstein-Barr nuclear antigen 6 (EBNA-6) at amino acid positions 741-779, where there is a triple repeat sequence (**Table 3**). Mimotope B had a higher predicted immune response specifically in patients with periodontitis and a background of smoking (**publication II**). EBV infection elicits the production of antibodies against many different EBV proteins, including glycoprotein 42 (gp42), Epstein Barr nuclear antigen 1 (EBNA-1), 2 (EBNA-2), and 6 (Bu et al. 2016; Henle et al. 1987; Lennette et al. 1993), however we could only identify the anti-VP26 immune response as differential between patients with periodontitis compared to gum-healthy patients, and anti-EBNA-6 response characteristic to patients with periodontitis and a history of smoking (**publication II**).

Herpesvirus infection, including EBV infection, has been associated with periodontal disease before (Gao, Lv, and Wang 2017). An interesting topic is the complex interplay between EBV and periodontal pathogenic bacteria (reviewed in Jørgen Slots 2015), however the exact molecular mechanisms behind the link are not yet fully elucidated. One possible mechanism describes EBV as the promoter of periodontal pathogen increase by inducing the release of pro-inflammatory cytokines which in turn activate osteoclasts and matrix metalloproteases (MMPs) to inhibit an adequate antibacterial immune response (reviewed in Jørgen Slots 2010). On the other hand, some studies suggest that *P. gingivalis* secretes butyric acid into its local environment, which in turn induces EBV re-activation (from latent form) in infected cells (Imai and Ogata 2020).

Taken together our findings show a strong specific immune response to EBV in patients with periodontitis and provide further and more detailed evidence for the causal link between oral EBV infection, oral microbiota composition, and lifestyle habits such as smoking. However, it would be important to gain more detailed knowledge about the

role of EBV in the context of chronic diseases. Failure to handle the primary infection of EBV correctly may cause lasting health problems in the coming years. Furthermore, future studies should focus on characterizing in detail how an individual's genetic background affects their interaction with EBV exposure and any resulting health risks. Mapping how mimicry of EBV epitopes with human self proteins affects the host-microbe interaction is also an important research question for future studies. Given the role of EBV in systemic diseases, herpesviruses could constitute as potential targets for new vaccines to mitigate or reduce some of the symptoms of systemic diseases. However, before that, diagnostic tools should be developed to map in detail the precise disease-associated herpesviral antigens and epitopes which could be used as inputs in vaccine or immune drug developments.

6.4 MVA and microbiome studies as valuable tools for exposome research

As the gut microbiome makes up most of the whole microbiome of a person, it is highly important to study the microbiome-host interaction in scope of exposome research. In the study with the young athlete, in addition to profiling the gut microbiome at certain time points, the athlete's training load was also monitored alongside with a dietary intervention. Even with this limited data, interesting insights could be gathered from the correlations of microbiome and lifestyle changes. However, it would have been more beneficial to have received the raw data from the microbiome profiling to analyze the patterns in more detail. Furthermore, and where exposome research is also moving towards, the study could have benefitted from more detailed health data from a wearable activity monitor that could have kept track of heart rate, activity load, and sleep patterns. Combined with a more detailed diet log, the study could have provided a more in-depth investigation into the interplay of athletic activity, diet, and the gut microbiome.

Turning to another powerful technology, MVA could be a valuable tool within the scope of exposome research due to multiple reasons. Firstly, MVA allows to generate a comprehensive snapshot of a person's individual immune response profile (**Figure 7**). The default MVA pipeline generates 3 million sequencing reads per individual sample. Sequencing reads are transformed into an MVA immunoprofile comprising of peptide sequences and their capturing frequency, where on average there are hundreds of thousands of different peptide sequences captured per sample. The high-throughput nature of MVA places it above other methods that allow to investigate the functionality of existing antibodies, such as microarrays or ELISA panels. Microarrays and ELISA panels usually require the directed synthesis of target molecules (for example, full proteins or protein fragments) and they can usually yield up to tens of thousands of data points, compared to hundreds of thousands of data points generated via MVA.

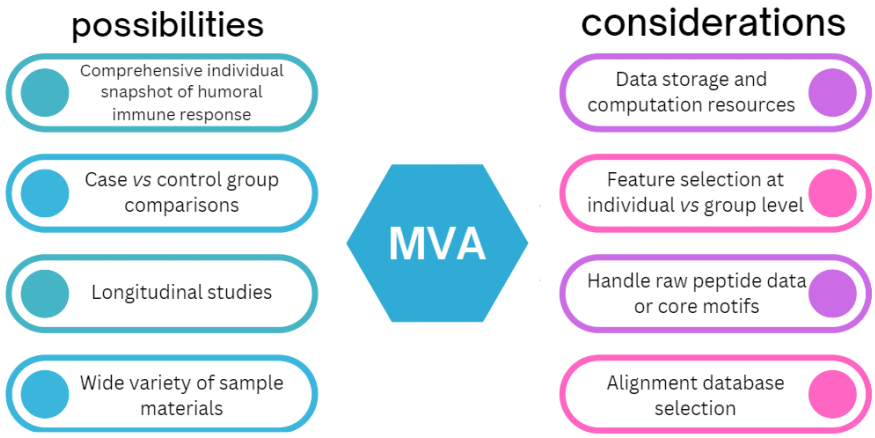


Figure 7. Possibilities MVA offers for exposome research and required considerations. *MVA allows to obtain comprehensive individual snapshots of the humoral immune response at a given time. MVA immunoprofiles can be used in case versus control group comparisons and longitudinal studies, allowing to delineate both health- and disease-specific immune response characteristics. MVA can be performed on a wide variety of sample materials. Before performing MVA, data storage options and computational resource availability should be considered. During MVA data analysis, decisions should be made about whether to perform feature selection at the individual or the group level, whether to handle and compare raw peptide data or condense peptide sequences into shorter core motifs. The selection of which data to use for alignment analysis is important in order to reduce noise and increase the chance of identifying biologically relevant results.*

Secondly, by analysing different disease cohorts to identify specific immune response features, MVA can prove useful in the discovery of early blood-based biomarkers. For example, based on the evidence presented above about the association between CAD and periodontitis, it becomes important to take the periodontal disease background into consideration when creating risk analyses or prognosis of patients. Knowledge of the differences in the antibody immune response to certain specific epitopes could be utilised in developing blood-based biomarkers to help evaluate the current health condition. Blood-based biomarkers offer great benefits as they are time- and cost-effective and non-invasive. Using only the MVA-predicted immune response to three different mimotopes P..T.PR, P[Y]..[Y]Q, and [P][W].P.SPF we could differentiate patients with periodontitis from gum-healthy controls with a balanced accuracy of 81% (**publication II**). Higher response to mimotopes P..T.PR and P[Y]..[Y]Q was characteristic to patients with periodontitis, however higher response to mimotope [P][W].P.SPF was associated with periodontally healthy status (**publication II**). Then, across the different periodontal disease groups, the immune response characteristics associating with CAD were different. In the group of no periodontal inflammation, high response to mimotope [P][W].P.SPF was observed in patients with an ACS diagnosis (Mann-Whitney U, p-value = 0.097, **publication II**). Patients with gingivitis and ACS exhibited a higher immune response to mimotope P..T.PR (Mann-Whitney U, * p-value < 0.05, **publication II**). On the other hand, in subjects with periodontitis, a high response to mimotope P..T.PR correlated with the no-CAD phenotype instead, together with an independent mimotope [G]P.HT.K (Mann-Whitney U, * p-value < 0.05, ** p-value < 0.01, **publication II**). Therefore, considering the periodontal disease background of the subjects first, and then assessing the MVA-predicted immune response to three independent mimotopes we could differentiate patients with

an ACS diagnosis from the no-CAD controls with sensitivity 71-89% and specificity 67–100% (**publication II**). Early detection of bacterial-dysbiosis-related biomarkers can shed light on pathogenic processes early on and therefore could result in more efficient treatment and better health outcomes. Considering the person's individual background, this approach could also facilitate a targeted pre-emptive approach in the scope of precision medicine.

Thirdly, when monitoring the human exposome over time, multiple snapshots of MVA-generated immune response profiles from a single person could be analyzed. This would enable to monitor the individual baseline antibody composition, the progression of infections, or notice the emerging signals of known biomarkers of disease. An example of a longitudinal study that analyzed MVA immune response profiles was by Rähni and colleagues (Rähni et al. 2022). There, melanoma-specific response was observed to emerge after treatment with a cancer vaccine and to change over time, correlating also with the disease symptoms (Rähni et al. 2022). Furthermore, MVA profiles could be correlated with the person's location history, medical background, familial medical history, and infectious disease exposure, to obtain a more comprehensive image.

Fourth, MVA technology can be employed to study various sampling materials, including blood serum, plasma, CSF, any synthetic samples containing antibodies, as well as biomaterial surfaces and biomaterial-exposed fluids (**publication II**; Sadam et al. 2021; Vrana, Palm, and Lavalle 2020). This allows to seamlessly plug MVA into various study designs without adding many complications to sample gathering and pre-processing procedures. Furthermore, MVA is rather insensitive to the exact time the sample was taken, because given optimal conditions, samples can be stored cryogenically for a long time before analysis with MVA is performed and samples can also endure multiple thaw-freeze cycles. This allows to gather samples on-site in clinics or emergency rooms, which is more convenient to the patient, and then analyze centrally at other locations, decreasing the need of special equipment and procedures in multiple places. Another beneficial implication is that MVA can be performed retroactively on already gathered samples to supplement the knowledge base and data. Moreover, MVA needs only a few microlitres of sample to yield results.

However, despite all the advantages MVA offers to exposome research, the data analysis portion must be thoroughly discussed and planned beforehand. As MVA yields a large amount of data per sample, generating even larger datasets per studied cohorts, data storage solution should be planned, allowing for the safe storage of data and convenient retrieval of data for analysis. Depending on the study size, either powerful personal workstations or computing cluster resources could be used for computational analysis. Next, the data analysis plan should decide whether to analyze the samples separately or converge them and retrieve shared features to analyze further. For example, in the annotation analysis in the Corogene cohort, individual peptide sequences were aligned onto EBV protein VP26 separately for each individual (**publication II**), allowing to preserve high individual detail in the result. On the other hand, feature selection process can highly benefit from analysing a group of samples as a whole because the data volume of initial features is great and any shared abundant characteristics would also provide robust and relevant biomarkers for further research (**publication II, publication III**). Careful thought should also be placed on which databases would be used for annotating the mimotope features from MVA. A truly unbiased approach would include too many database resources and thus yield many false positive results, offering no real benefit into describing the biological process. On the other hand, intelligent decision on which

alignment data to use would allow to better pinpoint biologically relevant results. For example, in the narcolepsy study, self proteins were included in the alignment analysis due to the autoimmune nature of the disease (**publication III**), whereas more focus was put on oral bacterial proteomes in the Corogene cohort study due to patients having a clinical background of periodontitis (**publication II**). Public dataset and database selection should be done mindfully, however, as they can be biased towards certain populations. Such biases have been observed when assessing the risk of hypertrophic cardiomyopathy in patients of African ancestry (Manrai et al. 2016). Furthermore, it should be made aware that having a strong hypothesis may lead to being blinded by the expected outcome even in the face of data revealing something else (discussed in Yanai and Lercher 2020). However, independent validation methods on other study cohorts/paradigms should overcome these biases.

Taken together, the studies undertaken have exemplified that MVA is a powerful method that could provide beneficial insights into the humoral immune system response in the context of exposome research. As biological sciences evolve towards methods yielding more comprehensive datasets at higher individual resolution, MVA is just another means towards future research.

7 Conclusion

Here we presented the beneficial use of metagenomic sequencing and MVA in exposome research to characterize the host-microbiome-environment relationship in more detail. We investigated the exposome interactions in the background of dietary intervention and physical activity, as well as in the background of chronic and autoimmune disease. The principal findings of the study are:

- Although antibody response profiles show great heterogeneity across individual patients, there exists common shared characteristics that can be further correlated with lifestyle or disease background.
- Different study design approaches, e.g. time-series, single case *versus* community, classical case *versus* control cohorts, result in both unique challenges in data analysis, but also opportunities for discovery. During MVA data analysis, feature selection can be performed at different analysis stages to gather a broad range of insights about the underlying study cohort.
- Metagenomic sequencing yields timeseries gut microbiota profiles which show significant compositional changes upon dietary fibre supplementation.
- Gut microbiota profiles reveal bacterial community characteristics associated with an athletic lifestyle and MVA reveals antibody response profiles associated with diabetes or a history of smoking. Patients with periodontitis have a higher antibody immune response to C-terminal epitope on EBV protein VP26.
- Patients with CVD have a different antibody response profile according to their periodontal background. A two-step classification of patients based on MVA-delineated biomarkers allows to differentiate patients with ACS from healthy controls. MVA-generated antibody response profiles allow to classify patients with NT1 from controls and identify differentially targeted viral and self epitopes, providing value for biomarker development and further research into the mechanism of disease onset and progression.
- Next generation phage display technology MVA is a useful tool for characterizing the comprehensive antibody profiles of individual patients in a high-throughput manner. MVA is a beneficial tool for exposome research as it can be used with many different sampling materials, can be used in group- or individual-based timeseries studies, and allows for different data analysis approaches for feature selection and refinement.

References

Abeles, Shira R., Refugio Robles-Sikisaka, Melissa Ly, Andrew G. Lum, Julia Salzman, Tobias K. Boehm, and David T. Pride. 2014. "Human Oral Viruses Are Personal, Persistent and Gender-Consistent." *The ISME Journal* 8 (9): 1753–67. https://doi.org/10.1038/ismej.2014.31.

Acharya, Aneesha, Yuki Chan, Supriya Kheur, Li Jian Jin, Rory M. Watt, and Nikos Mattheos. 2017. "Salivary Microbiome in Non-Oral Disease: A Summary of Evidence and Commentary." *Archives of Oral Biology* 83 (November): 169–73. https://doi.org/10.1016/j.archoralbio.2017.07.019.

Afik, Shaked, Kathleen B. Yates, Kevin Bi, Samuel Darko, Jernej Godec, Ulrike Gerdemann, Leo Swadling, et al. 2017. "Targeted Reconstruction of T Cell Receptor Sequence from Single Cell RNA-Seq Links CDR3 Length to T Cell Differentiation State." *Nucleic Acids Research* 45 (16): e148. https://doi.org/10.1093/nar/gkx615.

Agache, Ioana, Rachel Miller, James E. Gern, Peter W. Hellings, Marek Jutel, Antonella Muraro, Wanda Phipatanakul, Santiago Quirce, and David Peden. 2019. "Emerging Concepts and Challenges in Implementing the Exposome Paradigm in Allergic Diseases and Asthma: A Practall Document." *Allergy* 74 (3): 449–63. https://doi.org/10.1111/all.13690.

Aggarwala, Varun, Guanxiang Liang, and Frederic D. Bushman. 2017. "Viral Communities of the Human Gut: Metagenomic Analysis of Composition and Dynamics." *Mobile DNA* 8: 12. https://doi.org/10.1186/s13100-017-0095-y.

Al-Ghamdi, Abdullah. 2012. "Role of Herpes Simplex Virus-1, Cytomegalovirus and Epstein-Barr Virus in Atherosclerosis." *Pakistan Journal of Pharmaceutical Sciences* 25 (1): 89–97.

Allen, Jacob M., Lucy J. Mailing, Grace M. Niemiro, Rachel Moore, Marc D. Cook, Bryan A. White, Hannah D. Holscher, and Jeffrey A. Woods. 2018. "Exercise Alters Gut Microbiota Composition and Function in Lean and Obese Humans." *Medicine and Science in Sports and Exercise* 50 (4): 747–57. https://doi.org/10.1249/MSS.0000000000001495.

Arthur, Janelle C., Ernesto Perez-Chanona, Marcus Mühlbauer, Sarah Tomkovich, Joshua M. Uronis, Ting-Jia Fan, Barry J. Campbell, et al. 2012. "Intestinal Inflammation Targets Cancer-Inducing Activity of the Microbiota." *Science (New York, N.Y.)* 338 (6103): 120–23. https://doi.org/10.1126/science.1224820.

Atarashi, Koji, Wataru Suda, Chengwei Luo, Takaaki Kawaguchi, Iori Motoo, Seiko Narushima, Yuya Kiguchi, et al. 2017. "Ectopic Colonization of Oral Bacteria in the Intestine Drives TH1 Cell Induction and Inflammation." *Science (New York, N.Y.)* 358 (6361): 359–65. https://doi.org/10.1126/science.aan4526.

Bäckhed, Fredrik, Ruth E. Ley, Justin L. Sonnenburg, Daniel A. Peterson, and Jeffrey I. Gordon. 2005. "Host-Bacterial Mutualism in the Human Intestine." *Science (New York, N.Y.)* 307 (5717): 1915–20. https://doi.org/10.1126/science.1104816.

Bagaitkar, Juhi, Donald R. Demuth, Carlo Amorin Daep, Diane E. Renaud, Deanne L. Pierce, and David A. Scott. 2010. "Tobacco Upregulates P. Gingivalis Fimbrial Proteins Which Induce TLR2 Hyposensitivity." Edited by Adam J. Ratner. *PLoS ONE* 5 (5): e9323. https://doi.org/10.1371/journal.pone.0009323.

Bassetti, Claudio L. A., Antoine Adamantidis, Denis Burdakov, Fang Han, Steffen Gay, Ulf Kallweit, Ramin Khatami, et al. 2019. “Narcolepsy — Clinical Spectrum, Aetiopathophysiology, Diagnosis and Treatment.” *Nature Reviews Neurology* 15 (9): 519–39. https://doi.org/10.1038/s41582-019-0226-9.

Baunwall, Simon Mark Dahl, Mads Ming Lee, Marcel Kjærsgaard Eriksen, Benjamin H. Mullish, Julian R. Marchesi, Jens Frederik Dahlerup, and Christian Lodberg Hvas. 2020. “Faecal Microbiota Transplantation for Recurrent Clostridioides Difficile Infection: An Updated Systematic Review and Meta-Analysis.” *eClinicalMedicine* 29 (December). https://doi.org/10.1016/j.eclinm.2020.100642.

Belenguer, Alvaro, Sylvia H. Duncan, A. Graham Calder, Grietje Holtrop, Petra Louis, Gerald E. Lobley, and Harry J. Flint. 2006. “Two Routes of Metabolic Cross-Feeding between Bifidobacterium Adolescentis and Butyrate-Producing Anaerobes from the Human Gut.” *Applied and Environmental Microbiology* 72 (5): 3593–99. https://doi.org/10.1128/AEM.72.5.3593-3599.2006.

Belstrøm, Daniel, Palle Holmstrup, Claus H. Nielsen, Nikolai Kirkby, Svante Twetman, Berit L. Heitmann, Vanja Klepac-Ceraj, Bruce J. Paster, and Nils-Erik Fiehn. 2014. “Bacterial Profiles of Saliva in Relation to Diet, Lifestyle Factors, and Socioeconomic Status.” *Journal of Oral Microbiology* 6 (1): 23609. https://doi.org/10.3402/jom.v6.23609.

Bendinelli, M., M. Pistello, F. Maggi, C. Fornai, G. Freer, and M. L. Vatteroni. 2001. “Molecular Properties, Biology, and Clinical Implications of TT Virus, a Recently Identified Widespread Infectious Agent of Humans.” *Clinical Microbiology Reviews* 14 (1): 98–113. https://doi.org/10.1128/CMR.14.1.98-113.2001.

Béziat, Vivien, Franck Rapaport, Jiafen Hu, Matthias Titeux, Mathilde Bonnet Des Claustres, Mathieu Bourgey, Heather Griffin, et al. 2021. “Humans with Inherited T Cell CD28 Deficiency Are Susceptible to Skin Papillomaviruses but Are Otherwise Healthy.” *Cell* 184 (14): 3812-3828.e30. https://doi.org/10.1016/j.cell.2021.06.004.

Biazzo, Manuele, and Gabriele Deidda. 2022. “Fecal Microbiota Transplantation as New Therapeutic Avenue for Human Diseases.” *Journal of Clinical Medicine* 11 (14): 4119. https://doi.org/10.3390/jcm11144119.

Boisseau, Nathalie, Nicolas Barnich, and Christelle Koechlin-Ramonatxo. 2022. “The Nutrition-Microbiota-Physical Activity Triad: An Inspiring New Concept for Health and Sports Performance.” *Nutrients* 14 (5): 924. https://doi.org/10.3390/nu14050924.

Bolotin, Dmitriy A., Stanislav Poslavsky, Alexey N. Davydov, Felix E. Frenkel, Lorenzo Fanchi, Olga I. Zolotareva, Saskia Hemmers, et al. 2017. “Antigen Receptor Repertoire Profiling from RNA-Seq Data.” *Nature Biotechnology* 35 (10): 908–11. https://doi.org/10.1038/nbt.3979.

Bonder, Marc Jan, Alexander Kurilshikov, Ettje F. Tigchelaar, Zlatan Mujagic, Floris Imhann, Arnau Vich Vila, Patrick Deelen, et al. 2016. “The Effect of Host Genetics on the Gut Microbiome.” *Nature Genetics* 48 (11): 1407–12. https://doi.org/10.1038/ng.3663.

Bostanci, Nagihan, and Georgios N. Belibasakis. 2012. “Porphyromonas Gingivalis: An Invasive and Evasive Opportunistic Oral Pathogen.” *FEMS Microbiology Letters* 333 (1): 1–9. https://doi.org/10.1111/j.1574-6968.2012.02579.x.

Bradshaw, D. J., P. D. Marsh, G. K. Watson, and C. Allison. 1998. “Role of Fusobacterium Nucleatum and Coaggregation in Anaerobe Survival in Planktonic and Biofilm Oral Microbial Communities during Aeration.” *Infection and Immunity* 66 (10): 4729–32. https://doi.org/10.1128/IAI.66.10.4729-4732.1998.

Breitbart, Mya, Matthew Haynes, Scott Kelley, Florent Angly, Robert A. Edwards, Ben Felts, Joseph M. Mahaffy, et al. 2008. “Viral Diversity and Dynamics in an Infant Gut.” *Research in Microbiology* 159 (5): 367–73. https://doi.org/10.1016/j.resmic.2008.04.006.

Britton, Graham J., Eduardo J. Contijoch, Ilaria Mogno, Olivia H. Vennaro, Sean R. Llewellyn, Ruby Ng, Zhihua Li, et al. 2019. “Microbiotas from Humans with Inflammatory Bowel Disease Alter the Balance of Gut Th17 and RORγt+ Regulatory T Cells and Exacerbate Colitis in Mice.” *Immunity* 50 (1): 212-224.e4. https://doi.org/10.1016/j.immuni.2018.12.015.

Brodin, Petter, and Mark M. Davis. 2017. “Human Immune System Variation.” *Nature Reviews. Immunology* 17 (1): 21–29. https://doi.org/10.1038/nri.2016.125.

Brusic, Vladimir, Raphael Gottardo, Steven H. Kleinstein, Mark M. Davis, and HIPC steering committee. 2014. “Computational Resources for High-Dimensional Immune Analysis from the Human Immunology Project Consortium.” *Nature Biotechnology* 32 (2): 146–48. https://doi.org/10.1038/nbt.2777.

Bu, Wei, Gregory M. Hayes, Hui Liu, Lorraine Gemmell, David O. Schmeling, Pierce Radecki, Fiona Aguilar, et al. 2016. “Kinetics of Epstein-Barr Virus (EBV) Neutralizing and Virus-Specific Antibodies after Primary Infection with EBV.” *Clinical and Vaccine Immunology* 23 (4): 363–69. https://doi.org/10.1128/CVI.00674-15.

Budden, Kurtis F., Shaan L. Gellatly, David L. A. Wood, Matthew A. Cooper, Mark Morrison, Philip Hugenholtz, and Philip M. Hansbro. 2017. “Emerging Pathogenic Links between Microbiota and the Gut–Lung Axis.” *Nature Reviews Microbiology* 15 (1): 55–63. https://doi.org/10.1038/nrmicro.2016.142.

Buffie, Charlie G., Irene Jarchum, Michele Equinda, Lauren Lipuma, Asia Gobourne, Agnes Viale, Carles Ubeda, Joao Xavier, and Eric G. Pamer. 2012. “Profound Alterations of Intestinal Microbiota Following a Single Dose of Clindamycin Results in Sustained Susceptibility to Clostridium Difficile-Induced Colitis.” *Infection and Immunity* 80 (1): 62–73. https://doi.org/10.1128/IAI.05496-11.

Calero, C.D. Quero, E. Ortega Rincón, and P. Manonelles Marqueta. 2020. “Probiotics, Prebiotics and Synbiotics: Useful for Athletes and Active Individuals? A Systematic Review.” *Beneficial Microbes* 11 (2): 135–49. https://doi.org/10.3920/BM2019.0076.

Cani, Patrice D., Jacques Amar, Miguel Angel Iglesias, Marjorie Poggi, Claude Knauf, Delphine Bastelica, Audrey M. Neyrinck, et al. 2007. “Metabolic Endotoxemia Initiates Obesity and Insulin Resistance.” *Diabetes* 56 (7): 1761–72. https://doi.org/10.2337/db06-1491.

Cani, Patrice D., Rodrigo Bibiloni, Claude Knauf, Aurélie Waget, Audrey M. Neyrinck, Nathalie M. Delzenne, and Rémy Burcelin. 2008. “Changes in Gut Microbiota Control Metabolic Endotoxemia-Induced Inflammation in High-Fat Diet-Induced Obesity and Diabetes in Mice.” *Diabetes* 57 (6): 1470–81. https://doi.org/10.2337/db07-1403.

Canzar, Stefan, Karlynn E Neu, Qingming Tang, Patrick C Wilson, and Aly A Khan. 2017. "BASIC: BCR Assembly from Single Cells." *Bioinformatics* 33 (3): 425–27. https://doi.org/10.1093/bioinformatics/btw631.

CARDIoGRAMplusC4D Consortium, Panos Deloukas, Stavroula Kanoni, Christina Willenborg, Martin Farrall, Themistocles L. Assimes, John R. Thompson, et al. 2013. "Large-Scale Association Analysis Identifies New Risk Loci for Coronary Artery Disease." *Nature Genetics* 45 (1): 25–33. https://doi.org/10.1038/ng.2480.

Castellarin, Mauro, René L. Warren, J. Douglas Freeman, Lisa Dreolini, Martin Krzywinski, Jaclyn Strauss, Rebecca Barnes, et al. 2012. "Fusobacterium Nucleatum Infection Is Prevalent in Human Colorectal Carcinoma." *Genome Research* 22 (2): 299–306. https://doi.org/10.1101/gr.126516.111.

Cebriá-Mendoza, María, María A. Bracho, Cristina Arbona, Luís Larrea, Wladimiro Díaz, Rafael Sanjuán, and José M. Cuevas. 2021. "Exploring the Diversity of the Human Blood Virome." *Viruses* 13 (11): 2322. https://doi.org/10.3390/v13112322.

Chapela, María-José, Alejandro Garrido-Maestu, and Ana G. Cabado. 2015. "Detection of Foodborne Pathogens by qPCR: A Practical Approach for Food Industry Applications." Edited by Fatih Yildiz. *Cogent Food & Agriculture* 1 (1): 1013771. https://doi.org/10.1080/23311932.2015.1013771.

Chapple, Iain L. C., Robert Genco, and Working group 2 of joint EFP/AAP workshop. 2013. "Diabetes and Periodontal Diseases: Consensus Report of the Joint EFP/AAP Workshop on Periodontitis and Systemic Diseases." *Journal of Clinical Periodontology* 40 Suppl 14 (April): S106-112. https://doi.org/10.1111/jcpe.12077.

Chassaing, Benoit, Shreya M. Raja, James D. Lewis, Shanthi Srinivasan, and Andrew T. Gewirtz. 2017. "Colonic Microbiota Encroachment Correlates With Dysglycemia in Humans." *Cellular and Molecular Gastroenterology and Hepatology* 4 (2): 205–21. https://doi.org/10.1016/j.jcmgh.2017.04.001.

Chassaing, Benoit, Tom Van de Wiele, Jana De Bodt, Massimo Marzorati, and Andrew T. Gewirtz. 2017. "Dietary Emulsifiers Directly Alter Human Microbiota Composition and Gene Expression Ex Vivo Potentiating Intestinal Inflammation." *Gut* 66 (8): 1414–27. https://doi.org/10.1136/gutjnl-2016-313099.

Chen, Xingjian, Lingjing Liu, Weitong Zhang, Jianyi Yang, and Ka-Chun Wong. 2021. "Human Host Status Inference from Temporal Microbiome Changes via Recurrent Neural Networks." *Briefings in Bioinformatics* 22 (6): bbab223. https://doi.org/10.1093/bib/bbab223.

Chen, Yi-Ming, Li Wei, Yen-Shuo Chiu, Yi-Ju Hsu, Tsung-Yu Tsai, Ming-Fu Wang, and Chi-Chang Huang. 2016. "Lactobacillus Plantarum TWK10 Supplementation Improves Exercise Performance and Increases Muscle Mass in Mice." *Nutrients* 8 (4): 205. https://doi.org/10.3390/nu8040205.

Chen, Zhangling, Djawad Radjabzadeh, Lianmin Chen, Alexander Kurilshikov, Maryam Kavousi, Fariba Ahmadizar, M. Arfan Ikram, et al. 2021. "Association of Insulin Resistance and Type 2 Diabetes With Gut Microbial Diversity: A Microbiome-Wide Analysis From Population Studies." *JAMA Network Open* 4 (7): e2118811. https://doi.org/10.1001/jamanetworkopen.2021.18811.

Cho, Ilseung, Shingo Yamanishi, Laura Cox, Barbara A. Methé, Jiri Zavadil, Kelvin Li, Zhan Gao, et al. 2012. "Antibiotics in Early Life Alter the Murine Colonic Microbiome and Adiposity." *Nature* 488 (7413): 621–26. https://doi.org/10.1038/nature11400.

Clarke, James, Hai-Chen Wu, Lakmal Jayasinghe, Alpesh Patel, Stuart Reid, and Hagan Bayley. 2009. “Continuous Base Identification for Single-Molecule Nanopore DNA Sequencing.” *Nature Nanotechnology* 4 (4): 265–70. https://doi.org/10.1038/nnano.2009.12.

Conticini, Edoardo, Bruno Frediani, and Dario Caro. 2020. “Can Atmospheric Pollution Be Considered a Co-Factor in Extremely High Level of SARS-CoV-2 Lethality in Northern Italy?” *Environmental Pollution (Barking, Essex: 1987)* 261 (June): 114465. https://doi.org/10.1016/j.envpol.2020.114465.

Cronin, Peter, Susan A. Joyce, Paul W. O’Toole, and Eibhlís M. O’Connor. 2021. “Dietary Fibre Modulates the Gut Microbiota.” *Nutrients* 13 (5): 1655. https://doi.org/10.3390/nu13051655.

Cuevas-Ramos, Gabriel, Claude R. Petit, Ingrid Marcq, Michèle Boury, Eric Oswald, and Jean-Philippe Nougayrède. 2010. “Escherichia Coli Induces DNA Damage in Vivo and Triggers Genomic Instability in Mammalian Cells.” *Proceedings of the National Academy of Sciences of the United States of America* 107 (25): 11537–42. https://doi.org/10.1073/pnas.1001261107.

Das, Bhabatosh, and G. Balakrish Nair. 2019. “Homeostasis and Dysbiosis of the Gut Microbiome in Health and Disease.” *Journal of Biosciences* 44 (5): 117.

Davis, Hannah E., Lisa McCorkell, Julia Moore Vogel, and Eric J. Topol. 2023. “Long COVID: Major Findings, Mechanisms and Recommendations.” *Nature Reviews Microbiology* 21 (3): 133–46. https://doi.org/10.1038/s41579-022-00846-2.

De Filippis, Francesca, Manolo Laiola, Giuseppe Blaiotta, and Danilo Ercolini. 2017. “Different Amplicon Targets for Sequencing-Based Studies of Fungal Diversity.” *Applied and Environmental Microbiology* 83 (17): e00905-17. https://doi.org/10.1128/AEM.00905-17.

De Jager, Philip L., Nir Hacohen, Diane Mathis, Aviv Regev, Barbara E. Stranger, and Christophe Benoist. 2015. “ImmVar Project: Insights and Design Considerations for Future Studies of ‘Healthy’ Immune Variation.” *Seminars in Immunology* 27 (1): 51–57. https://doi.org/10.1016/j.smim.2015.03.003.

DeBord, D. Gayle, Tania Carreón, Thomas J. Lentz, Paul J. Middendorf, Mark D. Hoover, and Paul A. Schulte. 2016. “Use of the ‘Exposome’ in the Practice of Epidemiology: A Primer on -Omic Technologies.” *American Journal of Epidemiology* 184 (4): 302–14. https://doi.org/10.1093/aje/kwv325.

Denou, Emmanuel, Katarina Marcinko, Michael G. Surette, Gregory R. Steinberg, and Jonathan D. Schertzer. 2016. “High-Intensity Exercise Training Increases the Diversity and Metabolic Capacity of the Mouse Distal Gut Microbiota during Diet-Induced Obesity.” *American Journal of Physiology. Endocrinology and Metabolism* 310 (11): E982-993. https://doi.org/10.1152/ajpendo.00537.2015.

Dethlefsen, Les, Paul B. Eckburg, Elisabeth M. Bik, and David A. Relman. 2006. “Assembly of the Human Intestinal Microbiota.” *Trends in Ecology & Evolution* 21 (9): 517–23. https://doi.org/10.1016/j.tree.2006.06.013.

Dewhirst, Floyd E., Tuste Chen, Jacques Izard, Bruce J. Paster, Anne C. R. Tanner, Wen-Han Yu, Abirami Lakshmanan, and William G. Wade. 2010. “The Human Oral Microbiome.” *Journal of Bacteriology* 192 (19): 5002–17. https://doi.org/10.1128/JB.00542-10.

Ding, Enmin, Yu Wang, Juan Liu, Song Tang, and Xiaoming Shi. 2022. “A Review on the Application of the Exposome Paradigm to Unveil the Environmental Determinants of Age-Related Diseases.” *Human Genomics* 16 (1): 54. https://doi.org/10.1186/s40246-022-00428-6.

Domingo, José L., and Joaquim Rovira. 2020. “Effects of Air Pollutants on the Transmission and Severity of Respiratory Viral Infections.” *Environmental Research* 187 (August): 109650. https://doi.org/10.1016/j.envres.2020.109650.

Duncan, Sylvia H., Petra Louis, and Harry J. Flint. 2004. “Lactate-Utilizing Bacteria, Isolated from Human Feces, That Produce Butyrate as a Major Fermentation Product.” *Applied and Environmental Microbiology* 70 (10): 5810–17. https://doi.org/10.1128/AEM.70.10.5810-5817.2004.

Durk, Ryan P., Esperanza Castillo, Leticia Márquez-Magaña, Gregory J. Grosicki, Nicole D. Bolter, C. Matthew Lee, and James R. Bagley. 2019. “Gut Microbiota Composition Is Related to Cardiorespiratory Fitness in Healthy Young Adults.” *International Journal of Sport Nutrition and Exercise Metabolism* 29 (3): 249–53. https://doi.org/10.1123/ijsnem.2018-0024.

Eckburg, Paul B., Elisabeth M. Bik, Charles N. Bernstein, Elizabeth Purdom, Les Dethlefsen, Michael Sargent, Steven R. Gill, Karen E. Nelson, and David A. Relman. 2005. “Diversity of the Human Intestinal Microbial Flora.” *Science (New York, N.Y.)* 308 (5728): 1635–38. https://doi.org/10.1126/science.1110591.

Eckhardt, Christina M., Haotian Wu, Diddier Prada, Pantel S. Vokonas, David Sparrow, Lifang Hou, Joel Schwartz, and Andrea A. Baccarelli. 2022. “Predicting Risk of Lung Function Impairment and All-Cause Mortality Using a DNA Methylation-Based Classifier of Tobacco Smoke Exposure.” *Respiratory Medicine* 200: 106896. https://doi.org/10.1016/j.rmed.2022.106896.

Eltahla, Auda A., Simone Rizzetto, Mehdi R. Pirozyan, Brigid D. Betz-Stablein, Vanessa Venturi, Katherine Kedzierska, Andrew R. Lloyd, Rowena A. Bull, and Fabio Luciani. 2016. “Linking the T Cell Receptor to the Single Cell Transcriptome in Antigen-Specific Human T Cells.” *Immunology and Cell Biology* 94 (6): 604–11. https://doi.org/10.1038/icb.2016.16.

Estaki, Mehrbod, Jason Pither, Peter Baumeister, Jonathan P. Little, Sandeep K. Gill, Sanjoy Ghosh, Zahra Ahmadi-Vand, Katelyn R. Marsden, and Deanna L. Gibson. 2016. “Cardiorespiratory Fitness as a Predictor of Intestinal Microbial Diversity and Distinct Metagenomic Functions.” *Microbiome* 4 (1): 42. https://doi.org/10.1186/s40168-016-0189-7.

Faith, Jeremiah J., Janaki L. Guruge, Mark Charbonneau, Sathish Subramanian, Henning Seedorf, Andrew L. Goodman, Jose C. Clemente, et al. 2013. “The Long-Term Stability of the Human Gut Microbiota.” *Science (New York, N.Y.)* 341 (6141): 1237439. https://doi.org/10.1126/science.1237439.

Falk, K., A. Linde, D. Johnson, E. Lennette, I. Ernberg, and A. Lundkvist. 1995. “Synthetic Peptides Deduced from the Amino Acid Sequence of Epstein-Barr Virus Nuclear Antigen 6 (EBNA 6): Antigenic Properties, Production of Monoreactive Reagents, and Analysis of Antibody Responses in Man.” *Journal of Medical Virology* 46 (4): 349–57. https://doi.org/10.1002/jmv.1890460410.

Fecho, Karamarie, Anne E. Thessen, Sergio E. Baranzini, Chris Bizon, Jennifer J. Hadlock, Sui Huang, Ryan T. Roper, et al. 2022. “Progress toward a Universal Biomedical Data Translator.” *Clinical and Translational Science* 15 (8): 1838–47. https://doi.org/10.1111/cts.13301.

Fernandes, Melissa A., Sofia G. Verstraete, Tung G. Phan, Xutao Deng, Emily Stekol, Brandon LaMere, Susan V. Lynch, Melvin B. Heyman, and Eric Delwart. 2019. "Enteric Virome and Bacterial Microbiota in Children With Ulcerative Colitis and Crohn Disease." *Journal of Pediatric Gastroenterology and Nutrition* 68 (1): 30–36. https://doi.org/10.1097/MPG.0000000000002140.

Ferrante, Giuliana, Salvatore Fasola, Giovanna Cilluffo, Giorgio Piacentini, Giovanni Viegi, and Stefania La Grutta. 2022. "Addressing Exposome: An Innovative Approach to Environmental Determinants in Pediatric Respiratory Health." *Frontiers in Public Health* 10. https://www.frontiersin.org/articles/10.3389/fpubh.2022.871140.

Filippi, Massimo, Amit Bar-Or, Fredrik Piehl, Paolo Preziosa, Alessandra Solari, Sandra Vukusic, and Maria A. Rocca. 2018. "Multiple Sclerosis." *Nature Reviews Disease Primers* 4 (1): 1–27. https://doi.org/10.1038/s41572-018-0041-4.

Finnegan, Peter, Bruce Adams, Daniel Reyes, Sean Marrache, Payam Shahi, FuNien Tsai, Thomas Vollbrecht, and Micheal Stubbington. 2022. "924 Barcode Enabled Antigen Mapping (BEAM) and Single-Cell Sequencing Enables next-Generation Systems Immunology Analysis of the Post-COVID-19 Immune Landscape." *Journal for ImmunoTherapy of Cancer* 10 (Suppl 2). https://doi.org/10.1136/jitc-2022-SITC2022.0924.

Flint, Harry J. 2004. "Polysaccharide Breakdown by Anaerobic Microorganisms Inhabiting the Mammalian Gut." *Advances in Applied Microbiology* 56: 89–120. https://doi.org/10.1016/S0065-2164(04)56003-3.

Fons, Michel, Ana Gomez, and Tuomo Karjalainen. 2000. "Mechanisms of Colonisation and Colonisation Resistance of the Digestive Tract: Fons." *Microbial Ecology in Health & Disease* 12 (June). https://doi.org/10.3402/mehd.v12i2.8059.

Francis, Michael, Changwei Li, Yitang Sun, Jingqi Zhou, Xiang Li, J. Thomas Brenna, and Kaixiong Ye. 2021. "Genome-Wide Association Study of Fish Oil Supplementation on Lipid Traits in 81,246 Individuals Reveals New Gene-Diet Interaction Loci." *PLoS Genetics* 17 (3): e1009431. https://doi.org/10.1371/journal.pgen.1009431.

Frank, Daniel N., Allison L. St. Amand, Robert A. Feldman, Edgar C. Boedeker, Noam Harpaz, and Norman R. Pace. 2007. "Molecular-Phylogenetic Characterization of Microbial Community Imbalances in Human Inflammatory Bowel Diseases." *Proceedings of the National Academy of Sciences of the United States of America* 104 (34): 13780–85. https://doi.org/10.1073/pnas.0706625104.

Fujinami, Robert S., Matthias G. von Herrath, Urs Christen, and J. Lindsay Whitton. 2006. "Molecular Mimicry, Bystander Activation, or Viral Persistence: Infections and Autoimmune Disease." *Clinical Microbiology Reviews* 19 (1): 80–94. https://doi.org/10.1128/CMR.19.1.80-94.2006.

Gao, Zilong, Juan Lv, and Min Wang. 2017. "Epstein–Barr Virus Is Associated with Periodontal Diseases: A Meta-Analysis Based on 21 Case–Control Studies." *Medicine* 96 (6): e5980. https://doi.org/10.1097/MD.0000000000005980.

Geoghegan, Jemma L., Xiaoyun Ren, Matthew Storey, James Hadfield, Lauren Jelley, Sarah Jefferies, Jill Sherwood, et al. 2020. "Genomic Epidemiology Reveals Transmission Patterns and Dynamics of SARS-CoV-2 in Aotearoa New Zealand." *Nature Communications* 11 (1): 6351. https://doi.org/10.1038/s41467-020-20235-8.

Gibbons, R. J, S. S Socransky, W. C de Araujo, and J van Houte. 1964. “Studies of the Predominant Cultivable Microbiota of Dental Plaque.” *Archives of Oral Biology* 9 (3): 365–70. https://doi.org/10.1016/0003-9969(64)90069-X.

Gill, Steven R., Mihai Pop, Robert T. DeBoy, Paul B. Eckburg, Peter J. Turnbaugh, Buck S. Samuel, Jeffrey I. Gordon, David A. Relman, Claire M. Fraser-Liggett, and Karen E. Nelson. 2006. “Metagenomic Analysis of the Human Distal Gut Microbiome.” *Science (New York, N.Y.)* 312 (5778): 1355–59. https://doi.org/10.1126/science.1124234.

Gregory, Ann C., Olivier Zablocki, Ahmed A. Zayed, Allison Howell, Benjamin Bolduc, and Matthew B. Sullivan. 2020. “The Gut Virome Database Reveals Age-Dependent Patterns of Virome Diversity in the Human Gut.” *Cell Host & Microbe* 28 (5): 724-740.e8. https://doi.org/10.1016/j.chom.2020.08.003.

Griffiths, Paul, and Matthew Reeves. 2021. “Pathogenesis of Human Cytomegalovirus in the Immunocompromised Host.” *Nature Reviews. Microbiology* 19 (12): 759–73. https://doi.org/10.1038/s41579-021-00582-z.

Grossi, S. G., and R. J. Genco. 1998. “Periodontal Disease and Diabetes Mellitus: A Two-Way Relationship.” *Annals of Periodontology* 3 (1): 51–61. https://doi.org/10.1902/annals.1998.3.1.51.

Grundtman, Cecilia, and Georg Wick. 2011. “The Autoimmune Concept of Atherosclerosis.” *Current Opinion in Lipidology* 22 (5): 327–34. https://doi.org/10.1097/MOL.0b013e32834aa0c2.

Hajishengallis, G. 2014. “The Inflammophilic Character of the Periodontitis-Associated Microbiota.” *Molecular Oral Microbiology* 29 (6): 248–57. https://doi.org/10.1111/omi.12065.

Hajishengallis, George. 2014. “Immunomicrobial Pathogenesis of Periodontitis: Keystones, Pathobionts, and Host Response.” *Trends in Immunology* 35 (1): 3–11. https://doi.org/10.1016/j.it.2013.09.001.

Halenius, Anne, and Hartmut Hengel. 2014. “Human Cytomegalovirus and Autoimmune Disease.” *BioMed Research International* 2014: 472978. https://doi.org/10.1155/2014/472978.

Hall, Molly A., Scott M. Dudek, Robert Goodloe, Dana C. Crawford, SARAH A. Pendergrass, Peggy Peissig, Murray Brilliant, Catherine A. McCarty, and Marylyn D. Ritchie. 2014. “Environment-Wide Association Study (EWAS) for Type 2 Diabetes in the Marshfield Personalized Medicine Research Project Biobank.” *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 200–211.

Hallberg, Pär, Hans Smedje, Niclas Eriksson, Hugo Kohnke, Makrina Daniilidou, Inger Öhman, Qun-Ying Yue, et al. 2019. “Pandemrix-Induced Narcolepsy Is Associated with Genes Related to Immunity and Neuronal Survival.” *EBioMedicine* 40 (February): 595–604. https://doi.org/10.1016/j.ebiom.2019.01.041.

Han, Maozhen, Pengshuo Yang, Chaofang Zhong, and Kang Ning. 2018. “The Human Gut Virome in Hypertension.” *Frontiers in Microbiology* 9. https://www.frontiersin.org/articles/10.3389/fmicb.2018.03150.

Han, Y. W., and X. Wang. 2013. “Mobile Microbiome: Oral Bacteria in Extra-Oral Infections and Inflammation.” *Journal of Dental Research* 92 (6): 485–91. https://doi.org/10.1177/0022034513487559.

Harris, Steven A., and Elizabeth A. Harris. 2015. “Herpes Simplex Virus Type 1 and Other Pathogens Are Key Causative Factors in Sporadic Alzheimer’s Disease.” *Journal of Alzheimer’s Disease: JAD* 48 (2): 319–53. https://doi.org/10.3233/JAD-142853.

Hart, Bert A. 't, Rogier Q. Hintzen, and Jon D. Laman. 2009. "Multiple Sclerosis - a Response-to-Damage Model." *Trends in Molecular Medicine* 15 (6): 235–44. https://doi.org/10.1016/j.molmed.2009.04.001.

Haubek, Dorte, Oum-Keltoum Ennibi, Knud Poulsen, Michael Vaeth, Sven Poulsen, and Mogens Kilian. 2008. "Risk of Aggressive Periodontitis in Adolescent Carriers of the JP2 Clone of Aggregatibacter (Actinobacillus) Actinomycetemcomitans in Morocco: A Prospective Longitudinal Cohort Study." *Lancet (London, England)* 371 (9608): 237–42. https://doi.org/10.1016/S0140-6736(08)60135-X.

Hawley, John A. 2020. "Microbiota and Muscle Highway - Two Way Traffic." *Nature Reviews. Endocrinology* 16 (2): 71–72. https://doi.org/10.1038/s41574-019-0291-6.

Henle, W, G Henle, J Andersson, I Ernberg, G Klein, C A Horwitz, G Marklund, L Rymo, C Wellinder, and S E Straus. 1987. "Antibody Responses to Epstein-Barr Virus-Determined Nuclear Antigen (EBNA)-1 and EBNA-2 in Acute and Chronic Epstein-Barr Virus Infection." *Proceedings of the National Academy of Sciences* 84 (2): 570–74. https://doi.org/10.1073/pnas.84.2.570.

Henry, David A, Mark A Jones, Paulina Stehlik, and Paul P Glasziou. 2021. "Effectiveness of COVID-19 Vaccines: Findings from Real World Studies." *The Medical Journal of Australia* 215 (4): 149-151.e1. https://doi.org/10.5694/mja2.51182.

Holt, Stanley C., and Jeffrey L. Ebersole. 2005. "Porphyromonas Gingivalis, Treponema Denticola, and Tannerella Forsythia: The 'Red Complex', a Prototype Polybacterial Pathogenic Consortium in Periodontitis." *Periodontology 2000* 38: 72–122. https://doi.org/10.1111/j.1600-0757.2005.00113.x.

Holtz, Lori R., Song Cao, Guoyan Zhao, Irma K. Bauer, Donna M. Denno, Eileen J. Klein, Martin Antonio, et al. 2014. "Geographic Variation in the Eukaryotic Virome of Human Diarrhea." *Virology* 468–470 (November): 556–64. https://doi.org/10.1016/j.virol.2014.09.012.

"HOMD :: Human Oral Microbiome Database." n.d. Accessed March 12, 2023. https://www.homd.org/.

Hrncir, Tomas. 2022. "Gut Microbiota Dysbiosis: Triggers, Consequences, Diagnostic and Therapeutic Options." *Microorganisms* 10 (3): 578. https://doi.org/10.3390/microorganisms10030578.

Hsu, Yi Ju, Chien Chao Chiu, Yen Peng Li, Wen Ching Huang, Yen Te Huang, Chi Chang Huang, and Hsiao Li Chuang. 2015. "Effect of Intestinal Microbiota on Exercise Performance in Mice." *Journal of Strength and Conditioning Research* 29 (2): 552–58. https://doi.org/10.1519/JSC.0000000000000644.

