UNIVERSITY OF LJUBLJANA
BIOTECHNICAL FACULTY

Leon DEUTSCH

BIOINFORMATICS INTEGRATION OF MICROBIOME AND METABOLOMICS DATA IN THE TRANSLATIONAL CONTEXT

DOCTORAL DISSERTATION

Ljubljana, 2022

UNIVERSITY OF LJUBLJANA
BIOTECHNICAL FACULTY

Leon DEUTSCH

BIOINFORMATICS INTEGRATION OF MICROBIOME AND METABOLOMICS DATA IN THE TRANSLATIONAL CONTEXT

DOCTORAL DISSERTATION

Ljubljana, 2022

"It is better to fail aiming high than to succeed aiming low."

– Bill Nicholson

Based on the Statute of the University of Ljubljana and the decision of the Biotechnical Faculty senate, as well as the decision of the Commission for Doctoral Studies of the University of Ljubljana adopted on 15.02.2021 it has been confirmed that the candidate meets the requirements for pursuing a PhD in the interdisciplinary doctoral programme in Biosciences, Scientific Field Bioinformatics. Prof. Dr. Blaž Stres is appointed as supervisor.

Na podlagi Statuta Univerze v Ljubljani ter po sklepu Senata Biotehniške fakultete in sklepu Komisije za doktorski študij Univerze v Ljubljani z dne 15.2.2021 je bilo potrjeno, da kandidat izpolnjuje pogoje za opravljanje doktorata znanosti na Interdisciplinarnem doktorskem študijskem programu Bioznanosti znanstveno področje bioinformatika. Za mentorja je bil imenovan prof. dr. Blaž Stres.

Supervisor (mentor):	Prof. Blaž STRES, PhD, University of Ljubljana, Biotechnical Faculty, Faculty of Civil and Geodetic Engineering, Jožef Stefan Institute, Slovenia

Committee for the evaluation and the defense (Komisija za oceno in zagovor):

Chair (predsednik):	Prof. Andrej, BLEJEC, PhD University of Ljubljana, Biotechnical Faculty, Department of Biology, Slovenia
Member (član):	Prof. Gregor ANDERLUH, PhD National Institute of Chemistry, Department for Molecular Biology and Nanobiotechnology, Slovenia
Member (član):	Assoc. Prof. David GOMEZ CABRERO, PhD King Abdullah University of Science and Technology, Thuwal, Saudi Arabia Translational Bioinformatics Unit at Navarra biomed, University of Valencia, Valencia, Spain

Date of the defense (datum zagovora): 21.11.2022

Leon DEUTSCH

KEY WORDS DOCUMENTATION

ND   Dd
DC   UDC 579:004(043.3)
CX   bioinformatics, metabolomics, metagenomcis, physical inactivity, data integration, systems biology, microbiome, urine, nuclear magnetic spectrometry
AU   DEUTSCH, Leon
AA   STRES, Blaž (supervisor)
PP   SI-1000 Ljubljana, Jamnikarjeva 101
PB   University of Ljubljana, Biotechnical Faculty, Interdisciplinary Doctoral Programme in Biosciences, Scientific field Bioinformatics
PY   2022
TY   BIOINFORMATICS INTEGRATION OF MICROBIOME AND METABOLOMICS DATA IN THE TRANSLATIONAL CONTEXT
DT   Doctoral Dissertation
NO   XI, 205 p., 2 tab., 18 fig., 323 ref.
LA   EN
Al   en/sl

AB   Human metabolism was studied in three different projects focusing on different levels of inactivity. Urine, liquor and serum metabolomics were used to assess the impact of nusinersen treatment in patients with spinal muscular atrophy. Urine samples were contrasted with samples from matching healthy cohort. In the PreTerm project, metabolomics (fecal and urine) and fecal microbial metagenomics were used to assess the differences between preterm and full-term born adults. In the X-Adapt project, urinary metabolomics was used to evaluate the 10-day training regime and the differences between trained and untrained individuals. In all projects, classification models based on different data sets were developed as a proof of principle and to foster their use in future studies or possibly in medical diagnostics. In addition, two workflows (GUMPP and MAGO tool) and a method to study physicochemical parameters (minimum pressure of piercing strength) were developed to study the microbiome and its environment, respectively. The final work resulted in the creation of the first Slovenian urine ^{1}H-NMR database, which consists of 1200 urine metabolomes from different projects (PlanHab, Spinal Muscular Atrophy, X-Adapt, PreTerm, Healthy males and females) measured by ^{1}H-NMR, outlining the baseline for future extensions. The entire database can be used to build machine-learning models for classification between different diseases or levels of physical activity at a national level.