Huang, Steven K., Priya Tripathi, Lada A. Koneva, Raymond G. Cavalcante, Nathan Craig, Anne M. Scruggs, Maureen A. Sartor, Furong Deng, and Yahong Chen. 2021. "Effect of Concentration and Duration of Particulate Matter Exposure on the Transcriptome and DNA Methylome of Bronchial Epithelial Cells." *Environmental Epigenetics* 7 (1): dvaa022. https://doi.org/10.1093/eep/dvaa022.

Huang, Wen-Ching, Yi-Hsun Chen, Hsiao-Li Chuang, Chien-Chao Chiu, and Chi-Chang Huang. 2019. "Investigation of the Effects of Microbiota on Exercise Physiological Adaption, Performance, and Energy Utilization Using a Gnotobiotic Animal Model." *Frontiers in Microbiology* 10: 1906. https://doi.org/10.3389/fmicb.2019.01906.

Huang, Wen-Ching, Yi-Ju Hsu, HuaShuai Li, Nai-Wen Kan, Yi-Ming Chen, Jin-Seng Lin, Tien-Ken Hsu, Tsung-Yu Tsai, Yen-Shuo Chiu, and Chi-Chang Huang. 2018. "Effect of Lactobacillus Plantarum TWK10 on Improving Endurance Performance in Humans." *The Chinese Journal of Physiology* 61 (3): 163–70. https://doi.org/10.4077/CJP.2018.BAH587.

Ike, Anthony C., Chisom J. Onu, Chukwuebuka M. Ononugbo, Eleazar E. Reward, and Sophia O. Muo. 2020. "Immune Response to Herpes Simplex Virus Infection and Vaccine Development." *Vaccines* 8 (2): 302. https://doi.org/10.3390/vaccines8020302.

"Illumina | Next-Generation Sequencing (NGS)." 2023. March 3, 2023. https://emea.illumina.com/science/technology/next-generation-sequencing.html.

Imai, Kenichi, and Yorimasa Ogata. 2020. "How Does Epstein–Barr Virus Contribute to Chronic Periodontitis?" *International Journal of Molecular Sciences* 21 (6): 1940. https://doi.org/10.3390/ijms21061940.

"Ion Torrent | Thermo Fisher." 2023. March 3, 2023. https://www.thermofisher.com/tr/en/home/brands/ion-torrent.html.

Jaago, Mariliis, Nadežda Pupina, Annika Rähni, Arno Pihlak, Helle Sadam, Nihal Engin Vrana, Juha Sinisalo, Pirkko Pussinen, and Kaia Palm. 2022. "Antibody Response to Oral Biofilm Is a Biomarker for Acute Coronary Syndrome in Periodontal Disease." *Communications Biology* 5 (1): 1–10. https://doi.org/10.1038/s42003-022-03122-4.

Jaago, Mariliis, Annika Rähni, Nadežda Pupina, Arno Pihlak, Helle Sadam, Jürgen Tuvikene, Annela Avarlaid, et al. 2022. "Differential Patterns of Cross-Reactive Antibody Response against SARS-CoV-2 Spike Protein Detected for Chronically Ill and Healthy COVID-19 Naïve Individuals." *Scientific Reports* 12 (1): 16817. https://doi.org/10.1038/s41598-022-20849-6.

Jäger, Ralf, Kevin A. Shields, Ryan P. Lowery, Eduardo O. De Souza, Jeremy M. Partl, Chase Hollmer, Martin Purpura, and Jacob M. Wilson. 2016. "Probiotic Bacillus Coagulans GBI-30, 6086 Reduces Exercise-Induced Muscle Damage and Increases Recovery." *PeerJ* 4: e2276. https://doi.org/10.7717/peerj.2276.

Jaiswal, Shubham K., Shitij Manojkumar Agarwal, Parikshit Thodum, and Vineet K. Sharma. 2021. "SkinBug: An Artificial Intelligence Approach to Predict Human Skin Microbiome-Mediated Metabolism of Biotics and Xenobiotics." *iScience* 24 (1): 101925. https://doi.org/10.1016/j.isci.2020.101925.

Janda, J. Michael, and Sharon L. Abbott. 2007. "16S rRNA Gene Sequencing for Bacterial Identification in the Diagnostic Laboratory: Pluses, Perils, and Pitfalls." *Journal of Clinical Microbiology* 45 (9): 2761–64. https://doi.org/10.1128/JCM.01228-07.

Jasinski-Bergner, Simon, Ofer Mandelboim, and Barbara Seliger. 2020. "Molecular Mechanisms of Human Herpes Viruses Inferring with Host Immune Surveillance." *Journal for ImmunoTherapy of Cancer* 8 (2): e000841. https://doi.org/10.1136/jitc-2020-000841.

Jia, Peng, Weihua Dong, Shujuan Yang, Zhicheng Zhan, La Tu, and Shengjie Lai. 2020. "Spatial Lifecourse Epidemiology and Infectious Disease Research." *Trends in Parasitology* 36 (3): 235–38. https://doi.org/10.1016/j.pt.2019.12.012.

Jiang, Luxi, Hongyu Ren, Haijian Zhou, Tian Qin, and Yu Chen. 2017. "Simultaneous Detection of Nine Key Bacterial Respiratory Pathogens Using Luminex xTAG® Technology." *International Journal of Environmental Research and Public Health* 14 (3): 223. https://doi.org/10.3390/ijerph14030223.

Jie, Zhuye, Huihua Xia, Shi-Long Zhong, Qiang Feng, Shenghui Li, Suisha Liang, Huanzi Zhong, et al. 2017. "The Gut Microbiome in Atherosclerotic Cardiovascular Disease." *Nature Communications* 8 (1): 845. https://doi.org/10.1038/s41467-017-00900-1.

Karl, J. Philip, Lee M. Margolis, Elisabeth H. Madslien, Nancy E. Murphy, John W. Castellani, Yngvar Gundersen, Allison V. Hoke, et al. 2017. "Changes in Intestinal Microbiota Composition and Metabolism Coincide with Increased Intestinal Permeability in Young Adults under Prolonged Physiological Stress." *APSselect* 4 (5): G559–71. https://doi.org/10.1152/ajpgi.00066.2017@apsselect.2017.4.issue-5.

Kaufmann, Max, Hayley Evans, Anna-Lena Schaupp, Jan Broder Engler, Gurman Kaur, Anne Willing, Nina Kursawe, et al. 2021. "Identifying CNS-Colonizing T Cells as Potential Therapeutic Targets to Prevent Progression of Multiple Sclerosis." *Med (New York, N.Y.)* 2 (3): 296-312.e8. https://doi.org/10.1016/j.medj.2021.01.006.

Kawasaki, Eiji. 2014. "Type 1 Diabetes and Autoimmunity." *Clinical Pediatric Endocrinology: Case Reports and Clinical Investigations: Official Journal of the Japanese Society for Pediatric Endocrinology* 23 (4): 99–105. https://doi.org/10.1297/cpe.23.99.

Kaźmierczak-Siedlecka, Karolina, Agnieszka Daca, Mateusz Fic, Thierry van de Wetering, Marcin Folwarski, and Wojciech Makarewicz. 2020. "Therapeutic Methods of Gut Microbiota Modification in Colorectal Cancer Management – Fecal Microbiota Transplantation, Prebiotics, Probiotics, and Synbiotics." *Gut Microbes* 11 (6): 1518–30. https://doi.org/10.1080/19490976.2020.1764309.

Keum, Do Hee, Su-Kyoung Kim, Jahyun Koo, Geon-Hui Lee, Cheonhoo Jeon, Jee Won Mok, Beom Ho Mun, et al. 2020. "Wireless Smart Contact Lens for Diabetic Diagnosis and Therapy." *Science Advances* 6 (17): eaba3252. https://doi.org/10.1126/sciadv.aba3252.

Kim, Kyunghun, Ho Joong Kim, Haozhe Zhang, Woohyun Park, Dawn Meyer, Min Ku Kim, Bongjoong Kim, et al. 2021. "All-Printed Stretchable Corneal Sensor on Soft Contact Lenses for Noninvasive and Painless Ocular Electrodiagnosis." *Nature Communications* 12 (1): 1544. https://doi.org/10.1038/s41467-021-21916-8.

Kim, Tae Yeon, Sangbaie Shin, Hyunsik Choi, Sang Hoon Jeong, David Myung, and Sei Kwang Hahn. 2021. "Smart Contact Lenses with a Transparent Silver Nanowire Strain Sensor for Continuous Intraocular Pressure Monitoring." *ACS Applied Bio Materials* 4 (5): 4532–41. https://doi.org/10.1021/acsabm.1c00267.

Kobayashi, Kumiko, Hiroshi Yoshida, Tomoya Tanjo, and Kento Aida. 2022. "Cloud Service Checklist for Academic Communities and Customization for Genome Medical Research." *Human Genome Variation* 9 (1): 1–7. https://doi.org/10.1038/s41439-022-00214-9.

Kobiyama, Kouji, and Klaus Ley. 2018. "Atherosclerosis." *Circulation Research* 123 (10): 1118–20. https://doi.org/10.1161/CIRCRESAHA.118.313816.

Koenig, Jeremy E., Aymé Spor, Nicholas Scalfone, Ashwana D. Fricker, Jesse Stombaugh, Rob Knight, Largus T. Angenent, and Ruth E. Ley. 2011. "Succession of Microbial Consortia in the Developing Infant Gut Microbiome." *Proceedings of the National Academy of Sciences* 108 (supplement_1): 4578–85. https://doi.org/10.1073/pnas.1000081107.

Koonin, Eugene V, Valerian V Dolja, and Mart Krupovic. 2021. “The Healthy Human Virome: From Virus–Host Symbiosis to Disease.” *Current Opinion in Virology* 47 (April): 86–94. https://doi.org/10.1016/j.coviro.2021.02.002.

Kott, Katharine A., Stephen T. Vernon, Thomas Hansen, Macha de Dreu, Souvik K. Das, Joseph Powell, Barbara Fazekas de St Groth, Belinda A. Di Bartolo, Helen M. McGuire, and Gemma A. Figtree. 2020. “Single-Cell Immune Profiling in Coronary Artery Disease: The Role of State-of-the-Art Immunophenotyping With Mass Cytometry in the Diagnosis of Atherosclerosis.” *Journal of the American Heart Association* 9 (24): e017759. https://doi.org/10.1161/JAHA.120.017759.

Kountouras, Jannis, Emmanuel Gavalas, Christos Zavos, Christos Stergiopoulos, Dimitrios Chatzopoulos, Nikolaos Kapetanakis, and Dimitrios Gisakis. 2007. “Alzheimer’s Disease and Helicobacter Pylori Infection: Defective Immune Regulation and Apoptosis as Proposed Common Links.” *Medical Hypotheses* 68 (2): 378–88. https://doi.org/10.1016/j.mehy.2006.06.052.

Kristensen, David M., Arcady R. Mushegian, Valerian V. Dolja, and Eugene V. Koonin. 2010. “New Dimensions of the Virus World Discovered through Metagenomics.” *Trends in Microbiology* 18 (1): 11–19. https://doi.org/10.1016/j.tim.2009.11.003.

Lafaurie, Gloria Inés, Yineth Neuta, Rafael Ríos, Mauricio Pacheco-Montealegre, Roquelina Pianeta, Diana Marcela Castillo, David Herrera, et al. 2022. “Differences in the Subgingival Microbiome According to Stage of Periodontitis: A Comparison of Two Geographic Regions.” *PLoS ONE* 17 (8): e0273523. https://doi.org/10.1371/journal.pone.0273523.

Lagorio, Susanna, Maria Blettner, Dan Baaken, Maria Feychting, Ken Karipidis, Tom Loney, Nicola Orsini, Martin Röösli, Marilia Silva Paulo, and Mark Elwood. 2021. “The Effect of Exposure to Radiofrequency Fields on Cancer Risk in the General and Working Population: A Protocol for a Systematic Review of Human Observational Studies.” *Environment International* 157 (December): 106828. https://doi.org/10.1016/j.envint.2021.106828.

Lamont, Richard J., Hyun Koo, and George Hajishengallis. 2018. “The Oral Microbiota: Dynamic Communities and Host Interactions.” *Nature Reviews Microbiology* 16 (12): 745–59. https://doi.org/10.1038/s41579-018-0089-x.

Lamprecht, Manfred, Simon Bogner, Gert Schippinger, Kurt Steinbauer, Florian Fankhauser, Seth Hallstroem, Burkhard Schuetz, and Joachim F. Greilberger. 2012. “Probiotic Supplementation Affects Markers of Intestinal Barrier, Oxidation, and Inflammation in Trained Men; a Randomized, Double-Blinded, Placebo-Controlled Trial.” *Journal of the International Society of Sports Nutrition* 9 (1): 45. https://doi.org/10.1186/1550-2783-9-45.

Lamurias, Andre, Sofia Jesus, Vanessa Neveu, Reza M. Salek, and Francisco M. Couto. 2021. “Information Retrieval Using Machine Learning for Biomarker Curation in the Exposome-Explorer.” *Frontiers in Research Metrics and Analytics* 6 (August): 689264. https://doi.org/10.3389/frma.2021.689264.

Lang, Heather L. E., Helle Jacobsen, Shinji Ikemizu, Christina Andersson, Karl Harlos, Lars Madsen, Peter Hjorth, et al. 2002. “A Functional and Structural Basis for TCR Cross-Reactivity in Multiple Sclerosis.” *Nature Immunology* 3 (10): 940–43. https://doi.org/10.1038/ni835.

Larsen, David A., and Krista R. Wigginton. 2020. “Tracking COVID-19 with Wastewater.” *Nature Biotechnology* 38 (10): 1151–53. https://doi.org/10.1038/s41587-020-0690-1.

Lawley, Trevor D, and Alan W Walker. 2013. “Intestinal Colonization Resistance.” *Immunology* 138 (1): 1–11. https://doi.org/10.1111/j.1365-2567.2012.03616.x.

Lax, Simon, Daniel P. Smith, Jarrad Hampton-Marcell, Sarah M. Owens, Kim M. Handley, Nicole M. Scott, Sean M. Gibbons, et al. 2014. “Longitudinal Analysis of Microbial Interaction between Humans and the Indoor Environment.” *Science (New York, N.Y.)* 345 (6200): 1048–52. https://doi.org/10.1126/science.1254529.

Le Goallec, Alan, Braden T. Tierney, Jacob M. Luber, Evan M. Cofer, Aleksandar D. Kostic, and Chirag J. Patel. 2020. “A Systematic Machine Learning and Data Type Comparison Yields Metagenomic Predictors of Infant Age, Sex, Breastfeeding, Antibiotic Usage, Country of Origin, and Delivery Type.” *PLoS Computational Biology* 16 (5): e1007895. https://doi.org/10.1371/journal.pcbi.1007895.

Le, Vuong, Thomas P. Quinn, Truyen Tran, and Svetha Venkatesh. 2020. “Deep in the Bowel: Highly Interpretable Neural Encoder-Decoder Networks Predict Gut Metabolites from Gut Microbiome.” *BMC Genomics* 21 (Suppl 4): 256. https://doi.org/10.1186/s12864-020-6652-7.

Lee, Alison, Hsiao-Hsien Leon Hsu, Yueh-Hsiu Mathilda Chiu, Sonali Bose, Maria José Rosa, Itai Kloog, Ander Wilson, et al. 2018. “Prenatal Fine Particulate Exposure and Early Childhood Asthma: Effect of Maternal Stress and Fetal Gender.” *The Journal of Allergy and Clinical Immunology* 141 (5): 1880–86. https://doi.org/10.1016/j.jaci.2017.07.017.

Lee, Yun Kyung, Parpi Mehrabian, Silva Boyajian, Wei-Li Wu, Jane Selicha, Steven Vonderfecht, and Sarkis K. Mazmanian. 2018. “The Protective Role of Bacteroides Fragilis in a Murine Model of Colitis-Associated Colorectal Cancer.” *mSphere* 3 (6): e00587-18. https://doi.org/10.1128/mSphere.00587-18.

Lennette, E. T., L. Rymo, M. Yadav, G. Masucci, K. Merk, L. Timar, and G. Klein. 1993. “Disease-Related Differences in Antibody Patterns against EBV-Encoded Nuclear Antigens EBNA 1, EBNA 2 and EBNA 6.” *European Journal of Cancer (Oxford, England: 1990)* 29A (11): 1584–89. https://doi.org/10.1016/0959-8049(93)90299-u.

Li, Bo, Taiwen Li, Jean-Christophe Pignon, Binbin Wang, Jinzeng Wang, Sachet A. Shukla, Ruoxu Dou, et al. 2016. “Landscape of Tumor-Infiltrating T Cell Repertoire of Human Cancers.” *Nature Genetics* 48 (7): 725–32. https://doi.org/10.1038/ng.3581.

Li, Bolei, Yang Ge, Lei Cheng, Benhua Zeng, Jinzhao Yu, Xian Peng, Jianhua Zhao, et al. 2019. “Oral Bacteria Colonize and Compete with Gut Microbiota in Gnotobiotic Mice.” *International Journal of Oral Science* 11 (1): 1–9. https://doi.org/10.1038/s41368-018-0043-9.

Li, Yang, Marije Oosting, Sanne P. Smeekens, Martin Jaeger, Raul Aguirre-Gamboa, Kieu T. T. Le, Patrick Deelen, et al. 2016. “A Functional Genomics Approach to Understand Variation in Cytokine Production in Humans.” *Cell* 167 (4): 1099-1110.e14. https://doi.org/10.1016/j.cell.2016.10.017.

Liang, Guanxiang, Maire A. Conrad, Judith R. Kelsen, Lyanna R. Kessler, Jessica Breton, Lindsey G. Albenberg, Sarah Marakos, et al. 2020. "Dynamics of the Stool Virome in Very Early-Onset Inflammatory Bowel Disease." *Journal of Crohn's & Colitis* 14 (11): 1600–1610. https://doi.org/10.1093/ecco-jcc/jjaa094.

Liang, Guanxiang, Chunyu Zhao, Huanjia Zhang, Lisa Mattei, Scott Sherrill-Mix, Kyle Bittinger, Lyanna R. Kessler, et al. 2020. "Step-Wise Assembly of the Neonatal Virome Modulated by Breastfeeding." *Nature* 581 (7809): 470–74. https://doi.org/10.1038/s41586-020-2192-1.

Libertucci, Josie, Usha Dutta, Sandeep Kaur, Jennifer Jury, Laura Rossi, Michelle E. Fontes, M. Sharif Shajib, et al. 2018. "Inflammation-Related Differences in Mucosa-Associated Microbiota and Intestinal Barrier Function in Colonic Crohn's Disease." *American Journal of Physiology-Gastrointestinal and Liver Physiology* 315 (3): G420–31. https://doi.org/10.1152/ajpgi.00411.2017.

Lim, Efrem S., Yanjiao Zhou, Guoyan Zhao, Irma K. Bauer, Lindsay Droit, I. Malick Ndao, Barbara B. Warner, Phillip I. Tarr, David Wang, and Lori R. Holtz. 2015. "Early Life Dynamics of the Human Gut Virome and Bacterial Microbiome in Infants." *Nature Medicine* 21 (10): 1228–34. https://doi.org/10.1038/nm.3950.

Lindeman, Ida, Guy Emerton, Lira Mamanova, Omri Snir, Krzysztof Polanski, Shuo-Wang Qiao, Ludvig M. Sollid, Sarah A. Teichmann, and Michael J. T. Stubbington. 2018. "BraCeR: B-Cell-Receptor Reconstruction and Clonality Inference from Single-Cell RNA-Seq." *Nature Methods* 15 (8): 563–65. https://doi.org/10.1038/s41592-018-0082-3.

Lindesmith, Lisa, Christine Moe, Severine Marionneau, Nathalie Ruvoen, Xi Jiang, Lauren Lindblad, Paul Stewart, Jacques LePendu, and Ralph Baric. 2003. "Human Susceptibility and Resistance to Norwalk Virus Infection." *Nature Medicine* 9 (5): 548–53. https://doi.org/10.1038/nm860.

Liu, Lei, Tao Gong, Wanyin Tao, Bolong Lin, Cong Li, Xuesen Zheng, Shu Zhu, Wei Jiang, and Rongbin Zhou. 2019. "Commensal Viruses Maintain Intestinal Intraepithelial Lymphocytes via Noncanonical RIG-I Signaling." *Nature Immunology* 20 (12): 1681–91. https://doi.org/10.1038/s41590-019-0513-z.

Lockhart, Peter B., Ann F. Bolger, Panos N. Papapanou, Olusegun Osinbowale, Maurizio Trevisan, Matthew E. Levison, Kathryn A. Taubert, et al. 2012. "Periodontal Disease and Atherosclerotic Vascular Disease: Does the Evidence Support an Independent Association?" *Circulation* 125 (20): 2520–44. https://doi.org/10.1161/CIR.0b013e31825719f3.

Loebel, Madlen, Maren Eckey, Franziska Sotzny, Elisabeth Hahn, Sandra Bauer, Patricia Grabowski, Johannes Zerweck, et al. 2017. "Serological Profiling of the EBV Immune Response in Chronic Fatigue Syndrome Using a Peptide Microarray." *PloS One* 12 (6): e0179124. https://doi.org/10.1371/journal.pone.0179124.

Loesche, W. J. 1986. "Role of Streptococcus Mutans in Human Dental Decay." *Microbiological Reviews* 50 (4): 353–80. https://doi.org/10.1128/mr.50.4.353-380.1986.

Loesche, W. J., J. Rowan, L. H. Straffon, and P. J. Loos. 1975. "Association of Streptococcus Mutans with Human Dental Decay." *Infection and Immunity* 11 (6): 1252–60.

Loman, Nicholas J., and Mark J. Pallen. 2015. "Twenty Years of Bacterial Genome Sequencing." *Nature Reviews Microbiology* 13 (12): 787–94. https://doi.org/10.1038/nrmicro3565.

Lozupone, Catherine A., Jesse I. Stombaugh, Jeffrey I. Gordon, Janet K. Jansson, and Rob Knight. 2012. "Diversity, Stability and Resilience of the Human Gut Microbiota." *Nature* 489 (7415): 220–30. https://doi.org/10.1038/nature11550.

Luber, Jacob M., Braden T. Tierney, Evan M. Cofer, Chirag J. Patel, and Aleksandar D. Kostic. 2018. "Aether: Leveraging Linear Programming for Optimal Cloud Computing in Genomics." *Bioinformatics (Oxford, England)* 34 (9): 1565–67. https://doi.org/10.1093/bioinformatics/btx787.

Macgregor, I. D. 1989. "Effects of Smoking on Oral Ecology. A Review of the Literature." *Clinical Preventive Dentistry* 11 (1): 3–7.

Maitre, Léa, Mariona Bustamante, Carles Hernández-Ferrer, Denise Thiel, Chung-Ho E. Lau, Alexandros P. Siskos, Marta Vives-Usano, et al. 2022. "Multi-Omics Signatures of the Human Early Life Exposome." *Nature Communications* 13 (1): 7024. https://doi.org/10.1038/s41467-022-34422-2.

Malla, Muneer Ahmad, Anamika Dubey, Ashwani Kumar, Shweta Yadav, Abeer Hashem, and Elsayed Fathi Abd_Allah. 2019. "Exploring the Human Microbiome: The Potential Future Role of Next-Generation Sequencing in Disease Diagnosis and Treatment." *Frontiers in Immunology* 9. https://www.frontiersin.org/articles/10.3389/fimmu.2018.02868.

Man, Si Ming, Li Zhang, Andrew S. Day, Steven T. Leach, Daniel A. Lemberg, and Hazel Mitchell. 2010. "Campylobacter Concisus and Other Campylobacter Species in Children with Newly Diagnosed Crohn's Disease." *Inflammatory Bowel Diseases* 16 (6): 1008–16. https://doi.org/10.1002/ibd.21157.

Manichanh, C., L. Rigottier-Gois, E. Bonnaud, K. Gloux, E. Pelletier, L. Frangeul, R. Nalin, et al. 2006. "Reduced Diversity of Faecal Microbiota in Crohn's Disease Revealed by a Metagenomic Approach." *Gut* 55 (2): 205–11. https://doi.org/10.1136/gut.2005.073817.

Maqsood, Rabia, Rachel Rodgers, Cynthia Rodriguez, Scott A. Handley, I. Malick Ndao, Phillip I. Tarr, Barbara B. Warner, Efrem S. Lim, and Lori R. Holtz. 2019. "Discordant Transmission of Bacteria and Viruses from Mothers to Babies at Birth." *Microbiome* 7 (1): 156. https://doi.org/10.1186/s40168-019-0766-7.

Marcantonio, Matteo, Annapaola Rizzoli, Markus Metz, Roberto Rosà, Giovanni Marini, Elizabeth Chadwick, and Markus Neteler. 2015. "Identifying the Environmental Conditions Favouring West Nile Virus Outbreaks in Europe." *PLOS ONE* 10 (3): e0121158. https://doi.org/10.1371/journal.pone.0121158.

Margulies, Marcel, Michael Egholm, William E. Altman, Said Attiya, Joel S. Bader, Lisa A. Bemben, Jan Berka, et al. 2005. "Genome Sequencing in Microfabricated High-Density Picolitre Reactors." *Nature* 437 (7057): 376–80. https://doi.org/10.1038/nature03959.

Marsh, Phil D., David A. Head, and Deirdre A. Devine. 2015. "Ecological Approaches to Oral Biofilms: Control without Killing." *Caries Research* 49 Suppl 1: 46–54. https://doi.org/10.1159/000377732.

Marsh, Philip D., Thuy Do, David Beighton, and Deirdre A. Devine. 2016. "Influence of Saliva on the Oral Microbiota." *Periodontology 2000* 70 (1): 80–92. https://doi.org/10.1111/prd.12098.

Marsh, Philip D., David A. Head, and Deirdre A. Devine. 2014. "Prospects of Oral Disease Control in the Future - an Opinion." *Journal of Oral Microbiology* 6: 26176. https://doi.org/10.3402/jom.v6.26176.

Martínez, Inés, Catherine E. Muller, and Jens Walter. 2013. “Long-Term Temporal Analysis of the Human Fecal Microbiota Revealed a Stable Core of Dominant Bacterial Species.” *PLOS ONE* 8 (7): e69621. https://doi.org/10.1371/journal.pone.0069621.

Mazmanian, Sarkis K., Cui Hua Liu, Arthur O. Tzianabos, and Dennis L. Kasper. 2005. “An Immunomodulatory Molecule of Symbiotic Bacteria Directs Maturation of the Host Immune System.” *Cell* 122 (1): 107–18. https://doi.org/10.1016/j.cell.2005.05.007.

McCann, Angela, Feargal J. Ryan, Stephen R. Stockdale, Marion Dalmasso, Tony Blake, C. Anthony Ryan, Catherine Stanton, Susan Mills, Paul R. Ross, and Colin Hill. 2018. “Viromes of One Year Old Infants Reveal the Impact of Birth Mode on Microbiome Diversity.” *PeerJ* 6: e4694. https://doi.org/10.7717/peerj.4694.

McCoy, Lori, Ikuo Tsunoda, and Robert S. Fujinami. 2006. “Multiple Sclerosis and Virus Induced Immune Responses: Autoimmunity Can Be Primed by Molecular Mimicry and Augmented by Bystander Activation.” *Autoimmunity* 39 (1): 9–19. https://doi.org/10.1080/08916930500484799.

McGinnis, Denise P., John S. Brownstein, and Chirag J. Patel. 2016. “Environment-Wide Association Study of Blood Pressure in the National Health and Nutrition Examination Survey (1999–2012).” *Scientific Reports* 6 (1): 30373. https://doi.org/10.1038/srep30373.

Mehta, Supriya D., Dan Zhao, Stefan J. Green, Walter Agingu, Fredrick Otieno, Runa Bhaumik, Dulal Bhaumik, and Robert C. Bailey. 2020. “The Microbiome Composition of a Man’s Penis Predicts Incident Bacterial Vaginosis in His Female Sex Partner With High Accuracy.” *Frontiers in Cellular and Infection Microbiology* 10. https://www.frontiersin.org/articles/10.3389/fcimb.2020.00433.

Merchant, Anwar T., Waranuch Pitiphat, Eric B. Rimm, and Kaumudi Joshipura. 2003. “Increased Physical Activity Decreases Periodontitis Risk in Men.” *European Journal of Epidemiology* 18 (9): 891–98. https://doi.org/10.1023/a:1025622815579.

Mesia, Ruben, Fatemeh Gholami, Hong Huang, Michael Clare-Salzler, Ikramuddin Aukhil, Shannon M. Wallet, and Luciana M. Shaddox. 2016. “Systemic Inflammatory Responses in Patients with Type 2 Diabetes with Chronic Periodontitis.” *BMJ Open Diabetes Research and Care* 4 (1): e000260. https://doi.org/10.1136/bmjdrc-2016-000260.

Mesri, Enrique A., Mark Feitelson, and Karl Munger. 2014. “HUMAN VIRAL ONCOGENESIS: A CANCER HALLMARKS ANALYSIS.” *Cell Host & Microbe* 15 (3): 266–82. https://doi.org/10.1016/j.chom.2014.02.011.

Meyding-Lamadé, Uta, and Cornelia Strank. 2012. “Herpesvirus Infections of the Central Nervous System in Immunocompromised Patients.” *Therapeutic Advances in Neurological Disorders* 5 (5): 279–96. https://doi.org/10.1177/1756285612456234.

Meyle, Joerg, and Iain Chapple. 2015. “Molecular Aspects of the Pathogenesis of Periodontitis.” *Periodontology 2000* 69 (1): 7–17. https://doi.org/10.1111/prd.12104.

Miller, Melissa B., and Yi-Wei Tang. 2009. “Basic Concepts of Microarrays and Potential Applications in Clinical Microbiology.” *Clinical Microbiology Reviews* 22 (4): 611–33. https://doi.org/10.1128/CMR.00019-09.

Minot, Samuel, Rohini Sinha, Jun Chen, Hongzhe Li, Sue A. Keilbaugh, Gary D. Wu, James D. Lewis, and Frederic D. Bushman. 2011. "The Human Gut Virome: Inter-Individual Variation and Dynamic Response to Diet." *Genome Research* 21 (10): 1616–25. https://doi.org/10.1101/gr.122705.111.

Mohr, Alex E., Ralf Jäger, Katie C. Carpenter, Chad M. Kerksick, Martin Purpura, Jeremy R. Townsend, Nicholas P. West, et al. 2020. "The Athletic Gut Microbiota." *Journal of the International Society of Sports Nutrition* 17 (1): 24. https://doi.org/10.1186/s12970-020-00353-w.

Moirano, Giovenale, Antonio Gasparrini, Fiorella Acquaotta, Simona Fratianni, Franco Merletti, Milena Maule, and Lorenzo Richiardi. 2018. "West Nile Virus Infection in Northern Italy: Case-Crossover Study on the Short-Term Effect of Climatic Parameters." *Environmental Research* 167 (November): 544–49. https://doi.org/10.1016/j.envres.2018.08.016.

Momozawa, Yukihide, Julia Dmitrieva, Emilie Théâtre, Valérie Deffontaine, Souad Rahmouni, Benoît Charloteaux, François Crins, et al. 2018. "IBD Risk Loci Are Enriched in Multigenic Regulatory Modules Encompassing Putative Causative Genes." *Nature Communications* 9 (1): 2427. https://doi.org/10.1038/s41467-018-04365-8.

Montrose, David C., Ryohei Nishiguchi, Srijani Basu, Hannah A. Staab, Xi Kathy Zhou, Hanhan Wang, Lingsong Meng, et al. 2021. "Dietary Fructose Alters the Composition, Localization, and Metabolism of Gut Microbiota in Association With Worsening Colitis." *Cellular and Molecular Gastroenterology and Hepatology* 11 (2): 525–50. https://doi.org/10.1016/j.jcmgh.2020.09.008.

Mora, Camilo, Tristan McKenzie, Isabella M. Gaw, Jacqueline M. Dean, Hannah von Hammerstein, Tabatha A. Knudson, Renee O. Setter, et al. 2022. "Over Half of Known Human Pathogenic Diseases Can Be Aggravated by Climate Change." *Nature Climate Change* 12 (9): 869–75. https://doi.org/10.1038/s41558-022-01426-1.

Moreno-Navarrete, José María, and José Manuel Fernandez-Real. 2019. "The Gut Microbiota Modulates Both Browning of White Adipose Tissue and the Activity of Brown Adipose Tissue." *Reviews in Endocrine & Metabolic Disorders* 20 (4): 387–97. https://doi.org/10.1007/s11154-019-09523-x.

Mosca, Alexis, Marion Leclerc, and Jean P. Hugot. 2016. "Gut Microbiota Diversity and Human Diseases: Should We Reintroduce Key Predators in Our Ecosystem?" *Frontiers in Microbiology* 7. https://www.frontiersin.org/articles/10.3389/fmicb.2016.00455.

Mose, Lisle E., Sara R. Selitsky, Lisa M. Bixby, David L. Marron, Michael D. Iglesia, Jonathan S. Serody, Charles M. Perou, Benjamin G. Vincent, and Joel S. Parker. 2016. "Assembly-Based Inference of B-Cell Receptor Repertoires from Short Read RNA Sequencing Data with V'DJer." *Bioinformatics (Oxford, England)* 32 (24): 3729–34. https://doi.org/10.1093/bioinformatics/btw526.

Mu, Andre, William P. Klare, Sarah L. Baines, C. N. Ignatius Pang, Romain Guérillot, Nichaela Harbison-Price, Nadia Keller, et al. 2023. "Integrative Omics Identifies Conserved and Pathogen-Specific Responses of Sepsis-Causing Bacteria." *Nature Communications* 14 (1): 1530. https://doi.org/10.1038/s41467-023-37200-w.

Mushtaq, Hisham Ahmed, Anwar Khedr, Thoyaja Koritala, Brian N. Bartlett, Nitesh K. Jain, and Syed Anjum Khan. 2022. "A Review of Adverse Effects of COVID-19 Vaccines." *Le Infezioni in Medicina* 30 (1): 1–10. https://doi.org/10.53854/liim-3001-1.

Nakatsu, Geicho, Haokui Zhou, William Ka Kei Wu, Sunny Hei Wong, Olabisi Oluwabukola Coker, Zhenwei Dai, Xiangchun Li, et al. 2018. “Alterations in Enteric Virome Are Associated With Colorectal Cancer and Survival Outcomes.” *Gastroenterology* 155 (2): 529-541.e5. https://doi.org/10.1053/j.gastro.2018.04.018.

Nassar, Hamdy, Alpdogan Kantarci, and Thomas E. Van Dyke. 2007. “Diabetic Periodontitis: A Model for Activated Innate Immunity and Impaired Resolution of Inflammation.” *Periodontology 2000* 43: 233–44. https://doi.org/10.1111/j.1600-0757.2006.00168.x.

Navidad, Jose F., David J. Griswold, M. Stephen Gradus, and Sanjib Bhattacharyya. 2013. “Evaluation of Luminex xTAG Gastrointestinal Pathogen Analyte-Specific Reagents for High-Throughput, Simultaneous Detection of Bacteria, Viruses, and Parasites of Clinical and Public Health Importance.” *Journal of Clinical Microbiology* 51 (9): 3018–24. https://doi.org/10.1128/jcm.00896-13.

Nay, Kevin, Maxence Jollet, Benedicte Goustard, Narjes Baati, Barbara Vernus, Maria Pontones, Luz Lefeuvre-Orfila, et al. 2019. “Gut Bacteria Are Critical for Optimal Muscle Function: A Potential Link with Glucose Homeostasis.” *American Journal of Physiology. Endocrinology and Metabolism* 317 (1): E158–71. https://doi.org/10.1152/ajpendo.00521.2018.

Needleman, Ian, Paul Ashley, Peter Fine, Fares Haddad, Mike Loosemore, Akbar de Medici, Nikos Donos, et al. 2015. “Oral Health and Elite Sport Performance.” *British Journal of Sports Medicine* 49 (1): 3–6. https://doi.org/10.1136/bjsports-2014-093804.

Ni, Josephine, Gary D. Wu, Lindsey Albenberg, and Vesselin T. Tomov. 2017. “Gut Microbiota and IBD: Causation or Correlation?” *Nature Reviews. Gastroenterology & Hepatology* 14 (10): 573–84. https://doi.org/10.1038/nrgastro.2017.88.

Ning, Jie, Hairong Du, Yaling Zhang, Qingping Liu, Tao Jiang, Yaxian Pang, Xiaochen Tian, Liqun Yan, Yujie Niu, and Rong Zhang. 2022. “N6-Methyladenosine Modification of CDH1 mRNA Promotes PM2.5-Induced Pulmonary Fibrosis via Mediating Epithelial Mesenchymal Transition.” *Toxicological Sciences: An Official Journal of the Society of Toxicology* 185 (2): 143–57. https://doi.org/10.1093/toxsci/kfab133.

Nociti Jr, Francisco H., Marcio Z. Casati, and Poliana Mendes Duarte. 2015. “Current Perspective of the Impact of Smoking on the Progression and Treatment of Periodontitis.” *Periodontology 2000* 67 (1): 187–210. https://doi.org/10.1111/prd.12063.

Norman, Jason M., Scott A. Handley, Megan T. Baldridge, Lindsay Droit, Catherine Y. Liu, Brian C. Keller, Amal Kambal, et al. 2015. “Disease-Specific Alterations in the Enteric Virome in Inflammatory Bowel Disease.” *Cell* 160 (3): 447–60. https://doi.org/10.1016/j.cell.2015.01.002.

Ogasawara, Osamu. 2022. “Building Cloud Computing Environments for Genome Analysis in Japan.” *Human Genome Variation* 9 (1): 1–8. https://doi.org/10.1038/s41439-022-00223-8.

Okamoto, Takuya, Katsutaro Morino, Satoshi Ugi, Fumiyuki Nakagawa, Mengistu Lemecha, Shogo Ida, Natsuko Ohashi, Daisuke Sato, Yukihiro Fujita, and Hiroshi Maegawa. 2019. “Microbiome Potentiates Endurance Exercise through Intestinal Acetate Production.” *American Journal of Physiology. Endocrinology and Metabolism* 316 (5): E956–66. https://doi.org/10.1152/ajpendo.00510.2018.

Oliveira, Erick Prado de, Roberto Carlos Burini, and Asker Jeukendrup. 2014. "Gastrointestinal Complaints During Exercise: Prevalence, Etiology, and Nutritional Recommendations." *Sports Medicine* 44 (1): 79–85. https://doi.org/10.1007/s40279-014-0153-2.

Ott, S. J., M. Musfeldt, D. F. Wenderoth, J. Hampe, O. Brant, U. R. Fölsch, K. N. Timmis, and S. Schreiber. 2004. "Reduction in Diversity of the Colonic Mucosa Associated Bacterial Microflora in Patients with Active Inflammatory Bowel Disease." *Gut* 53 (5): 685–93. https://doi.org/10.1136/gut.2003.025403.

Oude Munnink, Bas B., and Lia van der Hoek. 2016. "Viruses Causing Gastroenteritis: The Known, The New and Those Beyond." *Viruses* 8 (2): 42. https://doi.org/10.3390/v8020042.

Pablo, Paola de, Iain L. C. Chapple, Christopher D. Buckley, and Thomas Dietrich. 2009. "Periodontitis in Systemic Rheumatic Diseases." *Nature Reviews. Rheumatology* 5 (4): 218–24. https://doi.org/10.1038/nrrheum.2009.28.

Palmer, Chana, Elisabeth M. Bik, Daniel B. DiGiulio, David A. Relman, and Patrick O. Brown. 2007. "Development of the Human Infant Intestinal Microbiota." *PLOS Biology* 5 (7): e177. https://doi.org/10.1371/journal.pbio.0050177.

Pascal, Victoria, Marta Pozuelo, Natalia Borruel, Francesc Casellas, David Campos, Alba Santiago, Xavier Martinez, et al. 2017. "A Microbial Signature for Crohn's Disease." *Gut* 66 (5): 813–22. https://doi.org/10.1136/gutjnl-2016-313235.

Pei, Yonggang, Shuvomoy Banerjee, Hem Chandra Jha, Zhiguo Sun, and Erle S. Robertson. 2017. "An Essential EBV Latent Antigen 3C Binds Bcl6 for Targeted Degradation and Cell Proliferation." *PLOS Pathogens* 13 (7): e1006500. https://doi.org/10.1371/journal.ppat.1006500.

Petersen, Lauren M., Eddy J. Bautista, Hoan Nguyen, Blake M. Hanson, Lei Chen, Sai H. Lek, Erica Sodergren, and George M. Weinstock. 2017. "Community Characteristics of the Gut Microbiomes of Competitive Cyclists." *Microbiome* 5 (1): 98. https://doi.org/10.1186/s40168-017-0320-4.

Peterson, Vanessa M., Kelvin Xi Zhang, Namit Kumar, Jerelyn Wong, Lixia Li, Douglas C. Wilson, Renee Moore, Terrill K. McClanahan, Svetlana Sadekova, and Joel A. Klappenbach. 2017. "Multiplexed Quantification of Proteins and Transcripts in Single Cells." *Nature Biotechnology* 35 (10): 936–39. https://doi.org/10.1038/nbt.3973.

Pleguezuelos-Manzano, Cayetano, Jens Puschhof, Axel Rosendahl Huber, Arne van Hoeck, Henry M. Wood, Jason Nomburg, Carino Gurjao, et al. 2020. "Mutational Signature in Colorectal Cancer Caused by Genotoxic Pks+ E. Coli." *Nature* 580 (7802): 269–73. https://doi.org/10.1038/s41586-020-2080-8.

Plichta, Damian R., Daniel B. Graham, Sathish Subramanian, and Ramnik J. Xavier. 2019. "Therapeutic Opportunities in Inflammatory Bowel Disease: Mechanistic Dissection of Host-Microbiome Relationships." *Cell* 178 (5): 1041–56. https://doi.org/10.1016/j.cell.2019.07.045.

Pollard, Andrew J., and Else M. Bijker. 2021. "A Guide to Vaccinology: From Basic Principles to New Developments." *Nature Reviews Immunology* 21 (2): 83–100. https://doi.org/10.1038/s41577-020-00479-7.

Pool-Zobel, Beatrice, Selvaraju Veeriah, and Frank-D. Böhmer. 2005. “Modulation of Xenobiotic Metabolising Enzymes by Anticarcinogens -- Focus on Glutathione S-Transferases and Their Role as Targets of Dietary Chemoprevention in Colorectal Carcinogenesis.” *Mutation Research* 591 (1–2): 74–92. https://doi.org/10.1016/j.mrfmmm.2005.04.020.

Pope, C. Arden, Nathan Coleman, Zachari A. Pond, and Richard T. Burnett. 2020. “Fine Particulate Air Pollution and Human Mortality: 25+ Years of Cohort Studies.” *Environmental Research* 183 (April): 108924. https://doi.org/10.1016/j.envres.2019.108924.

Preshaw, Philip M., and Susan M. Bissett. 2019. “Periodontitis and Diabetes.” *British Dental Journal* 227 (7): 577–84. https://doi.org/10.1038/s41415-019-0794-5.

Pupina, Nadežda, Annela Avarlaid, Helle Sadam, Arno Pihlak, Mariliis Jaago, Jürgen Tuvikene, Annika Rähni, et al. 2022. “Immune Response to a Conserved Enteroviral Epitope of the Major Capsid VP1 Protein Is Associated with Lower Risk of Cardiovascular Disease.” *eBioMedicine* 76 (February). https://doi.org/10.1016/j.ebiom.2022.103835.

Qin, Junjie, Ruiqiang Li, Jeroen Raes, Manimozhiyan Arumugam, Kristoffer Solvsten Burgdorf, Chaysavanh Manichanh, Trine Nielsen, et al. 2010. “A Human Gut Microbial Gene Catalogue Established by Metagenomic Sequencing.” *Nature* 464 (7285): 59–65. https://doi.org/10.1038/nature08821.

Qin, Junjie, Yingrui Li, Zhiming Cai, Shenghui Li, Jianfeng Zhu, Fan Zhang, Suisha Liang, et al. 2012. “A Metagenome-Wide Association Study of Gut Microbiota in Type 2 Diabetes.” *Nature* 490 (7418): 55–60. https://doi.org/10.1038/nature11450.

Quince, Christopher, Alan W. Walker, Jared T. Simpson, Nicholas J. Loman, and Nicola Segata. 2017. “Shotgun Metagenomics, from Sampling to Analysis.” *Nature Biotechnology* 35 (9): 833–44. https://doi.org/10.1038/nbt.3935.

Rähni, Annika, Mariliis Jaago, Helle Sadam, Nadežda Pupina, Arno Pihlak, Jürgen Tuvikene, Margus Annuk, et al. 2022. “Melanoma-Specific Antigen-Associated Antitumor Antibody Reactivity as an Immune-Related Biomarker for Targeted Immunotherapies.” *Communications Medicine* 2 (1): 1–11. https://doi.org/10.1038/s43856-022-00114-7.

Ramakrishna, Chandran, Maciej Kujawski, Hiutung Chu, Lin Li, Sarkis K. Mazmanian, and Edouard M. Cantin. 2019. “Bacteroides Fragilis Polysaccharide A Induces IL-10 Secreting B and T Cells That Prevent Viral Encephalitis.” *Nature Communications* 10 (1): 2153. https://doi.org/10.1038/s41467-019-09884-6.

Ramanathan, Aravind, Michael Weintraub, Natalie Orlovetskie, Raphael Serruya, Dhivakar Mani, Orly Marcu, Polina Stepensky, et al. 2020. “A Mutation in POLR3E Impairs Antiviral Immune Response and RNA Polymerase III.” *Proceedings of the National Academy of Sciences* 117 (36): 22113–21. https://doi.org/10.1073/pnas.2009947117.

Read, Emily, Michael A. Curtis, and Joana F. Neves. 2021. “The Role of Oral Bacteria in Inflammatory Bowel Disease.” *Nature Reviews Gastroenterology & Hepatology* 18 (10): 731–42. https://doi.org/10.1038/s41575-021-00488-4.

Readhead, Ben, Jean-Vianney Haure-Mirande, Cory C. Funk, Matthew A. Richards, Paul Shannon, Vahram Haroutunian, Mary Sano, et al. 2018. “Multiscale Analysis of Independent Alzheimer’s Cohorts Finds Disruption of Molecular, Genetic, and Clinical Networks by Human Herpesvirus.” *Neuron* 99 (1): 64-82.e7. https://doi.org/10.1016/j.neuron.2018.05.023.

Reed, Brian D., Michael J. Meyer, Valentin Abramzon, Omer Ad, Omer Ad, Pat Adcock, Faisal R. Ahmad, et al. 2022. “Real-Time Dynamic Single-Molecule Protein Sequencing on an Integrated Semiconductor Device.” *Science* 378 (6616): 186–92. https://doi.org/10.1126/science.abo7651.

Reese, Justin T., Deepak Unni, Tiffany J. Callahan, Luca Cappelletti, Vida Ravanmehr, Seth Carbon, Kent A. Shefchek, et al. 2021. “KG-COVID-19: A Framework to Produce Customized Knowledge Graphs for COVID-19 Response.” *Patterns* 2 (1). https://doi.org/10.1016/j.patter.2020.100155.

Regev, Aviv, Sarah A. Teichmann, Eric S. Lander, Ido Amit, Christophe Benoist, Ewan Birney, Bernd Bodenmiller, et al. 2017. “The Human Cell Atlas.” *eLife* 6 (December): e27041. https://doi.org/10.7554/eLife.27041.

Reyes, Alejandro, Matthew Haynes, Nicole Hanson, Florent E. Angly, Andrew C. Heath, Forest Rohwer, and Jeffrey I. Gordon. 2010. “Viruses in the Fecal Microbiota of Monozygotic Twins and Their Mothers.” *Nature* 466 (7304): 334–38. https://doi.org/10.1038/nature09199.

Ringelhan, Marc, Tracy O’Connor, Ulrike Protzer, and Mathias Heikenwalder. 2015. “The Direct and Indirect Roles of HBV in Liver Cancer: Prospective Markers for HCC Screening and Potential Therapeutic Targets.” *The Journal of Pathology* 235 (2): 355–67. https://doi.org/10.1002/path.4434.

Ronsmans, Steven, Karin Sørig Hougaard, Tim S. Nawrot, Michelle Plusquin, François Huaux, María Jesús Cruz, Horatiu Moldovan, et al. 2022. “The EXIMIOUS Project—Mapping Exposure-Induced Immune Effects: Connecting the Exposome and the Immunome.” *Environmental Epidemiology* 6 (1): e193. https://doi.org/10.1097/EE9.0000000000000193.

Rothschild, Daphna, Omer Weissbrod, Elad Barkan, Alexander Kurilshikov, Tal Korem, David Zeevi, Paul I. Costea, et al. 2018. “Environment Dominates over Host Genetics in Shaping Human Gut Microbiota.” *Nature* 555 (7695): 210–15. https://doi.org/10.1038/nature25973.

Roubaud-Baudron, Claire, Pierre Krolak-Salmon, Isabelle Quadrio, Francis Mégraud, and Nathalie Salles. 2012. “Impact of Chronic Helicobacter Pylori Infection on Alzheimer’s Disease: Preliminary Results.” *Neurobiology of Aging* 33 (5): 1009.e11-19. https://doi.org/10.1016/j.neurobiolaging.2011.10.021.

Rutledge, Jarod, Hamilton Oh, and Tony Wyss-Coray. 2022. “Measuring Biological Age Using Omics Data.” *Nature Reviews. Genetics* 23 (12): 715–27. https://doi.org/10.1038/s41576-022-00511-7.

Sadam, Helle, Arno Pihlak, Mariliis Jaago, Nadežda Pupina, Annika Rähni, Maarja Toots, Antti Vaheri, et al. 2021. “Identification of Two Highly Antigenic Epitope Markers Predicting Multiple Sclerosis in Optic Neuritis Patients.” *eBioMedicine* 64 (February). https://doi.org/10.1016/j.ebiom.2021.103211.

Sadam, Helle, Arno Pihlak, Anri Kivil, Susan Pihelgas, Mariliis Jaago, Priit Adler, Jaak Vilo, et al. 2018. “Prostaglandin D2 Receptor DP1 Antibodies Predict Vaccine-Induced and Spontaneous Narcolepsy Type 1: Large-Scale Study of Antibody Profiling.” *eBioMedicine* 29 (March): 47–59. https://doi.org/10.1016/j.ebiom.2018.01.043.

Sammut, Stephen-John, Mireia Crispin-Ortuzar, Suet-Feung Chin, Elena Provenzano, Helen A. Bardwell, Wenxin Ma, Wei Cope, et al. 2022. “Multi-Omic Machine Learning Predictor of Breast Cancer Therapy Response.” *Nature* 601 (7894): 623–29. https://doi.org/10.1038/s41586-021-04278-5.

Scheiman, Jonathan, Jacob M. Luber, Theodore A. Chavkin, Tara MacDonald, Angela Tung, Loc-Duyen Pham, Marsha C. Wibowo, et al. 2019. “Meta-Omics Analysis of Elite Athletes Identifies a Performance-Enhancing Microbe That Functions via Lactate Metabolism.” *Nature Medicine* 25 (7): 1104–9. https://doi.org/10.1038/s41591-019-0485-4.

Schnupf, Pamela, Valérie Gaboriau-Routhiau, and Nadine Cerf-Bensussan. 2013. “Host Interactions with Segmented Filamentous Bacteria: An Unusual Trade-off That Drives the Post-Natal Maturation of the Gut Immune System.” *Seminars in Immunology* 25 (5): 342–51. https://doi.org/10.1016/j.smim.2013.09.001.

Schroeder, Bjoern O., George M. H. Birchenough, Marcus Ståhlman, Liisa Arike, Malin E. V. Johansson, Gunnar C. Hansson, and Fredrik Bäckhed. 2018. “Bifidobacteria or Fiber Protects against Diet-Induced Microbiota-Mediated Colonic Mucus Deterioration.” *Cell Host & Microbe* 23 (1): 27-40.e7. https://doi.org/10.1016/j.chom.2017.11.004.

Schultheiß, Christoph, Lisa Paschold, Donjete Simnica, Malte Mohme, Edith Willscher, Lisa von Wenserski, Rebekka Scholz, et al. 2020. “Next-Generation Sequencing of T and B Cell Receptor Repertoires from COVID-19 Patients Showed Signatures Associated with Severity of Disease.” *Immunity* 53 (2): 442-455.e4. https://doi.org/10.1016/j.immuni.2020.06.024.

Schultze, Joachim L., and Philip Rosenstiel. 2018. “Systems Medicine in Chronic Inflammatory Diseases.” *Immunity* 48 (4): 608–13. https://doi.org/10.1016/j.immuni.2018.03.022.

“Search | HMP.” n.d. Accessed February 28, 2023. https://portal.hmpdacc.org/search/s?facetTab=cases.

Sedghi, Lea M., Margot Bacino, and Yvonne Lorraine Kapila. 2021. “Periodontal Disease: The Good, The Bad, and The Unknown.” *Frontiers in Cellular and Infection Microbiology* 11. https://www.frontiersin.org/articles/10.3389/fcimb.2021.766944.

Shen, Yue, Maria Letizia Giardino Torchia, Gregory W. Lawson, Christopher L. Karp, Jonathan D. Ashwell, and Sarkis K. Mazmanian. 2012. “Outer Membrane Vesicles of a Human Commensal Mediate Immune Regulation and Disease Protection.” *Cell Host & Microbe* 12 (4): 509–20. https://doi.org/10.1016/j.chom.2012.08.004.

Shi, Yu, and Osamu Tokunaga. 2002. “Herpesvirus (HSV-1, EBV and CMV) Infections in Atherosclerotic Compared with Non-Atherosclerotic Aortic Tissue.” *Pathology International* 52 (1): 31–39. https://doi.org/10.1046/j.1440-1827.2002.01312.x.

Shkoporov, Andrey N., Adam G. Clooney, Thomas D. S. Sutton, Feargal J. Ryan, Karen M. Daly, James A. Nolan, Siobhan A. McDonnell, et al. 2019. “The Human Gut Virome Is Highly Diverse, Stable, and Individual Specific.” *Cell Host & Microbe* 26 (4): 527-541.e5. https://doi.org/10.1016/j.chom.2019.09.009.

Shkoporov, Andrey N., and Colin Hill. 2019. “Bacteriophages of the Human Gut: The ‘Known Unknown’ of the Microbiome.” *Cell Host & Microbe* 25 (2): 195–209. https://doi.org/10.1016/j.chom.2019.01.017.

“Single-Cell Multiomics Guide.” n.d. Accessed June 4, 2023. https://www.biolegend.com/en-us/blog/cite-seq-and-totalseq-reagents.

Siqueira, W. L., W. Custodio, and E. E. McDonald. 2012. “New Insights into the Composition and Functions of the Acquired Enamel Pellicle.” *Journal of Dental Research* 91 (12): 1110–18. https://doi.org/10.1177/0022034512462578.

Siroux, Valérie, Lydiane Agier, and Rémy Slama. 2016. “The Exposome Concept: A Challenge and a Potential Driver for Environmental Health Research.” *European Respiratory Review* 25 (140): 124–29. https://doi.org/10.1183/16000617.0034-2016.

Slots, J., H. S. Reynolds, and R. J. Genco. 1980. “Actinobacillus Actinomycetemcomitans in Human Periodontal Disease: A Cross-Sectional Microbiological Investigation.” *Infection and Immunity* 29 (3): 1013–20. https://doi.org/10.1128/iai.29.3.1013-1020.1980.

Slots, Jørgen. 2010. “Human Viruses in Periodontitis.” *Periodontology 2000* 53 (June): 89–110. https://doi.org/10.1111/j.1600-0757.2009.00325.x.

———. 2015. “Periodontal Herpesviruses: Prevalence, Pathogenicity, Systemic Risk.” *Periodontology 2000* 69 (1): 28–45. https://doi.org/10.1111/prd.12085.

Smatti, Maria K., Farhan S. Cyprian, Gheyath K. Nasrallah, Asmaa A. Al Thani, Ruba O. Almishal, and Hadi M. Yassine. 2019. “Viruses and Autoimmunity: A Review on the Potential Interaction and Molecular Mechanisms.” *Viruses* 11 (8): 762. https://doi.org/10.3390/v11080762.

Smith, Aaron Asael, Rui Li, and Zion Tsz Ho Tse. 2023. “Reshaping Healthcare with Wearable Biosensors.” *Scientific Reports* 13 (1): 4998. https://doi.org/10.1038/s41598-022-26951-z.

Soares, Anne Danieli Nascimento, Samuel Penna Wanner, Elissa Stefane Silva Morais, Alexandre Sérvulo Ribeiro Hudson, Flaviano Santos Martins, and Valbert Nascimento Cardoso. 2019. “Supplementation with Saccharomyces Boulardii Increases the Maximal Oxygen Consumption and Maximal Aerobic Speed Attained by Rats Subjected to an Incremental-Speed Exercise.” *Nutrients* 11 (10): 2352. https://doi.org/10.3390/nu11102352.

Socransky, S. S., A. D. Haffajee, M. A. Cugini, C. Smith, and R. L. Kent. 1998. “Microbial Complexes in Subgingival Plaque.” *Journal of Clinical Periodontology* 25 (2): 134–44. https://doi.org/10.1111/j.1600-051x.1998.tb02419.x.

Socransky, Sigmund S., and Anne D. Haffajee. 2002. “Dental Biofilms: Difficult Therapeutic Targets.” *Periodontology 2000* 28: 12–55. https://doi.org/10.1034/j.1600-0757.2002.280102.x.

Sohail, Muhammad U., Hadi M. Yassine, Aaqib Sohail, and Asmaa A. Al Thani. 2019. “Impact of Physical Exercise on Gut Microbiome, Inflammation, and the Pathobiology of Metabolic Disorders.” *The Review of Diabetic Studies: RDS* 15: 35–48. https://doi.org/10.1900/RDS.2019.15.35.

Sokol, H., P. Seksik, J. P. Furet, O. Firmesse, I. Nion-Larmurier, L. Beaugerie, J. Cosnes, G. Corthier, P. Marteau, and J. Doré. 2009. “Low Counts of Faecalibacterium Prausnitzii in Colitis Microbiota.” *Inflammatory Bowel Diseases* 15 (8): 1183–89. https://doi.org/10.1002/ibd.20903.

Sokol, Harry, Bénédicte Pigneur, Laurie Watterlot, Omar Lakhdari, Luis G. Bermúdez-Humarán, Jean-Jacques Gratadoux, Sébastien Blugeon, et al. 2008. “Faecalibacterium Prausnitzii Is an Anti-Inflammatory Commensal Bacterium Identified by Gut Microbiota Analysis of Crohn Disease Patients.” *Proceedings of the National Academy of Sciences of the United States of America* 105 (43): 16731–36. https://doi.org/10.1073/pnas.0804812105.

Sonnenburg, Erica D., Samuel A. Smits, Mikhail Tikhonov, Steven K. Higginbottom, Ned S. Wingreen, and Justin L. Sonnenburg. 2016. “Diet-Induced Extinctions in the Gut Microbiota Compound over Generations.” *Nature* 529 (7585): 212–15. https://doi.org/10.1038/nature16504.

Sonnenburg, Justin L., and Fredrik Bäckhed. 2016. “Diet-Microbiota Interactions as Moderators of Human Metabolism.” *Nature* 535 (7610): 56–64. https://doi.org/10.1038/nature18846.

Sorlie, P. D., E. Adam, S. L. Melnick, A. Folsom, T. Skelton, L. E. Chambless, R. Barnes, and J. L. Melnick. 1994. “Cytomegalovirus/Herpesvirus and Carotid Atherosclerosis: The ARIC Study.” *Journal of Medical Virology* 42 (1): 33–37. https://doi.org/10.1002/jmv.1890420107.

Ståhl, Patrik L., Fredrik Salmén, Sanja Vickovic, Anna Lundmark, José Fernández Navarro, Jens Magnusson, Stefania Giacomello, et al. 2016. “Visualization and Analysis of Gene Expression in Tissue Sections by Spatial Transcriptomics.” *Science (New York, N.Y.)* 353 (6294): 78–82. https://doi.org/10.1126/science.aaf2403.

Stefan, Christopher P., Jeffrey W. Koehler, and Timothy D. Minogue. 2016. “Targeted Next-Generation Sequencing for the Detection of Ciprofloxacin Resistance Markers Using Molecular Inversion Probes.” *Scientific Reports* 6 (1): 25904. https://doi.org/10.1038/srep25904.

Stoeckius, Marlon, Christoph Hafemeister, William Stephenson, Brian Houck-Loomis, Pratip K. Chattopadhyay, Harold Swerdlow, Rahul Satija, and Peter Smibert. 2017. “Simultaneous Epitope and Transcriptome Measurement in Single Cells.” *Nature Methods* 14 (9): 865–68. https://doi.org/10.1038/nmeth.4380.

Strauss, Jaclyn, Gilaad G. Kaplan, Paul L. Beck, Kevin Rioux, Remo Panaccione, Rebekah Devinney, Tarah Lynch, and Emma Allen-Vercoe. 2011. “Invasive Potential of Gut Mucosa-Derived Fusobacterium Nucleatum Positively Correlates with IBD Status of the Host.” *Inflammatory Bowel Diseases* 17 (9): 1971–78. https://doi.org/10.1002/ibd.21606.

Stubbington, Michael J. T., Tapio Lönnberg, Valentina Proserpio, Simon Clare, Anneliese O. Speak, Gordon Dougan, and Sarah A. Teichmann. 2016. “T Cell Fate and Clonality Inference from Single-Cell Transcriptomes.” *Nature Methods* 13 (4): 329–32. https://doi.org/10.1038/nmeth.3800.

Subramanian, Indhupriya, Srikant Verma, Shiva Kumar, Abhay Jere, and Krishanpal Anamika. 2020. “Multi-Omics Data Integration, Interpretation, and Its Application.” *Bioinformatics and Biology Insights* 14 (January): 1177932219899051. https://doi.org/10.1177/1177932219899051.

Suez, Jotham, Tal Korem, David Zeevi, Gili Zilberman-Schapira, Christoph A. Thaiss, Ori Maza, David Israeli, et al. 2014. “Artificial Sweeteners Induce Glucose Intolerance by Altering the Gut Microbiota.” *Nature* 514 (7521): 181–86. https://doi.org/10.1038/nature13793.

Swidsinski, Alexander, Victoria Ung, Beate C. Sydora, Vera Loening-Baucke, Yvonne Doerffel, Hans Verstraelen, and Richard N. Fedorak. 2009. “Bacterial Overgrowth and Inflammation of Small Intestine after Carboxymethylcellulose Ingestion in Genetically Susceptible Mice.” *Inflammatory Bowel Diseases* 15 (3): 359–64. https://doi.org/10.1002/ibd.20763.

Sylvén, Christer, Eva Wärdell, Agneta Månsson-Broberg, Eugenio Cingolani, Konstantinos Ampatzis, Ludvig Larsson, Åsa Björklund, and Stefania Giacomello. 2023. “High Cardiomyocyte Diversity in Human Early Prenatal Heart Development.” *iScience* 26 (1): 105857. https://doi.org/10.1016/j.isci.2022.105857.

Tamayo-Trujillo, Rafael, Patricia Guevara-Ramírez, Santiago Cadena-Ullauri, Elius Paz-Cruz, Viviana A. Ruiz-Pozo, and Ana Karina Zambrano. 2023. "Human Virome: Implications in Cancer." *Heliyon* 9 (3): e14086. https://doi.org/10.1016/j.heliyon.2023.e14086.

Tao, Yue, Hui Yan, Yujie Liu, Fang Zhang, Lijuan Luo, Yajuan Zhou, Kang An, et al. 2022. "Diagnostic Performance of Metagenomic Next-Generation Sequencing in Pediatric Patients: A Retrospective Study in a Large Children's Medical Center." *Clinical Chemistry* 68 (8): 1031–41. https://doi.org/10.1093/clinchem/hvac067.

Ter Horst, Rob, Martin Jaeger, Sanne P. Smeekens, Marije Oosting, Morris A. Swertz, Yang Li, Vinod Kumar, et al. 2016. "Host and Environmental Factors Influencing Individual Human Cytokine Responses." *Cell* 167 (4): 1111-1124.e13. https://doi.org/10.1016/j.cell.2016.10.018.

Thissen, James B., Nicholas A. Be, Kevin McLoughlin, Shea Gardner, Paul G. Rack, Michael H. Shapero, Raymond R. R. Rowland, Tom Slezak, and Crystal J. Jaing. 2019. "Axiom Microbiome Array, the next Generation Microarray for High-Throughput Pathogen and Microbiome Analysis." *PLOS ONE* 14 (2): e0212045. https://doi.org/10.1371/journal.pone.0212045.

Thomas, Stéphanie, Vincent Rouilly, Etienne Patin, Cécile Alanio, Annick Dubois, Cécile Delval, Louis-Guillaume Marquier, et al. 2015. "The Milieu Intérieur Study - an Integrative Approach for Study of Human Immunological Variance." *Clinical Immunology (Orlando, Fla.)* 157 (2): 277–93. https://doi.org/10.1016/j.clim.2014.12.004.

Thukral, Ashwani. 2017. "A Review on Measurement of Alpha Diversity in Biology." *Agricultural Research Journal* 54 (January): 1. https://doi.org/10.5958/2395-146X.2017.00001.1.

Ticinesi, Andrea, Fulvio Lauretani, Claudio Tana, Antonio Nouvenne, Erminia Ridolo, and Tiziana Meschi. 2019. "Exercise and Immune System as Modulators of Intestinal Microbiome: Implications for the Gut-Muscle Axis Hypothesis." *Exercise Immunology Review* 25: 84–95.

Tilman, David. 2004. "Niche Tradeoffs, Neutrality, and Community Structure: A Stochastic Theory of Resource Competition, Invasion, and Community Assembly." *Proceedings of the National Academy of Sciences* 101 (30): 10854–61. https://doi.org/10.1073/pnas.0403458101.

Townsend, Jeremy R., David Bender, William C. Vantrease, Philip A. Sapp, Ann M. Toy, Clint A. Woods, and Kent D. Johnson. 2018. "Effects of Probiotic (Bacillus Subtilis DE111) Supplementation on Immune Function, Hormonal Status, and Physical Performance in Division I Baseball Players." *Sports (Basel, Switzerland)* 6 (3): 70. https://doi.org/10.3390/sports6030070.

Tran, Hoa Thi Nhu, Kok Siong Ang, Marion Chevrier, Xiaomeng Zhang, Nicole Yee Shin Lee, Michelle Goh, and Jinmiao Chen. 2020. "A Benchmark of Batch-Effect Correction Methods for Single-Cell RNA Sequencing Data." *Genome Biology* 21 (1): 12. https://doi.org/10.1186/s13059-019-1850-9.

Tsai, Ming-Horng, Miao-Ching Chi, Jen-Fu Hsu, I.-Ta Lee, Ko-Ming Lin, Mei-Ling Fang, Ming-Hsueh Lee, Chiang-Wen Lee, and Ju-Fang Liu. 2020. "Urban Particulate Matter Enhances ROS/IL-6/COX-II Production by Inhibiting MicroRNA-137 in Synovial Fibroblast of Rheumatoid Arthritis." *Cells* 9 (6): 1378. https://doi.org/10.3390/cells9061378.

Tsuchida, Sachio. 2020. “Proteome Analysis of Molecular Events in Oral Pathogenesis and Virus: A Review with a Particular Focus on Periodontitis.” *International Journal of Molecular Sciences* 21 (15): 5184. https://doi.org/10.3390/ijms21155184.

Turin, Christie G., and Theresa J. Ochoa. 2014. “The Role of Maternal Breast Milk in Preventing Infantile Diarrhea in the Developing World.” *Current Tropical Medicine Reports* 1 (2): 97–105. https://doi.org/10.1007/s40475-014-0015-x.

Turner, Michelle C., Mark Nieuwenhuijsen, Kim Anderson, David Balshaw, Yuxia Cui, Genevieve Dunton, Jane A. Hoppin, Petros Koutrakis, and Michael Jerrett. 2017. “Assessing the Exposome with External Measures: Commentary on the State of the Science and Research Recommendations.” *Annual Review of Public Health* 38 (March): 215–39. https://doi.org/10.1146/annurev-publhealth-082516-012802.

Tzeng, Nian-Sheng, Chi-Hsiang Chung, Fu-Huang Lin, Chien-Ping Chiang, Chin-Bin Yeh, San-Yuan Huang, Ru-Band Lu, et al. 2018. “Anti-Herpetic Medications and Reduced Risk of Dementia in Patients with Herpes Simplex Virus Infections-a Nationwide, Population-Based Cohort Study in Taiwan.” *Neurotherapeutics: The Journal of the American Society for Experimental NeuroTherapeutics* 15 (2): 417–29. https://doi.org/10.1007/s13311-018-0611-x.

Ubeda, Carles, Ying Taur, Robert R. Jenq, Michele J. Equinda, Tammy Son, Miriam Samstein, Agnes Viale, et al. 2010. “Vancomycin-Resistant Enterococcus Domination of Intestinal Microbiota Is Enabled by Antibiotic Treatment in Mice and Precedes Bloodstream Invasion in Humans.” *The Journal of Clinical Investigation* 120 (12): 4332–41. https://doi.org/10.1172/JCI43918.

Unni, Deepak R., Sierra A. T. Moxon, Michael Bada, Matthew Brush, Richard Bruskiewich, J. Harry Caufield, Paul A. Clemons, et al. 2022. “Biolink Model: A Universal Schema for Knowledge Graphs in Clinical, Biomedical, and Translational Science.” *Clinical and Translational Science* 15 (8): 1848–55. https://doi.org/10.1111/cts.13302.

Vaara, Satu, Markku S Nieminen, Marja-Liisa Lokki, Markus Perola, Pirkko J Pussinen, Jaakko Allonen, Olavi Parkkonen, and Juha Sinisalo. 2012. “Cohort Profile: The Corogene Study.” *International Journal of Epidemiology* 41 (5): 1265–71. https://doi.org/10.1093/ije/dyr090.

Valdes, Ana M., Jens Walter, Eran Segal, and Tim D. Spector. 2018. “Role of the Gut Microbiota in Nutrition and Health.” *BMJ (Clinical Research Ed.)* 361 (June): k2179. https://doi.org/10.1136/bmj.k2179.

Valouev, Anton, Jeffrey Ichikawa, Thaisan Tonthat, Jeremy Stuart, Swati Ranade, Heather Peckham, Kathy Zeng, et al. 2008. “A High-Resolution, Nucleosome Position Map of C. Elegans Reveals a Lack of Universal Sequence-Dictated Positioning.” *Genome Research* 18 (7): 1051–63. https://doi.org/10.1101/gr.076463.108.

Vanhoutte, Tom, Geert Huys, Evie Brandt, and Jean Swings. 2004. “Temporal Stability Analysis of the Microbiota in Human Feces by Denaturing Gradient Gel Electrophoresis Using Universal and Group-Specific 16S rRNA Gene Primers.” *FEMS Microbiology Ecology* 48 (3): 437–46. https://doi.org/10.1016/j.femsec.2004.03.001.

Vanuytsel, Tim, Jan Tack, and Ricard Farre. 2021. “The Role of Intestinal Permeability in Gastrointestinal Disorders and Current Methods of Evaluation.” *Frontiers in Nutrition* 8 (August): 717925. https://doi.org/10.3389/fnut.2021.717925.

Verzosa, Amanda L., Lea A. McGeever, Shun-Je Bhark, Tracie Delgado, Nicole Salazar, and Erica L. Sanchez. 2021. “Herpes Simplex Virus 1 Infection of Neuronal and Non-Neuronal Cells Elicits Specific Innate Immune Responses and Immune Evasion Mechanisms.” *Frontiers in Immunology* 12. https://www.frontiersin.org/articles/10.3389/fimmu.2021.644664.

Vieira-Silva, Sara, João Sabino, Mireia Valles-Colomer, Gwen Falony, Gunter Kathagen, Clara Caenepeel, Isabelle Cleynen, Schalk van der Merwe, Séverine Vermeire, and Jeroen Raes. 2019. “Quantitative Microbiome Profiling Disentangles Inflammation- and Bile Duct Obstruction-Associated Microbiota Alterations across PSC/IBD Diagnoses.” *Nature Microbiology* 4 (11): 1826–31. https://doi.org/10.1038/s41564-019-0483-9.

Vrana, Nihal Engin, Kaia Palm, and Philippe Lavalle. 2020. “Personalization of Medical Device Interfaces: Decreasing Implant-Related Complications by Modular Coatings and Immunoprofiling.” *Future Science OA* 6 (8): FSO607. https://doi.org/10.2144/fsoa-2020-0074.

Wang, George C., Wen Hong L. Kao, Peter Murakami, Qian-Li Xue, Roger B. Chiou, Barbara Detrick, John F. McDyer, et al. 2010. “Cytomegalovirus Infection and the Risk of Mortality and Frailty in Older Women: A Prospective Observational Cohort Study.” *American Journal of Epidemiology* 171 (10): 1144–52. https://doi.org/10.1093/aje/kwq062.

Wang, Jun, Louise B Thingholm, Jurgita Skiecevičienė, Philipp Rausch, Martin Kummen, Johannes R Hov, Frauke Degenhardt, et al. 2016. “Genome-Wide Association Analysis Identifies Variation in Vitamin D Receptor and Other Host Factors Influencing the Gut Microbiota.” *Nature Genetics* 48 (11): 1396–1406. https://doi.org/10.1038/ng.3695.

Wang, Minqiang, Yiran Yang, Jihong Min, Yu Song, Jiaobing Tu, Daniel Mukasa, Cui Ye, et al. 2022. “A Wearable Electrochemical Biosensor for the Monitoring of Metabolites and Nutrients.” *Nature Biomedical Engineering* 6 (11): 1225–35. https://doi.org/10.1038/s41551-022-00916-z.

Wang, Rong-Fu, Marjorie L. Beggs, Bruce D. Erickson, and Carl E. Cerniglia. 2004. “DNA Microarray Analysis of Predominant Human Intestinal Bacteria in Fecal Samples.” *Molecular and Cellular Probes* 18 (4): 223–34. https://doi.org/10.1016/j.mcp.2004.03.002.

Wang, Yan-Bin, Zhu-Hong You, Shan Yang, Hai-Cheng Yi, Zhan-Heng Chen, and Kai Zheng. 2020a. “A Deep Learning-Based Method for Drug-Target Interaction Prediction Based on Long Short-Term Memory Neural Network.” *BMC Medical Informatics and Decision Making* 20 (2): 49. https://doi.org/10.1186/s12911-020-1052-0.

———. 2020b. “A Deep Learning-Based Method for Drug-Target Interaction Prediction Based on Long Short-Term Memory Neural Network.” *BMC Medical Informatics and Decision Making* 20 (2): 49. https://doi.org/10.1186/s12911-020-1052-0.

Wickens, Kristin, Christine Barthow, Edwin A. Mitchell, Janice Kang, Noleen van Zyl, Gordon Purdie, Thorsten Stanley, Penny Fitzharris, Rinki Murphy, and Julian Crane. 2018. “Effects of Lactobacillus Rhamnosus HN001 in Early Life on the Cumulative Prevalence of Allergic Disease to 11 Years.” *Pediatric Allergy and Immunology: Official Publication of the European Society of Pediatric Allergy and Immunology* 29 (8): 808–14. https://doi.org/10.1111/pai.12982.

Wild, Christopher Paul. 2005a. “Complementing the Genome with an ‘Exposome’: The Outstanding Challenge of Environmental Exposure Measurement in Molecular Epidemiology.” *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology* 14 (8): 1847–50. https://doi.org/10.1158/1055-9965.EPI-05-0456.

———. 2005b. “Complementing the Genome with an ‘Exposome’: The Outstanding Challenge of Environmental Exposure Measurement in Molecular Epidemiology.” *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology* 14 (8): 1847–50. https://doi.org/10.1158/1055-9965.EPI-05-0456.

———. 2012. “The Exposome: From Concept to Utility.” *International Journal of Epidemiology* 41 (1): 24–32. https://doi.org/10.1093/ije/dyr236.

Winter, Sebastian E., Parameth Thiennimitr, Maria G. Winter, Brian P. Butler, Douglas L. Huseby, Robert W. Crawford, Joseph M. Russell, et al. 2010. “Gut Inflammation Provides a Respiratory Electron Acceptor for Salmonella.” *Nature* 467 (7314): 426–29. https://doi.org/10.1038/nature09415.

Woese, C R, and G E Fox. 1977. “Phylogenetic Structure of the Prokaryotic Domain: The Primary Kingdoms.” *Proceedings of the National Academy of Sciences of the United States of America* 74 (11): 5088–90.

Woo, P. C. Y., S. K. P. Lau, J. L. L. Teng, H. Tse, and K.-Y. Yuen. 2008. “Then and Now: Use of 16S rDNA Gene Sequencing for Bacterial Identification and Discovery of Novel Bacteria in Clinical Microbiology Laboratories.” *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases* 14 (10): 908–34. https://doi.org/10.1111/j.1469-0691.2008.02070.x.

Wook Kim, Ki, Digby W. Allen, Thomas Briese, Jennifer J. Couper, Simon C. Barry, Peter G. Colman, Andrew M. Cotterill, et al. 2019. “Distinct Gut Virome Profile of Pregnant Women With Type 1 Diabetes in the ENDIA Study.” *Open Forum Infectious Diseases* 6 (2): ofz025. https://doi.org/10.1093/ofid/ofz025.

Wu, Gary D., Jun Chen, Christian Hoffmann, Kyle Bittinger, Ying-Yu Chen, Sue A. Keilbaugh, Meenakshi Bewtra, et al. 2011. “Linking Long-Term Dietary Patterns with Gut Microbial Enterotypes.” *Science (New York, N.Y.)* 334 (6052): 105–8. https://doi.org/10.1126/science.1208344.

Wu, Haotian, Christina M. Eckhardt, and Andrea A. Baccarelli. 2023. “Molecular Mechanisms of Environmental Exposures and Human Disease.” *Nature Reviews Genetics* 24 (5): 332–44. https://doi.org/10.1038/s41576-022-00569-3.

Wu, Jing, Brandilyn A. Peters, Christine Dominianni, Yilong Zhang, Zhiheng Pei, Liying Yang, Yingfei Ma, et al. 2016a. “Cigarette Smoking and the Oral Microbiome in a Large Study of American Adults.” *The ISME Journal* 10 (10): 2435–46. https://doi.org/10.1038/ismej.2016.37.

———. 2016b. “Cigarette Smoking and the Oral Microbiome in a Large Study of American Adults.” *The ISME Journal* 10 (10): 2435–46. https://doi.org/10.1038/ismej.2016.37.

Wu, X., R. C. Nethery, M. B. Sabath, D. Braun, and F. Dominici. 2020. “Air Pollution and COVID-19 Mortality in the United States: Strengths and Limitations of an Ecological Regression Analysis.” *Science Advances* 6 (45): eabd4049. https://doi.org/10.1126/sciadv.abd4049.

Wu, Yu peng, Dan dan Sun, Yun Wang, Wen Liu, and Jun Yang. 2016. “Herpes Simplex Virus Type 1 and Type 2 Infection Increases Atherosclerosis Risk: Evidence Based on a Meta-Analysis.” *BioMed Research International* 2016: 2630865. https://doi.org/10.1155/2016/2630865.

Yanai, Itai, and Martin Lercher. 2020. “A Hypothesis Is a Liability.” *Genome Biology* 21 (1): 231. https://doi.org/10.1186/s13059-020-02133-w.

Yang, Jin-Young, Min-Soo Kim, Eugene Kim, Jae Hee Cheon, Yong-Soo Lee, Yeji Kim, Su-Hyun Lee, et al. 2016. “Enteric Viruses Ameliorate Gut Inflammation via Toll-like Receptor 3 and Toll-like Receptor 7-Mediated Interferon-β Production.” *Immunity* 44 (4): 889–900. https://doi.org/10.1016/j.immuni.2016.03.009.

Yang, Yaohua, Wei Zheng, Qiu-Yin Cai, Martha J. Shrubsole, Zhiheng Pei, Robert Brucker, Mark D. Steinwandel, et al. 2019. “Cigarette Smoking and Oral Microbiota in Low-Income and African-American Populations.” *Journal of Epidemiology and Community Health* 73 (12): 1108–15. https://doi.org/10.1136/jech-2019-212474.

Young, J. C., C. Chehoud, K. Bittinger, A. Bailey, J. M. Diamond, E. Cantu, A. R. Haas, et al. 2015. “Viral Metagenomics Reveal Blooms of Anelloviruses in the Respiratory Tract of Lung Transplant Recipients.” *American Journal of Transplantation: Official Journal of the American Society of Transplantation and the American Society of Transplant Surgeons* 15 (1): 200–209. https://doi.org/10.1111/ajt.13031.

Yu, Miao, Susan Teitelbaum, Georgia Dolios, Lam-Ha Dang, Peijun Tu, Mary Wolff, and Lauren Petrick. 2021. “Molecular Gatekeeper Discovery: Workflow for Linking Multiple Environmental Biomarkers to Metabolomics.” ChemRxiv. https://doi.org/10.26434/chemrxiv.14781498.v1.

Yu, Yanbo, Genxiang Mao, Jirong Wang, Liyue Zhu, Xiaoling Lv, Qian Tong, Yefei Fang, Yinxiang Lv, and Guofu Wang. 2018. “Gut Dysbiosis Is Associated with the Reduced Exercise Capacity of Elderly Patients with Hypertension.” *Hypertension Research: Official Journal of the Japanese Society of Hypertension* 41 (12): 1036–44. https://doi.org/10.1038/s41440-018-0110-9.

Zandian, Arash, Björn Forsström, Anna Häggmark-Månberg, Jochen M. Schwenk, Mathias Uhlén, Peter Nilsson, and Burcu Ayoglu. 2017. “Whole-Proteome Peptide Microarrays for Profiling Autoantibody Repertoires within Multiple Sclerosis and Narcolepsy.” *Journal of Proteome Research* 16 (3): 1300–1314. https://doi.org/10.1021/acs.jproteome.6b00916.

Zaura, Egija, Elena A. Nicu, Bastiaan P. Krom, and Bart J. F. Keijser. 2014. “Acquiring and Maintaining a Normal Oral Microbiome: Current Perspective.” *Frontiers in Cellular and Infection Microbiology* 4: 85. https://doi.org/10.3389/fcimb.2014.00085.

Zeller, Georg, Julien Tap, Anita Y. Voigt, Shinichi Sunagawa, Jens Roat Kultima, Paul I. Costea, Aurélien Amiot, et al. 2014. “Potential of Fecal Microbiota for Early-Stage Detection of Colorectal Cancer.” *Molecular Systems Biology* 10 (11): 766. https://doi.org/10.15252/msb.20145645.

Zeng, Xiangxiang, Siyi Zhu, Weiqiang Lu, Zehui Liu, Jin Huang, Yadi Zhou, Jiansong Fang, et al. 2020. “Target Identification among Known Drugs by Deep Learning from Heterogeneous Networks.” *Chemical Science* 11 (7): 1775–97. https://doi.org/10.1039/C9SC04336E.

Zhang, Chenhong, Menghui Zhang, Xiaoyan Pang, Yufeng Zhao, Linghua Wang, and Liping Zhao. 2012. “Structural Resilience of the Gut Microbiota in Adult Mice under High-Fat Dietary Perturbations.” *The ISME Journal* 6 (10): 1848–57. https://doi.org/10.1038/ismej.2012.27.

Zhang, Chenhong, Menghui Zhang, Shengyue Wang, Ruijun Han, Youfang Cao, Weiying Hua, Yuejian Mao, et al. 2010. “Interactions between Gut Microbiota, Host Genetics and Diet Relevant to Development of Metabolic Syndromes in Mice.” *The ISME Journal* 4 (2): 232–41. https://doi.org/10.1038/ismej.2009.112.

Zhao, Guoyan, Tommi Vatanen, Lindsay Droit, Arnold Park, Aleksandar D. Kostic, Tiffany W. Poon, Hera Vlamakis, et al. 2017. “Intestinal Virome Changes Precede Autoimmunity in Type I Diabetes-Susceptible Children.” *Proceedings of the National Academy of Sciences* 114 (30): E6166–75. https://doi.org/10.1073/pnas.1706359114.

Zhao, Lingling, Junjie Wang, Long Pang, Yang Liu, and Jun Zhang. 2020a. “GANsDTA: Predicting Drug-Target Binding Affinity Using GANs.” *Frontiers in Genetics* 10. https://www.frontiersin.org/articles/10.3389/fgene.2019.01243.

———. 2020b. “GANsDTA: Predicting Drug-Target Binding Affinity Using GANs.” *Frontiers in Genetics* 10. https://www.frontiersin.org/articles/10.3389/fgene.2019.01243.

Zhao, Liping. 2013. “The Gut Microbiota and Obesity: From Correlation to Causality.” *Nature Reviews. Microbiology* 11 (9): 639–47. https://doi.org/10.1038/nrmicro3089.

Zheng, Pei, and Weiying Zhou. 2015. “Relation between Periodontitis and Helicobacter Pylori Infection.” *International Journal of Clinical and Experimental Medicine* 8 (9): 16741–44.

Zhong, Ling, Wanlin Zhang, Claude Krummenacher, Yixin Chen, Qingbing Zheng, Qinjian Zhao, Mu-Sheng Zeng, et al. 2023. “Targeting Herpesvirus Entry Complex and Fusogen Glycoproteins with Prophylactic and Therapeutic Agents.” *Trends in Microbiology* 0 (0). https://doi.org/10.1016/j.tim.2023.03.001.

Zhong, Suijuan, Wenyu Ding, Le Sun, Yufeng Lu, Hao Dong, Xiaoying Fan, Zeyuan Liu, et al. 2020. “Decoding the Development of the Human Hippocampus.” *Nature* 577 (7791): 531–36. https://doi.org/10.1038/s41586-019-1917-5.

Zhong, Wenting, Xiaolan Lu, Haitao Shi, Gang Zhao, Yahua Song, Yonghua Wang, Jing Zhang, Yaofeng Jin, and Shenhao Wang. 2019. “Distinct Microbial Populations Exist in the Mucosa-Associated Microbiota of Diarrhea Predominant Irritable Bowel Syndrome and Ulcerative Colitis.” *Journal of Clinical Gastroenterology* 53 (9): 660–72. https://doi.org/10.1097/MCG.0000000000000961.

Zhu, J., A. A. Quyyumi, J. E. Norman, G. Csako, M. A. Waclawiw, G. M. Shearer, and S. E. Epstein. 2000. “Effects of Total Pathogen Burden on Coronary Artery Disease Risk and C-Reactive Protein Levels.” *The American Journal of Cardiology* 85 (2): 140–46. https://doi.org/10.1016/s0002-9149(99)00653-0.

Zmora, Niv, Jotham Suez, and Eran Elinav. 2019. “You Are What You Eat: Diet, Health and the Gut Microbiota.” *Nature Reviews Gastroenterology & Hepatology* 16 (1): 35–56. https://doi.org/10.1038/s41575-018-0061-2.

Zuo, Tao, Xiao-Juan Lu, Yu Zhang, Chun Pan Cheung, Siu Lam, Fen Zhang, Whitney Tang, et al. 2019. “Gut Mucosal Virome Alterations in Ulcerative Colitis.” *Gut* 68 (7): 1169–79. https://doi.org/10.1136/gutjnl-2018-318131.

Zuo, Tao, Yang Sun, Yating Wan, Yun Kit Yeoh, Fen Zhang, Chun Pan Cheung, Nan Chen, et al. 2020. “Human-Gut-DNA Virome Variations across Geography, Ethnicity, and Urbanization.” *Cell Host & Microbe* 28 (5): 741-751.e4. https://doi.org/10.1016/j.chom.2020.08.005.

Acknowledgements

This work has been carried out at Protobios LLC, Estonia.

I would like to thank my supervisor Dr. Kaia Palm for the immense support and mentorship during my PhD studies. Without her vast knowledge of molecular biology, her supportive guidance, and constructive questions and ideas, the research behind this thesis would not have come to fruition and the scientific world would have suffered a loss.

I am grateful to my supervisor prof. Tõnis Timmusk for guidance and support throughout these years, as well as the pleasure of working together on many interesting research topics.

I am grateful to all my colleagues I had the pleasure of working together with in Protobios and Neurobiology lab. Thank you, Helle Sadam, Arno Pihlak, Annika Rähni, Nadežda Pupina, Mirjam Luhakooder, Annela Avarlaid, Jürgen Tuvikene, Valentina Božok, Helen Verev, Susan Pihelgas, and Grete Rullinkov for teaching me the insights of MVA, sparking a love for data analysis, helping to overcome challenges, and providing a warm and productive lab environment.

I am thankful to all teachers and professors at the Department of Chemistry and Biology at Tallinn University of Technology with whom I had the pleasure of meeting and learning from throughout the entirety of my academic pursuit. Thank you for your love of chemistry and biology, your willingness of passing your knowledge onto new students, and your patience with dealing with us as we sometimes failed our commitments.

Finally, my gratitude goes towards family and friends, especially my partner Joonatan, who have supported and cheered me on during these years.

I thank all the funding agencies that have supported the research supporting this thesis: grants of Protobios from the Estonian Ministry of Education (5.1-4/20/17), from the Estonian Research Council (PRG573, PSG691, PRG1953), from the European Union programs of H2020-MSCA-RISE-2016 (EU734791) and H2020 PANBioRA (EU760921), the Helsinki University Hospital grants, Mary and Georg C. Ehrnrooth Foundation, and the Finnish Eye Foundation, and the Doctoral School of Clinical Medicine, supported by the European Union, European Regional Development Fund through Tallinn University of Technology ASTRA “TTÜ Institutional Development Program for 2016-2022”.

Abstract

Exploration of host-agent-environment interactions using tools of metagenomic sequencing and next generation phage display

The exposome refers to the totality of environmental exposures that an individual experiences throughout their lifetime, including both external and internal factors. It encompasses various elements such as air pollution, diet, lifestyle choices, chemicals, stress, and socioeconomic factors. Studying the exposome allows for a comprehensive understanding of how environmental factors contribute to disease development, progression, and response to treatment. Understanding the exposome can provide valuable insights into the underlying causes of diseases, help identify novel risk factors, improve public health policies, and guide the development of personalized prevention and treatment strategies. Advancements in computational technology have allowed to gather and analyze increasingly larger datasets. The technological background of exposome research involves the integration of diverse sampling procedures, advanced omics, bioinformatics tools, and systems biology approaches. Here we used metagenomic sequencing and next generation phage display-based Mimotope Variation Analysis (MVA) to characterize the exposome in relation to health and disease. Firstly, we identified compositional changes in the gut microbiome and discussed the links with dietary fibre supplementation and sports performance. Secondly, we characterized the immune response directed to commensal microbiota and based on these findings proposed a two-step classification based on MVA-delineated biomarkers to differentiate patients with severe heart conditions from healthy. Third, we characterized a positive association between diagnosis of Pandemrix vaccine-induced narcolepsy and adverse immune reaction to specific epitopes of self-antigens. Further, we discussed data analysis aspects of studying the exposome, reviewing the benefits and considerations of using MVA in exposome research. Overall, assessing the association between many exposures and health still raises many experimental issues and data challenges, however, despite its current limitations allows a more complete understanding of disease.

Lühikokkuvõte

Metagenoomi sekveneerimise ja järgmise põlvkonna faagidisplei kasutamine inimese eksposoomi kirjeldamisel

Inimene puutub oma eluea jooksul kokku paljude väliste keskkonna- ja sisemiste bioloogiliste mõjuritega, sealhulgas õhusaaste, toit, elustiilivalikud, kemikaalid, stress ja sotsiaalmajanduslikud mõjurid. Eluea jooksul kokku puututud mõjurite ehk eksposoomi koos uurimine võimaldab saada terviklikum ülevaade sellest, kuidas inimest ümbritsev elus ja eluta keskkond mõjutab haiguste avaldumist ja väljakujunemist ning ravile allumist. Erinevate mõjurite ja inimese omavaheliste interaktsioonide kirjeldamine võimaldab paremini välja selgitada haiguste tekkepõhjuseid, tuvastada uusi riskifaktoreid, parendada ühiskondlikku tervisepoliitikat, ja kiirendada personaalmeditsiini arengut. Arengud arvutustehnoloogias on võimaldanud analüüsida järjest suuremaid andmehulkasid. Eksposoomi uurimise raames kogutakse korraga mitmeid keemilisi ja bioloogilisi proove ning taustandmeid, mida seejärel analüüsitakse erinevate oomika meetodite (genoomika, transkriptoomika, proteoomika ja metaboloomika), bioinformaatiliste tööriistade (sealhulgas masinõpet) ning süsteemibioloogia analüüsimeetodite abil.

Käesolevate uuringute käigus kasutati metagenoomi sekveneerimist ja järgmise põlvkonna faagidisplei meetodil põhinevat mimotoobi variatsiooni analüüsi (MVA), et kirjeldada eksposoomi seoses inimese tervise ja haigusega. Esiteks tuvastati muutuseid soolestiku mikrofloora koosluses peale kiudainelisandi manustamist ning arutleti nende seose üle sportliku sooritusega. Teiseks kirjeldati antikehalist immuunvastust inimese kommensaalse mikrobioomi vastu. MVA-tuvastatud biomarkeritel põhinev kahe-etapiline hindamismeetodi võimaldas eristada raske südamehaigusega patsiente tervetest. Kolmandaks kirjeldati positiivne seos Pandemrix vaktsiin-seoselise narkolepsia ning kahjuliku, enese epitoopide vastu suunatud immuunvastuse vahel. Neljandaks arutleti erinevate andmeanalüüsimeetodite üle ning kirjeldati MVA meetodi eeliseid ja raskuskohti eksposoomi uurimisel. Bioloogiliste ja keskkondlike mõjurite koosmõju uurimiseks tervisele on vaja veel lahendada mitmeid tehnilisi ja andmeanalüüsiga seotud katsumusi, kuid hoolimata piirangutest võimaldab eksposoomi uurimine saavutada terviklikuma arusaamise inimese tervisest ja haigusest.

Appendix 1

PUBLICATION I

Mariliis Jaago, Timmusk, U. S., Timmusk, T., Palm, K. 2021. **Drastic Effects on the Microbiome of a Young Rower Engaged in High-Endurance Exercise After a Month Usage of a Dietary Fiber Supplement.** *Frontiers in Nutrition*, 8:654008. doi: 10.3389/fnut.2021.654008.

frontiers in Nutrition

BRIEF RESEARCH REPORT
published: 30 April 2021
doi: 10.3389/fnut.2021.654008

Drastic Effects on the Microbiome of a Young Rower Engaged in High-Endurance Exercise After a Month Usage of a Dietary Fiber Supplement

Mariliis Jaago[1,2], *Uku Siim Timmusk*[1], *Tõnis Timmusk*[1,2] *and Kaia Palm*[1,2*]

[1] Protobios Llc, Tallinn, Estonia, [2] Department of Chemistry and Biotechnology, Tallinn University of Technology, Tallinn, Estonia

Food supplements are increasingly used worldwide. However, research on the efficacy of such supplements on athlete's well-being and optimal sports performance is very limited. This study performed in junior academic rowing explores the effects of nutritional supplements to aid to the high energy requirements at periods of intense exercise. Herein, the effects of prebiotic fibers on the intestinal microbiome composition of an 18-year-old athlete exercising at high loads during an 8-month period in a "real-life" setting were examined using next-generation sequencing analysis. Results demonstrated that although the alpha diversity of the subject's microbiome drastically decreased [from 2.11 precompetition to 1.67 ($p < 0.05$)] upon fiber consumption, the *Firmicutes*/*Bacteroidetes* ratio increased significantly [from 3.11 to 4.55, as compared with population average ($p < 0.05$)]. Underlying these macrolevel microbial alterations were demonstrable shifts from acetate- to butyrate-producing bacteria, although with stable effects on the *Veillonella* species. To our knowledge, this a unique study that shows pronounced changes in the gut microbiome of the young athlete at the competition season and their favorable compensation by the dietary fiber intake. The data here expand the overall understanding of how the high energy needs in high-intensity sports like academic rowing could be supported by dietary fiber supplement consumption.

Keywords: microbiome, endurance sports, rowing, dietary fiber, athlete, junior, case study

OPEN ACCESS

Edited by:
Yan Huang,
University of Arkansas, United States

Reviewed by:
Feng Wang,
Emory University, United States
Xi Wu,
UC Davis Genome Center, United States

***Correspondence:**
Kaia Palm
kaia@protobios.com

Specialty section:
This article was submitted to Nutrition and Metabolism, a section of the journal Frontiers in Nutrition

Received: *02 February 2021*
Accepted: *22 March 2021*
Published: *30 April 2021*

Citation:
Jaago M, Timmusk US, Timmusk T and Palm K (2021) Drastic Effects on the Microbiome of a Young Rower Engaged in High-Endurance Exercise After a Month Usage of a Dietary Fiber Supplement. Front. Nutr. 8:654008. doi: 10.3389/fnut.2021.654008

INTRODUCTION

The microbiome contributes to thehomeostatic regulation of different tissues in our body (1) with the largest and most diverse cluster of microorganisms inhabiting the gut (2). These core functions are linked to the production of essential and extremely diverse metabolites such as vitamins (vitamin B_{12}, folic acid, or vitamin K), bile acids, neurotransmitters (serotonin, dopamine, acetylcholine), and short-chain fatty acids (SCFAs: acetic acid, propionic acid, and butyric acid) (3). Diet and the level of physical activity are the main determinates for altering the gut microbiota (4). Increases in bacterial diversity and a proliferation of taxa responsible for the production of SCFAs, such as butyrate, are among the most pervasively observed microbial alterations with exercise (5).