KLJUČNA DOKUMENTACIJSKA INFORMACIJA

ŠD   Dd

DK   UDK 579:004(043.4)

KG   bioinformatika, metabolomika, metagenomika, fizikalna inaktivnost, integracija podatkov, sistemska biologija, mikrobiom, urin, jedrska magnetna resonanca

AV   DEUTSCH, Leon

SA   STRES, Blaž (mentor)

KZ   SI-1000 Ljubljana, Jamnikarjeva 101

ZA   Univerza v Ljubljani, Biotehniška fakulteta, Interdisciplinarni doktorski študij Bioznanost, Znanstveno področje Bioinformatika

LI   2022

IN   BIOINFORMACIJSKA INTEGRACIJA MIKROBIOMSKIH IN METABOLOMSKIH PODATKOV V TRANSLACIJSKEM KONTEKSTU

TD   Doktorska disertacija

OP   XI, 205 str., 2 pregl., 18 sl., 323 vir.

IJ   sl

JI   sl/en

AI   V okviru večih projektov smo preučevali metabolome preiskovancev z različnimi stopnjami neaktivnosti. Za oceno učinka zdravljenja z zdravilom nusinersen pri bolnikih s spinalno mišično atrofijo smo uporabili metabolomiko urina, likvorja in seruma. Dodatne vzorce urina smo primerjali s tistimi iz zdrave kontrolne skupine. V projektu PreTerm smo s kombinacijo metabolomike fecesa in urina ter z metagenomiko fekalnega mikrobioma raziskovali razlike med predčasno in pravočasno rojenimi odraslimi. V projektu X-Adapt smo z metabolomiko urina ocenili učinke 10-dnevnega režima treninga in razlik med treniranimi in netreniranimi posamezniki. V vseh projektih smo na zbranih podatkih razvili modele za razvrščanje skupin z namenom razvoja analitskih poti ter prikaza možnostjo uporabe teh modelov v prihodnjih študijah ali morda v medicinski diagnostiki. Poleg tega sta bila razvita dva cevovoda (orodje GUMPP in MAGO) ter metoda za preučevanje fizikalno-kemijskih parametrov (minimalna prebodna sila) za preučevanje mikrobioma in njegovega okolja. Končni rezultat analize je bila izdelava prve slovenske metabolomske baze podatkov ^{1}H-NMR urina, ki jo sestavlja 1200 metabolomov urina iz različnih projektov (PlanHab, Spinalna mišična atrofija, X-Adapt, PreTerm, Zdravi moški in ženske), merjenih z ^{1}H-NMR, ki predstavljajo osnovo za prihodnje razširitve iz novih projektov. Celotno bazo podatkov je mogoče uporabiti za gradnjo modelov strojnega učenja za razvrščanje med različnimi boleznimi ali stopnjami telesne aktivnosti na nacionalni ravni.

TABLE OF CONTENTS

TABLE OF CONTENTS OF SCIENTIFIC WORKS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS AND SYMBOLS

16S rRNA	Subunit of ribosomal ribonucleic acid
^{1}H-NMR	Proton nuclear magnetic resonance
ANI	Average nucleotide identity
ASV	Amplicon sequence variants
AUC	The area under the curve
AutoML	Automatic machine learning
BMI	Body mass index
BSS	Bristol stool Scale
CHPOP INTEND	The Children's Hospital of Philadelphia Infant Test of Neuromuscular Disorders
DSS	Sodium trimethylsilylpropanesulfonate
EMA	The European Medicines Agency
FDA	The Food and Drug Administration
FiO$_2$	Inspired O$_2$ fraction
GenomesDB	The reference genome database
GUMPP	General Unified Microbiome Profiling Pipeline
HAmb	Hypoxic ambulation
HBR	Hypoxic bedrest
HCA	Hierarchical cluster analysis
HMDB	Human metabolome database
HMFS	Hammersmith Functional Motor Scale
HMFSE	Expanded Hammersmith Functional Motor Scale
HPC	high performance computing
Humann	The HMP Unified Metabolic Analysis Network
ID	Identity
JADBio	Just add bio data
JSpeciesWS	JSpecies Web Server
LPS	lipopolysaccharide
MAG	Metagenome assembled genome
MAGO	Metagenome-Assembled Genomes Orchestra
MetaPhlAn	Metagenomic Phylogenetic Analysis
MFM	Motor Function Measurement
MIMAG	Minimum information about a metagenome-assembled genome
ML	Machine learning
MP	Minimal pressure
MS	Mass spectrometry
NBR	Normoxic bedrest
NPMANOVA	Non-parametric multivariate analysis of variance
OTU	Opetrational taxonomic units
PCA	Principal component analysis
PCR	Polymerase chain raction