Athlete microbiomes have been found to contain distinct microbial compositions defined by elevated abundance of *Veillonellaceae*, *Bacteroides*, *Prevotella*, *Methanobrevibacter*, or *Akkermansia* (6). Cardiorespiratory fitness in exercising subjects was associated with higher abundance of butyrate-producing bacteria by the *Clostridiales*, *Erysipelotrichaceae*, *Lachnospiraceae*, and *Roseburia* families (7). Exercise type along with athlete diet patterns (bodybuilders: high protein, high fat, low carbohydrate, and low dietary fiber diet; distance runners: low carbohydrate and low dietary fiber diet) was significantly associated with the relative differential abundance of *Faecalibacterium*, *Sutterella*, *Clostridium*, *Haemophilus*, *Eisenbergiella*, *Bifidobacterium*, and *Parasutterella* in bodybuilders and distance runners (8). Variation in genera was suggested to be linked to the variance in species' composition across different types of sports (9). So, athletes participating in sports with high dynamic and static component like academic rowing displayed greater abundance of *Bacteroides caccae* (9). The effects of exercise on gut microbial microorganisms were concluded to depend significantly on its intensity and timing with the notion that the microbiota could also influence muscle mass, as reported by Ticinesi et al. (10). Excessive exercise among professional athletes disturbs the homeostasis of the gut microbiota [reviewed in (7)]. Physical exertion at a very high level for a prolonged time means that the whole body initiates a defense response because of oxidative stress, intestinal permeability, muscle damage, systemic inflammation, and immune responses (11). It has been observed that endurance athletes present a high prevalence of upper respiratory tract infections and gastrointestinal troubles, including a "leaky gut," disruption of mucous thickness, and higher rates of bacterial translocation (12, 13). Overall, all these studies suggest that the gut microbiome affects exercise performance and vice versa.

Diet has a major impact on gut microbiota composition, diversity, and richness. Dietary supplements that employ non-digestible dietary fibers have been developed for several decades as prebiotics to support growth of beneficial GI microbiota (14). Dietary fibers can be found in plants, bacteria, and fungi and can be chemically synthesized (15). The health effects of these dietary fibers have extensively been reviewed and accepted worldwide (14). It has been concluded that the extent by which different fiber types are utilized or fermented by the GI microbiota is structure dependent and relies on the metabolic capabilities of the individual's microbiome (16). Virtually, all fibers induce specific shifts in microbiota composition due to competitive interactions; however, which of these shifts contribute to health, or if at all, is not known (17, 18). The commensal bacteria ferment non-digestible fiber primarily into CO_2, H_2, and CH_4 and SCFAs (19). Most of the SCFAs produced in the intestine are absorbed by the host to contribute to energy and beneficial metabolites (20) that are also used as carbon and energy sources by other specialized bacteria including reductive acetogens, sulfate-reducing bacteria, and methanogens (21).

Multiple lines of evidence support the hypothesis that modification of the microbial community through diet could be an effective tool to improve athlete's health (14) performance and energy availability while controlling redox levels and inflammation (22). Endurance diets are rich in protein (1.2–1.6 g/kg/day), which produce a range of potentially harmful compounds in the intestine in addition to SCFA productions (22). There are only a few demonstrated studies in athletes consuming prebiotics (23). Research has suggested the validity of probiotics to improve training parameters and increase training capabilities (24).

Nutritional supplements are popular among athletes to improve performance and physical recovery. Long-term (10 weeks) protein supplement (whey isolate and beef hydrolysate) consumption by cross-country runners, however, decreased the presence of health-related taxa including *Roseburia*, *Blautia*, and *Bifidobacterium longum* and increased the abundance of the Bacteroidetes phylum (25). However, it appeared that protein overconsumption was an offset by a higher intake of indigestible polysaccharides (26).

Athletes with very high training and competition loads can have serious problems getting the necessary amount of energy from regular food. Clark and Mach (2016) reported that diets recommended for athletes likely influence gut microbiota by reducing diversity because these diets include insufficient dietary fiber (27). In addition, a recent study showed that *Veillonella atypica* has a beneficial impact on the performance of elite athletes (6). Based on these findings, we decided to look for dietary ways to increase beneficial bacteria for better athletic performance and faster recovery, in particular considering that nutritional supplements are popular among athletes. We hypothesized that microbial profiles of the young rower might share features of those previously described in endurance sports studies but that this might change in response to the dietary changes upon dietary fiber supplement intake.

MATERIALS AND METHODS

Case Presentation

At the time of the study, the male athlete was 18 years old and was studied over the course of a 31-week period during the 2019 race and training season preceding the world championship competition in U19 category of academic rowing. By this time, the subject had been undertaking rowing for 4 years and had not previously sought any conditioning or dietary advice. Furthermore, the athlete was not on any prescribed medication and was a non-smoker, and his usual diet was previously supplemented with whey protein [SiS (Science in Sport) Limited products] only. In the 7 months prior to the supplement intake, the subject held a normal diet that was alike on a daily basis, comprising mostly of meals that were high in carbohydrate and protein, medium in fat, and modest in dietary fiber. The athlete was fully informed of the study aims and confirmed participation in the study by signing a consent form, understanding that the parameters of health were not associated with the study and that the subject was not physically harmed by the study. This study was approved by Protobios (1-05/2019).

Goals of the Study

The primary goals of the support provided were to (a) get insights into the subject's intestinal microbial community during periods

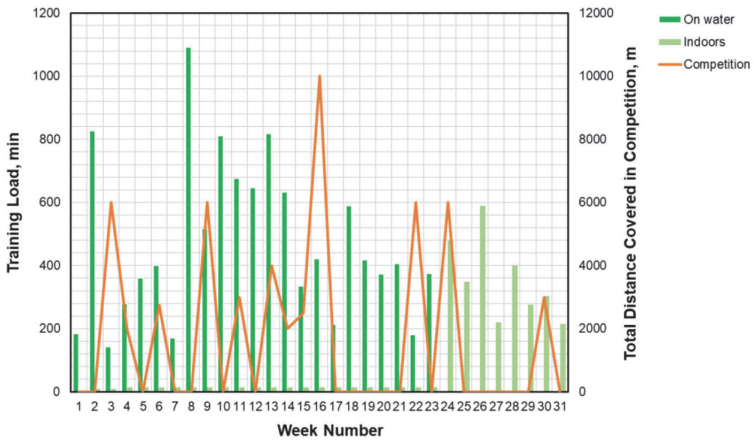


FIGURE 1 | Training program undertaken throughout the study period. Electronic exercise diary (www.sportlyzer.com) was kept for water and ergometer training hours throughout the study period. On the left *y*-axis—training load in minutes per week, min/week (Rowing on the *Water* and *Indoors*); on the right *y*-axis—distance in meters per week, m/week (*Competition/m*); on the *x*-axis—duration of the study period in weeks. Samples for microbiome analysis were taken at weeks 1, 27, and 31. *On water*, training on the water; *Indoors*, training on rowing ergometer; *Competition*, competition calendar of the rower.

of high exercise and (b) examine the effects of dietary fiber intake to the bacterial compositions associated with energy production.

Diet and Activity Recordings and Microbiome Sampling

The participant was informed to maintain the usual dietary habits throughout the study. Body composition estimates were made preseason (sample #1) and post study (sample #3) with no substantial change (BMI 23.3 ± 0.2, fat percentage of 8.4 ± 0.0) according to the medical sports health survey records. Based on diet recall, the usual macronutrient intake was assessed using image-based dietary assessment software (NutriData, National Institute for Health Development, Estonia). We used this tool to ascertain the usual eating patterns of the subject including type, frequency, and amount of foods consumed. Foods consumed were matched to the nutritional analysis for the specific menu items that had been coded in NutriData. If not consumed from the menu, the item was coded against the most appropriate matching food. On average, across the study period, the athlete consumed 2,560 ± 750 kcal/day, and the estimated macronutrient intake was 23 ± 7.12% protein, 52 ± 19.1% carbohydrate, 26 ± 19.3% fat, and 15 ± 4.5 g fiber. The consumption of the nutrient supplement (*Food, not only for thought*, Elsavie, Estonia) started on week 27 and continued on a daily basis until week 31 (altogether 30 days), with intake immediately after breakfast, and the recommended daily intake [1.5 tablespoons (1.5 tbsp/20 g) mixed with water] was not exceeded. The dietary supplement provided to the participant as the prebiotic mix (20 g) included dietary fiber (8.79 g), consisting of resistant starch (2.25 g), arabinoxylan (2.05 g), citrus fiber (2 g), beta-glucans (1.03 g), inulin (1.03 g), and rye fiber (0.57 g). The athlete kept the training diary in Sportlyzer (Sportlyzer, Estonia). The 31-week high-training and intensive competition program is presented in **Figure 1**. The mean physical activity (8 months duration) of the subject during the study was 472 min/week of water rowing, 354 min/week of indoor rowing, and 60 min/week of stretching exercises (**Figure 1**). Since the time the first microbial sample (#1) was taken, the subject was participating in a series of national and international competition activities including five international and 11 local competitions (53.25 km total of total race distance) following the planned program as in **Figure 1**. The second sample (#2) was taken at week 27, by the end of the water season and at early weeks of the indoor rowing season. Thereafter, the subject began to take the dietary fiber supplement as recommended for 30 days. A third sample (#3) was taken at week 31. The seasonality of sampling was spring (#1), autumn (#2), and winter (#3), respectively.

Microbiome Assessment

Microbiome composition was determined on three occasions (week 1, week 27, and week 31, **Figure 1**). Samples were self-collected from morning stool samples by using the commercially available kit (INTEST.pro, BIOMES NGS GmbH, Germany) in accordance with the specifications laid out by the manufacturer. The first two samples were taken at normal nutrition (at the start of and after the active competition period), followed by 1 month of dietary supplement intake to investigate the dynamics of the intestinal microbiome and the effects of the fiber supplement on the microbiota. Collected samples were transported to the lab according to the service provider's instructions where the microbiome composition was analyzed *via* 16S rRNA gene amplification and sequencing by BIOMES NGS GmbH (Germany). In brief, microbial genomic DNA from fecal material was extracted by the bead-beating technique,

the V3–V4 region of the 16S rRNA gene was amplified, and sequencing was performed on the Illumina MiSeq platform using a 2 × 300-bp paired-end protocol (Illumina, San Diego, CA, USA). These DNA sequencing techniques were then used to generate data outputs that provided a comprehensive bacterial taxonomic profile of the subject (28) in comparison with the average microbiome of the European population as the evidence indicates that microbiome may vary by geography (29).

Data Analysis and Statistics

Different packages of MS Excel (based on MS Excel 2011) and licensed MedCalc (version 19.1.6) statistical analysis programs were used for taxonomic, functional analysis, and visualization of bacterial composition data obtained from microbiome sample study reports. Spearman's correlation coefficients r_s were calculated for comparing abundances of genera at two time points. Statistical significance of differences of Shannon's indices (α-diversity) across time points was assessed using Hutcheson's modified t-test with a significance level of $p < 0.05$. Statistical significance of differences between the athlete's characteristics and the control group (general population) was assessed using the single mean t-test with a significance level of $p < 0.05$. Relative abundance values of bacteria on genus and species level at different time points were compared using non-parametric Wilcoxon rank sum tests (also named the Mann–Whitney U-test or Mann–Whitney–Wilcoxon test) with a significance level of $p < 0.05$.

RESULTS

General Diversity, Phylogenic Composition, and Core Gut Microbiota of the Athlete

The Shannon index indicating the diversity of bacterial families present in the samples of the subject ranged from 1.67 to 2.11 (**Figure 2**). At the beginning and end of the training period, the Shannon indices were similar (2.11 and 2.08, respectively), concluding that the microbiome diversity did not change significantly at times of high competition (Hutcheson's modified t-test, p-value > 0.05). After the dietary fiber mix intake, the Shannon index dropped to 1.67 with a significant decrease in community diversity (Hutcheson's modified t-test, p-value < 0.05). The Shannon index of the control group was 1.63 (data not shown).

The evenness of the distribution of species in communities showed that at high exercise, microbiome uniformity indices were similar (J1 = 0.54 vs. J2 = 0.53), whereas, after the fiber intake, J showed a substantial drop (to J3 = 0.42). Although, also the evenness values of the microbiota were the lowest after the fiber mix diet (changed from 0.53 to 0.42), the statistical significance of the reduction could not be concluded. The uniformity index for the control group was $J = 0.41$. These data allowed us to conclude that at the family level, the athlete's microbiome at high training and competition loads (samples #1 and #2) were more diverse and more balanced (even) than after the dietary fiber consumption (sample #3). Intense exercise accompanied by a high intake of dietary fiber did not lead to the increased diversity of gut microbiota as was initially expected.

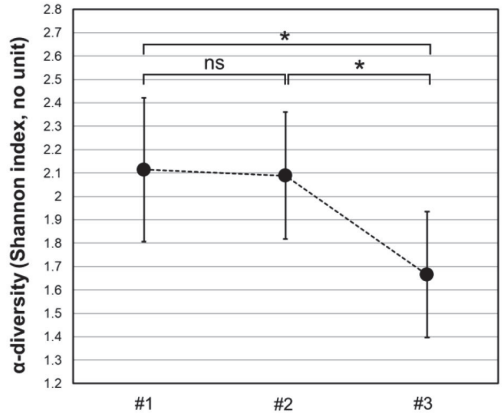


FIGURE 2 | Changes in the microbiome diversity over time and drastic effects on its relative diversity within bacterial families upon the dietary fiber consumption. The Shannon index expressed as a measure of the incidence and frequency of bacterial families in the test samples, ranged from 2.11 (#1) to 2.08 (#2), which was not a statistically significant difference (p-value > 0.05, Hutcheson's modified t-test, denoted *ns*—not significant). One month after consumption of the fiber mixture, the Shannon index dropped to 1.67, which was a significant decrease in the family diversity values (p-value < 0.05, Hutcheson's modified t-test, denoted by *). Error bars represent 95% CI for the calculated Shannon index. y-axis—Shannon index, without units. x-axis—sampling times. #1, time point week 1 (baseline); #2, time point week 27 (high training and competition period); #3, time point week 31 (after 30 days of dietary supplement intake).

As for the phylogenic composition, a total of nine phyla were present in all the samples with the dominance of *Firmicutes*, *Bacteroidetes*, *Actinobacteria*, and *Proteobacteria* (**Figure 3**). The *Firmicutes* [67.8 ± 6.2 (mean abundance across three time points ± SD)] outranked *Bacteroidetes* (18.0 ± 1.8), *Actinobacteria* (8.8 ± 1.7), *Proteobacteria* (2.7 ± 1.6), *Verrucomicrobia* (1.6 ± 1.2), and *Cyanobacteria* (1.0 ± 1.4) phyla. Dietary fiber consumption had a positive effect on the abundance of *Firmicutes* (+20%), whereas, it showed drastic negative effects on *Verrucomicrobia* and *Cyanobacteria*, whose drop in abundance was almost 100%, and on *Proteobacteria*, *Bacteroidetes*, and *Actinobacteria* that declined by 75, 18, and 13%, respectively (**Figure 3B**). Analysis of *Firmicutes*/*Bacteroidetes* (*F*/*B*) ratio values showed relative stability during the study, where during the intense competition period, the reduction in *Firmicutes* [changing the *F*/*B* ratio by 17.9% (from 3.78 to 3.11)] was rescued upon fiber consumption by increases in abundance by 20% and resulting in the *F*/*B* ratio of 4.55 (**Figure 3C**). Compared with the control group, the *F*/*B* ratio of the athlete was significantly higher in all time points (t-test of one mean, p-value < 0.05, **Figure 3D**). Altogether, these results showed that dietary fiber along with high exercise loads affected the phylogenic composition by the microbiome of the athlete becoming relatively poorer at the phylum level. Overall, these results were in good agreement with data showing that training promoted relative increases in

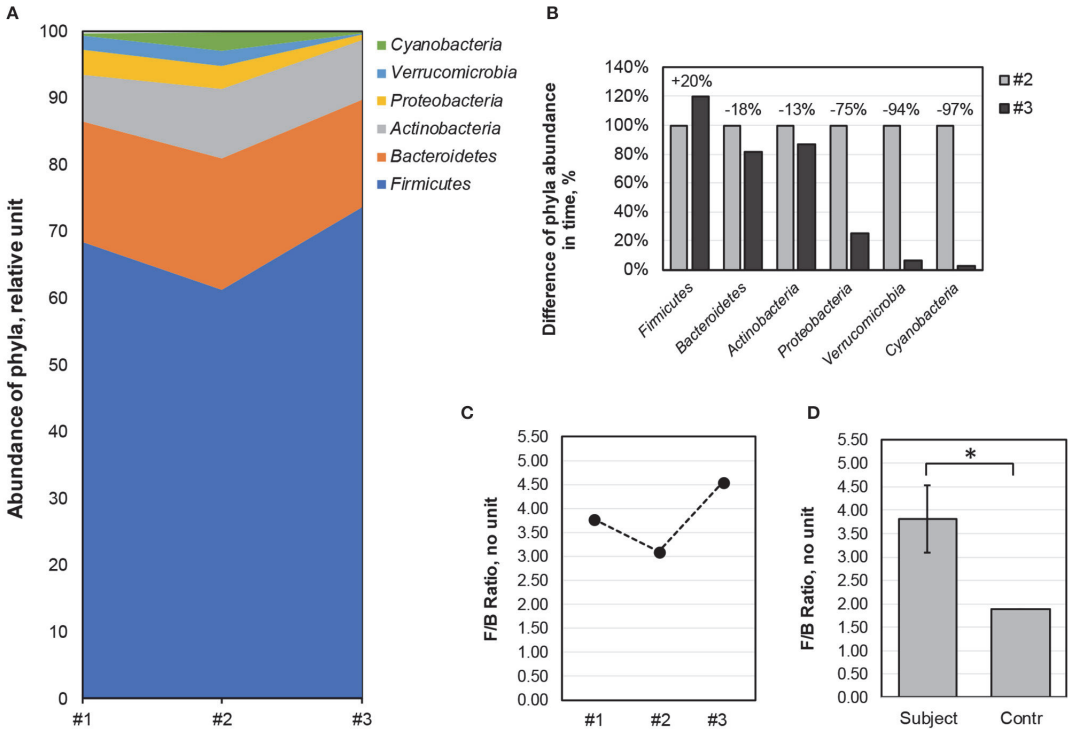


FIGURE 3 | Dynamic changes in bacterial phyla composition over the study period. **(A)** Changes in the abundance of the six most numerous phyla (*Firmicutes*, *Bacteroidetes*, *Actinobacteria*, *Proteobacteria*, *Verrucomicrobia*, and *Cyanobacteria*) in three serial samples (*x*-axis). *y*-axis—abundance of phyla, in relative units. **(B)** Proportional changes in the abundance of the six major phyla upon fiber consumption. The abundance of each phylum in sample #2 was equated to 100%, and the change in sample #3 relative to sample #2 is indicated above the columns. *y*-axis—relative abundance (%) of phylum in samples #2 and #3; *x*-axis—phyla. **(C)** Changes in *F/B* ratio over the 31-week study period, measured at three serial time points (*x*-axis). *y*-axis—*F/B* ratio. **(D)** The young athlete showed a high *F/B* ratio as compared with the control group (single mean *t*-test, p-value < 0.05, denoted by *). The bar graph represents the arithmetic mean of the *F/B* ratio in study samples and the error bars represent the standard deviation. *y*-axis—*F/B* ratio. #1, time point week 1 (baseline); #2, time point week 27 (high training and competition period); #3, time point week 31 (after 30 days of dietary supplement intake).

Firmicutes (30, 31) and that *F/B* ratio correlated significantly with cardiorespiratory fitness (32).

Opposite Dynamics of Butyrate-Producing Bacteria in Periods of Competition and Upon Dietary Fiber Intake

The relative abundance of 77 genera, of which 29 were shared across all samples, was significantly different between samples (p-values < 0.05, Wilcoxon rank sum test, **Figure 4**). At the genus level, *Prevotella* [11.7 ± 2.1 (mean abundance across three time points ± SD)], *Faecalibacter* (5.1 ± 1.8), *Blautia* (5.4 ± 1.2), *Ruminococcus* (3.8 ± 2.8), and *Bifidobacterium* (5.0 ± 3.8) were the most abundant genera (**Figure 4**). The predominance of *Prevotella* compared with the families of *Bacteroides* and *Ruminococcus* indicated that the subject had *Prevotella*-predominant enterotype, e.g., enterotype II (**Figure 4A**). *Prevotella*'s abundance was associated with long-term fiber intake (33).Similar trends were noted also in the current study whereupon dietary fiber intake resulted in enhanced abundance of *Prevotella* that became 41.7% more abundant as compared with the previous time point (**Figure 4A**). It is known that extreme dietary changes can lead to wide-ranging shifts in the gut bacterial community (34). Herein, the relative abundance of the acetate-producing bacteria (e.g., *Blautia*, *Bifidobacterium*, *Sutterella* groups) and the lactate-producing bacteria (e.g., *Bifidobacterium*, *Streptococcus*, *Lactococcus* groups) increased during high training and competition period (sample #1 vs. #2), but showed decreasing patterns (except *Blautia*) by the end of the dietary supplement intake period when acetate- and lactate-consuming and butyrate-producing genera including *Faecalibacterium* increased significantly (sample #2 vs. #3, p-value < 0.05, Wilcoxon rank sum test, **Figures 4B,C**). Interestingly, the propionate-producing genera (35) showed differential patterns.

Bacteroides and *Acidaminococcus* showed increasing trends in contrast to *Phascolarctobacterium* and *Veillonella* that decreased upon competition, with only the levels of *Veillonella* slightly rescued upon fiber consumption (**Figures 4B,C**).

The abundance of the shared largest 29 genera was better correlated in samples #1 and #2 (Spearman's correlation coefficient $r_s = 0.87$) as compared with that in samples #2 and #3 (rs = 0.76, **Figure 4D**). However, the strongest correlation in abundance of these major genera was observed between baseline and endpoint (samples #1 and #3, $r_s = 0.92$). This result was somewhat surprising. Similar to other studies (36), we also noted strong patterns of individuality of the response to exercise and diet, with a possible explanation that these activities supported the original (primary, "keystone") bacterial community dynamics of the subject.

Taken together, high exercise along with dietary fiber intake resulted in dynamic shifts in genera composition especially in the balance of lactate- and acetate/butyrate-producing bacteria.

Selective Effects of the Dietary Fiber Supplement on Individual Species of the Gut Microbiota

Next, we compared the mean relative abundance of 32 individual species to identify those that were the most affected by the dietary switch. Among the studied species, the six most abundant bacterial species (*Prevotella copri*, *Faecalibacterium prausnitzii*, *Akkermansia muciniphila*, *Bifidobacterium adolescentis*, *Coprococcus eutactus*, *Collinsella aerofaciens*) accounted for 92.5% on average of the abundance of the top species. For most of these analyzed species, a notable variation was associated both with the intense exercise loads and the dietary fiber consumption (**Figure 5**). Interestingly, *A. muciniphila* (*Verrucomicrobia*) that produces both propionate and acetate (37, 38) showed decreased proportions upon fiber consumption and was replaced by the abundance of the butyrate producer *C. aerofaciens* (**Figure 5A**). Another major shift upon fiber consumption was noted in the abundance of *C. eutactus* with known beneficial effects on butyrate production (39). However, the abundance of 12 species attributed with a protective role on the intestinal mucosa reduced significantly after dietary fiber intake (p-value < 0.05, Wilcoxon rank sum test), from on average 50.4 to 33.5% of abundance of the detected species (**Figure 5B**). Herein, two of the five species which were among the most significantly affected were the *Bacteroidetes* spp. Also, the species of *Akkermansia*, *Bacteroides*, *Bifidobacterium*, and *Ruminococcus* with protective functions on the intestinal mucosa showed a significant decrease upon fiber intake, whereas, *F. prausnitzii*, one of the major manufacturers of butyrate, showed increased abundance upon dietary fiber (**Figure 5B**). *Veillonella dispar* species were specifically monitored because of their potential impact on performance enhancement as a lactic acid-utilizing community (6). Although, changes in *V. dispar* abundance during periods of high exercise load and dietary fiber intake were detected, the statistical significance of these findings could not be determined (**Figure 5C**). It is noteworthy that the number of *V. dispar* in the microbiome of the young athlete was significantly higher than of the control group (**Figures 5C,D**, single mean t-test, p-value < 0.05). The observed changes in the abundance of bacteria producing SCFAs associated with energy consumption of the skeletal muscle (40) supported the initial work hypothesis that dietary fiber intake could facilitate athletic endurance by favorable shifts in microbial composition.

DISCUSSION

This work investigated the effects of dietary fiber supplement on the microbiome of the young rower at high-intensity exercising loads. Several consistent patterns in the gut microbiota were observed. First, the shifts induced by high exercise and dietary fibers were restricted to a limited number of phyla and genera, but were remarkable at the species level contributing to energy production. Second, the magnitude of change in microbial alpha diversity upon fiber consumption was drastic, constituting a 20.3% drop in diversity, by substantially enhancing the *Firmicutes*/*Bacteroidetes* ratio. Third, as suggested previously (41, 42) but now confirmed by this longitudinal case study, microbial response to dietary fiber consumption included the keystone species of the individual.

Our data showed that fiber consumption at high exercise loads led to decreased alpha diversity of the gut microbiota (**Figure 2**). Recent findings suggest a dynamic positive relationship between gut microbiota diversity and physical activity as professional athletes exhibit more diverse composition [ref (43)]. Paradoxically, in our study, there was a significant drop in alpha diversity upon dietary fiber consumption, most conceivably due to the rise in select advantageous bacterial species, such as those involved in butyrate production (**Figure 5**). Overall, these data support the view (44) that complex fibers of the dietary mix are highly selective for specific bacteria. It has become clear that animals can get by and often have high fitness with low-diversity microbiota (45, 46). Also, high gut microbial diversity has been linked to longer colonic transit time and systemic circulation of potentially harmful protein degradation products (47). Therefore, interpreting the health of the athlete's gut based on alpha diversity values of the microbiome implies a personal approach.

We observed a strong association between exercise loads, fiber intake, and *F*/*B* ratio (**Figure 3**), in good agreement with previous findings (48, 49). The proportional composition of the phyla clearly differentiated the subject from the matched control cohort with the abundance of *Firmicutes* and *Prevotella*. The *Prevotella* enterotype is supported by the diet of the subject in good support with earlier findings (50, 51). The dietary fiber intake has a limited influence on the communal stability of the latter two as compared with the baseline (**Figure 3**). However, dietary fiber intake resulted in enhanced abundance of *Prevotella* and *Roseburia* that became 41.7 and 4.2% more abundant, respectively, as compared with the time point of high competition (**Figure 4**). A recent *in vitro* study elucidating the mechanism of action of select dietary fibers on gut microbiota found that beta-glucan from oats induced the growth of *Prevotella* and *Roseburia* with a concomitant increase in SCFA propionate production (52).

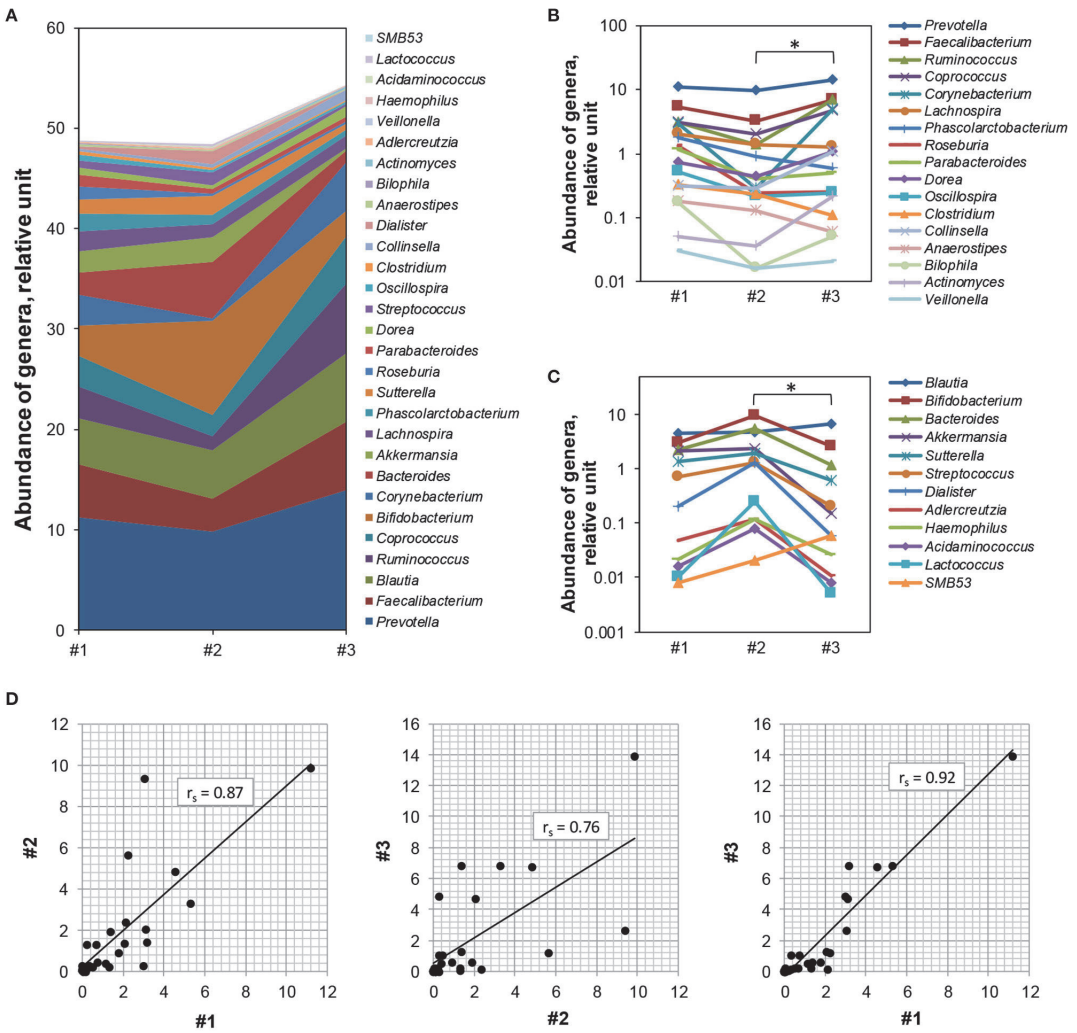


FIGURE 4 | Dynamic changes in bacterial genera composition during the study period. **(A)** The bacterial genera ($n = 29$) detected in all three samples at different time points (*x*-axis) are presented. *y*-axis—abundance of genera, in relative units. **(B,C)** Those species ($n = 17$) predominantly producing acetate had higher abundance in sample #2 compared with sample #1 but were significantly lower in sample #3 (Wilcoxon rank sum test, p-value < 0.05, denoted by *) **(B)**, whereas, butyrate-producing species were higher in sample #3 and low in samples #1 and #2 (Wilcoxon rank sum test, p-value < 0.05, denoted by *) **(C)**. *x*-axis—serial time points; *y*-axis—abundance of genera, in relative units. The number of bacterial genera is plotted in logarithmic scale. **(D)** Correlations between the differential abundance of large bacterial genera ($n = 29$) in the microbiome composition. The difference was high before vs. after the intake of fiber [sample #2 vs. #3, middle graph, Spearman's correlation coefficient $r_s = 0.76$ ($p < 0.0001$)], as compared with that at the beginning and the end of the intense training period [sample #1 vs. #2, left graph, Spearman's correlation coefficient $r_s = 0.87$ ($p < 0.0001$)], and also to that at the beginning and the end of the study [samples #1 and #3, Spearman's correlation coefficient $r_s = 0.92$ ($p < 0.0001$)]. *x*- and *y*-axes—abundance of genera, in relative units. #1, time point week 1 (baseline); #2, time point week 27 (high training and competition period); #3, time point week 31 (after 30 days of dietary supplement intake).

This study also showed that non-digestible sugars like inulin and oligosaccharides increase SCFA levels (52). Thus, our data allow concluding that non-digestible carbohydrates of the dietary fiber supplement promoted the growth of beneficial microorganisms for the performance of the athlete. In addition, there was less *Actinobacteria* (**Figure 3**), in harmony with previous studies (53). Interestingly, *P*/*A* was found to be <1 for all samples (0.5, 0.3, and 0.1), suggesting that decreased abundance of the *Proteobacteria*

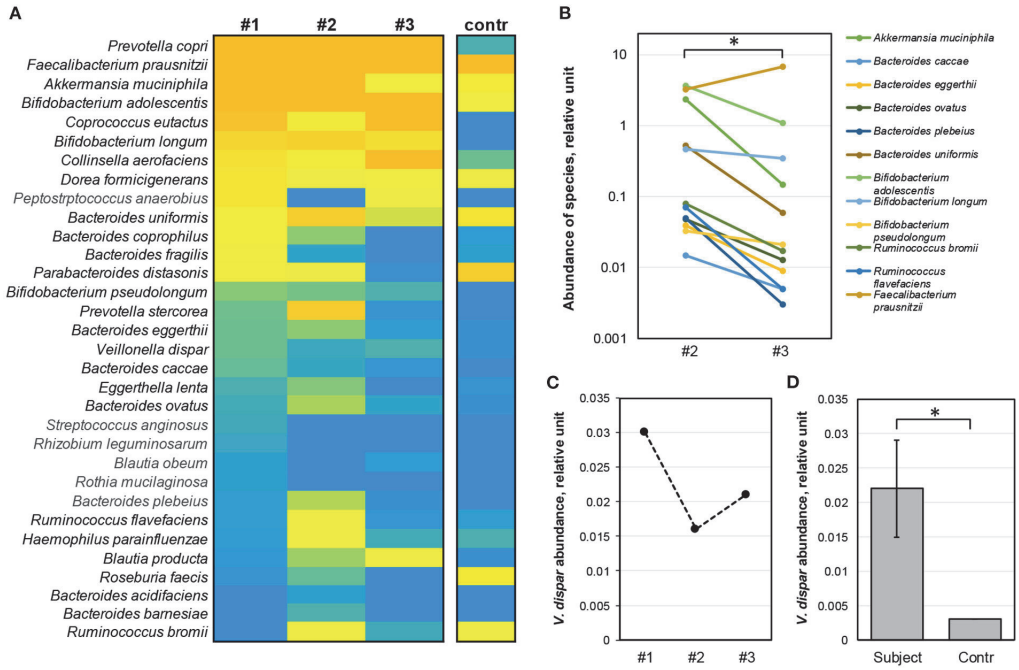


FIGURE 5 | Changes in composition of the 32 most abundant bacterial species during the study period. **(A)** Dynamic changes in the abundance of observed species in different samples. Relative abundance of specific species (*rows*) in samples (*columns*) are shown in color scale with orange depicting abundance values >1 RUs, yellow >0.07, green >0.03, and blue = 0. **(B)** Abundance of species with a protective role on the intestinal mucosa ($n = 12$) showed a significant reduction after monthly fiber intake (sample #2 vs. sample #3, Wilcoxon rank sum test, p-value < 0.05, denoted by *); *y*-axis—number of bacterial species, in relative units. **(C)** Differential changes in the level of abundance of *Veillonella dispar* (in relative units) before the competition season (#1), after the end of the intensive competition period (#2) and after the fiber intake (#3). *x*-axis—serial time points; *y*-axis—abundance, in RU. **(D)** High abundance of *V. dispar* in the microbiome under study (*Subject*) was observed as compared with the control group (p-value < 0.05, single t-test, denoted by *). Bar graph expressed as arithmetic mean of abundance of *V. dispar* in samples and error bars depicted by standard deviation. *y*-axis—abundance of *V. dispar*, in RU. #1, time point week 1 (baseline); #2, time point week 27 (high training and competition period); #3, time point week 31 (after 30 days of dietary supplement intake).

was linked to high exercise loads [see also (54)]. Of note, *Proteobacteria* are a major group behind the gut's metagenome functional variability (55).

Furthermore, we observed opposite dynamics of lactate- and acetate/butyrate-producing bacteria in periods of competition and upon dietary fiber intake, supporting the mechanism where during exercise the gut supplies lactate (56) and acetate (57) as fuel energy source (58). Though, generally fiber tends to increase SCFA-producing bacteria such as *Bacteroidetes* and *Actinobacteria* and decrease *Firmicutes* (59) as also observed in our study, whether it is a direct cause or because of changes in training and competition routine (outdoor vs. indoors rowing) along with dietary fiber intake that might have changed the metabolism needs further evaluation.

Our data show that high-endurance exercise and a prebiotic fiber-supplemented diet resulted in significant shifts across the key genera. We found that seven genera, namely *Prevotella*, *Parabacteroides* (Bacteroidetes), *Faecalibacterium*, *Ruminococcus*, *Coprococcus*, *Lachnospira* (Firmicutes), and *Corynebacterium* (Actinobacteria), were reduced upon high exercise loads, but the levels of these were restored upon dietary fiber consumption. These findings suggested that these seven genera affected primarily the levels of acetate and propionate available to the host. Both of these SCFAs are the known substrates for energy production, as well as in skeletal muscle (60, 61). In contrast, six genera, namely *Streptococcus* and *Dialister* (Firmicutes), *Bacteroides* (*Bacteroidetes*), *Bifidobacterium* (*Actinobacteria*), *Akkermansia* (*Verrucomicrobia*), and *Sutterella* (*Proteobacteria*), were specifically stimulated at high exercise loads, but inhibited by dietary fiber intake. These findings suggested higher butyrate production upon dietary fiber consumption with also potentially ameliorative effects on gut mucosal inflammation and oxidative status (62, 63). Elite athletes' dietary plans are based on the consumption of certain micronutrients, but the health of the gut microbiota is rarely considered (27). Here, we show that despite

the individual features of the microbiota composition of the athlete, exercise-driven prevalence of acetate- and propionate-producing species was flexibly switched to butyrate producers after dietary fiber intake.

Baseline bacterial composition has repeatedly been observed to be a key factor of changes in the gut microbiota following dietary interventions (17, 36). What was noticeable upon dietary fiber consumption by the athlete was the increase in taxa, such as *F. prausnitzii*, with known beneficial effects on muscle function (15). However, in contrast to the published studies, we found that while having a positive effect on bacterial families associated with athletic excellence, fiber intake had detrimental contracting effects on the overall microbial community. This was surprising although in concordance with the notions that the host might lack keystone species (64) or lack strains to utilize specific dietary fiber (65). It is expected that the microbiome reverts to its original state after short-term dietary interventions (33, 34), although, positive impacts on the gut microbiota could be maintained for at least a year (12).

Strengths

The major strength of the study was that it was conducted in "real-life" scenario as the temporal dynamics of the athlete's microbiota was explored by combining high-endurance exercise specifically with the athlete-designed dietary fiber supplement. The benefits of this study may lead to new insights into the cumulative effects a particular physiological interference has on the gut microbiome, that is on the role of the host's enterotype (with defined keystone species) has on the covariation of microbial communities upon dietary shifts and at high exercise loads. Finally, this is an individual athlete's study, and as such, it does not allow to draw solid and supportive conclusions. The value of the study lies in the aspect that overall variability in the physiological response of athletes to training and nutrition has not yet been adequately explored.

Limitations

Firstly, it was not possible to fully control dietary intake, although, the participant was instructed to maintain normal habits. Secondly, the study did not examine the causal relationship between exercise performance and the gut microbiota of the athlete, although, a high number of intestinal bacteria of the *Veillonella* family that would be of advantage to the athlete were observed. Also, recording of metrics inflammation and immunosuppression could have helped to examine the microbial gut stress levels of the athlete. Finally, given that microbiome sequencing had a limited capacity to resolve taxa to the species level, therefore, the study focused on the proportionally largest reservoir of multiple species' diversity and functionality.

CONCLUSIONS

Testing the microbiome of young athletes is necessary to obtain information on the dynamics and composition of this activity during the training process and at competition. At high levels of endurance exercise, athletes may have serious problems getting the amount of energy they need, so taking supplements to increase or to recover gut microbiota diversity at times of physical exertion is highly recommended. Our observations suggest that the dietary fiber-supplemented diet produces pronounced changes in the gut microbiota of the subject with high fractions of *Bacteriodetes* (*Prevotella*). This fact solely could be used to stratify athletes by their baseline gut bacterial composition before assigning such a fiber-supplemented diet. We evidenced that high dietary fiber intake at high exercise loads might produce profound changes beneficial to human health. Establishing the causal role of the GI microbiota and the underlying mechanisms would remain essential for the development of improved next-generation personal nutritional strategies. Only this type of in-depth understanding will allow for the selection of dietary fibers (or) mixtures thereof, to systematically target specific features of the GI microbiome (i.e., specific taxa, diversity, metabolites) with the goal of alleviating the immunometabolic features (frequently dysbiotic) that are characteristic of athletes' gut.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MJ, UT, TT, and KP collected the samples, carried out the analysis, wrote the paper, and reviewed and accepted the final version of the manuscript. KP (corresponding) is responsible for the integrity of the work as a whole. All authors have read and agreed to the published version of the manuscript.

FUNDING

This study was supported by funding grants of Protobios (5.1-4/20/170 and PRG573) from the Estonian Ministry of Education and Estonian Research Council, respectively, and H2020 PANBioRA (EU760921) projects from the European Union.

ACKNOWLEDGMENTS

We would like to thank the athlete's coach and the team of Tallinna Sõudeklubi (www.soudeklubi.ee) for their contributions to the athlete's performance. The authors gratefully acknowledge Helle Sadam, Arno Pihlak, Nadezda Pupina, and Annika Rähni of Protobios for their kind support.

REFERENCES

1. Schroeder BO, Backhed F. Signals from the gut microbiota to distant organs in physiology and disease. *Nat Med.* (2016) 22:1079–89. doi: 10.1038/nm.4185
2. Koh A, De Vadder F, Kovatcheva-Datchary P, Backhed F. From dietary fiber to host physiology: short-chain fatty acids as key bacterial metabolites. *Cell.* (2016) 165:1332–45. doi: 10.1016/j.cell.2016.05.041
3. Precup G, Vodnar DC. Gut prevotella as a possible biomarker of diet and its eubiotic versus dysbiotic roles: a comprehensive literature review. *Br J Nutr.* (2019) 122:131–40. doi: 10.1017/S0007114519000680
4. Das B, Nair GB. Homeostasis and dysbiosis of the gut microbiome in health and disease. *J Biosci.* (2019) 44:117. doi: 10.1007/s12038-019-9926-y
5. Mailing LJ, Allen JM, Buford TW, Fields CJ, Woods JA. Exercise and the gut microbiome: a review of the evidence, potential mechanisms, and implications for human health. *Exerc Sport Sci Rev.* (2019) 47:75–85. doi: 10.1249/JES.0000000000000183
6. Scheiman J, Luber JM, Chavkin TA, MacDonald T, Tung A, Pham LD, et al. Meta-omics analysis of elite athletes identifies a performance-enhancing microbe that functions *via* lactate metabolism. *Nat Med.* (2019) 25:1104–9. doi: 10.1038/s41591-019-0485-4
7. Sohail MU, Yassine HM, Sohail A, Al Thani AA. Impact of physical exercise on gut microbiome, inflammation, and the pathobiology of metabolic disorders. *Rev Diabet Stud.* (2019) 15:35–48. doi: 10.1900/RDS.2019.15.35
8. Jang LG, Choi G, Kim SW, Kim BY, Lee S, Park H. The combination of sport and sport-specific diet is associated with characteristics of gut microbiota: an observational study. *J Int Soc Sports Nutr.* (2019) 16:21. doi: 10.1186/s12970-019-0290-y
9. O'Donovan CM, Madigan SM, Garcia-Perez I, Rankin A, Sullivan OO', Cotter PD. Distinct microbiome composition and metabolome exists across subgroups of elite Irish athletes. *J Sci Med Sport.* (2020) 23:63–8. doi: 10.1016/j.jsams.2019.08.290
10. Ticinesi A, Lauretani F, Tana C, Nouvenne A, Ridolo E, Meschi T. Exercise and immune system as modulators of intestinal microbiome: implications for the gut-muscle axis hypothesis. *Exerc Immunol Rev.* (2019) 25:84–95.
11. Mach N, Clark A. Micronutrient deficiencies and the human gut microbiota. *Trends Microbiol.* (2017) 25:607–10. doi: 10.1016/j.tim.2017.06.004
12. Hughes RL. A review of the role of the gut microbiome in personalized sports nutrition. *Front Nutr.* (2019) 6:191. doi: 10.3389/fnut.2019.00191
13. Lamprecht M, Frauwallner A. Exercise, intestinal barrier dysfunction, and probiotic supplementation. *Med Sport Sci.* (2012) 59:47–56. doi: 10.1159/000342169
14. Carlson JL, Erickson JM, Lloyd BB, Slavin JL. Health effects and sources of prebiotic dietary fiber. *Curr Dev Nutr.* (2018) 2:nzy005. doi: 10.1093/cdn/nzy005
15. Deehan EC, Duar RM, Armet AM, Perez-Munoz ME, Jin M, Walter J. Modulation of the gastrointestinal microbiome with nondigestible fermentable carbohydrates to improve human health. *Microbiol Spectr.* (2017) 5:1–24. doi: 10.1128/9781555819705.ch20
16. Hamaker BR, Tuncil YE. A perspective on the complexity of dietary fiber structures and their potential effect on the gut microbiota. *J Mol Biol.* (2014) 426:3838–50. doi: 10.1016/j.jmb.2014.07.028
17. Bindels LB, Delzenne NM, Cani PD, Walter J. Towards a more comprehensive concept for prebiotics. *Nat Rev Gastroenterol Hepatol.* (2015) 12:303–10. doi: 10.1038/nrgastro.2015.47
18. Holmes E, Li JV, Marchesi JR, Nicholson JK. Gut microbiota composition and activity in relation to host metabolic phenotype and disease risk. *Cell Metab.* (2012) 16:559–64. doi: 10.1016/j.cmet.2012.10.007
19. Patterson E, Cryan JF, Fitzgerald GF, Ross RP, Dinan TG, Stanton C. Gut microbiota, the pharmabiotics they produce and host health. *Proc Nutr Soc.* (2014) 73:477–89. doi: 10.1017/S0029665114001426
20. Tremaroli V, Backhed F. Functional interactions between the gut microbiota and host metabolism. *Nature.* (2012) 489:242–9. doi: 10.1038/nature11552
21. Marchesi JR, Adams DH, Fava F, Hermes GD, Hirschfield GM, Hold G, et al. The gut microbiota and host health: a new clinical frontier. *Gut.* (2016) 65:330–9. doi: 10.1136/gutjnl-2015-309990
22. Mach N, Fuster-Botella D. Endurance exercise and gut microbiota: a review. *J Sport Health Sci.* (2017) 6:179–97. doi: 10.1016/j.jshs.2016.05.001
23. Calero CDQ, Rincon EO, Marqueta PM. Probiotics, prebiotics and synbiotics: useful for athletes and active individuals? A systematic review. *Benef Microbes.* (2020) 11:135–49. doi: 10.3920/BM2019.0076
24. Przewlocka K, Folwarski M, Kazmierczak-Siedlecka K, Skonieczna-Zydecka K, Kaczor JJ. Gut-Muscle axisexists and may affect skeletal muscle adaptation to training. *Nutrients.* (2020) 12:1451. doi: 10.3390/nu12051451
25. Moreno-Perez D, Bressa C, Bailen M, Hamed-Bousdar S, Naclerio F, Carmona M, et al. Effect of a protein supplement on the gut microbiota of endurance athletes: a randomized, controlled, double-blind pilot study. *Nutrients.* (2018) 10:337. doi: 10.3390/nu10030337
26. Karlund A, Gomez-Gallego C, Turpeinen AM, Palo-Oja OM, El-Nezami H, Kolehmainen M. Protein supplements and their relation with nutrition, microbiota composition and health: is more protein always better for sportspeople? *Nutrients.* (2019) 11:829. doi: 10.3390/nu11040829
27. Clark A, Mach N. Exercise-induced stress behavior, gut-microbiota-brain axis and diet: a systematic review for athletes. *J Int Soc Sports Nutr.* (2016) 13:43. doi: 10.1186/s12970-016-0155-6
28. Lilja S, Stoll C, Krammer U, Hippe B, Duszka K, Debebe T, et al. Five days periodic fasting elevates levels of longevity related christensenella and sirtuin expression in humans. *Int J Mol Sci.* (2021) 22:2331. doi: 10.3390/ijms22052331
29. Nasidze I, Li J, Quinque D, Tang K, Stoneking M. Global diversity in the human salivary microbiome. *Genome Res.* (2009) 19:636–43. doi: 10.1101/gr.084616.108
30. Allen JM, Mailing LJ, Niemiro GM, Moore R, Cook MD, White BA, et al. Exercise alters gut microbiota composition and function in lean and obese humans. *Med Sci Sports Exerc.* (2018) 50:747–57. doi: 10.1249/MSS.0000000000001495
31. Estaki M, Pither J, Baumeister P, Little JP, Gill SK, Ghosh S, et al. Cardiorespiratory fitness as a predictor of intestinal microbial diversity and distinct metagenomic functions. *Microbiome.* (2016) 4:42. doi: 10.1186/s40168-016-0189-7
32. Durk RP, Castillo E, Marquez-Magana L, Grosicki GJ, Bolter ND, Lee CM, et al. Gut microbiota composition is related to cardiorespiratory fitness in healthy young adults. *Int J Sport Nutr Exerc Metab.* (2019) 29:249–53. doi: 10.1123/ijsnem.2018-0024
33. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, Keilbaugh SA, et al. Linking long-term dietary patterns with gut microbial enterotypes. *Science.* (2011) 334:105–8. doi: 10.1126/science.1208344
34. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, et al. Diet rapidly and reproducibly alters the human gut microbiome. *Nature.* (2014) 505:559–63. doi: 10.1038/nature12820
35. Shimizu J, Kubota T, Takada E, Takai K, Fujiwara N, Arimitsu N, et al. Propionate-producing bacteria in the intestine may associate with skewed responses of IL10-producing regulatory T cells in patients with relapsing polychondritis. *PLoS ONE.* (2018) 13:e0203657. doi: 10.1371/journal.pone.0203657
36. Korpela K, Flint HJ, Johnstone AM, Lappi J, Poutanen K, Dewulf E, et al. Gut microbiota signatures predict host and microbiota responses to dietary interventions in obese individuals. *PLoS ONE.* (2014) 9:e90702. doi: 10.1371/journal.pone.0090702
37. Louis P. Different substrate preferences help closely related bacteria to coexist in the gut. *mBio.* (2017) 8:e01824–17. doi: 10.1128/mBio.01824-17
38. Derrien M, Vaughan EE, Plugge CM, de Vos WM. Akkermansia muciniphila gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. *Int J Syst Evol Microbiol.* (2004) 54(Pt. 5):1469–76. doi: 10.1099/ijs.0.02873-0
39. Louis P, Duncan SH, McCrae SI, Millar J, Jackson MS, Flint HJ. Restricted distribution of the butyrate kinase pathway among butyrate-producing bacteria from the human colon. *J Bacteriol.* (2004) 186:2099–106. doi: 10.1128/JB.186.7.2099-2106.2004
40. Canfora EE, Jocken JW, Blaak EE. Short-chain fatty acids in control of body weight and insulin sensitivity. *Nat Rev Endocrinol.* (2015) 11:577–91. doi: 10.1038/nrendo.2015.128
41. Davis LM, Martinez I, Walter J, Goin C, Hutkins RW. Barcoded pyrosequencing reveals that consumption of galactooligosaccharides results in a highly specific bifidogenic response in humans. *PLoS ONE.* (2011) 6:e25200. doi: 10.1371/journal.pone.0025200

42. Martinez I, Kim J, Duffy PR, Schlegel VL, Walter J. Resistant starches types 2 and 4 have differential effects on the composition of the fecal microbiota in human subjects. *PLoS ONE.* (2010) 5:e15046. doi: 10.1371/journal.pone.0015046
43. Mohr AE, Jager R, Carpenter KC, Kerksick CM, Purpura M, Townsend JR, et al. The athletic gut microbiota. *J Int Soc Sports Nutr.* (2020) 17:24. doi: 10.1186/s12970-020-00353-w
44. Cantu-Jungles TM, Hamaker BR. New view on dietary fiber selection for predictable shifts in gut microbiota. *mBio.* (2020) 11:e02179–19. doi: 10.1128/mBio.02179-19
45. Foster JA, Rinaman L, Cryan JF. Stress & the gut-brain axis: regulation by the microbiome. *Neurobiol Stress.* (2017) 7:124–36. doi: 10.1016/j.ynstr.2017.03.001
46. Reese AT, Cho EH, Klitzman B, Nichols SP, Wisniewski NA, Villa MM, et al. Antibiotic-induced changes in the microbiota disrupt redox dynamics in the gut. *Elife.* (2018) 7:e35987. doi: 10.7554/eLife.35987
47. Roager HM, Hansen LB, Bahl MI, Frandsen HL, Carvalho V, Gobel RJ, et al. Colonic transit time is related to bacterial metabolism and mucosal turnover in the gut. *Nat Microbiol.* (2016) 1:16093. doi: 10.1038/nmicrobiol.2016.93
48. Choi JJ, Eum SY, Rampersaud E, Daunert S, Abreu MT, Toborek M. Exercise attenuates PCB-induced changes in the mouse gut microbiome. *Environ Health Perspect.* (2013) 121:725–30. doi: 10.1289/ehp.1306534
49. Queipo-Ortuno MI, Seoane LM, Murri M, Pardo M, Gomez-Zumaquero JM, Cardona F, et al. Gut microbiota composition in male rat models under different nutritional status and physical activity and its association with serum leptin and ghrelin levels. *PLoS ONE.* (2013) 8:e65465. doi: 10.1371/journal.pone.0065465
50. Tomova A, Bukovsky I, Rembert E, Yonas W, Alwarith J, Barnard ND, et al. The effects of vegetarian and vegan diets on gut microbiota. *Front Nutr.* (2019) 6:47. doi: 10.3389/fnut.2019.00047
51. Hjorth MF, Blaedel T, Bendtsen LQ, Lorenzen JK, Holm JB, Kiilerich P, et al. Prevotella-to-Bacteroides ratio predicts body weight and fat loss success on 24-week diets varying in macronutrient composition and dietary fiber: results from a post-hoc analysis. *Int J Obes.* (2019) 43:149–57. doi: 10.1038/s41366-018-0093-2
52. Fehlbaum S, Prudence K, Kieboom J, Heerikhuisen M, van den Broek T, Schuren FHJ, et al. *In vitro* fermentation of selected prebiotics and their effects on the composition and activity of the adult gut microbiota. *Int J Mol Sci.* (2018) 19:3097. doi: 10.3390/ijms19103097
53. Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, et al. Diversity of the human intestinal microbial flora. *Science.* (2005) 308:1635–8. doi: 10.1126/science.1110591
54. Mobeen F, Sharma V, Tulika P. Enterotype variations of the healthy human gut microbiome in different geographical regions. *Bioinformation.* (2018) 14:560–73. doi: 10.6026/97320630014560
55. Bradley PH, Pollard KS. Proteobacteria explain significant functional variability in the human gut microbiome. *Microbiome.* (2017) 5:36. doi: 10.1186/s40168-017-0244-z
56. Brooks GA. Lactate as a fulcrum of metabolism. *Redox Biol.* (2020) 35:101454. doi: 10.1016/j.redox.2020.101454
57. Okamoto T, Morino K, Ugi S, Nakagawa F, Lemecha M, Ida S, et al. Microbiome potentiates endurance exercise through intestinal acetate production. *Am J Physiol Endocrinol Metab.* (2019) 316:E956–66. doi: 10.1152/ajpendo.00510.2018
58. Bergman BC, Horning MA, Casazza GA, Wolfel EE, Butterfield GE, Brooks GA. Endurance training increases gluconeogenesis during rest and exercise in men. *Am J Physiol Endocrinol Metab.* (2000) 278:E244–51. doi: 10.1152/ajpendo.2000.278.2.E244
59. Sharma P, Bhandari C, Kumar S, Sharma B, Bhadwal P, Agnihotri N. Dietary fibers: a way to a healthy microbiome. *Diet Microbiome Health.* (2018) 6:299–345. doi: 10.1016/B978-0-12-811440-7.00011-9
60. Samuel BS, Shaito A, Motoike T, Rey FE, Backhed F, Manchester JK, et al. Effects of the gut microbiota on host adiposity are modulated by the short-chain fatty-acid binding G protein-coupled receptor, Gpr41. *Proc Natl Acad Sci U S A.* (2008) 105:16767–72. doi: 10.1073/pnas.0808567105
61. Wong JM, de Souza R, Kendall CW, Emam A, Jenkins DJ. Colonic health: fermentation and short chain fatty acids. *J Clin Gastroenterol.* (2006) 40:235–43. doi: 10.1097/00004836-200603000-00015
62. Clark A, Mach N. The crosstalk between the gut microbiota and mitochondria during exercise. *Front Physiol.* (2017) 8:319. doi: 10.3389/fphys.2017.00319
63. Donohoe DR, Collins LB, Wali A, Bigler R, Sun W, Bultman SJ. The Warburg effect dictates the mechanism of butyrate-mediated histone acetylation and cell proliferation. *Mol Cell.* (2012) 48:612–26. doi: 10.1016/j.molcel.2012.08.033
64. Ze X, Duncan SH, Louis P, Flint HJ. Ruminococcus bromii is a keystone species for the degradation of resistant starch in the human colon. *ISME J.* (2012) 6:1535–43. doi: 10.1038/ismej.2012.4
65. Zhao L, Zhang H, O'Gorman EJ, Tian W, Ma A, Moore JC, et al. Weighting and indirect effects identify keystone species in food webs. *Ecol Lett.* (2016) 19:1032–40. doi: 10.1111/ele.12638

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Appendix 2

PUBLICATION II

Mariliis Jaago, Pupina, N., Rähni, A., Pihlak, A., Sadam, H., Vrana, N. E., Sinisalo, J., Pussinen, P., Palm, K. 2022. **Antibody response to oral biofilm is a biomarker for acute coronary syndrome in periodontal disease.** *Communications Biology*, 5 (1): 205, ARTN 205. doi: 10.1038/s42003-022-03122-4.

communications biology

ARTICLE

https://doi.org/10.1038/s42003-022-03122-4 OPEN

Antibody response to oral biofilm is a biomarker for acute coronary syndrome in periodontal disease

Mariliis Jaago[1,2], Nadežda Pupina[1], Annika Rähni[1,2], Arno Pihlak[1], Helle Sadam[1,2], Nihal Engin Vrana[3], Juha Sinisalo[4], Pirkko Pussinen[5] & Kaia Palm[1,2]✉

Cumulative evidence over the last decades have supported the role of gum infections as a risk for future major cardiovascular events. The precise mechanism connecting coronary artery disease (CAD) with periodontal findings has remained elusive. Here, we employ next generation phage display mimotope-variation analysis (MVA) to identify the features of dysfunctional immune system that associate CAD with periodontitis. We identify a fine molecular description of the antigenic epitope repertoires of CAD and its most severe form - acute coronary syndrome (ACS) by profiling the antibody reactivity in a patient cohort with invasive heart examination and complete clinical oral assessment. Specifically, we identify a strong immune response to an EBV VP26 epitope mimicking multiple antigens of oral biofilm as a biomarker for the no-CAD group. With a 2-step biomarker test, we stratify subjects with periodontitis from healthy controls (balanced accuracy 84%), and then assess the risk for ACS with sensitivity 71–89% and specificity 67–100%, depending on the oral health status. Our findings highlight the importance of resolving the immune mechanisms related to severe heart conditions such as ACS in the background of oral health. Prospective validation of these findings will support incorporation of these non-invasive biomarkers into clinical practice.

[1] Protobios Llc, Mäealuse 4, 12618 Tallinn, Estonia. [2] Department of Chemistry and Biotechnology, Tallinn University of Technology, Akadeemia tee 15, 12618 Tallinn, Estonia. [3] Spartha Medical, 14B Rue de la Canardiere, 67100 Strasbourg, France. [4] Heart and Lung Center, Helsinki University Hospital, and Helsinki University, Helsinki, Finland. [5] Oral and Maxillofacial Diseases, University of Helsinki, FI-00014 Helsinki, Finland. ✉email: kaia@protobios.com

Coronary artery disease (CAD) is the leading cause of morbidity and mortality worldwide[1] caused by metabolic disorders in lipid oxidation promoting inflammatory alterations on the endothelium[2,3] and culminating in plaque rupture[4,5]. The heritability of CAD and its familial clustering are well established[6]. Genome-wide association studies (GWAS) have identified a number of causal CAD-associated genes and loci[7]. These findings highlight the largely polygenic nature of the inheritability of CAD[8,9], rendering some individuals more susceptible or resilient to developing atherosclerosis[10]. CAD has been associated with the unhealthy lifestyle placing it among "immunoinflammatory" diseases[11]. While remarkable progress has been made in understanding the mechanisms of atherogenesis as robust methods of identifying high-risk atherosclerosis via genomics and imaging are at hand, highly sensitive and specific biomarkers for CAD have remained elusive. Importantly, as much as 30% of control populations are thought to unknowingly include subjects with CAD, impacting power and accuracy of clinical biomarker studies[12–14].

Periodontitis, a major oral dysbiosis-driven inflammatory disease, is associated with increased risk of atherosclerotic cardiovascular diseases[15]. Up to 700 bacterial species have been identified in the oral cavity (Human Microbiome Project Consortium). Intriguingly, DNA of periodontal pathogens (e.g., *Porphyromonas gingivalis*) and live bacteria have been detected in atherosclerotic lesions[16–18]. The microbial composition of gut microflora of patients with CAD has been found to be more inflammatory than in healthy patients[19]. Similarly, the oral microbiome of CAD patients may be altered[20]. Host-microbe interaction in the periodontium can initiate or even aggravate atherosclerotic processes through the activation of innate immunity, bacteremia, and direct involvement of cytokines and inflammatory proteins of oral microbiota[21–23].

Recent research suggests that abnormal changes to the gut microbiota flora may also contribute extensively to the progression of CAD[24]. As the microbiome plays a central role in the balance between immune activation and immune tolerance[25] and in the light of dysbiosis in microflora, it is no surprise that atherosclerosis has a strong autoimmune component[26]. First, CAD risk locus includes the major histocompatibility complex (MHC) containing a dense cluster of genes involved in inflammation, immunity, and self-recognition[21,27]. Furthermore, a depletion of T or B cells leads to an attenuation of atherosclerosis[22]. Although T cells seem to be essential, B cells and antibodies play an accelerating and perpetuating role[23]. Thus, atherosclerosis is a chronic inflammatory disease with an autoimmune component[26]. Antibodies against oxidized low-density lipoproteins (oxLDL) positively correlated with the disease[28]. Besides oxLDL/ApoB, heat shock proteins (HSPs) and some foreign peptides from pathogens including cytomegalovirus (CMV), hepatitis C virus (HCV), HIV, human papillomavirus (HPV), and others have been proposed as atherosclerosis-relevant antigens[29]. However, the relation between antibodies and atherosclerotic disease burden and progression has remained unclear.

Here we used MVA[30,31], an unbiased, high-throughput, comprehensive approach based on next-generation phage display, to identify biomarkers of periodontal conditions associated with stable coronary artery disease and progression to acute coronary syndrome.

Results

Shared immunoreactivity to epitopes linked to periodontal pathogens. We used MVA immunoprofiling analysis of sera samples of 96 individuals from the Corogene cohort[32] to identify peptide antigens related to antibody immune response in periodontal disease and CAD. Characteristics of the subjects according to their CAD and periodontal status are presented in Fig. 1 (Fig. 1a, Supplementary Table S1). Within the cohort, the proportion of ex- or active smokers was significantly higher in patients with periodontitis than in the rest of the cohort (Chi^2, $p < 0.01$, Supplementary Table S1, Fig. 1b, Supplementary Fig. S2).

Altogether we identified 14.5 million distinct peptide epitopes from the MVA immunoprofiles across the whole study cohort and converged these to 8088 most abundant and shared antigenic epitopes by clustering analysis (Supplementary Fig. S3). Given the gum disease background of the samples[33,34], we first examined whether we could detect from the immunoprofiles of the study samples the immunoreactivity to the 7 most common periodontal pathogens (*Porphyromonas gingivalis*, *Tannerella forsythia*, *Prevotella intermedia*, *Fusobacterium nucleatum*, *Campylobacter rectus*, *Aggregatibacter actinomycetemcomitans*, and *Porphyromonas endodontalis*). Data analysis showed that antibody response

a

CAD diagnosis	no-CAD			s-CAD			ACS		
periodontal disease	H	G	P	H	G	P	H	G	P
N	7	9	16	7	9	16	7	9	16
Total	*32*			*32*			*32*		

b

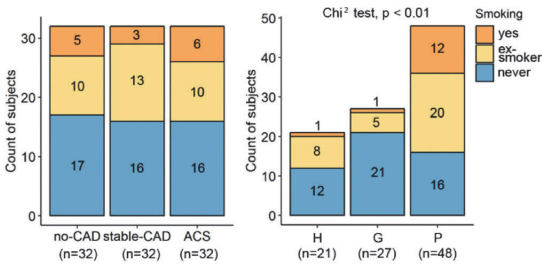
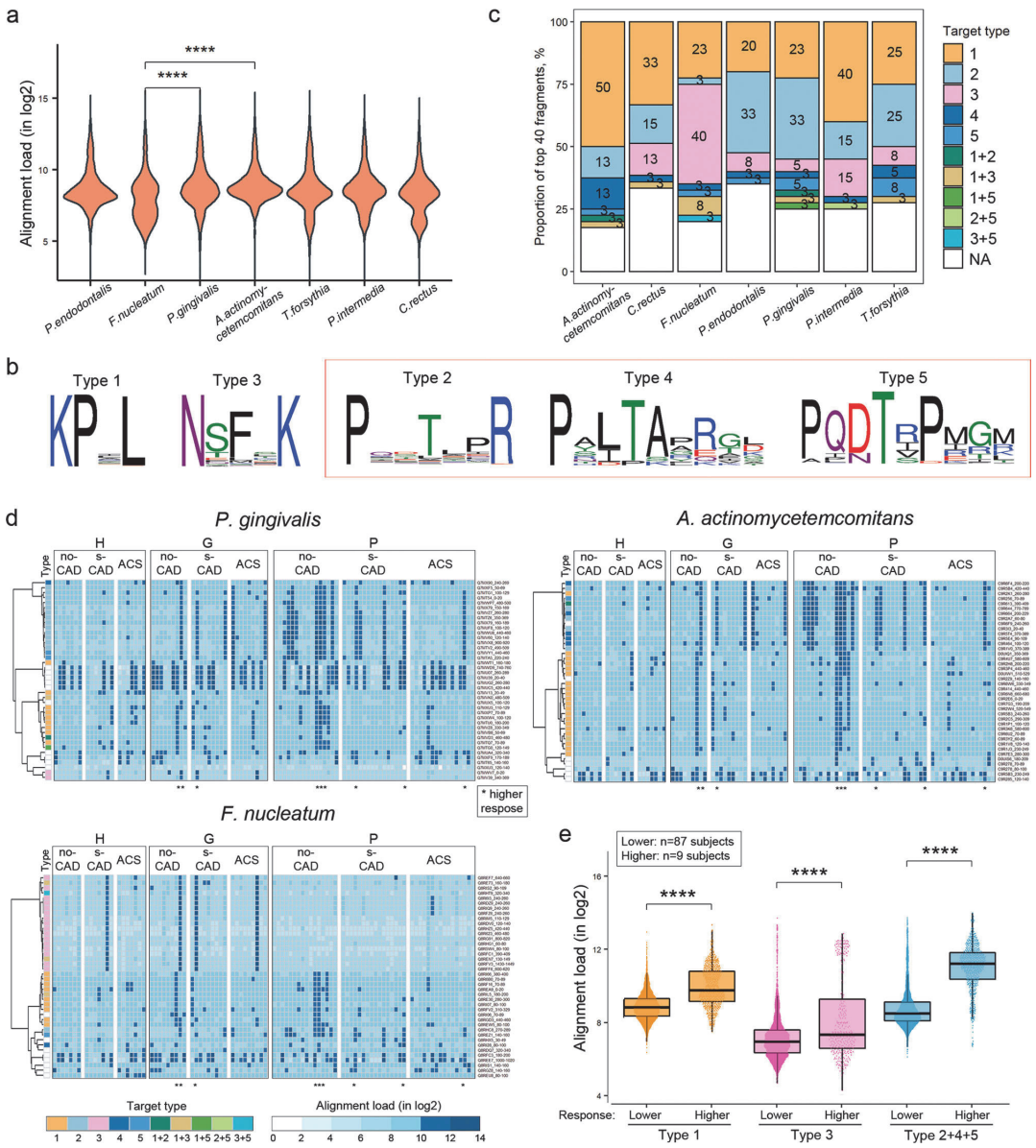


Fig. 1 Characteristics of the clinical cohort. a Subjects divided in three groups based on CAD diagnosis (no-CAD, stable-CAD (s-CAD), or ACS). *Total*, size of the group. Each of the three CAD groups included subjects with different periodontal health diagnosis: periodontally healthy (H), patients with gingivitis (G) – a transient gum inflammation –, or patients with periodontitis (P). *N*, size of the sub-group. **b** Significant association between periodontal diagnosis and smoking is observed in the study cohort. Frequency distribution graphs of subjects ($n = 96$) in groups with coronary artery health condition (no-CAD, stable-CAD, or ACS) or periodontal condition (H, G, P). Statistically significant difference in prevalence of cigarette smoking was observed in periodontal condition groups, being highest among the periodontitis group subjects, but not within CAD groups (Chi^2 test, p value < 0.05, $n = 96$ independent subjects). *x-axes – clinical subgroups; y-axes – number of subjects; color-fill – yes*, active cigarette smoker *(orange); ex-smoker*, has quit cigarette smoking *(yellow); never*, no history of cigarette smoking *(blue)*.



to antigens of *P. gingivalis* and *A. actinomycetemcomitans* was the highest, whereas *F. nucleatum* antigens were on average significantly less recognized (Mann–Whitney *U*, $^{****}p < 0.0001$, two-sided) (Fig. 2a). By analyzing the sequences of the bacterial antigens, several dominant core epitopes shared by these antigens were identified, among which multiple types were derivatives of a common signature P.T.[P][R] (*Types 2, 4 and 5*), Fig. 2b, Supplementary Table S2). Interestingly, P..T.[P][R] patterns (i.e., Type 2, 4, or 5) were present in 45% and 30% of *P. gingivalis* and *A. actinomycetemcomitans* immunodominant antigens, respectively, whereas only 10% among *F. nucleatum* antigens (Fig. 2c). Shared antibody response patterns to bacterial antigen epitopes (*target types*, *clustering on left*) were observed as specific for certain clinical heart and dental conditions (*grouping on top*) (Fig. 2d, Supplementary Fig. S4). Notably, the majority of subjects (5/9) with high immune reactivity to these pathogenic periodontal bacteria (*indicated with dots under intensity plot*) belonged to the no-CAD group (Fig. 2d, Supplementary Fig. S5). Overall, in subjects ($n = 9$) with the highest immunoreactivity to these pathogens, P..T.[P][R] containing-epitopes (Types 2, 4, and 5) were the most prevalent and on average recognized at higher levels as compared to epitopes of Type 1 and 3 (Fig. 2e). Altogether, these data showed that immune response against a dominant core epitope P..T.[P][R], which was associated with common pathogenic oral bacteria, correlated with a differential risk of ACS.

Fig. 2 Shared immunoreactivity to epitopes linked to periodontal pathogens detected by MVA from immunoprofiles of the study cohort. Group-wide top 8088 epitopes from the MVA immunoprofiles were aligned onto proteomes of the most common periodontal pathogens ($n = 7$). Based on how many epitopes aligned and how enriched were the epitopes in MVA immunoprofiles for individual sample separately, top 40 fragments with highest alignment loads were selected per pathogen. **a** Antibody response to top 40 antigens for *P. gingivalis* and *A. actinomycetemcomitans* was the highest, whereas antibody response to *F. nucleatum* antigens was found to be low. Pair-wise Mann–Whitney *U*, two-sided, $^{****}p < 0.0001$, *p* values not adjusted for multiple comparisons. *y-axis* – alignment load, representing how many MVA immunoprofile epitopes aligned onto protein sequences and how abundantly were they seen in MVA immunoprofiles (in log2). **b** Unsupervised clustering identified most abundant epitopes with consensus sequences from alignments on target fragments. *Type 1*: KP.L in 1033 fragments; *type 2*: P..T.[P]R in 833 fragments; *type 3*: N[ST]F.K in 421 fragments; *type 4*: P[AYS][LI]TA.[REQ][GT][LDK] in 150 fragments; *type 5*: PQ[DN]T[RIV]P[MIR][GRT][MRK] in 107 fragments. *Outline* – type 2, 4, and 5 epitopes share similar core pattern P..T.[P][R]. **c** Proportions of antigenic epitope types across the antigens of the seven periodontal pathogens. *y-axis* – proportion (%) among top 40 antigen fragments (cumulative), *data labels on bars* - proportion (%), *x-axis* – oral pathogen species, *target type (fill color)* – epitope pattern type (from sequences in **b**). **d** Shared antibody response patterns to bacterial antigen epitopes (*target types, clustering on left*) were observed as specific for certain clinical heart and dental conditions (*grouping on top*). Immunoreactivity profiles against top 40 antigens in *F. nucleatum*, *P. gingivalis*, and *A. actinomycetemcomitans* are shown. Profiles were clustered row-wise according to epitope types (*type*, sequences in **b**). Subjects (in *lanes*) with highest immunoreactivity to top antigens (Supplementary Fig. S5) are marked with asterisks (*) under intensity plots. *Vertical lanes* – individual samples ($n = 96$), categorized based on CAD (no-CAD, s-CAD, ACS) and periodontal diagnoses (periodontally healthy (H), gingivitis (G), periodontitis (P)); *rows* – each row represents a distinct 20-aa antigenic region of protein primary sequence; *blue color-scale* – intensity of blue represents the alignment load of individual sample; *target type* – epitope type (sequences in **b**). **e** Distribution of subjects with low ($n = 87$) or high ($n = 9$) immune response (by Supplementary Fig. S5) to different targeted epitopes. Taken together, Types 2, 4, and 5 were more common to high-response subjects, as compared to other target types. Pair-wise Mann–Whitney *U*, two-sided, $^{****}p < 0.0001$, *p* values not adjusted for multiple comparisons. *y-axis* – alignment load in log2, *response* – high-response subjects ($n = 9$) or low-response subjects ($n = 87$), *type* – target epitope *type* (sequences in **b**).

Different clinical groups share common features in the immunoprofiles. Given the findings that the immune response against oral pathogenic bacteria was associated with the clinical diagnosis, we next analyzed the difference of immune response to each of the 8088 epitopes in CAD or periodontal disease groups using ROC analysis with specific criteria: sensitivity ≥ 50% and specificity ≥ 70%, Kruskal–Wallis test $p < 0.05$. Clustered analysis based on sequence similarity arrived at 62 group-differentially targeted epitope clusters with shared core patterns (Supplementary Fig. S6, Supplementary Table S3). When we correlated the 62 epitope clusters based on average abundances in clinical diagnosis groups, clusters with similar response patterns were seen grouping together (Fig. 3a). Pearson correlation-based analysis united the 62 clusters further into five major epitopes (A to E), where the largest epitope A shared the core pattern P..T.PR (Fig. 3b, c, Supplementary Fig. S7). The epitope pattern P..T.PR includes P..T.[P][R], the one also observed as predominant among periodontal pathogens (Fig. 2b, Supplementary Table S2). We found that differential antibody response patterns against epitopes A to E were diagnosis group-specific (Fig. 3a, b, Supplementary Figs. S7 and S8). Specifically, stronger response to epitope A was specific to periodontitis and no-CAD cohorts (*red outline*), whereas stronger response to epitope B was detected in periodontitis and smoking subgroups (*green outline*) (Fig. 3a, b, Supplementary Fig. S8a, b). Stronger antibody response to epitope C, on the other hand, was characteristic to periodontally healthy patients but with an ACS diagnosis (*blue outline*), epitope D was more targeted in subjects with gingivitis (*yellow outline*), and epitope E in subjects with gingivitis but not in CAD (*pink outline*) (Fig. 3a, b, Supplementary Fig. S8c–e). In conclusion, these five major epitopes (A-E) were targeted by the strongest and differential antibody response across diagnosis groups.

Microbial mimicry of the P..T.PR core epitope that is common to periodontitis encompasses the highly antigenic epitope of EBV VP26. Our analyses highlighted a strong response to epitope A with the core pattern P..T.PR (Figs. 2d, 3c), which we have previously mapped to EBV VP26 protein encompassing 153-176aa[30,31]. When aligning all individual MVA immunoprofile-derived peptides of the current study cohort to the primary sequence of EBV VP26, we observed that subjects within the periodontitis group exhibited high immunoreactivity to this C-terminal epitope (Mann–Whitney *U*, $^{*}p < 0.05$, two-sided, Fig. 4a, b). Independent validation was performed using dot ELISA analysis, where sera samples from the current clinical cohort were exposed to phage particles displaying the C-terminal VP26 epitope sequence (Fig. 4c). Patients predicted as seropositive against the P..T.PR epitope by MVA (MVA+) were observed with significantly higher immunoreactivity to the displayed EBV VP26 epitope in the dot ELISA analysis (Fig. 4c, Mann–Whitney *U*, $^{****}p < 0.0001$, two-sided). Therefore, these results further confirmed the MVA findings of specific seroreactivity mapping of epitope A (with the core pattern P..T.PR) to EBV VP26 (Fig. 4c). As we found that many antigens of the periodontal bacteria shared features of the epitope A (Fig. 2b, d, Supplementary Table S2) and could thus mimic the epitope of EBV VP26 antigen, we determined using annotation analysis that these could include a transmembrane protein signal peptidase I (100-129 aa, Uniprot accesssion Q7MTG1) and a cytosolic transcription termination factor Rho (160-189 aa, Q7MX79) of *P. gingivalis*, and isoleucine-tRNA ligase (770-789aa, C9R644) in *A. actinomycetemcomitans* (Supplementary Table S2). Also, epitope B was annotated to the tandem repeat (3 × 13 aa) in the C-terminal part (741-779 aa) of EBV nuclear antigen 6 (EBNA6) protein (Supplementary Fig. S9), in harmony with other studies reporting this region as highly immunogenic[35,36]. In conclusion, strong antibody response to EBV, in particular to VP26 and sequence-mimicking bacterial antigens, was distinguishing subjects with periodontitis from periodontally healthy controls.

Biomarkers to predict ACS risk from periodontal disease. To examine whether the delineated five epitopes (A-E) could act as biomarkers stratifying periodontitis and/or CAD conditions, multi-variable models were built and fitted using antibody response to epitopes as predictive biomarkers. Firstly, it was confirmed that immune response to epitopes A-E had no significant correlations with age or gender (Supplementary Fig. S10). The optimal model including immune response to epitopes A, B, and C to differentiate periodontitis from healthy showed a balanced accuracy of 81% for the training subset (80% of samples) and 84% for the validation subset (20% of samples) (Fig. 5a and Supplementary Fig. S11). As a whole, strong response to epitopes A and B was characteristic to the periodontitis group, whereas response to epitope C was more common

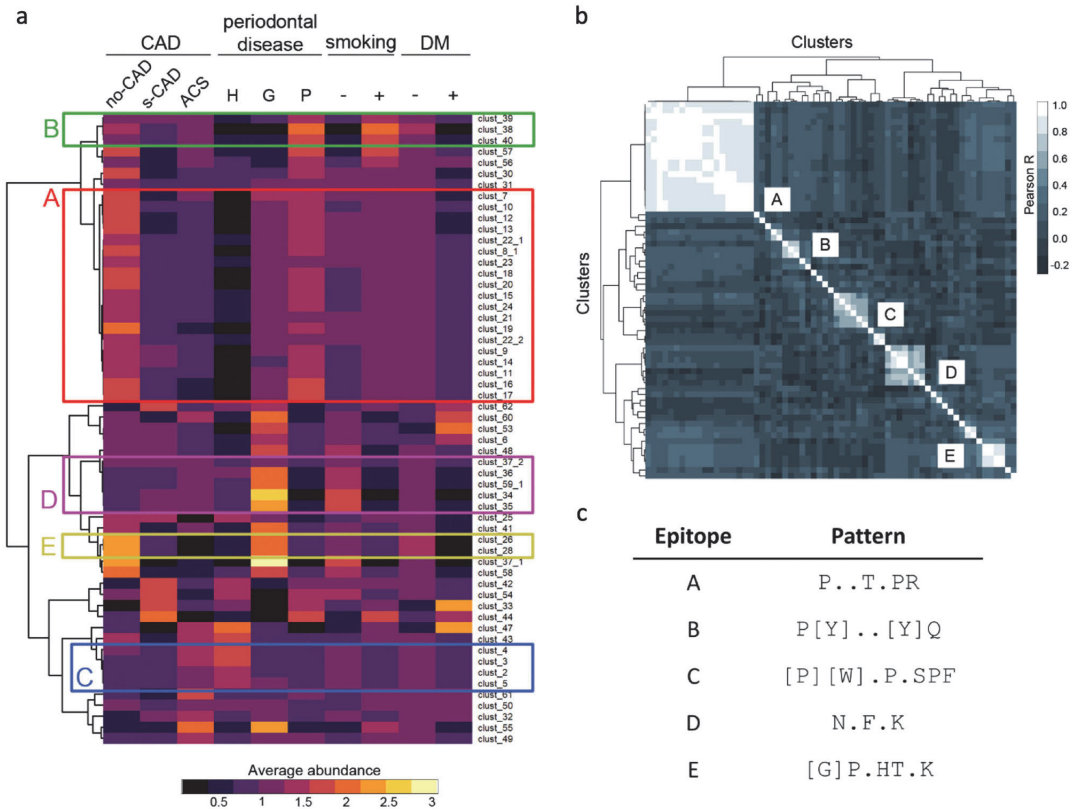


Fig. 3 MVA immunoprofiles of subjects with different CAD conditions, periodontal disease severity, cigarette smoking behavior, and diabetes diagnosis. a 62 epitope clusters with group-specific MVA immunoprofile features across clinical classifiers. Although 19% of the subjects were diagnosed with diabetes (either type I or type II, not specified further), no significant association of diabetes with either CAD or periodontal diagnosis was found (Supplementary Fig. S2). *Average abundance*, calculated as mean of peptides containing the epitope within a given group and normalized with the mean values across all groups (color-coded from purple to yellow). Colored outlines with capital letters refer to epitopes in panel **c**. Clustering distance: Pearson correlation coefficient; clustering method: ward.D2. 57/62 clusters with high mean abundance (>150) across groups are shown. "−" under smoking designates subjects without any exposure to smoking, "+" depicts subjects with a history of exposure (ex-smoker) or currently actively smoking. CAD - coronary artery disease, no-CAD - no CAD diagnosis, s-CAD - stable CAD diagnosis, ACS - acute coronary syndrome, H -periodontally healthy controls, G - patients with gingivitis diagnosis, P - patients with periodontitis diagnosis, DM - diabetes mellitus. **b** Similar behavior-based clustering of 62 epitope clusters (*x- and y-axis*) using peptide abundance values (in log10) across the study cohort. Pearson R correlation indices (*color-scale*) were calculated and visualized in a correlation matrix. Five distinct large clusters were identified and defined as epitopes A to E. **c** Core consensus patterns of epitopes A to E, identified in Supplementary Fig. S7.

in the periodontally healthy group (Fig. 5a). Response to epitopes D and E did not provide additional useful information when discriminating between periodontitis diagnosis groups (Supplementary Fig. S12a).

Next, by characterizing immune features that were CAD-specific, we discovered that periodontally healthy subjects shared a strong response against epitope C that was high in ACS subgroup (with statistically significant trends, Mann–Whitney U, $p = 0.097$, two-sided, Fig. 5b). In subjects with gingivitis, a transient inflammation condition, response against epitope A was differentiating between CAD groups, being high in ACS as compared to no-CAD (Mann–Whitney U, $^{*}p < 0.05$, two-sided, Fig. 5b). In the periodontitis group of subjects with chronic periodontal inflammation, antibody response to two independent markers (epitopes A and E) was identified as significant for the no-CAD cohort (Mann–Whitney U, $^{*}p < 0.05$, $^{**}p < 0.01$, two-sided, Fig. 5b). Response to epitopes B and D was not differentially linked to no-CAD or ACS diagnosis (Supplementary Fig. S12b). ROC analysis was used to set thresholds for each of the 3 epitopes (A, C and E) in discriminating between clinical subgroups (Fig. 5c). As a result, when first classifying subjects based on their periodontal findings (epitopes A, B and C) and then assessing response to epitopes A, C, and E which differentiate between CAD diagnoses, it was possible to predict the ACS risk in periodontal disease and ascertain the no-CAD phenotype (Fig. 5c). In conclusion, MVA immunoprofiling provided useful blood-biomarkers to predict ACS risk from periodontal disease.

Discussion

Here, we report the detailed antibody epitope delineation study identifying a set of immunogenic features associated with periodontal pathogens and CAD.

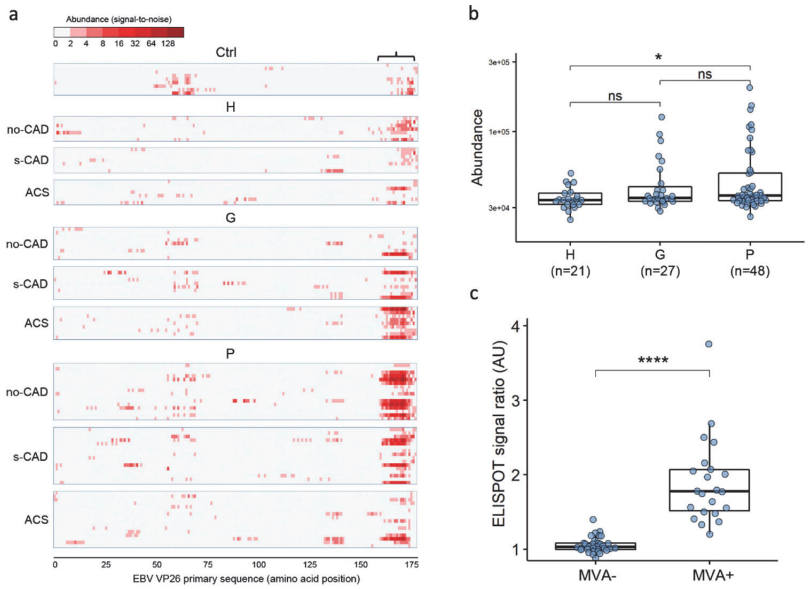
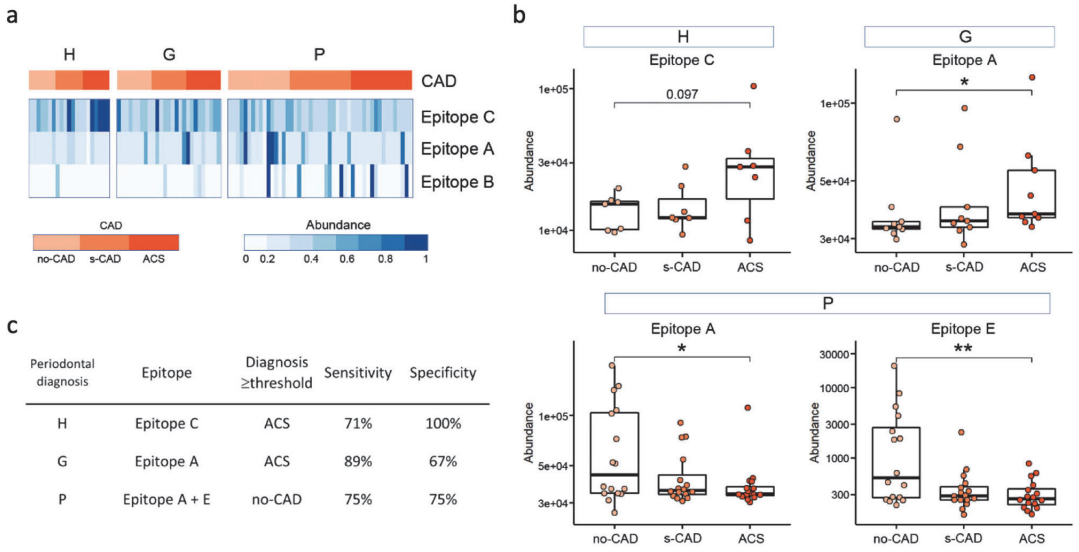


Fig. 4 Immunoreactivity to epitope on VP26 EBV and alike mimicking features stratifies patients with periodontitis. a Individual immunoreactivity profiles ($n = 96$) of the study cohort against EBV protein VP26, shown for different clinical groups: periodontally healthy (H), gingivitis (G), or periodontitis (P). All peptides from individual immunoprofiles were aligned (with ≥ 6 matching amino acid positions) to primary sequence of EBV VP26 protein along with random reference. Peptide epitope abundance (as signal-to-noise ratio, *red color-scale*) is visualized per subject (*in rows*, separated into periodontal diagnosis groups) and per amino acid position (on *x-axis*). *Ctrl* – samples with negative EBV-viral capsid antigen serology ($n = 9$). **b** Highly antigenic P..T.PR epitope is differentially targeted across periodontal groups shown as box plots of abundance of peptides containing the epitope (Mann–Whitney *U*, two-sided, *p* values not corrected for multiple comparisons, $^{*}p < 0.05$, ns $p > 0.05$, $n = 96$ independent patients), Supplementary File S6. **a**, **b** Abbreviations and group sizes: H periodontally healthy ($n = 21$ patients (3×7)), G gingivitis ($n = 27$ (3×9)), P periodontitis ($n = 48$ (3×16)), no-CAD no CAD diagnosis ($n = 32$), s-CAD stable CAD ($n = 32$), ACS acute coronary syndrome ($n = 32$). **c** Anti-P.DT.PR epitope-like immune response identified by MVA was in high correlation with data from independent dot ELISA validations. Phage particles displaying peptides with P.DT.PR or mutant P.DA.PR sequences were used for dot ELISA. Of the 52 randomly tested samples, those with a positive signal to anti-P.DT.PR response detected by MVA (>1200 abundance of P.DT.PR containing peptides, *MVA+*) showed significantly higher signals in dot ELISA (*y-axis, signal-to-background arbitrary unit (AU)*) compared to samples with a negative response from MVA data (*MVA-*, Mann–Whitney *U* test, two-sided, $^{****}p < 0.0001$, $n = 52$ independent subjects).

The remarkable heterogeneity in antigenic immune response between individuals has been noted previously, also by our recent studies[30,31]. One of the factors shaping the individual heterogeneity of immune response is associated with microbial symbiosis with the host and their antagonistic to mutualistic associations[37]. Among oral bacteria, these include health-associated early-stage, moderately pathogenic medium-stage and highly pathogenic late-stage colonisers of periodontal biofilm[37,38]. Our data reveal that many strongly targeted epitopes could potentially mimic and the antibodies could cross-react with the antigens of periodontal bacteria (Fig. 2, Supplementary Table S2, Supplementary Fig. S4). Other factors influencing individual variability in immune response to pathogens are age, lifestyle (diet, smoking, exercise etc), previous immune history (viral, bacterial), and the specific HLA-alleles that affect the presentation of major antigenic epitopes[39]. As some of these periodontal bacteria belong to orange or red complex groups of pathogenic oral species, the observed potential cross-reactivity of the antibody immune response (Fig. 2, Supplementary Fig. S4) could contribute to the dysbiosis of oral microbiota and thereby periodontal health.

We identified immunoprofiles stratifying the individuals with ACS from individuals with no CAD or stable CAD (Fig. 3). Our dataset has the advantage that we could identify and compare the presence of potential immunological markers in different CAD types, even in a limited study cohort. When the host's B cell response is inefficient against re-surfacing of latent infections, this may cause endothelial inflammation which in turn can contribute to the formation of atherosclerotic plaques and their instability[40]. Furthermore, differences were also observed in antibody response against epitopes in specific pathogens, including herpesviruses, in gingivitis, smoking or diabetes subgroups (Fig. 3, Supplementary Fig. S8). Diabetes along with smoking are two big risk factors for periodontal disease[41]. Thus, despite the small scale of the study and varied pathology background, our results provide proof of principle that different stages of CAD may be identifiable by different features of systemic immunoprofiles, which also include antibody response against highly antigenic epitopes of oral microbiota and common viruses.

Our data show that patients with progressing gingivitis through to periodontitis have increasing levels of antibodies to the highly antigenic epitope mimicking EBV VP26 (Figs. 3c and 4a, b). On the other hand, within subjects with periodontitis, a strong response to the EBV VP26 epitope is characteristic to healthy subjects, but not to ACS patients, suggesting its protective role against ACS (Fig. 5b). A few studies have directly addressed the role of herpesvirus infections in susceptibility to secondary oral bacterial infections. Relatedly, it is known that *A. actinomycetemcomitans* and *P. gingivalis* might require support from active herpesviruses for periodontal destruction, whereas the stable periodontal lesions may be devoid of viruses[42]. The herpesviral–bacterial hypothesis of periodontitis proposes that the herpesvirus infection triggers a release of proinflammatory



Periodontal diagnosis	Epitope	Diagnosis ≥threshold	Sensitivity	Specificity
H	Epitope C	ACS	71%	100%
G	Epitope A	ACS	89%	67%
P	Epitope A + E	no-CAD	75%	75%

Fig. 5 Immune response to epitope biomarkers to predict the ACS risk in periodontal disease and ascertain the no-CAD phenotype. a Strength of immune response to three epitopes (in *rows*) across study subjects ($n = 96$, in *lanes*), grouped by periodontal (H, G, or P) and CAD diagnosis (no-CAD, s-CAD, or ACS) shown as intensity plots. Relative abundance of immunoprofile features is shown in blue color-scale. These epitope biomarkers in a 3-biomarker generalized logistic model differentiate P group from H (Supplementary Fig. S11). *Vertical lanes* – subjects ($n = 96$); *rows* – epitopes; *blue color-scale* – normalized relative epitope-containing peptide abundance values (using 97.5th percentile values per feature and capped at value 1). **b** Different immunoprofile features (*above boxplots*) stratified ACS with periodontal diagnosis. *y-axes*: immunoreactivity to epitopes in individual immunoprofiles, shown as abundance of peptides containing the epitope. Mann–Whitney *U* test, two-sided, *p* values not corrected for multiple comparisons, $^{**}p < 0.01$, $^{*}p < 0.05$. H: $n = 21$ independent patients; G: $n = 27$ independent patients; P: $n = 48$ independent patients. **c** Sensitivity and specificity measures of using biomarkers from B in separate periodontal groups (*sub-group*) to predict CAD diagnoses (either no-CAD or ACS). *Diagnosis ≥ threshold* – diagnosis group into which patient was classified if immune response was over threshold. **a**–**c** Abbreviations and group sizes: H -periodontally healthy ($n = 21$), G -gingivitis ($n = 27$), *P* -periodontitis ($n = 48$), *no-CAD* -no CAD diagnosis ($n = 32$), *s-CAD* - stable CAD ($n = 32$), *ACS* - acute coronary syndrome ($n = 32$). Data in File S7.

cytokines to activate osteoclasts and matrix metalloproteinases (MMPs) to impair antibacterial immune mechanisms, causing an upgrowth of periodontopathic bacteria[42]. Given that the extracellular matrix (ECM) breakdown represents a crucial factor in the periodontal pathophysiology, periodontitis confounded diagnostics of ACS has been proposed by measuring serum levels of MMP-9[43], while serum MMP-8 and TIMP-1 levels were found to be associated with incident[44] and especially fatal cardiovascular events[45]. Overall, this indicates immune hyperresponsiveness to dysbiosis, typical to the pathogenesis of periodontitis[46]. On the other hand, periodontal herpesviruses themselves may disseminate via the systemic circulation to non-oral sites (including arteries) and thus represent a major link between periodontitis and cardiovascular diseases[41]. A recent meta-report described a world-wide association of EBV infection with periodontal disease[47]. Here we have investigated this observation further and fine-mapped the EBV VP26 epitope, against which the immunologic response could link periodontitis and associated CAD conditions. However, herein we show that it mimics antigens of periodontal bacteria (Fig. 2d, Supplementary Table S2). Microbial epitope similarity with pathogens, allergens and auto-antigens has been reported before, as it can elicit tolerogenic or inflammatory immune reactivity[48]. For example, different periodontopathogenic species share mimicry in their GroEL antigen, which in turn results in cross-reactivity with the human heat-shock proteins (HSP) expressed on the endothelial cell leading to endothelial dysfunction[49] and atherosclerosis[50]. This suggests that further studies of the herpesviral-bacterial-host epitope mimicry are warranted, in particular for improved diagnostics and therapy of dental health associated heart conditions.

Here we built a two-step biomarker model based on immune response to 4 epitopes that allows a) to classify subjects based on their periodontal diagnosis and b) to predict ACS risk and establish the no-CAD phenotype with 71–89% specificity (Fig. 5). Specific periodontitis-associated biomarkers for CAD could be beneficial in discerning ACS from other cardiac events. These data indicate that periodontitis, and ultimately putative progressing to ASC due to periodontitis, is the result of a partial response or lack of an efficient combined response against viral and bacterial infections landing on self-proteins.

Taken together, our findings clearly illustrate the power of MVA for the immunopathological analysis of oral health-related cardiac conditions, and we predict that the widespread use of this technology at scale will enhance the current understanding of chronic disease mechanisms, in particular cardiovascular diseases, and can lead to improved diagnostic accuracy and new markers.

Limitations of the study. We could not evaluate the prognostic value of these predicted biomarkers due to the clinical study design. Because of genetic and socioeconomic status variabilities in different study populations, it is hard to extrapolate the findings. This was a study on Finnish adults. New studies with other cohorts are needed. Also, the pathogenetic importance of the specific oral bacterial microbiota antigens remains firmly to be established.

Methods

Ethics statement. The study was conducted in accordance with the guiding principles of the Declaration of Helsinki and the study participants gave written informed consent before enrollment. The study was approved by the ethics

committees of the Helsinki University Central Hospital (approval reference number 106/2007) and The National Institute for Health Development, Estonia (approval number 1045).

Clinical cohort description. The cohort ($n = 96$) was selected from the initial Corogene study ($n = 5294$) and divided into 3 subgroups: periodontally healthy (H, $n = 21$), gingivitis positive (G, $n = 27$), and periodontitis (P, $n = 48$). The diagnostic features on clinical and radiographic findings have been described in detail earlier[51,52]. Periodontally healthy patients had no alveolar bone loss (ABL) and bleeding on probing (BOP) did not exceed 10%. Gingivitis was registered in patients without ABL but with BOP > 10%. Patients were diagnosed with periodontitis when the ABL exceeded the cervical third of the root. Coronary artery disease diagnosis (no coronary artery disease (no-CAD, $n = 32$), stable coronary artery disease (stable-CAD, $n = 32$), or acute coronary syndrome (ACS, $n = 32$)) was based on the degree of stenosis in the coronary arteries during the angiography, typical electrocardiographic changes, chest pain, and levels of cardiac biomarkers[32]. The age and gender proportions, along with other relevant clinical history, are in Supplementary Table S1.

Statistics and reproducibility. The statistical analyses performed during the study were accompanied by measures of statistical significance. The study was non-blinded and non-randomized and included $n = 96$ independent study subjects. Group-wise parameters, such as median values, were visualized alongside intra-group range using violin- or boxplots. Reproducibility of Mimotope Variation analysis was confirmed by establishing the correlation coefficient of two replicates as $R = 0.87$ ($p < 0.0001$).