Picrust2	Phylogenetic Investigation of Communities by Reconstruction of Unobserved States
PiO_2	Partial pressure of inspired O_2
PlanHab	The Planetary Habitat simulation project
PLSDA	Partial least squares regression discriminant analysis
PPAR-α	Peroxisome proliferator-activated receptor alpha
ppm	Parts per million
PreTerm	The physiological responses at adulthood as a result of preterm delivery
QC	Quality check
ROC	Receiver operating characteristic curve
ROS	Reactive oxygen subtances
SMA	Spinal Muscle Atrophy
SMN	The survival motor neuron
TCA	The tricarboxylic acid
TMAO	Trimethylamine N-oxide
TSP	Trimethylsilylpropanoic acid
VIP	The Variable importance in projection
VO_2max	Maximal oxygen output
WHO	World heatlh organization
W_{peak}	Maximal pedaling power output
X-Adapt	Cross-adaptation between heat and hypoxia - novel strategy for performance and work-ability enhancement in various environments

1 INTRODUCTION

This study built upon our past work in the field of metagenomics and metabolomics and significantly extended our analytical approaches developed for the Planetary Habitat simulation project (PlanHab) (Debevec et al., 2014; Sket et al., 2017a; Sket et al., 2017b; Sket et al., 2018), which was used to study the short-term and reversible effects of human host physical inactivity. The effects of short-term inactivity resulted in maladjustments in physiology, intestinal microbiota, and metabolomic profiles giving rise to increased inflammation, depression, and insulin resistance, resembling metabolic syndrome and type 2 diabetes symptoms. In contrast, the effects of long-term physical inactivity, the lack of oxygenation (e.g., cardiovascular fitness) and signals from large body muscles (e.g., lower limbs) are not well understood despite their direct and widespread biomedical relevance for people delivered preterm and/or genetic disorders, such as Spinal Muscle Atrophy (SMA), obesity, cardiovascular deconditioning, chronic obstructive pulmonary disease, and many other noncommunicable diseases.

To extend our understanding in the field of human physiology in relation to human gut microbiome, a diverse range of samples was collected within the following three major projects: i) the physiological responses at adulthood as a result of preterm delivery (PreTerm project; ARRS J3-7536; EU project https://recap-preterm.eu/); ii) the Spinal Muscular Atrophy (project at the University Clinical Centre Ljubljana) as an extreme case of physical inactivity, and iii) cross-adaptation between heat and hypoxia: a novel strategy for performance and work-ability enhancement in various environments (X-Adapt; research project ARRS J5-9350). The SMA and PreTerm projects dealt with the lifelong exposure to systemic effects of reduced physical activity that can be summarized as following: i) intermittent episodes of systemic hypoxia at rest/sleep (PreTerm), and ii) continuous systemic hypoxia due to reduced physical activity of the host and the alleviation of hypoxia after therapy. The X-Adapt project dealt with the influence of a standard 10-day training regime on the physiology of healthy trained and untrained individuals. The biochemical characterization of bodily fluids collected within the three projects was used to explore the biochemical makeup (metabolites) and their interactions (metabolic pathways) next to the differences between studied groups. The PreTerm and X-Adapt projects contained healthy baseline data collection for the SMA project to determine the different metabolic pathways between the healthy and affected groups, as well as before and after SMA genetic treatment. Additionally, samples from healthy individuals and their children (father and sons, mothers and daughters) were collected to match those of SMA group and in addition to provide a baseline healthy cohort for metabolomic database. As a result, a national Slovenian urine nuclear magnetic resonance (NMR) database was established with the intent to enable distinction between various "diseased" groups of participant form "healthy" group of participants based on urine metabolites in the future. The extended inclusion of novel samples is planned.