Statistical analysis of clinical characteristics of samples. Differences in proportions of genders, diabetes condition and smoking status were assessed in clinical sub-groups using Chi^2 test (MedCalc, 19.7.2, MedCalc Software Ltd, Ostend, Belgium; https://www.medcalc.org). For statistical analysis, two-sided Mann–Whitney U test was used for comparing two groups or two-sided Kruskal–Wallis test for >2 groups using R package "ggpubr"[53] and "ggplot2"[54].

Mimotope variation analysis. Peptide antigens were selected from random peptide phage modified library (PhD12, NEB) with 10^9 different 12-mer peptide sequences[30,31]. Two µl of serum/plasma samples, previously precleared to plastic and E. coli/wt M13 phage lysates were incubated with 2.5 µl library (~5×10^{11} phage particles) and immunoglobulin G (IgG) fraction was recovered using protein G-coated magnetic beads (S1430S, NEB). Captured phage DNA was analyzed by Illumina HiSeq sequencing of 50-bp single end reads using barcoded primers for sample multiplexing. Peptide abundance correlation coefficient (R) in two replicates by Pearson analysis was 0.87 ($p < 0.0001$) (Supplementary Fig. S1) (R package "ggpubr"[53]). For further data analysis, sequencing errors and known artefacts were eliminated.

Selecting peptides. Group-enriched peptides (TopPeptide sets) were selected for clinical sub-groups (Supplementary Table S1). Peptides were selected for no-CAD ($n = 292{,}667$), stable-CAD ($n = 279{,}020$), and ACS ($n = 308{,}445$), and periodontally healthy ($n = 450{,}531$ peptides), gingivitis ($n = 450{,}590$), and periodontitis ($n = 342{,}261$) groups, using the criteria that these were to be identified in ≥10% individual samples of the group with abundance threshold ≥10 sequence counts in at least one sample.

Sequence-based unsupervised clustering of peptide antigens. Exhaustive sequence pattern search tool SPEXS2 was used (http://egonelbre.github.io/spexs2/) for sequence-based unsupervised clustering of peptide antigens. Starting from the 292,667 group-enriched peptides (TopPeptide set) identified for no-CAD group, all were used as input to SPEXS2 in random order with the search criteria: peptide coverage threshold: ≥4; motif coverage threshold: ≥4 fixed amino acid positions; hyper-geometric P value $<10^{-5}$). Sequence pattern searches were performed in 2 iterative runs, where peptides from which a consensus was identified in the first run were excluded from the subsequent run. As a result, 4366 distinct motif consensus sequences were identified, which were contained in 29.0% or 84,873 of the original 292,667 peptides. Therefore, 4366 unique consensus motifs were identified for no-CAD (covering 29.0% of input peptides), similarly 2771 motifs were calculated for stable-CAD (27.0%), 4405 for ACS (30.5%), 6275 for periodontally healthy (4 iterative SPEXS2 runs due to greater starting peptide set) (42.8%), 9560 for gingivitis (4 runs) (44.4%), and 5936 for periodontitis group (4 runs) (27.7%) were defined.

Selecting for group-differentiating motifs. The epitope motifs contained high degree of redundancy, therefore stricter criteria (hyper-geometric P value $<10^{-8}$ or query/reference ratio ≥10) were imposed to select for characteristics with high significance and statistical power. Altogether 8088 unique motif sequences fit those criteria were selected for further analyses as the TopMotif set. Of 8088 distinct motif features, 995 were designated as group-differential. The 995 motifs satisfied all the criteria that 1) the average abundance value in each clinical sub-group was >3 greater than in another relevant sub-group, 2) the abundance-based separation of relevant clinical sub-groups was with ≥50% sensitivity and ≥70% specificity, and 3) group-separation was statistically significant (Mann–Whitney U test, p value < 0.05, two-sided).

Alignment profiles on periodontal bacteria. The TopMotif set of 8088 motifs was aligned to proteomes of the periodontal pathogen species: *Porphyromonas gingivalis* (UniProt accession: UP000000588), *Tannerella forsythia* (UP000005436), *Prevotella intermedia* (UP000010099), *Fusobacterium nucleatum* (UP000002521), *Campylobacter rectus* (UP000003082), *Aggregatibacter actinomycetemcomitans* (UP000002569), and *Porphyromonas endodontalis* (UP000004295) (accessed 20-21.12.2019). Only exact alignments where all fixed amino acid positions of a motif (minimum of 4 positions) to match with the target were allowed. The pathogen database comprised of 15,928 proteins, of which 15,116 matched at least with one epitope motif. For each pathogen, the individual alignment load was calculated per 20-amino acid fragment of the protein, in two frameshifts (0 and −10 aa), yielding in 10 aa overlaps between considered fragments. Altogether we analyzed ~480,000 distinct 20 amino acid fragments. The top 40 fragments for each bacterial species, from any frameshifts, with the highest total alignment loads were selected based on the formula:

$$Alignment\ load\ (per\ 20aa\ fragment) = \frac{sum\ of\ abundance\ of\ motifs\ aligned}{count\ of\ motifs\ aligned}$$

If higher alignment loads were encompassing two side-by-side regions (one with a frameshift), a longer 30-amino acid fragment was considered. Individual alignment loads (Supplementary Data 1) were compared across pathogen species using Kruskal–Wallis test for comparing all groups and two-sided Mann–Whitney U test for pair-wise comparisons. R packages "ggpubr" and "ggplot2" were used for calculation and visualization.

Identifying target types of potential epitopes of periodontal bacteria. Using the top 40 antigen fragments of periodontal pathogens, 12mer substrings were extracted for each aligned epitope motif (of the TopMotif set), whereas substrings shorter than 12mer were discarded (33 were found). Altogether 1691 substrings were selected, position weight matrices were built and the resulting enriched amino acid positions were visualized as sequence logos, all using a custom in-house tool with WebLogo[55,56] integration (parameters: no counts, distance cut-off 9, 10 minimum unique substrings in a cluster, similarity index 15). The analysis yielded 26 target types with consensus sequences, of which top 5 most prevalent are shown on Fig. 2b.

Amino acid-based clustering of group-differential epitopes. Sequence homology clustering analysis reduced 705 out of 995 epitope motifs into 62 clusters of similar epitopes (≥3 identical amino acid positions) whereas 290 out of 995 epitope motifs either formed too small clusters (<3 motifs in a cluster) or were not similar enough with any other epitope (<3 identical amino acid positions). Additional R packages used in data analysis and visualization were: "readr", "dplyr"[57].

Based on the selected peptides, position weight matrices were built and the resulting enriched amino acid positions were visualized as sequence logos, all using a custom in-house tool with WebLogo[55,56] integration.

Group-specific analysis of 62 clusters. The average abundance for each of the 62 clusters with the defined sequence logo was calculated across clinical sub-groups. Abundance values in clinical sub-groups were normalized with average abundance across all groups (Supplementary Data 2). Clusters with average abundance values of <150 were left out of further analysis. Next, hierarchical clustering was performed based on the normalized abundance values using Pearson correlation coefficients for clustering distance and ward.D2 clustering method (R package "pheatmap")[58]. R packages used in data analysis and visualization included "readr", "dplyr", "forcats", "reshape2"[54,57], and "viridis"[59].

Individual-based clustering of 62 clusters. Abundance of peptides in individual immunoprofiles was calculated for each cluster. Using log10 values of the abundances, Pearson correlation coefficients R were calculated pair-wise for all 62 clusters (Supplementary Data 3). The clusters were subsequently grouped using R package "pheatmap" with ward.D2 as clustering method and Pearson correlation coefficients as clustering distance. Five distinct larger groups were identified and defined as epitopes A-E. For each of the epitopes A-E, the containing clusters were compared to determine consensus sequences describing the core epitopes (Supplementary Data 7).

Alignment on EBV VP26. Epitope A has been previously mapped to Epstein-Barr virus (EBV) protein VP26[30]. To validate the mapping in the current clinical cohort, epitopes from individual immunoprofiles ($n = 96$) were aligned to the primary sequence of EBV protein VP26, a component of the viral capsid antigen (VCA) (UniProt accession code Q3KSU9, EBV strain GD-1, date accessed: 25.11.2020). Signal-to-random ratio for each sample was calculated across primary sequence of EBV VP26, where signal represented the count of aligned (≥6 matching amino acid positions) unique peptides defined by MVA and random represented the count of aligned random peptides (≥6 matching), generated by sequence scrambling of

peptides (Supplementary Data 4). As controls (Ctrl), peptides from EBV CA seronegative subjects ($n = 9$) were analyzed similarly.

ELISA. CMV and EBV serostatus was measured from serum samples with ISO/IEC 17025 accredited methods. In brief, serological analyses were performed with anti-CMV IgG ELISA method (EUROIMMUN EI 2570-9601G) and with anti-EBV CA (capsid antigen) IgG ELISA method (EUROIMMUN EI 2791-9601G) according to manufacturer's specifications. Absorbance was measured at 450 nm with SpectraMax Paradigm (Molecular Devices).

Dot ELISA. MVA analysis findings were validated by dot ELISA analysis. For that, M13K phages displaying peptides with either the EBV VP26 epitope-containing sequence TLPMDTSPRAHW or the mutant sequence TLPMDASPRAHW as parts of the pIII protein were printed onto nitrocellulose (NC) slides (ArrayIt, US). Unspecific binding was reduced by blocking for 1 h at room temperature with 5% non-fat dried milk in 1xPBS-Tween20-0.05%. Human serum samples were precleared to reduce unspecific binding to M13K phages, to *E. coli* bacterial proteins and to plastic. Preclearing step was performed in mix of 60 µl 2.5% skimmed milk-1xPBS-0.05% Tween20 + 30 µl Preabsorption Solution + 1:50 serum, at 4 °C overnight. Following this, the slides were incubated with either precleared serum 1:50 solutions ($n = 54$) or with 1:5000 mouse anti-M13 antibody (#27-9420-01, GE healthcare) in 2.5% skimmed milk in 1xPBS-0.05% Tween20 (GE Healthcare) for 1 h at room temperature. The anti-M13K antibodies were used to quantify phages printed on NC slides. Multiple washes were performed with 5% skimmed milk in 1xPBS-0.05% Tween20. For visualization, the secondary antibodies used were 1:1000 rabbit anti-human HRP-conjugated antibody (#ab6759, Abcam) (for human serum samples) in 2.5% skimmed milk in 1xPBS-0.05% Tween20 or 1:1000 rabbit anti-mouse HRP-conjugated antibody (#ab6728, Abcam) (for anti-M13K) in 2.5% skimmed milk in 1xPBS-0.05% Tween20, incubated for 1 h at room temperature. After multiple washes, the presence of bound human sera/plasma IgG antibodies was detected via reaction with HRP substrate DAB chromogen diluted in substrate buffer (1:100). The slides were digitally scanned and the signals were quantified using ImageQuantTL (version 8.1) (Supplementary Data 5). Results of 27/54 samples were validated in an independent similar experiment (in total $n = 2$ experiments) (including all samples which showed high signal-to-background ratios).

Predicting periodontitis and CAD diagnoses. Generalized linear model was fit to 80% of subjects' data (with 5x cross-validation) to classify different case-subgroups based on their immunoprofile features. Using model's prediction probabilities for subjects in the training set, receiver operating characteristic (ROC) analysis was performed. The model's area under the receiver operating characteristic curve (AUROC) was 0.843 with 95% CI (0.738…0.948). The model was validated on the validation subset (20% of samples).

Data visualization. Box and whisker plots were generated in the style of Tukey with R packages "ggpubr" or "ggplot2". Upper, middle and lower boxplot lines represent the 75th, 50th and 25th percentiles, while whiskers represent the largest or smallest value within 1.5 times interquartile range above the 75th percentile or below the 25th percentile, respectively. Individual data points without outliers are visualized.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The source data analyzed during the study was generated by MVA analysis. We provided the relevant data underlying the main findings in the Supplementary data. The whole datasets generated and/or analyzed during the current study are not publicly available due to containing sensitive clinical information but are available from the corresponding author on reasonable request.

Code availability

The code used during the current study is not publicly available due to its proprietary nature, but detailed explanation of the analysis approaches is available from the corresponding author on reasonable request.

Received: 2 September 2021; Accepted: 8 February 2022;
Published online: 04 March 2022

References

1. Benjamin, E. J. et al. Heart disease and stroke statistics-2019 update: a report from the American Heart Association. *Circulation* **139**, e56–e528 (2019).
2. Madamanchi, N. R., Hakim, Z. S. & Runge, M. S. Oxidative stress in atherogenesis and arterial thrombosis: the disconnect between cellular studies and clinical outcomes. *J. Thromb. Haemost.* **3**, 254–267 (2005).
3. Pant, S. et al. Inflammation and atherosclerosis-revisited. *J. Cardiovasc Pharm. Ther.* **19**, 170–178 (2014).
4. Hansson, G. K. Inflammation, atherosclerosis, and coronary artery disease. *N. Engl. J. Med* **352**, 1685–1695 (2005).
5. Catapano, A. L., Pirillo, A. & Norata, G. D. Vascular inflammation and low-density lipoproteins: is cholesterol the link? A lesson from the clinical trials. *Br. J. Pharm.* **174**, 3973–3985 (2017).
6. Voight, B. F. et al. The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. *PLoS Genet* **8**, e1002793 (2012).
7. Miller, C. L., Assimes, T. L., Montgomery, S. B. & Quertermous, T. Dissecting the causal genetic mechanisms of coronary heart disease. *Curr. Atheroscler. Rep.* **16**, 406 (2014).
8. McPherson, R. & Hegele, R. A. The sum of its parts: the polygenic basis of coronary artery disease. *Can. J. Cardiol.* **32**, 1372–1374 (2016).
9. Nikpay, M. et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet* **47**, 1121–1130 (2015).
10. Vernon, S. T. et al. Utilizing state-of-the-art "omics" technology and bioinformatics to identify new biological mechanisms and biomarkers for coronary artery disease. *Microcirculation* **26**, e12488 (2019).
11. Kott, K. A. et al. Single-cell immune profiling in coronary artery disease: the role of state-of-the-art immunophenotyping with mass cytometry in the diagnosis of atherosclerosis. *J. Am. Heart Assoc.* **9**, e017759 (2020).
12. Fernandez-Alvira, J. M. et al. Predicting subclinical atherosclerosis in low-risk individuals: ideal cardiovascular health score and fuster-BEWAT score. *J. Am. Coll. Cardiol.* **70**, 2463–2473 (2017).
13. Nelson, C. P. et al. Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat. Genet.* **49**, 1385–1391 (2017).
14. Silverman, M. G. et al. Baseline subclinical atherosclerosis burden and distribution are associated with frequency and mode of future coronary revascularization: multi-ethnic study of atherosclerosis. *JACC Cardiovasc. Imaging* **7**, 476–486 (2014).
15. Lockhart, P. B. et al. Periodontal disease and atherosclerotic vascular disease: does the evidence support an independent association?: a scientific statement from the American Heart Association. *Circulation* **125**, 2520–2544 (2012).
16. Yumoto, H. et al. Sensitization of human aortic endothelial cells to lipopolysaccharide via regulation of Toll-like receptor 4 by bacterial fimbria-dependent invasion. *Infect. Immun.* **73**, 8050–8059 (2005).
17. Kozarov, E. V., Dorn, B. R., Shelburne, C. E., Dunn, W. A. Jr. & Progulske-Fox, A. Human atherosclerotic plaque contains viable invasive Actinobacillus actinomycetemcomitans and Porphyromonas gingivalis. *Arterioscler Thromb. Vasc. Biol.* **25**, e17–e18 (2005).
18. Gaetti-Jardim, E., Marcelino, S. L., Feitosa, A. C. R., Romito, G. A. & Avila-Campos, M. J. Quantitative detection of periodontopathic bacteria in atherosclerotic plaques from coronary arteries. *J. Med. Microbiol.* **58**, 1568–1575 (2009).
19. Jie, Z. et al. The gut microbiome in atherosclerotic cardiovascular disease. *Nat. Commun.* **8**, 845 (2017).
20. Acharya, A. et al. Salivary microbiome in non-oral disease: a summary of evidence and commentary. *Arch. Oral. Biol.* **83**, 169–173 (2017).
21. Consortium, C. A. D. et al. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat. Genet* **45**, 25–33 (2013).
22. Grundtman, C. & Wick, G. The autoimmune concept of atherosclerosis. *Curr. Opin. Lipido.* **22**, 327–334 (2011).
23. Garcia-Rivas, G. et al. The role of B cells in heart failure and implications for future immunomodulatory treatment strategies. *ESC Heart Fail.* **7**, 1387–1399 (2020).
24. Troseid, M., Andersen, G. O., Broch, K. & Hov, J. R. The gut microbiome in coronary artery disease and heart failure: current knowledge and future directions. *EBioMedicine* **52**, 102649 (2020).
25. Lee, N. & Kim, W. U. Microbiota in T-cell homeostasis and inflammatory diseases. *Exp. Mol. Med.* **49**, e340 (2017).
26. Kobiyama, K. & Ley, K. Atherosclerosis. *Circ. Res.* **123**, 1118–1120 (2018).
27. Davies, R. W. et al. A genome-wide association study for coronary artery disease identifies a novel susceptibility locus in the major histocompatibility complex. *Circ. Cardiovasc. Genet* **5**, 217–225 (2012).
28. Tsiantoulas, D., Diehl, C. J., Witztum, J. L. & Binder, C. J. B cells and humoral immunity in atherosclerosis. *Circ. Res.* **114**, 1743–1756 (2014).
29. Wolf, D. & Ley, K. Immunity and inflammation in atherosclerosis. *Herz* **44**, 107–120 (2019).
30. Sadam, H. et al. Identification of two highly antigenic epitope markers predicting multiple sclerosis in optic neuritis patients. *EBioMedicine* **64**, 103211 (2021).
31. Sadam, H. et al. Prostaglandin D2 receptor DP1 antibodies predict vaccine-induced and spontaneous narcolepsy type 1: large-scale study of antibody profiling. *EBioMedicine* **29**, 47–59 (2018).

32. Vaara, S. et al. Cohort profile: the Corogene study. *Int J. Epidemiol.* **41**, 1265–1271 (2012).
33. Liljestrand, J. M. et al. Association of endodontic lesions with coronary artery disease. *J. Dent. Res.* **95**, 1358–1365 (2016).
34. Liljestrand, J. M. et al. Immunologic burden links periodontitis to acute coronary syndrome. *Atherosclerosis* **268**, 177–184 (2018).
35. Falk, K. et al. Synthetic peptides deduced from the amino acid sequence of Epstein-Barr virus nuclear antigen 6 (EBNA 6): antigenic properties, production of monoreactive reagents, and analysis of antibody responses in man. *J. Med. Virol.* **46**, 349–357 (1995).
36. Loebel, M. et al. Serological profiling of the EBV immune response in Chronic Fatigue Syndrome using a peptide microarray. *PLoS One* **12**, e0179124 (2017).
37. Lamont, R. J., Koo, H. & Hajishengallis, G. The oral microbiota: dynamic communities and host interactions. *Nat. Rev. Microbiol.* **16**, 745–759 (2018).
38. Socransky, S. S. & Haffajee, A. D. Dental biofilms: difficult therapeutic targets. *Periodontol. 2000* **28**, 12–55 (2002).
39. Brodin, P. & Davis, M. M. Human immune system variation. *Nat. Rev. Immunol.* **17**, 21–29 (2017).
40. Campbell, L. A. & Rosenfeld, M. E. Infection and atherosclerosis development. *Arch. Med Res* **46**, 339–350 (2015).
41. Tsuchida, S. Proteome analysis of molecular events in oral pathogenesis and virus: a review with a particular focus on periodontitis. *Int. J. Mol. Sci.* **21**, https://doi.org/10.3390/ijms21155184 (2020).
42. Slots, J. Human viruses in periodontitis. *Periodontol. 2000* **53**, 89–110 (2010).
43. Liukkonen, J. *et al.* Immunological and microbiological profiling of cumulative risk score for periodontitis. *Diagnostics* **10**, https://doi.org/10.3390/diagnostics10080560 (2020).
44. Tuomainen, A. M. et al. Serum tissue-degrading proteinases and incident cardiovascular disease events. *Eur. J. Prev. Cardiol.* **21**, 806–812 (2014).
45. Kormi, I. et al. Matrix metalloproteinase-8 and tissue inhibitor of matrix metalloproteinase-1 predict incident cardiovascular disease events and all-cause mortality in a population-based cohort. *Eur. J. Prev. Cardiol.* **24**, 1136–1144 (2017).
46. Kinane, D. F., Stathopoulou, P. G. & Papapanou, P. N. Periodontal diseases. *Nat. Rev. Dis. Prim.* **3**, 17038 (2017).
47. Gao, Z., Lv, J. & Wang, M. Epstein-Barr virus is associated with periodontal diseases: a meta-analysis based on 21 case-control studies. *Medicine* **96**, e5980 (2017).
48. Carrasco Pro, S. et al. Microbiota epitope similarity either dampens or enhances the immunogenicity of disease-associated antigenic epitopes. *PLoS One* **13**, e0196551 (2018).
49. Wang, C. et al. Characterization of a natural mouse monoclonal antibody recognizing epitopes shared by oxidized low-density lipoprotein and chaperonin 60 of Aggregatibacter actinomycetemcomitans. *Immunol. Res.* **64**, 699–710 (2016).
50. Goulhen, F., Grenier, D. & Mayrand, D. Oral microbial heat-shock proteins and their potential contributions to infections. *Crit. Rev. Oral. Biol. Med.* **14**, 399–412 (2003).
51. Buhlin, K. et al. Periodontitis is associated with angiographically verified coronary artery disease. *J. Clin. Periodontol.* **38**, 1007–1014 (2011).
52. Hyvarinen, K. et al. A common periodontal pathogen has an adverse association with both acute and stable coronary artery disease. *Atherosclerosis* **223**, 478–484 (2012).
53. Kassambra, A. ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.4.0. https://CRAN.R-project.org/package=ggpubr (2020).
54. H. Wickham. ggplot2: Elegant Graphics for Data Analysis. https://ggplot2.tidyverse.org. (Springer-Verlag New York, 2016).
55. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
56. Schneider, T. D. & Stephens, R. M. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* **18**, 6097–6100 (1990).
57. Wickham, H., Averick, M., Bryan, J., Chang, W. & McGowan, L. Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 43 (2019).
58. pheatmap: Pretty Heatmaps. R package version 1.0.12. https://doi.org/10.21105/joss.01686 (2019).
59. viridis: Default Color Maps from 'matplotlib'. R package version 0.5.1. https://sjmgarnier.github.io/viridis/ (2018).

Acknowledgements

We thank A.-H. Pool (Department of Neuroscience, University of Texas Southwestern Medical Center) for his expertise that supported this work. The members of Protobios team are thanked for their excellent technical assistance. This study was supported by research funding grants of Protobios (5.1-4/20/170, and PRG573) from the Estonian Ministry of Education and Estonian Research Council, respectively, and H2020-MSCA-RISE-2016 (EU734791) and H2020 PANBioRA (EU760921) projects from the European Union. K.P. was partially supported by the European Union through the European Regional Development Fund (Project No. 2014-2020.4.01.15-0012). N.E.V. was supported by H2020 PANBioRA (EU760921) project from the European Union. P.P. was supported by the Sigrid Jusélius Foundation.

Author contributions

M.J., N.P., A.R., A.P., H.S., and K.P. conceived and designed the study. J.S. and P.P. collected samples and acquired the clinical data. M.J., N.P., A.R., A.P., H.S., and K.P. analyzed and interpreted the data. N.E.V., J.S., P.P., and K.P. supervised the study. M.J. and K.P. drafted the manuscript. All authors revised and approved the final manuscript for submission.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-022-03122-4.

Correspondence and requests for materials should be addressed to Kaia Palm.

Peer review information *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editor: Anam Akhtar.

Reprints and permission information is available at http://www.nature.com/reprints

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

Appendix 3

PUBLICATION III

Sadam, H., Pihlak, A., Kivil, A., Pihelgas, S., **Mariliis Jaago**, Adler, P., Vilo, J., Vapalahti, O., Neuman, T., Lindholm, D., Partinen, M., Vaheri, A., Palm, K. 2018. **Prostaglandin D2 Receptor DP1 Antibodies Predict Vaccine-induced and Spontaneous Narcolepsy Type 1: Large-scale Study of Antibody Profiling.** *EBioMedicine*, 29, 47−59. doi: 10.1016/j.ebiom.2018.01.043.

EBioMedicine 29 (2018) 47–59

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

Research Paper

Prostaglandin D2 Receptor DP1 Antibodies Predict Vaccine-induced and Spontaneous Narcolepsy Type 1: Large-scale Study of Antibody Profiling

Helle Sadam [a,b,1], Arno Pihlak [a,b,1], Anri Kivil [a,1], Susan Pihelgas [a], Mariliis Jaago [a], Priit Adler [c,k], Jaak Vilo [c,k], Olli Vapalahti [e,f,g], Toomas Neuman [a,h], Dan Lindholm [i,j], Markku Partinen [d], Antti Vaheri [e], Kaia Palm [a,b,*]

[a] Protobios Llc, Mäealuse 4, 12618 Tallinn, Estonia
[b] Department of Gene Technology, Tallinn University of Technology, Akadeemia Tee 15, 12618 Tallinn, Estonia
[c] Institute of Computer Science, University of Tartu, Liivi 2-314, 50409 Tartu, Estonia
[d] Finnish Narcolepsy Research Center, Helsinki Sleep Clinic, Vitalmed Research Center, Valimotie 21, 00380, Helsinki, Finland
[e] Department of Virology, Medicum, Haartmaninkatu 3, 00014 University of Helsinki, Finland
[f] Department of Veterinary Biosciences, University of Helsinki, Agnes Sjöbergin Katu 2, 00014 University of Helsinki, Finland
[g] Virology and Immunology, HUSLAB, Helsinki University Hospital, 00290 Helsinki, Finland
[h] IPDx Immunoprofiling Diagnostics GmbH, Deutscher Platz 5e, 04103 Leipzig, Germany
[i] Department of Biochemistry and Developmental Biology, Medicum, Haartmaninkatu 8, 00014 University of Helsinki, Finland
[j] Minerva Foundation Medical Research Institute, Tukholmankatu 8, 00290 Helsinki, Finland
[k] Quretec LLC, Ülikooli 6a, 51003 Tartu, Estonia

ARTICLE INFO

Article history:
Received 22 December 2017
Received in revised form 23 January 2018
Accepted 31 January 2018
Available online 2 February 2018

Keywords:
Narcolepsy type 1
H1N1
Pandemrix
Antibody
Prostaglandin
DP1

ABSTRACT

Background: Neuropathological findings support an autoimmune etiology as an underlying factor for loss of orexin-producing neurons in spontaneous narcolepsy type 1 (narcolepsy with cataplexy; sNT1) as well as in Pandemrix influenza vaccine-induced narcolepsy type 1 (Pdmx-NT1). The precise molecular target or antigens for the immune response have, however, remained elusive.

Methods: Here we have performed a comprehensive antigenic repertoire analysis of sera using the next-generation phage display method - mimotope variation analysis (MVA). Samples from 64 children and adolescents were analyzed: 10 with Pdmx-NT1, 6 with sNT1, 16 Pandemrix-vaccinated, 16 H1N1 infected, and 16 unvaccinated healthy individuals. The diagnosis of NT1 was defined by the American Academy of Sleep Medicine international criteria of sleep disorders v3.

Findings: Our data showed that although the immunoprofiles toward vaccination were generally similar in study groups, there were also striking differences in immunoprofiles between sNT1 and Pdmx-NT1 groups as compared with controls. Prominent immune response was observed to a peptide epitope derived from prostaglandin D2 receptor (DP1), as well as peptides homologous to B cell lymphoma 6 protein. Further validation confirmed that these can act as true antigenic targets in discriminating NT1 diseased along with a novel epitope of hemagglutinin of H1N1 to delineate exposure to H1N1.

Interpretation: We propose that DP1 is a novel molecular target of autoimmune response and presents a potential diagnostic biomarker for NT1. DP1 is involved in the regulation of non-rapid eye movement (NREM) sleep and thus alterations in its functions could contribute to the disturbed sleep regulation in NT1 that warrants further studies. Together our results also show that MVA is a helpful method for finding novel peptide antigens to classify human autoimmune diseases, possibly facilitating the design of better therapies.

1. Introduction

Narcolepsy type 1 (NT1) is a chronic neurological disease characterized by irresistible daytime sleepiness, disturbed nocturnal sleep, and cataplexy associated with the inadequate function of the hypothalamus (Peyron et al., 2000; Thannickal et al., 2000; Partinen et al., 2014). The major neuropathological features of NT1 are loss of orexinergic neurons and an increased gliosis in the posterior hypothalamic nuclei (Partinen et al., 2014). Increased levels of pro-inflammatory cytokines have been associated with (spontaneously occurring) idiopathic (sNT1) and Pandemrix vaccine-induced narcolepsy (Pdmx-NT1) close to disease onset (Lecendreux et al., 2015). Pandemrix (Pdmx) is an influenza vaccine used during the H1N1 2009 swine influenza A(H1N1) pandemic and was distributed to over 30 million people in EU/EEA countries

* Corresponding author at: Protobios Llc, Mäealuse 4, 12618 Tallinn, Estonia.
E-mail address: kaia@protobios.com (K. Palm).
[1] These authors contributed equally to the work.

https://doi.org/10.1016/j.ebiom.2018.01.043

during the A(H1N1) outbreak. As of January 2015, >1300 cases of vaccine-associated NT1 had been reported to the European Medicines Agency. Epidemiologic and clinical studies conducted in different countries including Finland, Sweden, Ireland, England, Norway, and France have confirmed the association of NT1 in children and adolescents with the AS03-adjuvanted Pdmx (Partinen et al., 2014; Sarkanen et al., 2017). Subsequently, wild-type influenza A(H1N1) infections in China were associated with narcolepsy (Han et al., 2013, 2011). Along with the pandemic A(H1N1) infection, seasonality and post-infectious priming by upper respiratory tract viruses and streptococci have been suggested as triggers of autoimmune response that leads to NT1 in genetically susceptible individuals (Aran et al., 2009; Longstreth Jr et al., 2009).

Genome-wide association studies have revealed a strong association of narcolepsy with the T-cell receptor alpha locus (Hallmayer et al., 2009) and especially with Major Histocompatibility Complex (MHC) class II *DQB1*06:02* alleles (Bonvalet et al., 2017; Tafti et al., 2014). *DQB1*06:02* is present in approximately 30% of Finnish and Swedish populations (Bomfim et al., 2017). In Finland, all patients with Pdmx-NT1 have been positive for *DQB1*06:02* (Partinen et al., 2014). The latter immune haplotype is also strongly associated with the Pdmx-NT1 in Sweden (Bomfim et al., 2017). In another series of 522 patients with narcolepsy and cataplexy from different countries, only 9 patients (1.7%) with low levels of orexin (OX) in cerebrospinal fluid (CSF) were *DQB1*06:02* negative (Han et al., 2014). It was also suggested that cross-reactive epitopes to Pdmx vaccine antigens may exist in NT1 diseased as a significant proportion of HLA-*DQB1*0602*–positive Finns diagnosed with NT1 and with a history of H1N1 vaccination were immunoreactive to OX receptors (Ahmed et al., 2015). However, it still is unclear whether OX-positive neurons and/or their neighboring cells express OX receptors that could be targets for the immune response in NT1 (Valko et al., 2013; Vassalli et al., 2015). The antibody levels to viral nucleoprotein (NP), a Pdmx vaccine antigen, were increased in NT1-diseased carrying the HLA *DQB1*06:02* allele (Vaarala et al., 2014), whereas the role of this and other circulating (including intrathecal) autoantibodies in NT1 pathogenesis is not fully understood (see list of previously identified antigens in Table S1). Although NT1-related autoantibodies are found in some patients, the clinical response to intravascular immunoglobulin (IVIG) has been hard to predict (Knudsen et al., 2012). Likewise, use of the drug rituximab might have only short-lasting beneficial effects in NT1 (Sarkanen et al., 2016).

Recent advances in proteomics (immunomics) have made it possible to study the adaptive immune response in various diseases in great detail and at a high resolution (lately reviewed in: (Ayoglu et al., 2016, Wu et al., 2016). We and others have suggested a strategy of high-throughput sequencing-assisted epitope mapping directly on serum for biomarker discovery and disease detection based on the idea that self- and environmental (exposome) antigens are reflected in the immune response profiles (immunoprofiles) (Anastasina et al., 2017; Christiansen et al., 2015; Ionov, 2010; Xu et al., 2015). Hence, the profiling of antibody response repertoire with high-density random peptide/polypeptide display methods could be a novel mean to characterize and classify human diseases in an unbiased manner according to the molecular/cellular targets relevant for the disease.

In the present study, we have used the mimotope-variation analysis (MVA) method to immunoprofile autoantibody repertoires in patients afflicted by NT1 and in controls. We had access to the clinical cohorts composed of 16 NT1 (sNT1 ($n = 6$) and Pdmx-NT1 ($n = 10$)) cases, where all NT1-diseased subjects carried the HLA *DQB1*06:02* allele, and apart from 2 sNT1 patients, all had been vaccinated with Pdmx. For reference, we used three well-defined control groups: 16 Pandemrix-vaccinated healthy controls (Pdmx-HC), 16 H1N1-infected Finnish subjects (H1N1-HC), and 16 healthy Estonian donors (HC – healthy controls) (Table 1). Our data revealed complex patterns of immune response in all patient groups including novel epitope sequences present in sera of Pdmx-NT1 and H1N1-HC. One such peptide epitope was identified as belonging to the prostaglandin D2 receptor (DP1) that together with its ligand prostaglandin D2 (PGD2) is involved in sleep regulation in humans and experimental animal models (see ref. in Urade and Hayaishi (2011)).

2. Materials and Methods

2.1. Vaccines

Pandemrix vaccine is derived from X-179A, a reassortant of hemagglutinin (HA), neuraminidase (NA) and polymerase acidic protein (PA) of A/California/07/2009 and X-157 H3N2 in a PR8 backbone (Jacob et al., 2015; Nicolson et al., 2012; Robertson et al., 2011). The vaccine composition can be found summarized by European Medicines Agency and GlaxoSmithKline plc (European Medicines Agency, 2009).

2.2. Study Population

The present study comprises a total of 64 individuals (Table 1). Altogether, 16 serum samples of H1N1-infected military servicemen (H1N1-HC), 16 serum samples of age/sex-matched Pandemrix-vaccinated healthy controls (Pdmx-HC) were kindly provided by National Institute of Health and Welfare, Finland. 16 serum samples were collected from patients with H1N1-induced (Pdmx-NT1) and sporadic narcolepsy (sNT1). Four out of 6 sNT1 patients were vaccinated with Pdmx after they had been diagnosed with NT1. Narcolepsy patients were diagnosed at the

Table 1
Description of samples studied.

Characteristics	Narcolepsy (NT1) patients		Healthy controls (HC)		
	Pdmx-NT1	sNT1	Pdmx-HC	H1N1-HC	Other HC
Group size (*n*)	10	6	16	16	16
Gender (female/male)	5/5	5/1	12/2[a]	0/16	10/6
Pandemrix vaccination	11/2009–1/2010	11/2009-1/2010[b]	11/2009-1/2010	–	–
Sample collection	2011	2011	2011	2011	2009
Median age at onset (y)	13	18	–	–	–
Median age at sampling (y)	14	22	NA	21	34.5
Unambiguous cataplexy	10/10 (100%)	6/6 (100%)	–	–	–
MSLT mean SL (range)	2.0 (0.4–4.3)	2.6 (0–7.5)	NA	NA	NA
SOREMPS mean (range)	3.7 (2–5)	2.7 (2–4)	NA	NA	NA
HLA *DQB1*0602* (%)	10/10 (100%)	6/6 (100%)	NA	NA	NA
CSF-orexin < 150 pg/mL (lower 1/3 limit in Finland)	7/7 (100%)	5/5 (100%)	NA	NA	NA

HC- healthy control, *H1N1-HC* – H1N1 infected, *Pdmx-HC*– Pandemrix-vaccinated, *NT1*- narcolepsy type 1 (including 10 Pdmx-induced NT1 samples (Pdmx-NT1) and 6 sporadic NT1 (sNT1) samples), *NA* – not available, *SL* – sleep latency, *MSLT* – Multiple sleep latency test, *SOREMPS* – Sleep onset REM periods as defined by the American Academy of Sleep Medicine.

[a] Gender of two Pdmx-HC is unknown.

[b] Four out of 6 sNT1 patients were vaccinated after they had been diagnosed with NT1.

Finnish Narcolepsy Research Center (Helsinki Sleep Clinic, Vitalmed Research Center) by experienced neurologists with a special competence in sleep medicine. All narcolepsy patients had NT1 as defined by the American Academy of Sleep Medicine international criteria of sleep disorders version 3. All excessively sleepy patients with NT1 had unambiguous cataplexy and an abnormal MSLT (sleep latency <8 min and at least 2 sleep onset REM periods; Multiple Sleep Latency Test) recording after a polysomnography. Twelve of the 16 patients had their CSF-orexin levels measured using the standardized Phoenix RIA method with Stanford reference. All except one had CSF-orexin levels <110 pg/mL (one woman with clearly abnormal MSLT and unambiguous cataplexy had CSF-orexin level of 127 pg/mL).

Control serum specimen for the study included 16 serum samples from Blood Centre, North Estonian Regional Hospital, Estonia, collected in 2009 prior to the swine influenza pandemic (other HC). Sera were stored at −135 °C until use.

2.3. Ethical Permissions

The patients have participated in the NARPANord narcolepsy study (Academy of Finland, grant nr. 260603), and they have given a written informed consent. The serum samples of H1N1-infected military servicemen and serum samples of the Pdmx-vaccinated healthy controls were provided by the National Institute of Health and Welfare, Finland. The ethical permissions were approved by the Ethics Committee of the Hospital District of Helsinki and Uusimaa, Finland.

2.4. Mimotope-Variation Analysis

For qualitative and quantitative characterization of humoral immune response from sera samples, we used an in-house developed mimotope-variation analysis (MVA) method. Fig. 1A provides an overview of the process. In brief, a random 12-mer peptide phage library (Ph.D.-12, NEB, UK) was used according to the manufacturer's protocol. 2 μL of serum sample was incubated with 2.5 μL library (~5×10^{11}) and immunoglubulin G (IgG) fraction was recovered using protein G-coated magnetic beads (Thermo Fisher Scientific). The unbound phage particles were removed by extensive washes with TBS-T (TBS + 0.1% [v/v] Tween-20). Selectively captured phage DNA was analyzed by using next generation (Illumina) sequencing with barcoding primers (Islam et al., 2014). For that, DNA was extracted by using standard NaI/EtOH precipitation method and enriched by PCR amplification using primers enriched with adapters for the sequencing reaction that flank the variable region at the end of pIII in M13KE vector (Fw: 5′-AATGATACGGCGACCACCGAGATCTACACTGATCTAGTGGTACCTTTCTATTCTCA*C*T*C*T-3′ Rv: 5′-CAAGCAGAAGACGGCATACGAGAT**NNNN(NN)**CCCTCATAGTTAGCGTAACG-3′). PCR products were purified using the QIAquick PCR Purification Kit (QIAGEN), and the concentration of DNA with Illumina adapters was estimated by Qubit Fluorimeter (Invitrogen) according to the manufacturer's protocol. Sequencing was performed using Illumina HiSeq and 50-bp single end reads. Samples were analyzed at least in duplicates to ensure reproducibility. To evaluate the data reproducibility, we compared peptide abundance in two replicates using Pearson's correlation coefficient test (r value higher than 0.90, $p < .0001$).

2.5. Sequence Data Analysis

Every read that was considered valid by the Illumina HiSeq control software was prepared for further analysis by trimming. In brief, sequence reads of 50 bp were accompanied with a 4-bp tag to assign each read uniquely to one out of 48 multiplexed samples. After demultiplexing there were millions of sequence reads for each sample. Each read consisted of a 36-bp random insert region and a constant region. We discarded the reads with mismatches in flanking 4 bases of the constant region. Sequences from wild-type phages with no random insert were excluded. Next, we translated all the random insert regions in reads into peptide sequences of length 12aa. All non-translatable sequences were discarded. To reduce the effect of amplification and sequencing errors, only those peptide sequences were kept that had at least two copies sequenced per sample. In order to compensate for the different numbers of reads per sample normalization of read counts was performed. All samples were trimmed to 3 million reads (RPM units). The resulting data was represented as a cross-table where each row corresponded to a different 12mer peptide, each column corresponded to a different sample, and each cell showed the read count of the peptide in the respective sample measured in RPM-units. According to the manufacturer (NEB), naïve library contained up to 10^9 different sequences. For reasoned cost purposes, the estimated outcome of sequence data represented 0.1% of the initial library input containing up to 2.8×10^6 different peptide sequences per sample. Complete analysis of sequence diversities obtained by MVA remains out of the scope of the current study.

2.6. Clustering Workflow

The main assumption was that every obtained peptide sequence mimics the target of an antibody. The sequence reads of one sample often included many copies of the same peptide sequence. The read counts of a peptide could range from 1 to thousands. To reveal recognition patterns (epitope motifs) which were enriched in the cases compared to controls, we used SPEXS2 software (https://github.com/egonelbre/spexs2; (Vilo, 2002, Brazma et al., 1998)). For clustering the peptides with motifs and generating mimotope regular expression and sequence logos, the "motifTree" tool was used (Kruup, 2013). The Multiple EM for Motif Elicitation (MEME-MAST) algorithm (Bailey and Elkan, 1994; Bailey and Gribskov, 1998) was used to align peptides to proteins. For B cell epitope mapping IEDB 3.0 database was used (Vita et al., 2015).

2.7. Statistical Analysis

All statistical analyses (ANOVA, *t*-Test, correlation analyses, Chi-square test) were done using MedCalc software (MedCalc Statistical Software version 17.0.4 (MedCalc Software bvba, Ostend, Belgium; https://www.medcalc.org; 2017)). For visualization of peptide abundance across samples, peptide frequency values were converted to heatmap images (Tagged Image File) with Excel Visual Basic for Applications (VBA) scripts. For visualization of selected peptide set alignment profile on proteins of interest Excel VBA script was used. The protein sequence was scanned with every peptide and at every position where the peptide aligned with it in at least four perfectly matching positions, one was added with its frequency. For random reference profile, amino acid sequence of each peptide was randomized and scanned using the same rules over the target sequence.

2.8. Influenza Virus Serology

Levels of influenza-specific IgG antibodies were determined by the enzyme-linked immunosorbent Vir-ELISA anti-H1N1/H3N2 IgG assay (Influenza virus type A IgG ELISA test system, Euroimmun), carried out in accordance with the manufacturer's specifications. Absorbance was measured at 450 nm with SpectraMax Paradigm.

2.9. Peptide ELISPOT

For peptide ELISPOT the following peptides were designed:
peptide #1 - RVLAPALDSWGTGGGDYKDDD{LYS(BIOTIN)} (Genescript)
peptide #2 - LPKFSAPSASGPGGGDYKDDD{LYS(BIOTIN)} (Genescript)
peptide #3 - ESTRYQLWLPHQGGGDYKDDD{LYS(BIOTIN)} (Genescript)
control peptide - AVLAAALASWGTGGGDYKDDD{LYS(BIOTIN)} (Genescript)

In brief, 110 pg biotin-conjugated peptides were printed on nitrocelluose coated slides (10485323, Whatman) by SpotBot® 4 (Arrayit). For primary antibody human precleared serum (1:100) was used, for secondary antibody rabbit anti-human IgG (H&L) (HRP) (Abcam) was used. All incubations were done for 1 h at room temperature. Results were scanned using Ettan DigeImager (GE Healthcare Life Sciences) and images calculated using ImageQuant software version 8.1 (GE Healthcare Life Sciences).

2.10. Cancer Cells, Human Mesenchymal Stem Cells (hMSC) and Post-Mortem Tissues

Immortalized glioblastoma multiforme cells (human glioma cells - hGC) (kind gift of Prof. Aavo-Valdur Mikelsaar, Estonia), human neuroblastoma cell line Kelly (ATCC) and human mesenchymal stem cells (hMSC) (isolated from human subcutaneous adipose tissue as described (Jaager and Neuman, 2011)) were grown in Dulbecco's modified Eagle's medium (DMEM (PAA)) containing 10% fetal bovine serum (PAA), 1 mg mL^{-1} penicillin (PAA) and 0.1 mg mL^{-1} streptomycin (PAA). All cells were cultured at 37 °C in 5% CO_2. The identity of hMSC was confirmed by using cell morphology and flow cytometry methods for analysis of cell surface markers: CD73+/CD90+/CD105+/CD45−/CD34− (Kauts et al., 2013). For treatments, hMSCs were grown with media containing IL-1β (1 ng/mL), IFNγ (2 ng/mL) for 8 h, or PGD2 (10 μM) for 1 h.

Human post-mortem tissues were procured from the North-Estonian Regional Hospital, Tallinn, Estonia. All experiments with human tissues were done with the approval of the local ethical committee (license no. 2234, date of issue 09.12.2010).

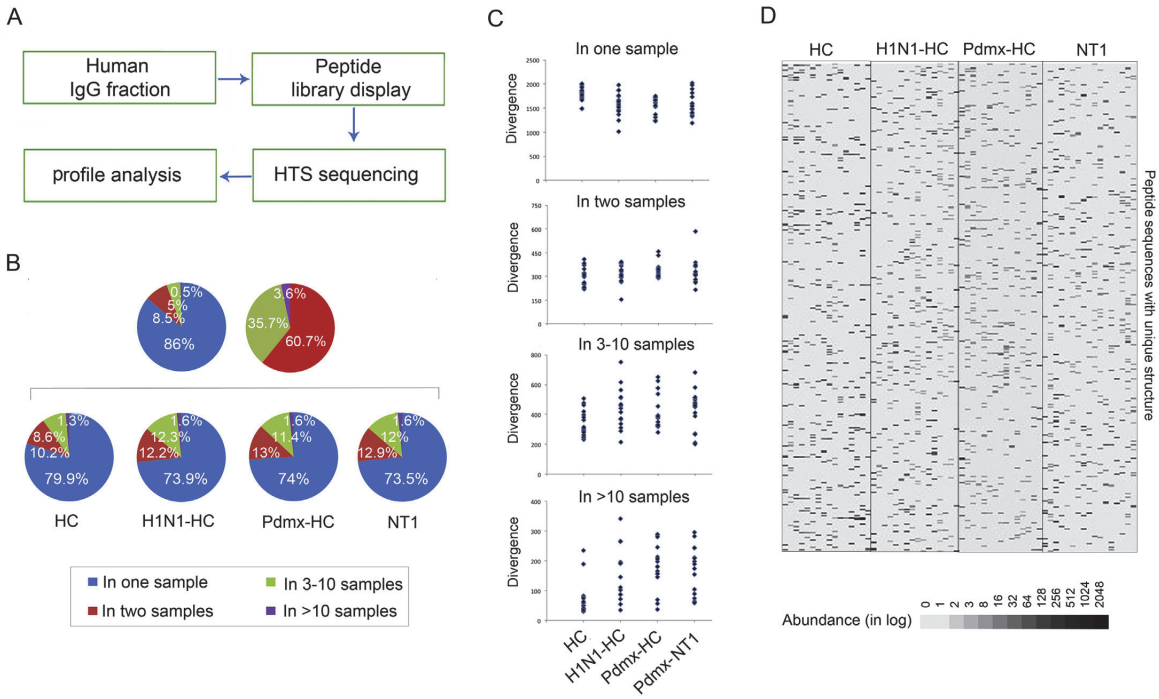


Fig. 1. Humoral immune response studied using the mimotope-variation analysis (MVA) method. A. Schematic drawing of the workflow in MVA. MVA is a high-throughput random peptide phage display analysis. A random peptide display library (PhD12) was used which contained 10^9 different 12-mer peptide sequences introduced to the N-terminus of the phage major coat protein pIII (NEB). For MVA, sample-specific IgG proteins (antibodies, *Human IgG fraction*) present in human sera of interest are allowed to interact with the phage-displayed peptides and the IgG-phage complexes were captured to protein G magnetic beads, while the unbound phages were washed away (*Peptide library display*). Captured phages were lysed and DNA amplified with primer sequences containing a tag with a unique barcode sequence and the final amplicons were pooled for NGS analysis (*HTS sequencing*). The primer set homologous to the M13KE vector sequences that flank the random peptide coding sequence was used to amplify a 50-bp fragment. Data analysis to classify peptides that were specific to Pmdx-infected, -vaccinated and NT1-diseased individuals was carried out by comparing the profiles of peptides (mimotopes) from diseased to those from non-diseased (*Peptide profile analysis*). On average, MVA generated 1.8 million peptide sequences with unique structure (divergence) totaling 2.8 million peptide sequences in abundance (total abundance; number of reads) per sample. Altogether, a peptide data set with >16 million sequences (Totpep) with unique structure was generated. B. Analysis of peptides revealed highly divergent patterns (immunoprofiles) across study cohorts. The fraction of top 2500 peptides with unique structure and highest values of abundance - reflecting the peaking immune reactivity of each sample - was analyzed for variance. Top2500 peptide dataset contained altogether 160,000 sequences out of which 121,142 were unique. Pie charts display the sequence distribution of unique peptides across all samples analyzed. The left pie (*blue*) displays the proportion of shared vs. unique peptides: ~86% were unique to one individual whereas ~14% of the peptide sequences were shared between samples, out of these ~8.5% were common to 2 samples, 5% to 3–10 samples and 0.5% were detected in >10 samples. The right pie (*red*) displays the distribution of shared 16,844 peptide sequences out of which ~60.7% were common to 2 samples, 35.7% to 3–10 samples and 3.6% were seen in >10 samples. The four pie charts (below) exemplify the peptide profile structures in different clinical cohorts. The size of each pie piece is proportional to the number of unique peptides common to one or more samples of a clinical cohort. *Blue* - represents unique peptides, *red* - the most shared. C. Individual variation in peptide divergence is characteristic to all immunoprofiles. Top 2500 peptides were analyzed to assess the range of individual peptide variation across study cohorts. Blue dots mark peptide divergence in a single sample. As indicated, between one to two thousand peptides were individual-specific, whilst the most common peptides (shared by >10 individuals) ranged in divergence from tens to 350 across samples. Range of unique peptide variations was similar across all study samples. D. Heat map image of a random fragment of MVA profile encompassing 400 peptides across study samples. Peptide profiles were individual-specific with a highly varying abundance. Each column represents the peptide profile of a single individual, and each line represents a peptide with a unique primary structure. Abundance is presented as counts in logarithmic scale (*in log*); black colour depicts peptides captured at higher abundance, and white those at lower abundance. Shown are peptide profiles that were common to 3–10 individuals. Abbreviations: *Abundance* – peptide frequency; *Divergence* – all unique peptides; *HC*- healthy control; *H1N1-HC* – H1N1 infected; *Pdmx-HC*- Pandemrix-vaccinated; *NT1*- narcolepsy type 1 (including 10 Pandemrix-induced NT1 samples).