In addition, little is known about the existence of differences in the human-gut microbiome relationship due to the lifelong exposure to hypoxic episodes in the preterm (compared to full-term born adolescents (The PreTerm project)) that could affect the functionalities and metabolism of microbiomes within such hosts and be linked to the various physiological differences observed globally between the two groups in previous research (Martin et al., 2018).

In short, a high number of wet-lab measurements was conducted on a large number of parameters, utilizing makeup of three major projects in the field of biomedical science, utilizing metagenomics, metabolomics, bioinformatics, and data integration approaches. The data generated within each 'omics technology was analysed and finally integrated to gain better understanding of humans as (holobiont) systems.

1.1 HUMAN SYSTEMS BIOLOGY AND HEALTH

Systems medicine or systems biology is a relatively new term (even as of 2022) that combines the application of systems biology concepts, methods, and analytical tools to scientific research and medical practice. The main goal of systems medicine is to integrate data from different levels of research into biomedical models that can predict the behaviour of a system, enhance our understanding of it, and ultimately be used in the prevention, cure, or treatment of disease. These approaches are utilized to study the daunting complexity of chronic (noncommunicable) and acute diseases, be they in humans, animals, or plants. Noncommunicable diseases were shown to develop slowly over prolonged periods of time (years to decades; Alzheimer's disease, type 2 diabetes, metabolic syndrome, insulin resistance, psychological disorders, etc.). Multiple factors were shown to contribute to the development of particular disease types, making them even more complex to study and understand. These factors range from host gene variants, epigenetic regulation of expression, to the microbiome and its metabolic activities, all in response to detrimental environmental factors (e.g., sedentary lifestyle, diet, stress, hydration, circadian rhythm, etc.) (Craig, 2008; Bousquet et al., 2011; Mizeranschi et al., 2016)). The term "system" has thus far been used depending on the scale of the study domain to describe behaviour at selected chemical compounds at the molecular level, extending to its reaction (an enzyme bound to a ligand) or a microbe or complex microbiome at the level of a single human gut or the population globally. Consequently, a system can be observed at different time and size scales (a few milliseconds and a few micrometres compared to the entire human body and 70 years) (Noble, 2002; Hunter and Nielsen, 2005). At the same time, the surrounding short- and long-term environment with its physical and chemical parameters exerts significant multivariate effects on the entire system of observation (diet, level of activity, use of medications, society, etc.). From this point of view, the microbiome is only one subsystem of the many present in human body, which interacts in many directions over various 'omic layers, thus generating a complex network of interfering signals acting differently over time and space (Figure 1, (Stres and Kronegger, 2019)). The first step in the systems medicine approach is to identify the key structuring variables important for systems functioning out of all that are measured in relation to the nature of the disease. In the

design of 'omics studies, each layer of 'omics data provides a list of differences associated with the disease state relative to previous time point or healthy state. Analysing a single type of 'omics data in the absence of other datasets has the potential to generate oversimplified conclusions; therefore, researchers should integrate various types of 'omics data from large cohorts (Hasin et al., 2017).
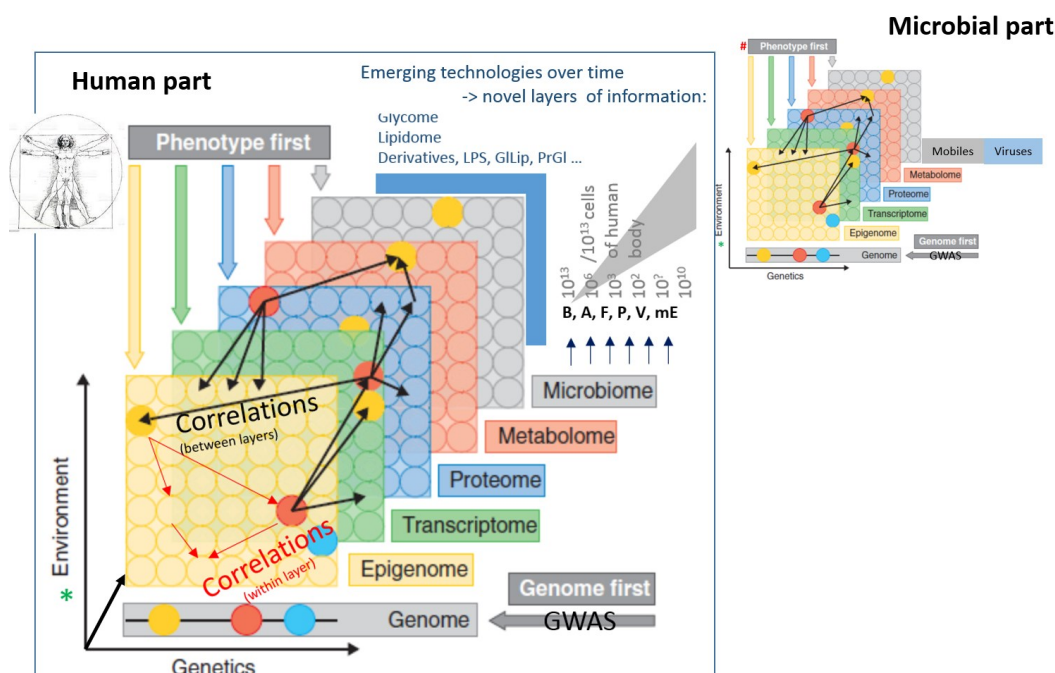
Figure 1:   Interactions between biological systems (Kronegger and Stres, 2019, Hasin et al., 2017).
Bidirectional interaction between different biological systems (e.g. Microbiome and human), between 'omics layers ((meta)genomics, transcriptomics, proteomics, metabolomics) and within individual 'omic layers.