2.11. Immunofluorescence and Western Blot Analysis

For immunofluorescence analysis, cells grown on glass inserts were fixed using 4% PFA (Scharlau) for 15 min and blocking of the unspecific reactivity was done with 5% BSA. The antibodies used included: anti-DP1 (Abnova; 1:500), precleared human sera (1:400), and the secondary Alexa Fluor 488 and 647 (Invitrogen, 1:2000) antibodies. For epitope blocking peptide #1 was used in final concentration 6.6 μg/mL. Hoechst 33342 (Invitrogen) was used to detect cell nuclei. Imaging was done using Nikon Eclipse 80i microscope.

Sequences (RVLAPALDSWGT and DYKDDDDK (flag)) were inserted at the N-terminus of the pIII of the M13KE phage by in vitro mutagenesis PCR using primers s1 5′GCTGGATAGTTGGGGAACCGGTGGAGGTTCG GCCGAAAC3′, as1 5′GCCGGAGCTAGTACACGAGAGTGGGAGTAAAACG GTACC3′,

s2 5′GCTGGATAGTTGGGGAACC3′, as2 5`GCCGGAGCTAGTACACG3′; s3 5′GATGATGATAAAGGTGGAGGTTCGGCCGAAAC3′, as3 5′ATCTTTATAAT CAGAGTGGGAGTAAAACGGTACC3′; s4 5′GATGATGATAAAGGTGG3′, as4 5′ATCTTTATAATCAGAGTGG3′. PCR reactions were carried out with phusion Hot Start II High-Fidelity DNA Polymerase (ThermoScientific). Constructs were verified by sequencing. For Western blot analysis, 30 μg of protein lysate or 1×10^{13} phage particles were resolved on 10% SDS–polyacrylamide gels and transferred onto PVDF membranes (Amersham) for 1.5 h using BioRad wet blotter in standard Towbin buffer. The membrane was blocked with 5% nonfat milk (AppliChem), incubated overnight with the following primary antibodies: anti-DP1 (St. John Laboratory, 1:1000), anti-GAPDH (Sigma, 1:10,000), precleared human sera (1:500). The epitope blocking peptide #1 was used in final concentration 6.6 μg/mL. The membrane was incubated for 1 h at room temperature with the secondary anti-mouse, anti-rabbit, or anti-human IgG antibodies (Abcam; dilution 1:10,000). The ECLfemto kit (Amersham) was used for detection of immunoblotted target proteins.

2.12. RNA Extraction, RT-PCR and qRT-PCR

Total RNA from human brain parts was extracted using RNAWiz (Ambion) as recommended by the manufacturer. Total RNA from cells was isolated using TRIzol® Reagent (Invitrogen) according to the manufacturer's instructions. One microgram of RNA was reverse transcribed into cDNA using SuperScript III first strand cDNA synthesis kit (Invitrogen) according to the manufacturer's instructions. The resulting cDNAs were used as templates for subsequent RT-PCR reactions. RT-PCR was carried out using FIREPol® DNA polymerase (Solis Biodyne), 40 amplification cycles and an annealing temperature of 58 °C. Amplification of the housekeeping gene *GAPDH* was performed for 25 cycles using FIREPol® DNA polymerase (Solis Biodyne) and used as an internal control. Used primer sequences: *PTGDR* sense 5′ATGAAGTCGCCGTTCTAC C3′, *PTGDR* antisense 5′CATGAAGAAGGCGAAGGCTTG3′, *GAPDH* sense 5′ GAAGGTGAAGGTCGGAGT3′, *GAPDH* antisense 5′GCATGGACTGTGGTCA TGAG3′. *IL-1β* sense 5′GGGCCTCAAGGAAAAGAATC3′: *IL-1β* antisense 5′ TTCTGCTTGAGAGGTGCTGA3′, *IFNγ* sense 5′CTGTTACTGCCAGGACCCA T3′, *IFNγ* antisense 5′TTTCTGTCACTCTCCTCTTTCCA3′.

3. Results

3.1. Autoimmune Response Profiles Across Cohorts are Highly Heterogeneous

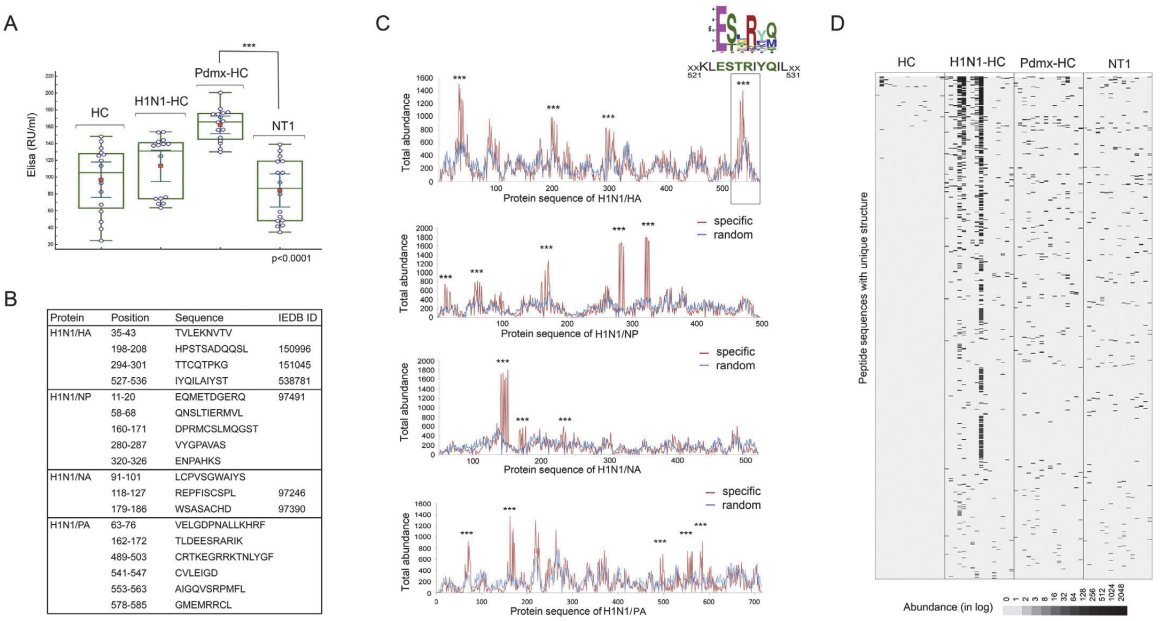
We performed MVA by selecting peptide antigens from random phage library (PhD12, NEB) with 10^9 different 12-mer peptide sequences based on their high avidity of interaction to antibodies in sera (Fig. 1A). A total dataset of 16 million peptides with unique sequences was generated. The data structure analysis of Top2500 peptide dataset (the most abundant peptides across individual samples) revealed that although these peptides were largely individual-specific (Fig. 1B and D), the study cohorts shared a fraction of common characteristics across Top2500 features (Fig. 1B). The remarkable heterogeneity of antigenic reactivity between individuals has also been noted previously (Zandian et al., 2017). However, the distribution of peptides according to the frequency was found to be similar in the different clinical subsets (Fig. 1C).

3.2. H1N1-specific Immunoprofiles are Largely Shared Between Pdmx-Vaccinated and Subjects Infected With H1N1

To evaluate the extent to which the presence of H1N1-specific peptides was restricted to specific clinical subsets, we assayed responses to H1N1 infection and Pdmx-vaccination using type A influenza ELISA (Quantum) diagnostic tests. High-titer responses to influenza A virus major antigens (including H1N1) were evident for both Pdmx-vaccinated and H1N1 naturally infected individuals (Fig. 2A). The humoral response to seasonal flu (A/H1N1 and A/H3N2) was relatively weaker in NT1-diseased as compared with Pdmx-HC individuals as determined by using a commercial ELISA test ($p < .001$). This was in slight contrast to earlier findings reporting that Pdmx-NT1 patients had higher median levels of anti-H1N1 antibodies than controls (Lind et al., 2014), and may reflect the characteristics of the samples collected (see Table 1, Materials and methods). Next, we assessed the reactivity of the sera to protein fragments representing the four major antigens of H1N1 virus proteome (strain A/California/7/2009). MVA data analyses of Top2500 peptide data set revealed 4 antigenic regions for hemagglutinin (H1N1/HA, C4RUW8), 5 for neuraminidase (H1N1/NA, C3W6G3), 3 for nucleoprotein (H1N1/NP, B4URE0), and 6 for polymerase acidic protein (H1N1/PA, I6THC5), some of which corresponded to known immunogenic epitopes from IEDB (http://www.iedb.org/; Fig. 2B). Statistically distinct coverage profiles with different peaks on H1N1 HA, NA, NP, and PA antigens were obtained from analysis of Top2500 peptide data sets of H1N1-HC, Pdmx-HC and NT1 samples (Fig. 2C). Data showed that the most commonly shared epitopes raised by the anti-Pdmx/anti-H1N1 immune response were found in the C-terminal region of H1N1/HA (C4RUW8) locating between amino acids 521 to 531 (Fig. 2B–C), directly before a proven T cell-antigenic region in HA between amino acids 527–541 of A/California/04/2009 (H1N1, (Schanen et al., 2011)). About 700 peptides from the total peptide dataset clustering to motif with sequence consensus E[ST].R.[QM] were highly abundant in H1N1-HC, and relatively infrequent in Pdmx-HC and NT1 samples as compared with HCs (Fig. 2D).

3.3. Examination of Identified NT1-specific Autoantigens in MVA Dataset

Next we determined peptides that were different between the clinical study groups to examine whether they were consistent with the prior knowledge of Pdmx-NT1-specific immunogenic epitopes. For the study, we used an exhaustive sequence pattern search (SPEXS - https://github.com/egonelbre/spexs2; (Vilo, 2002, Brazma et al., 1998)) gene ontology analysis, combined with the interrogation of the presence of known autoantigens previously identified in Pdmx-NT1 disease (Table S1(Ahmed et al., 2015, Bergman et al., 2014, Cvetkovic-Lopes et al., 2010, De La Herran-Arita et al., 2013, Haggmark-Manberg et al., 2016, Katzav et al., 2013, Zandian et al., 2017)). Thus, we were able to confirm statistically significant patterns of epitope recognition in the samples. Particularly, we identified epitopes resembling those in the N-termini of OX (4/16) and OX1R/2R (2/16; 4/16), in mitogen-activated protein kinase 7 (MAP3K7) (amino acids 318–328; 3/16) and in 5′-nucleotidase cytosolic IA (NT5C1A) (amino acids 35–48; 2/16), as well as in B-cell lymphoma 6 protein (BCL6), encompassing amino acids 279–288 in 6 out of 16 sera samples of NT1 diseased (Fig. 3, Table S1). According to MVA data, none of the previously identified antigens was prominently detected across NT1 diseased and were also common also to HC if less stringent statistical power criteria were used (Fig. 3). In contrast, we found no evidence of stratifying peptides with consensus sequences mimicking tribbles pseudokinase 2 (TRIB2), neuropeptide glutamic acid- isoleucine/α-melanocyte-stimulating hormone (NEI/aMSH), or others that were reported by earlier studies



Protein	Position	Sequence	IEDB ID
H1N1/HA	35-43	TVLEKNVTV	
	198-208	HPSTSADQQSL	150996
	294-301	TTCQTPKG	151045
	527-536	IYQILAIYST	538781
H1N1/NP	11-20	EQMETDGERQ	97491
	58-68	QNSLTIERMVL	
	160-171	DPRMCSLMQGST	
	280-287	VYGPAVAS	
	320-326	ENPAHKS	
H1N1/NA	91-101	LCPVSGWAIYS	
	118-127	REPFISCSPL	97246
	179-186	WSASACHD	97390
H1N1/PA	63-76	VELGDPNALLKHRF	
	162-172	TLDEESRARIK	
	489-503	CRTKEGRRKTNLYGF	
	541-547	CVLEIGD	
	553-563	AIGQVSRPMFL	
	578-585	GMEMRRCL	

Fig. 2. A novel common epitope of HA antigen of A/H1N1 in seasonal infection carriers and Pdmx-vaccinated individuals, encompassing a proven T cell- antigenic region. A. The humoral response to seasonal flu (A/H1N1 and A/H3N2) was relatively weaker in NT1-diseased as compared with Pdmx-HC individuals as determined by using a commercial ELISA test ($p < .001$). The presence of IgG antibodies against HA and H3N2 was assessed in sera samples of *HC*, *H1N1-HC*, *Pdmx-HC* and *Pdmx-NT1* by ELISA (Influenza virus type A IgG ELISA, Euroimmun). The HC samples were collected prior to A/H1N1 outburst in Estonia, before the fall 2009. *Blue circles* mark individuals of the study cohorts; *red dots* mark the mean values; *lines* depict median values; *inner whiskers* mark confidence interval for the mean; *boxes* mark upper and lower quartiles; *outer whiskers* mark the maximum and minimum values (excluding the outliers). *P*-values were calculated by ANOVA and are marked with asterisks. The cut-off value for ELISA was 16 RU/mL. Labels at the top of box plots demark the clinical origin of the sample. B. MVA predicted H1N1 epitopes partially overlapped with previously described H1N1 (A/California/08/2009(H1N1)) B cell specific epitopes from IEDB (http://www.iedb.org/). Top2500 peptide dataset containing 121,142 unique sequences was used to delineate the predominant epitopes of H1N1/HA (GI: 238,623,304), H1N1/NA (NA, GI:758899360), H1N1/NP (NP, GI:229891180) and H1N1/PA proteins in study samples. 9657 peptides from the studied dataset satisfied the selection criteria that these were not present in the HC samples. Specific alignment profiles for each of the A/H1N1 protein antigens were calculated with the criterion that the abundance of a peptide was to be 2-fold higher over random. C. MVA immunoprofiles predicted a novel epitope in the C-terminal region of HA encompassing amino acids 521–531 and with the sequence ESxRxQ that was common to both seasonal infection carriers and Pdmx-vaccinated individuals. The graphs show antigen-specific profiles of overall peptide abundance where the number of peptides were counted for each amino acid position for the following proteins: hemagglutinin (H1N1/HA, C4RUW8), neuraminidase (H1N1/NA, C3W6G3), nucleoprotein (H1N1/NP, B4URE0) and polymerase acidic protein (H1N1/PA_I6THC5). Amino acid sequence of the proteins is depicted on the x-axis. Marked with asterisks are regions where set calculation criteria were satisfied. Detailed analysis of immunoprofiles of H1N1 antigens revealed a novel immunogenic region of HA encompassing amino acids 521–531 that corresponds to the earlier experimentally determined A/Puerto Rico/8/1934 (H1N1) HA520–530 CTL epitope (Gianfrani et al., 2000) and is partially overlapping with broadly reactive CD4+ T cell epitope: HA527–541 of A/California/04/2009(H1N1) (Schanen et al., 2011). Peptides aligning to 521–531 of HA cluster to a minimal consensus sequence E[ST].R.[QM] by sequence homology alignment. D. Heat map image of immunoprofiles of peptides with consensus E[ST].R.[QM] across study samples. The total peptide data set was examined for the peptides with unique structure clustering to E[ST].R.[QM] motif. About 700 peptides with enriched abundance in H1N1-HC, Pdmx-HC or Pdmx-NT1 samples were found to cluster to the motif. The data of 700 peptides is presented on the heat map image. Each line represents peptides with unique sequence structure. The colour intensity of each cell corresponds to the peptide abundance (presented in log value). Black represents peptides captured at higher abundance whereas white represents peptides captured at lower abundance. Each column represents a peptide profile from a single sample. Labels at the top of the panels indicate the clinical origin of the sample. Abbreviations: *Random alignment* – amino acid sequences of peptides under analysis were randomized and aligned to respective protein coding sequence; *Total abundance* - the number of peptides counted for defined amino acid positions; *HC*– healthy control; *H1N1-HC* – H1N1 infected; *Pdmx-HC*- Pandemrix-vaccinated; *NT1*- narcolepsy type 1 (including 10 Pandemrix-induced NT1 samples).

(Table S1). These data allowed concluding that apart from the BCL6 related subset, peptides corresponding to previously identified autoantigens had relatively little discriminative power, suggesting also that these antigens were either rare or recognized promiscuously in patient groups with a clinical and ethnical heterogeneous background.

3.4. A Defined Set of Peptides Derived From DP1 Acts as Antigenic Epitopes in NT1

In analyses of the peptides that were unique among the disease groups, we observed that the Top2500 dataset contained >1300 peptides with a high enrichment in NT1 (Fig. S1A–B). The most abundant peptide having the sequence RVLAPALDSWGT showed a high sequence homology within the second extracellular loop region in the human prostaglandin D2 receptor DP1 (Q13258). This region in DP1 is predicted to function in ligand recognition (Avlani et al., 2007; Nagata et al., 2017) and is not conserved in mouse and rat (Fig. 4A). Extraction of all peptides from the total dataset having the highest homology to RVLAPALDSWGT and to DP1 revealed a set of 4428 unique peptides containing the RxxxPxxD (RPD) consensus sequence that discriminated the NT1 samples from controls ($p < .0001$, ANOVA, Fig. 4B–C). We then also determined that the 2157 RPD-containing peptides out of 4428 (Fig. 4B) had a high sequence homology to DP1 protein where the bona fide immunodominant epitope with sequence RVLAPALD encompassed amino acids 94 to 101 in DP1 (Fig. 4D). Interestingly, according to the IEBD database (www.iedb.org/), four MHC-I binding epitope regions of DP1 encompassing amino acids 132–140 (ID: 716767), 145–156 (ID: 637966), 195–203 (ID: 727099) and 303–311 (ID: 697995) have been defined. The latter (303—311) encoded another extracellular domain of DP1 that was also defined by us a potential target of B cell response (Fig. 4D).

To validate the data, we employed different methods and measured the serologic response to peptides carrying the RPD consensus sequence using sera samples of the study (Fig. 5–6). In line with previous reports

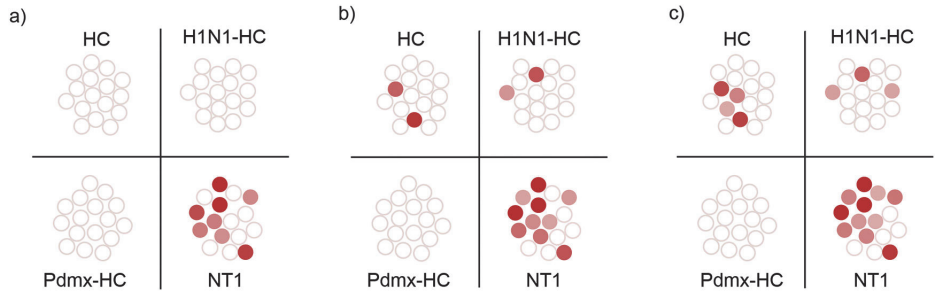


Fig. 3. Heterogeneity of the immune response to delineated antigens is apparent at the individual level. The total peptide data set was examined for peptides with homology to different antigens (see Table S1). Homology alignment analysis resulted in delineating six minimal consensus motifs with homologies to epitopes defined by previous studies. Namely, LPxTNxS (HCRT, O43612), RDxxYP (HCRTR1, O43613), SxLNxTxN (HCRTR2, O43614), KxxxPSAS (BCL6, P41182), STNxS (MAP3K7, O43318), LAxSxKP (NT5C1A, Q9BXI3), (Table S1). Datasets from different study cohorts were analyzed for the presence of peptides containing the 6 motifs and ROC analyze was performed at threshold values: a) specificity 100% (sensitivity 50%), b) specificity 90%, (sensitivity 63%), c) specificity 85% (sensitivity 69%). Note that the ensemble snapshots from MVA data did not distinguish between the possible scenarios of each individual antigen motif in single individuals. Individual samples are colour coded. White circles indicate that peptides with abundance values clustering to one or more aforementioned motifs were below threshold. Colored circles indicate peptides with abundance values clustering to one or more aforementioned motifs were above threshold. None of the motifs displayed a statistically significant association with many other known antigens (see the list of antigens in Table S1). This negative result may be due to a limited statistical power of delineated antigens, but also due to the heterogeneity of autoantibody repertoires in different individuals. Abbreviations: *HC*- healthy control; *H1N1-HC* – H1N1 infected; *Pdmx-HC*- Pandemrix-vaccinated; *NT1*- narcolepsy type 1 (including 10 Pandemrix-induced NT1 samples).

(Urade and Hayaishi, 2011), we observed a broad expression of *DP1* mRNA across different regions of human brain, in human glioma (hGC), and normal mesenchymal stem cells (hMSCs) (Fig. 5A). Treatments of hMSCs with the ligand prostaglandin D2 increased *DP1* expression, whilst the pro-inflammatory cytokines interleukin 1β and interferon γ (IL-1β and IFN-γ) either slightly increased or decreased its expression, respectively (Fig. 5B). In contrast, PGD2 strongly reduced *IL-1β* and *IFN-γ* expression suggesting that these cells recapitulate the intact PGD2-DP1 signaling pathway by inducing anti-inflammatory responses in the studied cells (Fig. 5B). Performing phage Western blot assay we confirmed that MVA predicted DP1-positive Pdmx-NT1 sera showed IgG reactivity to phages that displayed RVLAPALDSWGT peptides (RVLAPALD-pIII, Fig. 5C). No specific reactivity was detected using DP1-negative sera (Fig. 5C). This peptide target specificity was further confirmed by Western blot analysis using Pdmx-NT1 sera where the interactions between human IgGs and antigen expressing phages were blocked by **RVLAPALD**SWGT synthetic peptides (Fig. 5C, Fig. S2). Immunoblot analysis using commercial anti-DP1 polyclonal sera and clinical sera of Pdmx-NT1, confirmed the presence of DP1 expression in hGC_1 and not in hGC_2 glial cells and also here specific blocking effects to the seroreactivity in the presence of **RVLAPALD**SWGT peptides were confirmed (Fig. 5D), but not in case of control peptides or irrelevant sera (Fig. 5C, Fig. S2). Immunocytochemical analyses showed that DP1 was predominantly localized on the cell surface of hMSC and hGC cells, and more importantly, was equally well-detected by immunocytochemistry using commercial anti-DP1 polyclonal sera and Pdmx-NT1 clinical sera (Fig. 5E). Furthermore, the synthetic peptide **RVLAPALD**SWGT competed for the binding of anti-DP1 antibodies present in sera of Pdmx-NT1 diseased (Fig. 5E). Data combined from Western blot and immunocytochemistry analysis suggested that peptide **RVLAPALD**SWGT could embed a structural as well as a linear epitope given that upon competition it interfered with DP1-specific serorecognition of globular as well as denatured epitopes (Fig. 5D and E).

We next studied whether the peptides identified here could be developed to an ELISPOT assay to discriminate sera in different disease groups. ELISPOT analysis data showed that peptides containing H1N1/HA-specific sequence ESTRYQL (peptide_3) discriminated between naturally H1N1 infected and healthy samples with no earlier H1N1 infection (ANOVA $p < .001$, Fig. 6A). RVLAPALD (epitope on DP1) and KAPSAS (epitope on BCL6) (peptide_1 and _2) peptides that were selected upon MVA data, correctly assigned upon ELISPOT analysis the NT1 group from HC samples (ANOVA $p < .001$, Fig. 6B). Combined ELISPOT analysis using all 3 peptides, could correctly classify 11 out of 16 NT1 (specifically - 7 Pdmx-NT1 and 3 NT1) samples across all controls ($p < .001$ Chi-squared test, Fig. 6C). Notably, majority of the DP1 and BCL6-peptide-positive NT1 samples had undetectable OX findings from the related CSFs (with average values of 6.1 pg/mL), whereas those 4 that were negative by our ELISPOT measurements, had OX levels in respective CSF samples still low but in detectable range (with average values of 77 pg/mL (Fig. 6C and see Materials and methods). Unlike the IgG response, the IgM levels in response to the tested peptides were low or absent in all studied individuals (data not shown). These findings confirmed that peptides carrying the epitope motifs identified in the study could be used in ELISPOT analysis to develop a novel multi-biomarker diagnostic assay for NT1.

4. Discussion

Despite extensive research using biomarker and neurophysiological approaches, known heterogeneity among NT1 diseased is not always consistent with serologic marker-based subtype classification schemes. Using an unbiased analysis of serum samples from single individuals, we detected a high variance in humoral immune response profiles, both in healthy and diseased people. We found that variance in immunoprofiles representing multifactorial heterogeneity of NT1 clearly determined distinct disease-specific serological profiles. We focused our analysis on peptides specific to Pdmx-immunized and -NT1 diseased subjects, which encompassed vaccine antigens and autoantigens in order to have a full coverage of potential triggers of the disease. Our results show that patients with NT1 exhibit a specific immune response to epitopes of receptor DP1. This finding highlights the importance of the PGD2-DP1 pathway in the functioning of sleep-wake homeostasis as suggested by the role of DP1 in slow-wave sleep (Terao et al., 1998). However, the precise mechanism by which PGD2-DP1 signaling may influence orexinergic neurons and immune regulation in NT1 requires further studies. In addition, using MVA-based immunoprofiling, we discovered epitopes, such as those of the protein BCL6 specific for Pdmx-NT1 and sNT1 patient group. This underscores the complexity of NT1 with different molecular targets and pathways involved and contributing to the immune response. Enhanced inflammation due to immune system malfunction has been detected in human narcoleptics in the regions of OX cell loss (Bassetti et al., 2010; John et al., 2013; Nishino, 2011; Thannickal et al., 2000, 2003,

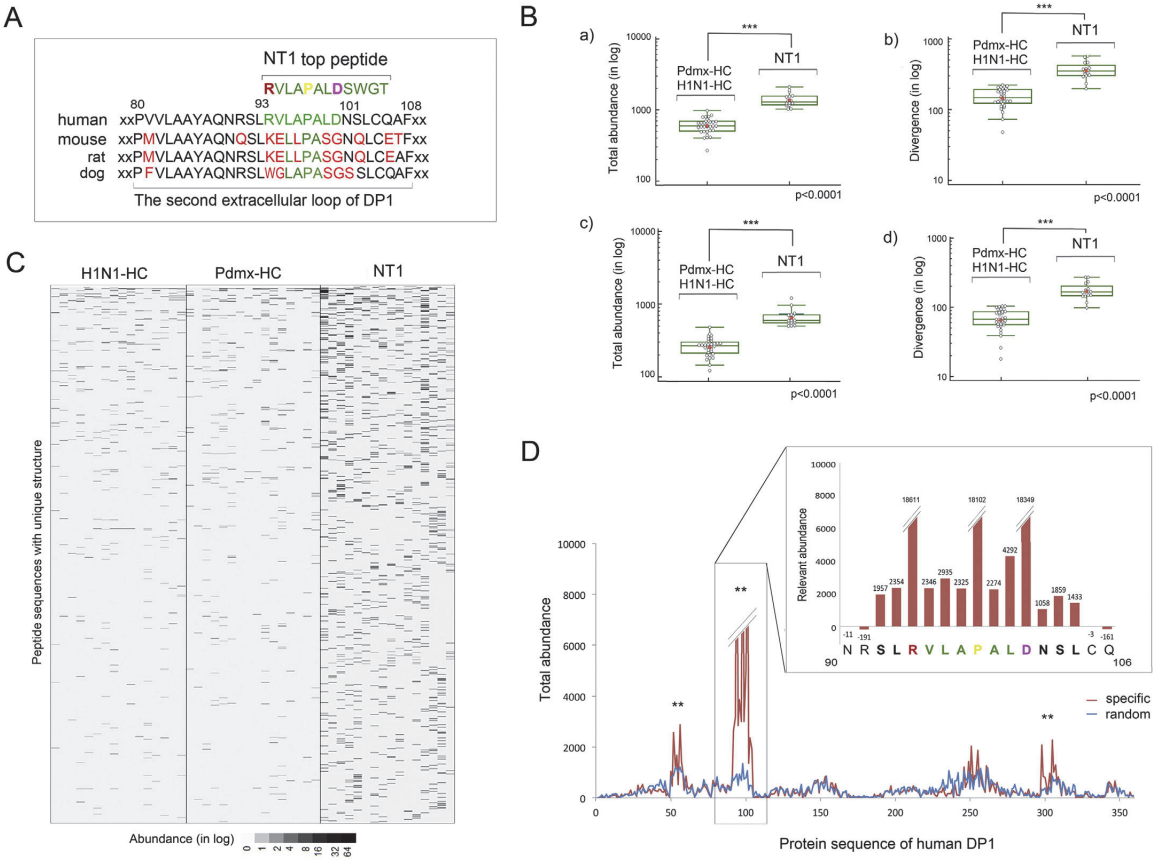


Fig. 4. Peptides with minimal epitope motif RxxxPxxD (RPD) are enriched in samples of NT1 and align to the extracellular loop of human DP1. A. Peptide RVLAPALDSWGT found as the most abundant in Pdmx-NT1 was highly homologous to the second extracellular loop region of DP1 protein. More than 1,300 peptides from the selection of the Top2500 peptide dataset were enriched in NT1 samples (Fig. S1A–B) where the peptide RVLAPALDSWGT had the highest values of abundance across the Pdmx-NT1 cohort. BLAST analysis revealed that RVLAPALDSWGT region of DP1 is highly dissimilar in mammals. To identify peptides from Totpep dataset with the highest homology to DP1 and to RVLAPALDSWGT, SPEXS2 software (https://github.com/egonelbre/spexs2.) was used and the top motif that fulfilled both criteria was found to be RxxxPxxD (RPD, a motif defined by three shared amino acids). B. Peptide sequences containing RPD motif were enriched in immunoprofiles of NT1 samples. Analyzing the Totpep library, 4428 peptides were identified from NT1 (including 10 Pdmx-induced NT1 samples and 6 sNT1 samples) data sets that contained the RPD motif. Box plot depicting that these 4428 peptides (with RVLAPALDSWGT eliminated beforehand as dominant) contained sufficient information to discriminate between HC and NT1 samples in a statistically significant manner a) by their abundance (in log, student *t*-test *p*-value < .0001) and b) divergence (in log, student *t*-test *p*-value < .0001). SPEXS2 analysis resulted in 2157 peptides that out of 4428 were highly homologous (with at least 4 consecutive amino acid matches) to RVLAPALD of human DP1 and discriminated NT1 samples (including 10 Pdmx-induced NT1 samples and 6 sNT1 samples) in a statistically significant manner c) by their abundance (in log, student *t*-test *p*-value < .0001) and d) divergence (in log, student *t*-test *p*-value < .0001). Six NT1 samples that were not Pdmx-induced were similarly to Pdmx-NT1 samples discriminated by the set of 4428 peptides from HC. In box plots - *blue circles* mark the single individuals of study cohorts; *red dots* mark the mean values; *line* marks the median values; *inner whiskers* mark confidence intervals for the mean; *boxes* mark the upper and lower quartiles; *outer whiskers* mark the maximum and the minimum values (excluding outliers). C. Heatmap images depicting the immunoprofiles of the top one thousand NT1-specific peptides out of 4428 across the clinical study-groups. The data are presented as heat map image generated via conditional formatting in MS Excel. Each column represents a peptide profile from a single individual. Each line represents peptides with unique sequence structure. The colour intensity of each cell corresponds to the peptide abundance (counts of sequences in log). Black represents peptides captured at higher abundance whereas white represents peptides captured at lower abundance. D. Epitope mapping of anti-peptide response to human DP1 in sera of Pdmx-NT1 diseased. SPEXS2 analysis resulted in 2157 peptides carrying RxxPxxxxD motif that were aligned to DP1 protein sequence (Q13258) with a 2-fold higher abundance over random as a chosen criterion. In addition, MEME-MAST algorithm (Bailey and Elkan, 1994) aligned these 2157 peptides to the region encompassing 94-101aa of DP1, E-value = 0.0078. Each bar on the x-axis corresponds to one of the overlapping peptides required to cover the antigen, and the height of the profiles shows the relative abundance. Three potentially immunogenic regions were described with the predominant alignment containing the sequence RVLAPALD and encompassing amino acids 94 to 101. Zoomed in the box is an extract of the immunoprofile of DP1 in positions 90–106. Calculated relative abundance values are marked above each amino acid position. Note that the weaker immunogenic regions of DP1 are expanded toward both N- and C-termini. Abbreviations: *Abundance* – peptide frequency; *Divergence* – all unique peptides; *HC*- healthy control; *H1N1-HC* – H1N1 infected; *Pdmx-HC*- Pandemrix-vaccinated; *NT1*- narcolepsy type 1 (including 10 Pandemrix-induced NT1 samples).

2009), but the exact factors or mediators leading to the ultimate death of OX neurons are yet unknown.

Here we show that the DP1 receptor is linked to NT1 by acting as a possible antigen in the disease process. Prostaglandins play a key role in the inflammatory response and their synthesis is significantly increased after tissue injury and cell stress (Ricciotti and Fitzgerald, 2011). PGD2 is a major eicosanoid both in the Central Nervous System (CNS) and peripheral tissues with a role in inflammation as well as homeostasis (Jowsey et al., 2001; Vijay et al., 2017). PGD2 is abundantly produced by mast cells and Th2 cells, and among a wide range of

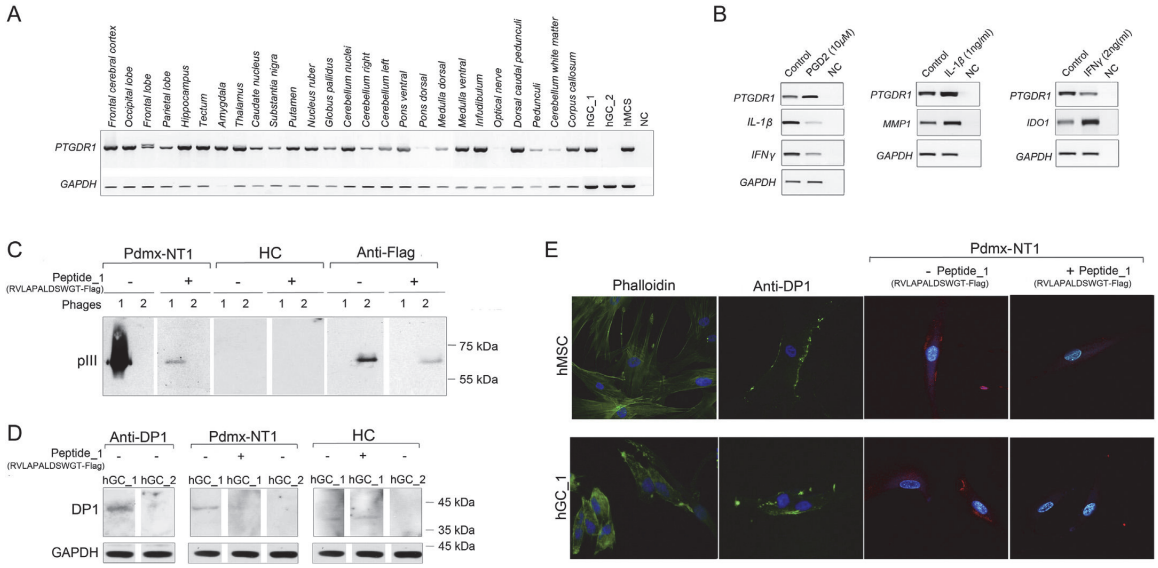


Fig. 5. Validation of DP1 as a true antigenic target in NT1 disease. A. Human *DP1* is expressed widely in different brain regions, and by mesenchymal and cancer stem cells. Cerebral cortex: *Frontal cerebral cortex*, *Occipital lobe*, *Frontal lobe*, *Parietal lobe*, *Hippocampus*; brain nuclei: *Tectum*, *Amygdala*, *Thalamus*, *Caudate nucleus*, *Substantia nigra*, *Putamen*, *Nucleus ruber*, *Globus pallidus*; *Cerebellum*: *Cerebellar nuclei*, *Cerebellum right*, *Cerebellum left*; Brainstem: *Ventral pons*, *Dorsal pons*, *Dorsal medulla*, *Ventral medulla*; Axonal tracts: *Infundibulum*, *Optical nerve*, *Dorsal caudal pedunculi*, *Pedunculi*, *Cerebellum white matter*, *Corpus callosum*; human glioma cells (hGC_1, hGC_2); human mesenchymal stem cells (hMSC); NC – negative control. PCR analyses were done using specific primers for human *DP1*. *GAPDH* mRNA expression was used to normalize the data across samples. B. PGD2/DP1 signaling is associated with inflammation regulation. Expressions of human *DP1*, *IL-1β*, *MMP1*, *IFNγ* and *IDO-1* were analyzed by PCR in human mesenhymal stem cells treated with PGD2 (10 μM) or cytokines IL-1β (1 ng/mL) and IFNγ (2 ng/mL). NC – negative control. *GAPDH* mRNA expression was used to normalize the data across samples. C. RVLAPALD was identified as a target antigen sequence for NT1-specific polyclonal IgG response. Western blot analysis of phage particles containing the RVLAPALD-pIII (phage 1) or FLAG-pIII fusion proteins (phage 2) show that human Pdmx-NT1 serum (dilution 1:500) reacted specifically with the pIII protein containing the peptide RVLAPALDSWGT sequence, but not with the phage backbone or FLAG-pIII fusion protein. Duplicate membranes were incubated with Pdmx-NT1 sera treated with the synthetic peptide (RVLAPALDSWGTGGGDYKDDD: final conc 6.6 μg/mL) that significantly blocked the interaction between phage #1 and human IgG similarly to anti-FLAG antibody (dilution 1:2000) and phage #2. RVLAPALDSWGT-pIII fusion protein was not detected by HC sera (1: 500). Protein size markers are indicated at the right side of blot. D. NT1-specific seroreactivity to DP1 protein is specifically blocked by RVLAPALD peptide. Western blot analysis of endogenous levels of DP1 protein (MW 40 kDa) in human glioma hGC_1 and hGC_2 cells using anti-PTGDR1 (DP1) polyclonal antibodies (1:500) (*left*, first panels). Note that, hGC_2 cells were negative of DP1 expression. The use of the Pdmx-NT1 serum (1:500) showed similar pattern of DP1 reactivity in hGC_1 and hGC_2 cells, and the DP1-specific signal was attenuated by pre-treatments of Pdmx-NT1 sera with a synthetic peptide #1 (final conc 6.6 μg/mL). Anti-GAPDH monoclonal antibody (1:10,000) was used as a control for immunoblots. E. DP1 expressed by hMSCs and cancer was specifically blocked by RVLAPALD peptide. IF analysis of DP1 in hMSC and glioma cells. The antibodies used included: anti-PTGDR1 (1:500; green), Pdmx-NT1 serum (1:400; red) and the secondary Alexa Flour 488 and 647 (Invitrogen, 1:2000) antibodies. For antibody-blocking, Pdmx-NT1 sera (1:400) and synthetic peptide #1 (final conc 6.6 μg/mL) were used. Cells were analyzed for phalloidin-labelled cytoskeleton proteins (green, left) and nuclear structures (Hoechst 33342, blue). Abbreviations: *Pdmx-NT1*- Pandermix-induced narcolepsy type 1; *HC* – healthy control.

other body cells (see ref. in Farhat et al. (2011)). PGD2 elicits its downstream effects by activating DP1 and DP2 receptors with opposing effects on cyclic AMP (cAMP) production, and/or phosphoinositol turnover and intracellular Ca2+ mobilization (Liang et al., 2005). In the brain, PGD2 regulates sleep, body temperature, and nociception and its levels exhibit marked changes in different neuropathologies (reviewed in (Liang et al., 2005, Mohri et al., 2006, Urade and Hayaishi, 2011)). The microglial PGD2-DP1 pathway is also known to mediate neuronal damage through microglial activation (Bate et al., 2006; Vijay et al., 2017).

Among cells expressing DP1, mast cells (MCs) can release histamine and other factors that affect sleep and the immune response in the brain. Accumulating evidence shows that MCs play a role in the regulation of sleep and behavior (Chikahisa et al., 2013). MCs are most abundant in young individuals under the age of 19, after which their counts decline with age (Porzionato et al., 2004; Turygin et al., 2005). Most significantly, the maturation of MCs is influenced by PGD2 and the receptor DP1 (Taketomi et al., 2013). In addition to histamine, DP1 signaling may also influence the levels of adenosine that is known to regulate sleep (Urade and Hayaishi, 2011). The precise role of MCs and their released factors such as histamine in NT1 warrants further studies.

In conclusion, the present study shows that anti-DP1 antibodies are autoimmune agents in the course of NT1 prompting more studies on the role of PDG2-DP1 signaling in OX-signaling and in the disease. Currently DP1-selective agonist/antagonist therapies are considered in treating autoimmune disorders such as asthma (Maicas et al., 2012; Santini et al., 2016; Santus and Radovanovic, 2016). Our data also indicate that, depending on the antibody concentrations and affinities, anti-DP1 antibodies may modify the function of pharmaceutical compounds targeting PGD2-DP1–signaling pathways (Narumiya and Fitzgerald, 2001) that need to be taken into account in clinical studies.

Supporting evidence that the humoral response in the CNS is derived from different peripheral tissue antigens is provided by the findings that sera from NT1 diseased can bind brain and muscle structures (Ahmed et al., 2014; Smith et al., 2004). There is a plethora of data that genetic or experimental alterations of the OX system are associated with NT1, however, OXs are not restricted to the CNS and together with their receptors OX1R and OX2R are widely expressed in peripheral tissues (see ref. in (Voisin et al., 2003)). BCL6, another NT1 antigen, is a master regulator required in mature B-cells during germinal center (GC) reaction (Ref in Pei et al. (2017)). NRXN1-α has been isolated from brain and heart tissues suggesting a role also in heart development (Nagase et al., 1998). TRIB2 is present in many

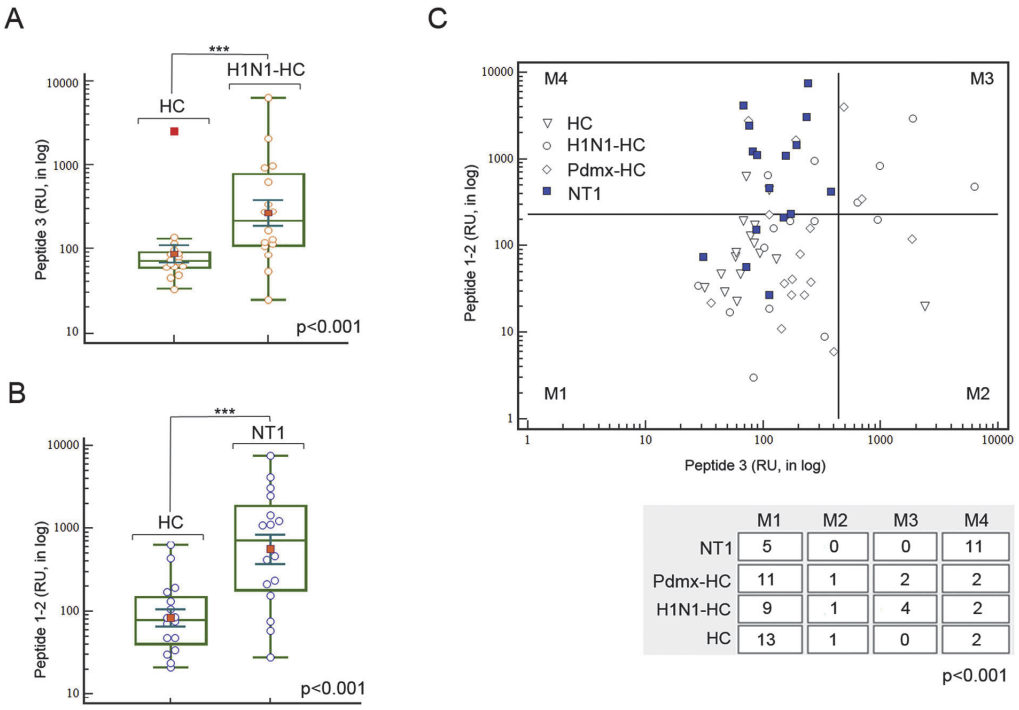


Fig. 6. Use of MVA-defined peptides for immunodiagnostic applications of NT1. A. ELISPOT analysis of peptides containing H1N1/HA-specific sequence ESTRYQL (peptide_3) confirmed the power of the peptide to discriminate samples of natural H1N1 infection from HC (ANOVA $p < .001$). In box plots - *yellow circles* mark the single individuals of study cohorts; *red dots* mark the mean values; *line* marks the median values; *inner whiskers* mark confidence intervals for the mean; *boxes* mark the upper and lower quartiles; *outer whiskers* mark the max and the min values (excluding outliers). B. ELISPOT analysis of peptides containing RVLAPALD (epitope of DP1) and KAPSAS (epitope of BCL6) (peptide_1 and _2, respectively) resulted in correct assignment of 10 Pdmx-NT1 samples and 6 sNT1 samples to NT1 group (ANOVA $p < .001$). On box plots - *blue circles* mark the single individuals of study cohorts; *red dots* mark the mean values; *line* marks the median values; *inner whiskers* mark confidence intervals for the mean; *boxes* mark the upper and lower quartiles; *outer whiskers* mark the max and the min values (excluding outliers). C. ELISPOT analysis data combined revealed the power of peptides 1, 2, and 3 to discriminate 11 (7 Pdmx-NT1 and 4 sNT1) out of 16 NT1 samples across all samples ($p < .001$ Chi-squared test). Thresholds were calculated by using ROC curve analysis and results were visualised using scatter plot analysis with divided threshold values (*M1-M4*). Statistical significance of differences was calculated by two-way classification Chi-square test (Chi-squared *p*-value > .001). Abbreviations: *HC*– healthy control; *H1N1-HC* – H1N1 infected; *Pdmx-HC*- Pandemrix-vaccinated; *NT1*- narcolepsy type 1 (including 10 Pandemrix-induced NT1 samples).

cell populations both in and outside the nervous system, including the immune cells (Eder et al., 2008; Sung et al., 2006). Gangliosides (anti-GM3) are abundant in the brain, but in extraneural tissues, relatively high concentrations of ganglio-series GMs were found in bone marrow, erythrocytes, intestine, liver, spleen, testis, kidney, and in embryonic stem cells (Kolter, 2012). NEI-MCH has mostly been detected in peripheral organs (Viale et al., 1997). NT5C1A is highly abundant in skeletal muscle tissue (Hunsucker et al., 2001). GLS2 is expressed specifically in the liver, but also in extrahepatic tissues, like the brain, pancreas, cells of the immune system (ref in Martin-Rufian et al. (2012)). However, it remains elusive what pathogenic roles these antibodies against the above-mentioned proteins may exert within the periphery.

Our data of immunoprofiling support the existence of immune defects in multiple pathways associating NT1 to a) DP1 and PDG2/histamine associated disorders, b) BCL6 and the chronic status of latent herpesviruses (such as EBV), c) orexin/OX1/2R–related dysfunctions, d) stress and inflammation-associated mitogen-activated pathways (such as MAP3K7, also known as transforming growth factor (TGF)-β-activated kinase 1 (TAK1)), and e) adenosine-deficiency linked dysfunctions (involving NT5C1A) (Table S1, Fig. 7). Together these results provide a comprehensive map of potential molecular targets contributing to NT1 that can be of help in designing future strategies for the diagnostics and treatment of the disease. More broadly, our study demonstrates the usefulness of MVA as a method for disease classification and for the discovery of novel biomarkers that can be applicable to any human disease.

Funding Sources

The research leading to these results was mainly funded by Protobios's institutional research funding from the Estonian Research Council (contract no. 5.1-4/1373). KP was partially supported by the institutional research grant IUT19-18 of TUT from the Estonian Research Council. MP was partially supported by NARPA funding from the Academy of Finland (grant no. 260603), Sigrid Jusélius Foundation and Magnus Ehrnrooth Foundation. The EU Research and Innovation Staff Exchange (RISE) program (SZ_TEST H2020-MSCA-RISE-2016 (EU734791)) provided additional funding. The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the manuscript.

Conflicts of Interest

AV, OV, MP, KP, TN, and KP are co-applicants on a pending patent application related to diagnosing of narcolepsy (WO 2017/203106). All other authors declare no competing interests.

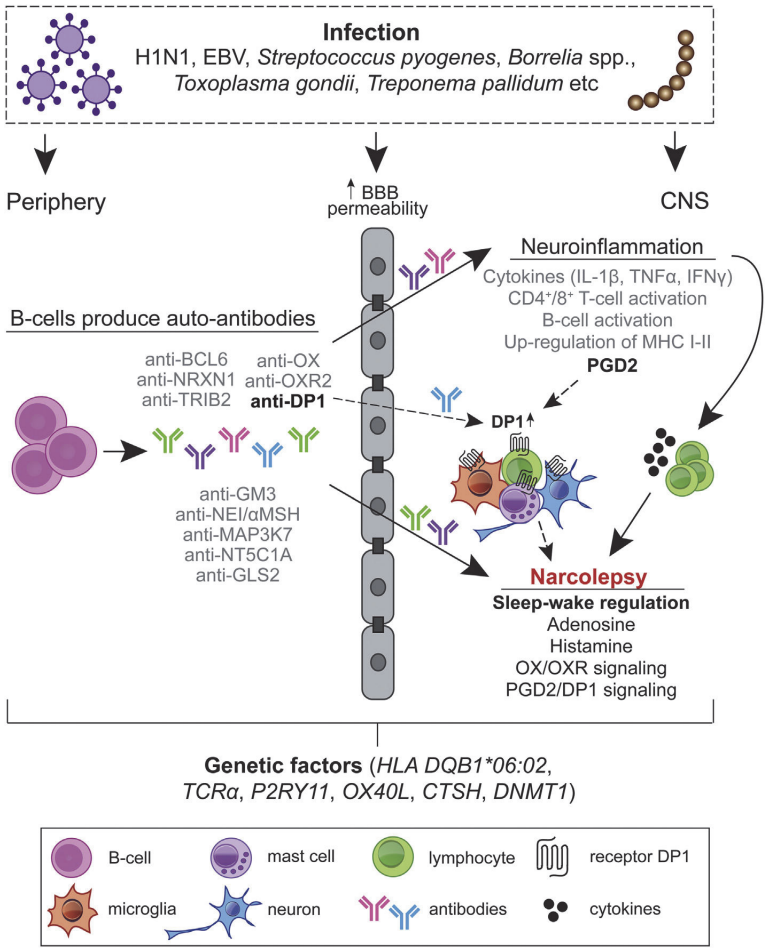


Fig. 7. Hypothetical model for the aggravation of autoimmune response in Pdmx-associated and spontaneous NT1. The immune response in NT1 is highly heterogenic with different pathways affected during the disease progression. We favor the idea that the lifelong risk for NT1 or for disease aggravation in pre-disposed individuals is increased following inflammatory triggers upon breaching of the blood–brain barrier (BBB) and with activation of preexisting auto-reactive antibodies (Ab) and cells reaching brain. Consequently, an immune response to A/H1N1 (and subsequent molecular mimicry) or a generalized stimulation of the immune system mediated by the Pdmx vaccine as AS03-adjuvanted vaccine can act as the inflammatory trigger (Morel et al., 2011; Carmona et al., 2010; Meyer et al., 2011). The inflammation triggers include i) *infections* (examples of pathogens are shown), ii) *genetic factors*, or iii) chronic *inflammation* (Kornum et al., 2011). The polyclonal Ab response from peripheral tissue may initiate disease by concentrating antigens in the brain to presentation-competent cells (Getahun et al., 2004). Recent data show further that peripherally produced human anti-CNS reactive antibodies are capable of opsonizing human CNS antigens (Kinzel et al., 2016). The entry of immune cells (T cells, B cells, macrophages, microglia and mast cells) cause *neuroinflammation* with the release of cytokines that damage neurons including HCRT+ neurons involved in sleep/wake regulation. Production of auto-reactive antibodies as a secondary response to cell death of HCRT-or other brain-resident cells can occur via antigen presenting cells. Prostaglandins are part of the inflammatory response in the brain acting via specific receptors. In particular, DP1 is produced by astrocytes, oligodendrocytes, neurons, microglia and meningeal cells (Liang et al., 2005; Mohri et al., 2007; Beuckmann et al., 2000). *PGD2* signaling is known to prevent excessive inflammasome activation and may act as an anti-inflammatory pathway in the brain. Additionally, in brain residing mast cells, DP1 activity promotes maturation and histamine release (Taketomi et al., 2013). The latter is of particular interest given that histamine levels in the CSF of NT1-diseased are reduced (Nishino et al., 2009). Thus, our findings suggest that the anti-DP1 immune response, whether causal or sequel, can interfere with PGD2 signaling in the brain. The results provide also evidence that the dysfunctional DP1 network can be a target for diagnosis and intervention of NT1, a conclusion that warrants further investigations.

Author Contributions

HS, AP, TN, MP, AV, and KP contributed to the design of the study. MP, OV, and AV were members of NARPANord consortium, MP was the chairman. HS, AP, AK, SP, MJ, MP, and AV contributed to data collection. HS, AP, AK, SP, MJ, DL, and KP contributed to the development and implementation of the data analysis plan. JV and PA were responsible for data management and pattern recognition analysis of the MVA data. All authors were involved in data interpretation, drafting, review, and approval of the report, and the decision to submit for publication.

Acknowledgements

Our deep gratitude goes to Lagle Kasak, Urmas Liivas, Grete Rullinkov, Veronika Tammekivi, Helen Verev, Mirjam Luhakooder, and Maila Rähn (Protobios, Estonia) as well as to Anne Huutoniemi, Kirsi Aaltonen, Sanna Mäki, and Mira Utriainen (UH, Finland).

Appendix A. Supplementary Data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ebiom.2018.01.043.

References

Ahmed, S.S., Schur, P.H., Macdonald, N.E., Steinman, L., 2014. Narcolepsy, 2009 A(H1N1) pandemic influenza, and pandemic influenza vaccinations: what is known and unknown about the neurological disorder, the role for autoimmunity, and vaccine adjuvants. J. Autoimmun. 50, 1–11.

Ahmed, S.S., Volkmuth, W., Duca, J., Corti, L., Pallaoro, M., Pezzicoli, A., Karle, A., Rigat, F., Rappuoli, R., Narasimhan, V., Julkunen, I., Vuorela, A., Vaarala, O., Nohynek, H., Pasini, F.L., Montomoli, E., Trombetta, C., Adams, C.M., Rothbard, J., Steinman, L., 2015. Antibodies to influenza nucleoprotein cross-react with human hypocretin receptor 2. Sci. Transl. Med. 7, 294ra105.

Anastasina, M., Domanska, A., Palm, K., Butcher, S., 2017. Human picornaviruses associated with neurological diseases and their neutralization by antibodies. J. Gen. Virol. 98, 1145–1158.

Aran, A., Lin, L., Nevsimalova, S., Plazzi, G., Hong, S.C., Weiner, K., Zeitzer, J., Mignot, E., 2009. Elevated anti-streptococcal antibodies in patients with recent narcolepsy onset. Sleep 32, 979–983.

Avlani, V.A., Gregory, K.J., Morton, C.J., Parker, M.W., Sexton, P.M., Christopoulos, A., 2007. Critical role for the second extracellular loop in the binding of both orthosteric and allosteric G protein-coupled receptor ligands. J. Biol. Chem. 282, 25677–25686.

Ayoglu, B., Schwenk, J.M., Nilsson, P., 2016. Antigen arrays for profiling autoantibody repertoires. Bioanalysis 8, 1105–1126.

Bailey, T.L., Elkan, C., 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc. Int. Conf. Intell. Syst. Mol. Biol. 2, 28–36.

Bailey, T.L., Gribskov, M., 1998. Combining evidence using p-values: application to sequence homology searches. Bioinformatics 14, 48–54.

Bassetti, C.L., Baumann, C.R., Dauvilliers, Y., Croyal, M., Robert, P., Schwartz, J.C., 2010. Cerebrospinal fluid histamine levels are decreased in patients with narcolepsy and excessive daytime sleepiness of other origin. J. Sleep Res. 19, 620–623.

Bate, C., Kempster, S., Williams, A., 2006. Prostaglandin D2 mediates neuronal damage by amyloid-beta or prions which activates microglial cells. Neuropharmacology 50, 229–237.

Bergman, P., Adori, C., Vas, S., Kai-Larsen, Y., Sarkanen, T., Cederlund, A., Agerberth, B., Julkunen, I., Horvath, B., Kostyalik, D., Kalmar, L., Bagdy, G., Huutoniemi, A., Partinen, M., Hokfelt, T., 2014. Narcolepsy patients have antibodies that stain distinct cell populations in rat brain and influence sleep patterns. Proc. Natl. Acad. Sci. U. S. A. 111, E3735–44.

Beuckmann, C.T., Lazarus, M., Gerashchenko, D., Mizoguchi, A., Nomura, S., Mohri, I., Uesugi, A., Kaneko, T., Mizuno, N., Hayaishi, O., Urade, Y., 2000. Cellular localization of lipocalin-type prostaglandin D synthase (beta-trace) in the central nervous system of the adult rat. J. Comp. Neurol. 428, 62–78.

Bomfim, I.L., Lamb, F., Fink, K., Szakacs, A., Silveira, A., Franzen, L., Azhary, V., Maeurer, M., Feltelius, N., Darin, N., Hallbook, T., Arnheim-Dahlstrom, L., Kockum, I., Olsson, T., 2017. The immunogenetics of narcolepsy associated with A(H1N1)pdm09 vaccination (Pandemrix) supports a potent gene-environment interaction. Genes Immun. 18, 75–81.

Bonvalet, M., Ollila, H.M., Ambati, A., Mignot, E., 2017. Autoimmunity in narcolepsy. Curr. Opin. Pulm. Med. 23, 522–529.

Brazma, A., Jonassen, I., Vilo, J., Ukkonen, E., 1998. Predicting gene regulatory elements in silico on a genomic scale. Genome Res. 8, 1202–1215.

Carmona, A., Omenaca, F., Tejedor, J.C., Merino, J.M., Vaman, T., Dieussaert, I., Gillard, P., Aristegui, J., 2010. Immunogenicity and safety of AS03-adjuvanted 2009 influenza A H1N1 vaccine in children 6-35 months. Vaccine 28, 5837–5844.

Chikahisa, S., Kodama, T., Soya, A., Sagawa, Y., Ishimaru, Y., Sei, H., Nishino, S., 2013. Histamine from brain resident MAST cells promotes wakefulness and modulates behavioral states. PLoS One 8, e78434.

Christiansen, A., Kringelum, J.V., Hansen, C.S., Bogh, K.L., Sullivan, E., Patel, J., Rigby, N.M., Eiwegger, T., Szepfalusi, Z., DE Masi, F., Nielsen, M., Lund, O., Dufva, M., 2015. High-throughput sequencing enhanced phage display enables the identification of patient-specific epitope motifs in serum. Sci. Rep. 5, 12913.

Cvetkovic-Lopes, V., Bayer, L., Dorsaz, S., Maret, S., Pradervand, S., Dauvilliers, Y., Lecendreux, M., Lammers, G.J., Donjacour, C.E., DU Pasquier, R.A., Pfister, C., Petit, B., Hor, H., Muhlethaler, M., Tafti, M., 2010. Elevated Tribbles homolog 2-specific antibody levels in narcolepsy patients. J. Clin. Invest. 120, 713–719.

De La Herran-Arita, A.K., Kornum, B.R., Mahlios, J., Jiang, W., Lin, L., Hou, T., Macaubas, C., Einen, M., Plazzi, G., Crowe, C., Newell, E.W., Davis, M.M., Mellins, E.D., Mignot, E., 2013. CD4+ T cell autoimmunity to hypocretin/orexin and cross-reactivity to a 2009 H1N1 influenza A epitope in narcolepsy. Sci. Transl. Med. 5, 216ra176 (*retracted*).

Eder, K., Guan, H., Sung, H.Y., Ward, J., Angyal, A., Janas, M., Sarmay, G., Duda, E., Turner, M., Dower, S.K., Francis, S.E., Crossman, D.C., Kiss-Toth, E., 2008. Tribbles-2 is a novel regulator of inflammatory activation of monocytes. Int. Immunol. 20, 1543–1550.

European Medicines Agency, 2009. Annex I Summary of Product Characteristics. Pandemrix: European Public Assessment Report, pp. 1–20 (first published).

Farhat, A., Philibert, P., Sultan, C., Poulat, F., Boizet-Bonhoure, B., 2011. Hematopoietic-prostaglandin D2 synthase through PGD2 production is involved in the adult ovarian physiology. J. Ovarian Res. 4, 3.

Getahun, A., Dahlström, J., Wernersson, S., Heyman, B., 2004. IgG2a-mediated enhancement of antibody and T cell responses and its relation to inhibitory and activating Fc gamma receptors. J. Immunol. 172, 5269–5276.

Gianfrani, C., Oseroff, C., Sidney, J., Chesnut, R.W., Sette, A., 2000. Human memory CTL response specific for influenza A virus is broad and multispecific. Hum. Immunol. 61, 438–452.

Haggmark-Manberg, A., Zandian, A., Forsstrom, B., Khademi, M., Lima Bomfim, I., Hellstrom, C., Arnheim-Dahlstrom, L., Hallbook, T., Darin, N., Lundberg, I.E., Uhlen, M., Partinen, M., Schwenk, J.M., Olsson, T., Nilsson, P., 2016. Autoantibody targets in vaccine-associated narcolepsy. Autoimmunity 49, 421–433.

Hallmayer, J., Faraco, J., Lin, L., Hesselson, S., Winkelmann, J., Kawashima, M., Mayer, G., Plazzi, G., Nevsimalova, S., Bourgin, P., Hong, S.C., Honda, Y., Honda, M., Högl, B., Longstreth, W.T.J., Montplaisir, J., Kemlink, D., Einen, M., Chen, J., Musone, S.L., Akana, M., Miyagawa, T., Duan, J., Desautels, A., Erhardt, C., Hesla, P.E., Poli, F., Frauscher, B., Jeong, J.H., Lee, S.P., Ton, T.G., Kvale, M., Kolesar, L., Dobrovolná, M., Nepom, G.T., Salomon, D., Wichmann, H.E., Rouleau, G.A., Gieger, C., Levinson, D.F., Gejman, P.V., Meitinger, T., Young, T., Peppard, P., Tokunaga, K., Kwok, P.Y., Risch, N., Mignot, E., 2009. Narcolepsy is strongly associated with the T-cell receptor alpha locus. Nat. Genet. 41, 708–711.

Han, F., Lin, L., Warby, S.C., Faraco, J., Li, J., Dong, S.X., An, P., Zhao, L., Wang, L.H., Li, Q.Y., Yan, H., Gao, Z.C., Yuan, Y., Strohl, K.P., Mignot, E., 2011. Narcolepsy onset is seasonal and increased following the 2009 H1N1 pandemic in China. Ann. Neurol. 70, 410–417.

Han, F., Lin, L., Li, J., Dong, X.S., Mignot, E., 2013. Decreased incidence of childhood narcolepsy 2 years after the 2009 H1N1 winter flu pandemic. Ann. Neurol. 73, 560.

Han, F., Lin, L., Schormair, B., Pizza, F., Plazzi, G., Ollila, H.M., Nevsimalova, S., Jennum, P., Knudsen, S., Winkelmann, J., Coquillard, C., Babrzadeh, F., Strom, T.M., Wang, C., Mindrinos, M., Fernandez Vina, M. & Mignot, E., 2014. HLA DQB1*06:02 negative narcolepsy with hypocretin/orexin deficiency. Sleep 37, 1601–1608.

Hunsucker, S.A., Spychala, J., Mitchell, B.S., 2001. Human cytosolic 5′-nucleotidase I: characterization and role in nucleoside analog resistance. J. Biol. Chem. 276, 10498–10504.

Ionov, Y., 2010. A high throughput method for identifying personalized tumor-associated antigens. Oncotarget 1, 148–155.

Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., Lönnerberg, P., Linnarsson, S., 2014. Quantitative single-cell RNA-seq with unique molecular identifiers. Nat. Methods. 11, 163–166.

Jaager, K., Neuman, T., 2011. Human dermal fibroblasts exhibit delayed adipogenic differentiation compared with mesenchymal stem cells. Stem Cells Dev. 20, 1327–1336.

Jacob, L., Leib, R., Ollila, H.M., Bonvalet, M., Adams, C.M., Mignot, E., 2015. Comparison of Pandemrix and Arepanrix, two pH1N1 AS03-adjuvanted vaccines differentially associated with narcolepsy development. Brain Behav. Immun. 47, 44–57.

John, J., Thannickal, T.C., Mcgregor, R., Ramanathan, L., Ohtsu, H., Nishino, S., Sakai, N., Yamanaka, A., Stone, C., Cornford, M., Siegel, J.M., 2013. Greatly increased numbers of histamine cells in human narcolepsy with cataplexy. Ann. Neurol. 74, 786–793.

Jowsey, I.R., Thomson, A.M., Flanagan, J.U., Murdock, P.R., Moore, G.B., Meyer, D.J., Murphy, G.J., Smith, S.A., Hayes, J.D., 2001. Mammalian class Sigma glutathione S-transferases: catalytic properties and tissue-specific expression of human and rat GSH-dependent prostaglandin D2 synthases. Biochem. J. 359, 507–516.

Katzav, A., Arango, M.T., Kivity, S., Tanaka, S., Givaty, G., Agmon-Levin, N., Honda, M., Anaya, J.M., Chapman, J., Shoenfeld, Y., 2013. Passive transfer of narcolepsy: anti-TRIB2 autoantibody positive patient IgG causes hypothalamic orexin neuron loss and sleep attacks in mice. J. Autoimmun. 45, 24–30.

Kauts, M.L., Pihelgas, S., Orro, K., Neuman, T., Piirsoo, A., 2013. CCL5/CCR1 axis regulates multipotency of human adipose tissue derived stromal cells. Stem Cell Res. 10, 166–178.

Kinzel, S., Lehmann-Horn, K., Torke, S., Hausler, D., Winkler, A., Stadelmann, C., Payne, N., Feldmann, L., Saiz, A., Reindl, M., Lalive, P.H., Bernard, C.C., Bruck, W., Weber, M.S., 2016. Myelin-reactive antibodies initiate T cell-mediated CNS autoimmune disease by opsonization of endogenous antigen. Acta Neuropathol. 132, 43–58.

Knudsen, S., Biering-Sorensen, B., Kornum, B.R., Petersen, E.R., Ibsen, J.D., Gammeltoft, S., Mignot, E., Jennum, P.J., 2012. Early IVIg treatment has no effect on post-H1N1 narcolepsy phenotype or hypocretin deficiency. Neurology 79, 102–103.

Kolter, T., 2012. Ganglioside biochemistry. ISRN Biochem. 2012, 506160.

Kornum, B.R., Faraco, J., Mignot, E., 2011. Narcolepsy with hypocretin/orexin deficiency, infections and autoimmunity of the brain. Curr. Opin. Neurobiol. 21, 897–903.

Kruup, M. (2013) Finding motifs from short peptides.

Lecendreux, M., Libri, V., Jaussent, I., Mottez, E., Lopez, R., Lavault, S., Regnault, A., Arnulf, I., Dauvilliers, Y., 2015. Impact of cytokine in type 1 narcolepsy: role of pandemic H1N1 vaccination? J. Autoimmun. 60, 20–31.

Liang, X., Wu, L., Hand, T., Andreasson, K., 2005. Prostaglandin D2 mediates neuronal protection via the DP1 receptor. J. Neurochem. 92, 477–486.

Lind, A., Ramelius, A., Olsson, T., Arnheim-Dahlstrom, L., Lamb, F., Khademi, M., Ambati, A., Maeurer, M., Nilsson, A.L., Bomfim, I.L., Fink, K., Lernmark, A., 2014. A/H1N1 antibodies and TRIB2 autoantibodies in narcolepsy patients diagnosed in conjunction with the Pandemrix vaccination campaign in Sweden 2009-2010. J. Autoimmun. 50, 99–106.

Longstreth Jr., W.T., Ton, T.G., Koepsell, T.D., 2009. Narcolepsy and streptococcal infections. Sleep 32, 1548.

Maicas, N., Ibanez, L., Alcaraz, M.J., Ubeda, A., Ferrandiz, M.L., 2012. Prostaglandin D2 regulates joint inflammation and destruction in murine collagen-induced arthritis. Arthritis Rheum. 64, 130–140.

Martin-Rufian, M., Tosina, M., Campos-Sandoval, J.A., Manzanares, E., Lobo, C., Segura, J.A., Alonso, F.J., Mates, J.M., Marquez, J., 2012. Mammalian glutaminase Gls2 gene encodes two functional alternative transcripts by a surrogate promoter usage mechanism. PLoS One 7, e38380.

Meyer, S., Adam, M., Schweiger, B., Ilchmann, C., Eulenburg, C., Sattinger, E., Runte, H., Schluter, M., Deuse, T., Reichenspurner, H., Costard-Jackle, A., 2011. Antibody response after a single dose of an AS03-adjuvanted split-virion influenza A (H1N1) vaccine in heart transplant recipients. Transplantation 91, 1031–1035.

Mohri, I., Taniike, M., Taniguchi, H., Kanekiyo, T., Aritake, K., Inui, T., Fukumoto, N., Eguchi, N., Kushi, A., Sasai, H., Kanaoka, Y., Ozono, K., Narumiya, S., Suzuki, K., Urade, Y., 2006. Prostaglandin D2-mediated microglia/astrocyte interaction enhances astrogliosis and demyelination in twitcher. J. Neurosci. 26, 4383–4393.

Mohri, I., Kadoyama, K., Kanekiyo, T., Sato, Y., Kagitani-Shimono, K., Saito, Y., Suzuki, K., Kudo, T., Takeda, M., Urade, Y., Murayama, S., Taniike, M., 2007. Hematopoietic prostaglandin D synthase and DP1 receptor are selectively upregulated in microglia and astrocytes within senile plaques from human patients and in a mouse model of Alzheimer disease. J. Neuropathol. Exp. Neurol. 66, 469–480.

Morel, S., Didierlaurent, A., Bourguignon, P., Delhaye, S., Baras, B., Jacob, V., Planty, C., Elouahabi, A., Harvengt, P., Carlsen, H., Kielland, A., Chomez, P., Garcon, N., VAN Mechelen, M., 2011. Adjuvant System AS03 containing alpha-tocopherol modulates innate immune response and leads to improved adaptive immunity. Vaccine 29, 2461–2473.

Nagase, T., Ishikawa, K., Suyama, M., Kikuno, R., Hirosawa, M., Miyajima, N., Tanaka, A., Kotani, H., Nomura, N., Ohara, O., 1998. Prediction of the coding sequences of unidentified human genes. XII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro. DNA Res. 5, 355–364.

Nagata, N., Iwanari, H., Kumagai, H., Kusano-Arai, O., Ikeda, Y., Aritake, K., Hamakubo, T., Urade, Y., 2017. Generation and characterization of an antagonistic monoclonal antibody against an extracellular domain of mouse DP2 (CRTH2/GPR44) receptors for prostaglandin D2. PLoS One 12, e0175452.

Narumiya, S., Fitzgerald, G.A., 2001. Genetic and pharmacological analysis of prostanoid receptor function. J. Clin. Invest. 108, 25–30.

Nicolson, C., Harvey, R., Johnson, R., Guilfoyle, K., Engelhardt, O.G., Robertson, J.S., 2012. An additional oligosaccharide moiety in the HA of a pandemic influenza H1N1 candidate vaccine virus confers increased antigen yield in eggs. Vaccine 30, 745–751.

Nishino, S., 2011. Hypothalamus, hypocretins/orexin, and vigilance control. Handb. Clin. Neurol. 99, 765–782.

Nishino, S., Sakurai, E., Nevsimalova, S., Yoshida, Y., Watanabe, T., Yanai, K., Mignot, E., 2009. Decreased CSF histamine in narcolepsy with and without low CSF hypocretin-1 in comparison to healthy controls. Sleep 32, 175–180.

Partinen, M., Kornum, B.R., Plazzi, G., Jennum, P., Julkunen, I., Vaarala, O., 2014. Narcolepsy as an autoimmune disease: the role of H1N1 infection and vaccination. Lancet Neurol. 13, 600–613.

Pei, Y., Banerjee, S., Jha, H.C., Sun, Z., Robertson, E.S., 2017. An essential EBV latent antigen 3C binds Bcl6 for targeted degradation and cell proliferation. PLoS Pathog. 13, e1006500.

Peyron, C., Faraco, J., Rogers, W., Ripley, B., Overeem, S., Charnay, Y., Nevsimalova, S., Aldrich, M., Reynolds, D., Albin, R., Li, R., Hungs, M., Pedrazzoli, M., Padigaru, M., Kucherlapati, M., Fan, J., Maki, R., Lammers, G.J., Bouras, C., Kucherlapati, R., Nishino, S., Mignot, E., 2000. A mutation in a case of early onset narcolepsy and a generalized absence of hypocretin peptides in human narcoleptic brains. Nat. Med. 6, 991–997.

Porzionato, A., Macchi, V., Parenti, A., De Caro, R., 2004. The distribution of mast cells in the human area postrema. J. Anat. 204, 141–147.

Ricciotti, E., Fitzgerald, G.A., 2011. Prostaglandins and inflammation. Arterioscler. Thromb. Vasc. Biol. 31, 986–1000.

Robertson, J.S., Nicolson, C., Harvey, R., Johnson, R., Major, D., Guilfoyle, K., Roseby, S., Newman, R., Collin, R., Wallis, C., Engelhardt, O.G., Wood, J.M., Le, J., Manojkumar, R., Pokorny, B.A., Silverman, J., Devis, R., Bucher, D., Verity, E., Agius, C., Camuglia, S., Ong, C., Rockman, S., Curtis, A., Schoofs, P., Zoueva, O., Xie, H., Li, X., Lin, Z., Ye, Z., Chen, L.M., O'neill, E., Balish, A., Lipatov, A.S., Guo, Z., Isakova, I., Davis, C.T., Rivailler, P., Gustin, K.M., Belser, J.A., Maines, T.R., Tumpey, T.M., Xu, X., Katz, J.M., Klimov, A., Cox, N.J., Donis, R.O., 2011. The development of vaccine viruses against pandemic A (H1N1) influenza. Vaccine 29, 1836–1843.

Santini, G., Mores, N., Malerba, M., Mondino, C., Macis, G., Montuschi, P., 2016. Investigational prostaglandin D2 receptor antagonists for airway inflammation. Expert Opin. Investig. Drugs 25, 639–652.

Santus, P., Radovanovic, D., 2016. Prostaglandin D2 receptor antagonists in early development as potential therapeutic options for asthma. Expert Opin. Investig. Drugs 25, 1083–1092.

Sarkanen, T., Alen, R., Partinen, M., 2016. Transient Impact of Rituximab in H1N1 Vaccination-associated Narcolepsy With Severe Psychiatric Symptoms. Neurologist 21, 85–86.

Sarkanen, T.O., Alakuijala, A., Dauvilliers, Y., Partinen, M.M., 2017. Incidence of narcolepsy after H1N1 influenza and vaccinations: Systematic review and meta-analysis. Sleep Med. Rev. https://doi.org/10.1016/j.smrv.2017.06.006.

Schanen, B.C., De Groot, A.S., Moise, L., Ardito, M., Mcclaine, E., Martin, W., Wittman, V., Warren, W.L., Drake 3rd, D.R., 2011. Coupling sensitive in vitro and in silico techniques to assess cross-reactive CD4(+) T cells against the swine-origin H1N1 influenza virus. Vaccine 29, 3299–3309.

Smith, A.J., Jackson, M.W., Neufing, P., Mcevoy, R.D., Gordon, T.P., 2004. A functional autoantibody in narcolepsy. Lancet 364, 2122–2124.

Sung, H.Y., Francis, S.E., Crossman, D.C., Kiss-Toth, E., 2006. Regulation of expression and signalling modulator function of mammalian tribbles is cell-type specific. Immunol. Lett. 104, 171–177.

Tafti, M., Hor, H., Dauvilliers, Y., Lammers, G.J., Overeem, S., Mayer, G., Javidi, S., Iranzo, A., Santamaria, J., Peraita-Adrados, R., Vicario, J.L., Arnulf, I., Plazzi, G., Bayard, S., Poli, F., Pizza, F., Geisler, P., Wierzbicka, A., Bassetti, C.L., Mathis, J., Lecendreux, M., Donjacour, C.E., Van Der Heide, A., Heinzer, R., Haba-Rubio, J., Feketeova, E., Hogl, B., Frauscher, B., Beneto, A., Khatami, R., Canellas, F., Pfister, C., Scholz, S., Billiard, M., Baumann, C.R., Ercilla, G., Verduijn, W., Claas, F.H., Dubois, V., Nowak, J., Eberhard, H.P., Pradervand, S., Hor, C.N., Testi, M., Tiercy, J.M., Kutalik, Z., 2014. DQB1 locus alone explains most of the risk and protection in narcolepsy with cataplexy in Europe. Sleep 37, 19–25.

Taketomi, Y., Ueno, N., Kojima, T., Sato, H., Murase, R., Yamamoto, K., Tanaka, S., Sakanaka, M., Nakamura, M., Nishito, Y., Kawana, M., Kambe, N., Ikeda, K., Taguchi, R., Nakamizo, S., Kabashima, K., Gelb, M.H., Arita, M., Yokomizo, T., Watanabe, K., Hirai, H., Okayama, Y., Ra, C., Aritake, K., Urade, Y., Morimoto, K., Sugimoto, Y., Shimizu, T., Narumiya, S., Hara, S., Murakami, M., 2013. Mast cell maturation is driven via a group III phospholipase A2-prostaglandin D2-DP1 receptor paracrine axis. Nat. Immunol. 14, 554–563.

Terao, A., Matsumura, H., Saito, M., 1998. Interleukin-1 induces slow-wave sleep at the prostaglandin D2-sensitive sleep-promoting zone in the rat brain. J. Neurosci. 18, 6599–6607.

Thannickal, T.C., Moore, R.Y., Nienhuis, R., Ramanathan, L., Gulyani, S., Aldrich, M., Cornford, M., Siegel, J.M., 2000. Reduced number of hypocretin neurons in human narcolepsy. Neuron 27, 469–474.

Thannickal, T.C., Siegel, J.M., Nienhuis, R., Moore, R.Y., 2003. Pattern of hypocretin (orexin) soma and axon loss, and gliosis, in human narcolepsy. Brain Pathol. 13, 340–351.

Thannickal, T.C., Nienhuis, R., Siegel, J.M., 2009. Localized loss of hypocretin (orexin) cells in narcolepsy without cataplexy. Sleep 32 (8), 993.

Turygin, V.V., Babik, T.M., Boyakov, A.A., 2005. Characteristics of mast cells in the choroid plexus of the ventricles of the human brain in aging. Neurosci. Behav. Physiol. 35, 909–911.

Urade, Y., Hayaishi, O., 2011. Prostaglandin D2 and sleep/wake regulation. Sleep Med. Rev. 15, 411–418.

Vaarala, O., Vuorela, A., Partinen, M., Baumann, M., Freitag, T.L., Meri, S., Saavalainen, P., Jauhiainen, M., Soliymani, R., Kirjavainen, T., Olsen, P., Saarenpaa-Heikkila, O., Rouvinen, J., Roivainen, M., Nohynek, H., Jokinen, J., Julkunen, I., Kilpi, T., 2014. Antigenic differences between AS03 adjuvanted influenza A (H1N1) pandemic vaccines: implications for pandemrix-associated narcolepsy risk. PLoS One 9, e114361.

Valko, P.O., Gavrilov, Y.V., Yamamoto, M., Reddy, H., Haybaeck, J., Mignot, E., Baumann, C.R., Scammell, T.E., 2013. Increase of histaminergic tuberomammillary neurons in narcolepsy. Ann. Neurol. 74, 794–804.

Vassalli, A., Li, S., Tafti, M., 2015. Comment on "Antibodies to influenza nucleoprotein cross-react with human hypocretin receptor 2". Sci. Transl. Med. 7, 314le2.

Viale, A., Zhixing, Y., Breton, C., Pedeutour, F., Coquerel, A., Jordan, D., Nahon, J.L., 1997. The melanin-concentrating hormone gene in human: flanking region analysis, fine chromosome mapping, and tissue-specific expression. Brain Res. Mol. Brain Res. 46, 243–255.

Vijay, R., Fehr, A.R., Janowski, A.M., Athmer, J., Wheeler, D.L., Grunewald, M., Sompallae, R., Kurup, S.P., Meyerholz, D.K., Sutterwala, F.S., Narumiya, S., Perlman, S., 2017. Virus-induced inflammasome activation is suppressed by prostaglandin D2/DP1 signaling. Proc. Natl. Acad. Sci. U. S. A. 114, E5444–E5453.

Vilo, J. (2002) Pattern discovery from biosequences.

Vita, R., Overton, J.A., Greenbaum, J.A., Ponomarenko, J., Clark, J.D., Cantrell, J.R., Wheeler, D.K., Gabbard, J.L., Hix, D., Sette, A., Peters, B., 2015. The immune epitope database (IEDB) 3.0. Nucleic Acids Res. 43 (Database issue), D405–12.

Voisin, T., Rouet-Benzineb, P., Reuter, N., Laburthe, M., 2003. Orexins and their receptors: structural aspects and role in peripheral tissues. Cell. Mol. Life Sci. 60, 72–87.

Wu, C.H., Liu, I.J., Lu, R.M., Wu, H.C., 2016. Advancement and applications of peptide phage display technology in biomedical science. J. Biomed. Sci. 23, 8.

Xu, G.J., Kula, T., Xu, Q., Li, M.Z., Vernon, S.D., Ndung'u, T., Ruxrungtham, K., Sanchez, J., Brander, C., Chung, R.T., O'connor, K.C., Walker, B., Larman, H.B., Elledge, S.J., 2015. Viral immunology. Comprehensive serological profiling of human populations using a synthetic human virome. Science 348, aaa0698.

Zandian, A., Forsstrom, B., Haggmark-Manberg, A., Schwenk, J.M., Uhlen, M., Nilsson, P., Ayoglu, B., 2017. Whole-proteome peptide microarrays for profiling autoantibody repertoires within multiple sclerosis and narcolepsy. J. Proteome Res. 16, 1300–1314.

Curriculum vitae

Personal data

Name:	Mariliis Jaago
Date of birth:	23.12.1992
Location:	Tallinn, Estonia
E-mail:	mariliis@protobios.com
ORCID:	0000-0003-4453-6440

Education

2018-	Tallinn University of Technology, doctoral studies
2015–2017	Tallinn University of Technology, MSc in gene technology, *cum laude*
2012–2015	Tallinn University of Technology, BSc in gene technology, *cum laude*

Work experience

2017–	OÜ Protobios, data analyst, research and development

Trainings

2018	EU project SZTest (No 734791) consortium meeting at Biomedicum at Helsinki University
2018	EU project SZTest (No 734791) mini-symposium "Gene expression in health and disease" in Tallinn
2018	Course "Expert Level Data Protection Officer", Open University of Tallinn University of Technology
2018	EASME Business Innovation Coaching by Dr. Stefan Blarer, DRIMARCH GmbH
2017	Course on "Performing Management System Internal Audits", AS Metrosert
2017	EU project SZTest (No 734791) conference "Scizophrenia: today's clinical need, emerging opportunities" in Tallinn

Creative work

2022	Oral presentation at Department of Chemistry and Biotechnology at Tallinn University of Technology, titled "Antibody response to oral biofilm is a biomarker for acute coronary syndrome in periodontal disease" (Tallinn)
2020	Poster presentation at 11th World Biomaterial Congress, titled "Immunoprofiling predictions of personalised immune response to biomaterials in clinical use". Additionally, an oral presentation on the same topic at the PanBioRa workshop organized within the conference. (Online) https://wbc2020.org/.
2019	Oral presentation at symposium 5th Finnish Biological Psychiatry Symposium, titled "Immune response heterogeneity in schizophrenia to neuronal antigens and to common pathogens" (Helsinki, University of Helsinki, University of Helsinki).

2019 Oral presentation at 8th International Conference on Computational Bioengineering, titled “Computational immunoprofiling predictions of immune response promiscuity to Epstein Barr Virus epitopes and oral biofilm in acute coronary syndrome” (Belgrade, University of Kragujevac, University of Belgrade, Bioengineering Research and Development Centre BioIRC, Faculty of Engineering Kragujevac) http://www.iccb2019.kg.ac.rs/.

2019 Oral presentation at EU project SZTest (No. 734791) event SZTest Week, titled “Immunoprofiling of multiple sclerosis: cerebrospinal fluid viral epitope profiles predict poor prognosis” (Tallinn, Tartu, Protobios OÜ, Tartu Ülikool).

Academic thesis and industrial property

Karmen Kalamees, Bachelor’s degree, 2019, (sup) Kaia Palm, Mariliis Jaago. Characterisation of potential humoral immune response against biomaterials hyaluronic acid and silver. Department of Chemistry and Biotechnology, Tallinn University of Technology

Kadri Kiho, Bachelor’s degree, 2019, (sup) Kaia Palm, Mariliis Jaago. Immunoprofiling predictions of serologic response against silver and gelatin as biomaterials. Department of Chemistry and Biotechnology, Tallinn University of Technology

Invention: Methods and systems for monitoring cardiac episode status using antibody profiling of biological samples using immunocomplexes; Owners: Protobios OÜ; Authors: Helle Sadam, Arno Pihlak, Mariliis Jaago, Kaia Palm; Priority number: 62/566,366; Priority date: 30.09.2017.

Publications

Jaago, M.*, Rähni, A.*, Pupina, N., Pihlak, A., Sadam, H., Tuvikene, J., Avarlaid, A., Planken, A., Planken, M., Haring, L., Vasar, E., Baćević, M., Lambert, F., Kalso, E., Pussinen, P., Tienari, P. J., Vaheri, A., Lindholm, D., Timmusk, T.,; Ghaemmaghami, A. M., Palm, K. 2022. Differential patterns of cross-reactive antibody response against SARS-CoV-2 spike protein detected for chronically ill and healthy COVID-19 naïve individuals. *Scientific Reports*, 12 (1): 16817. doi: 10.1038/s41598-022-20849-6.

Jaago, M., Pupina, N., Rähni, A., Pihlak, A., Sadam, H., Vrana, N. E., Sinisalo, J., Pussinen, P., Palm, K. 2022. Antibody response to oral biofilm is a biomarker for acute coronary syndrome in periodontal disease. *Communications Biology*, 5 (1), ARTN 205. doi: 10.1038/s42003-022-03122-4.

Pupina, N., Avarlaid, A., Sadam, H., Pihlak, A., **Jaago, M.**, Tuvikene, J., Rähni, A., Planken, A., Planken, M., Kalso, E., Tienari, P.J., Nieminen, J.K., Seppänen, M.R.J., Vaheri, A., Lindholm, D., Sinisalo, J., Pussinen, P., Timmusk, T., Palm, K. 2022. Immune response to a conserved enteroviral epitope of the major capsid VP1 protein is associated with lower risk of cardiovascular disease. *EBioMedicine*, 76, 103835. doi: 10.1016/j.ebiom.2022.103835.

Rähni, A., **Jaago, M.**, Sadam, H., Pupina, N., Pihlak, A., Tuvikene, J., Annuk, M., Mägi, A., Timmusk, T., Ghaemmaghami, A.M., Palm, K. 2022. Melanoma-specific antigen-associated antitumor antibody reactivity as an immune-related biomarker for targeted immunotherapies. *Commun Med 2*, 48 (2022). doi: 10.1038/s43856-022-00114-7.

Sadam, H., Pihlak, A., **Jaago, M.**, Pupina, N., Rähni, A., Toots, M., Vaheri, A., Nieminen, J. K., Siuko, M., Tienari, P. J., Palm, K. 2021. Identification of two highly antigenic epitope markers predicting multiple sclerosis in optic neuritis patients. *EBioMedicine*, 64, 103211. doi: 10.1016/j.ebiom.2021.103211.

Jaago, M., Timmusk, U. S., Timmusk, T., Palm, K. 2021. Drastic Effects on the Microbiome of a Young Rower Engaged in High-Endurance Exercise After a Month Usage of a Dietary Fiber Supplement. *Frontiers in Nutrition*, 8: 654008. doi: 10.3389/fnut.2021.654008.

Tamberg, L., **Jaago, M.**, Saalik, K., Sirp, A., Tuvikene, J., Shubina, A., Kiir, C. S., Nurm, K., Sepp, M., Timmusk, T., Palgi, M. 2020. Daughterless, the Drosophila orthologue of TCF4, is required for associative learning and maintenance of the synaptic proteome. *Disease Models & Mechanisms*, 13 (7): dmm042747. doi: 10.1242/dmm.042747.

Sadam, H., Pihlak, A., Kivil, A., Pihelgas, S., **Jaago, M.**, Adler, P., Vilo, J., Vapalahti, O., Neuman, T., Lindholm, D., Partinen, M., Vaheri, A., Palm, K. 2018. Prostaglandin D2 Receptor DP1 Antibodies Predict Vaccine-induced and Spontaneous Narcolepsy Type 1: Large-scale Study of Antibody Profiling. *EBioMedicine*, 29, 47−59. doi: 10.1016/j.ebiom.2018.01.043.

* equal contribution

Elulookirjeldus

Isikuandmed

Nimi:	Mariliis Jaago
Sünniaeg:	23.12.1992
Elukoht:	Tallinn, Eesti
E-mail:	mariliis@protobios.com
ORCID:	0000-0003-4453-6440

Haridustee

2018-	Tallinna Tehnikaülikool, doktoriõpe
2015–2017	Tallinna Tehnikaülikool, magistrikraad (geenitehnoloogia), *cum laude*
2012–2015	Tallinna Tehnikaülikool, bakalaureusekraad (geenitehnoloogia), *cum laude*

Töökogemus

2017–	OÜ Protobios, andmetöötleja, teadusarendus

Erialased kursused ja õpisündmused

2018	EL projekti SZTest (No 734791) konsortsiumi partnerite koosolek Helsinki Ülikooli Biomedicum'is
2018	EL projekti SZTest (No 734791) mini-sümpoosium "Gene expression in health and disease" Tallinnas.
2018	Koolitus "Eksperdi tasandil andmekaitsespetsialist", TTÜ Avatud Ülikool, 2018
2018	EASME Business Innovation Coaching by Dr. Stefan Blarer, DRIMARCH GmbH
2017	Täiendkoolitus "Juhtimissüsteemide siseauditite läbiviimine", AS Metrosert
2017	EL projekti SZTest (No 734791) konverents "Scizophrenia: today's clinical need, emerging opportunities" 29-30.08.2017 Tallinnas

Postrid ja ettekanded

2022	Suuline ettekanne Tallinna Tehnikaülikooli Keemia ja Biotehnoloogia instituudis, pealkirjaga "Antibody response to oral biofilm is a biomarker for acute coronary syndrome in periodontal disease" (Tallinn)
2020	Posterettekanne konverentsil 11th World Biomaterial Congress pealkirjaga "Immunoprofiling predictions of personalised immune response to biomaterials in clinical use". Lisaks samateemaline suuline ettekanne konverentsi raames korraldatud PanBioRa töökojas. (Online) https://wbc2020.org/.
2019	Suuline ettekanne sümpoosiumil 5th Finnish Biological Psychiatry Symposium pealkirjaga "Immune response heterogeneity in schizophrenia to neuronal antigens and to common pathogens" (Helsinki, University of Helsinki).

2019 Suuline ettekanne konverentsil 8th International Conference on Computational Bioengineering pealkirjaga "Computational immunoprofiling predictions of immune response promiscuity to Epstein Barr Virus epitopes and oral biofilm in acute coronary syndrome" (Belgrade, University of Kragujevac, University of Belgrade, Bioengineering Research and Development Centre BioIRC, Faculty of Engineering Kragujevac) http://www.iccb2019.kg.ac.rs/.

2019 Suuline ettekanne EL projekti SZTest (No. 734791) ürituse SZTest Week raames, pealkirjaga "Immunoprofiling of multiple sclerosis: cerebrospinal fluid viral epitope profiles predict poor prognosis" (Tallinn, Tartu, Protobios OÜ, Tartu Ülikool).

Juhendatud lõputööd ja tööstusinnovatsioon

Karmen Kalamees, bakalaureusekraad, 2019, (juh) Kaia Palm, Mariliis Jaago. Hüaluroonhappe, hõbeda ioonide ja hõbeda nanoosakeste kui biomaterjalide vastase humoraalse immuunvastuse profileerimine. Keemia ja biotehnoloogia instituut, Tallinna Tehnikaülikool

Kadri Kiho, bakalaureusekraad, 2019, (juh) Kaia Palm, Mariliis Jaago. Immunoprofiling predictions of serologic response against silver and gelatin as biomaterials. Keemia ja biotehnoloogia instituut, Tallinna Tehnikaülikool

Patentne leiutis: METHODS AND SYSTEMS FOR MONITORING CARDIAC EPISODE STATUS USING ANTIBODY PROFILING OF BIOLOGICAL SAMPLES USING IMMUNOCOMPLEXES; Omanikud: Protobios OÜ; Autorid: Helle Sadam, Arno Pihlak, Mariliis Jaago, Kaia Palm; Prioriteedi number: 62/566,366; Prioriteedi kuupäev: 30.09.2017.

Publikatsioonid

Jaago, M.*, Rähni, A.*, Pupina, N., Pihlak, A., Sadam, H., Tuvikene, J., Avarlaid, A., Planken, A., Planken, M., Haring, L., Vasar, E., Baćević, M., Lambert, F., Kalso, E., Pussinen, P., Tienari, P. J., Vaheri, A., Lindholm, D., Timmusk, T.,; Ghaemmaghami, A. M., Palm, K. 2022. Differential patterns of cross-reactive antibody response against SARS-CoV-2 spike protein detected for chronically ill and healthy COVID-19 naïve individuals. *Scientific Reports*, 12 (1): 16817. doi: 10.1038/s41598-022-20849-6.

Jaago, M., Pupina, N., Rähni, A., Pihlak, A., Sadam, H., Vrana, N. E., Sinisalo, J., Pussinen, P., Palm, K. 2022. Antibody response to oral biofilm is a biomarker for acute coronary syndrome in periodontal disease. *Communications Biology*, 5 (1), ARTN 205. doi: 10.1038/s42003-022-03122-4.

Pupina, N., Avarlaid, A., Sadam, H., Pihlak, A., **Jaago, M.**, Tuvikene, J., Rähni, A., Planken, A., Planken, M., Kalso, E., Tienari, P.J., Nieminen, J.K., Seppänen, M.R.J., Vaheri, A., Lindholm, D., Sinisalo, J., Pussinen, P., Timmusk, T., Palm, K. 2022. Immune response to a conserved enteroviral epitope of the major capsid VP1 protein is associated with lower risk of cardiovascular disease. *EBioMedicine*, 76, 103835. doi: 10.1016/j.ebiom.2022.103835.

Rähni, A., **Jaago, M.**, Sadam, H., Pupina, N., Pihlak, A., Tuvikene, J., Annuk, M., Mägi, A., Timmusk, T., Ghaemmaghami, A.M., Palm, K. 2022. Melanoma-specific antigen-associated antitumor antibody reactivity as an immune-related biomarker for targeted immunotherapies. *Commun Med 2*, 48 (2022). doi: 10.1038/s43856-022-00114-7.

Sadam, H., Pihlak, A., **Jaago, M.**, Pupina, N., Rähni, A., Toots, M., Vaheri, A., Nieminen, J. K., Siuko, M., Tienari, P. J., Palm, K. 2021. Identification of two highly antigenic epitope markers predicting multiple sclerosis in optic neuritis patients. *EBioMedicine*, 64, 103211. doi: 10.1016/j.ebiom.2021.103211.

Jaago, M., Timmusk, U. S., Timmusk, T., Palm, K. 2021. Drastic Effects on the Microbiome of a Young Rower Engaged in High-Endurance Exercise After a Month Usage of a Dietary Fiber Supplement. *Frontiers in Nutrition*, 8: 654008. doi: 10.3389/fnut.2021.654008.

Tamberg, L., **Jaago, M.**, Saalik, K., Sirp, A., Tuvikene, J., Shubina, A., Kiir, C. S., Nurm, K., Sepp, M., Timmusk, T., Palgi, M. 2020. Daughterless, the Drosophila orthologue of TCF4, is required for associative learning and maintenance of the synaptic proteome. *Disease Models & Mechanisms*, 13 (7): dmm042747. doi: 10.1242/dmm.042747.

Sadam, H., Pihlak, A., Kivil, A., Pihelgas, S., **Jaago, M.**, Adler, P., Vilo, J., Vapalahti, O., Neuman, T., Lindholm, D., Partinen, M., Vaheri, A., Palm, K. 2018. Prostaglandin D2 Receptor DP1 Antibodies Predict Vaccine-induced and Spontaneous Narcolepsy Type 1: Large-scale Study of Antibody Profiling. *EBioMedicine*, 29, 47−59. doi: 10.1016/j.ebiom.2018.01.043.

* võrdne panus

ISSN 2585-6901 (PDF)
ISBN 978-9916-80-070-6 (PDF)