Slika 1:   Interakcija med biološkimi sistemi (Kronegger in Stres, 2019; Hasin in sod., 2017).
Obojestranska interakcija med različnimi sistemi (npr. Mikrobiomom in človekom), med 'omskimi nivoji ((meta)genomiko, transkriptomiko, proteomiko, metabolomiko) in znotraj posameznih 'omskih nivojev.

The rise of 'omics technologies has enabled researchers to measure thousands of data points; this ability is now at the heart of systems biology and medicine. These high throughput technologies (genomics, transcriptomics, metabolomics, proteomics) enabled the discovery of complex sets of biomarkers describing healthy and disease states that can be objectively measured and evaluated. These variables can be used as indicators of a biological process (healthy versus diseased, active versus inactive, pre-treatment versus post-treatment) in the data-driven top-down research coupled to multivariate statistics and machine learning/artificial intelligence. Using high-throughput methods for analysis that capture the properties of systemic homeostasis and dysregulation, we can examine a large number of 'omic markers (also called "biochemical entities") simultaneously (Biomarkers working group, 2001; Holmes et al., 2008b; Fanos, 2016; Tebani et al., 2016; Apweiler et al., 2018; Gallo Cantafio et al., 2018). In this work, metabolomics and microbial metagenomics were used as the 'omics methods of choice.

1.1.1 Metabolomics

Metabolites are small molecules (< 1 kDa) in body fluids such as urine and serum. All metabolites detected in samples are part of the metabolome, which a quantitative description of low molecular weight molecules in a biological sample above the detection threshold of analytical approach. The metabolome is controlled partially by the host genome (primary metabolome), but it also depends on the microbiome metabolic activity (co-metabolome) (Holmes et al., 2008a; Vignoli et al., 2019) in response to changing local and outer environments. Using different spectroscopy methods (nuclear magnetic resonance, mass spectrometry (MS)), metabolic profiles can be analysed at a precise time point. This provides a top-down view of the biochemical processes that occur due to physiological status or environmental exposure (Barr, 2018). The genome is (as currently accepted) unchanged throughout the life span of an individual compared to the high responsiveness and fluidity of metabolome. The latter is heavily influenced by environmental factors such as gender, age, diet, physical activity, health status, and microbiome, to name the most relevant. Metabolome-wide association studies are improving the understanding of the relationship between metabolic profiles and disease risk factors in the general population (Holmes et al., 2008b; Elliott et al., 2015; Vignoli et al., 2019). Metabolomics complements functional metagenomics by mapping the complex metabolic interactions between the host and microbiota via metabolic profiles, compound identity and quantity, characterization of unknown small molecules produced by microbes, and defining the biochemical pathways of metabolites and biochemical reactions (Peisl et al., 2018).

In the previous two decades, NMR has become one of the most important methods for measuring metabolites in different samples (liquid or solid) (Emwas et al., 2019; Wishart, 2019). It is based on the quantum mechanical property (spin) of each nucleus in the molecule. When a nucleus is excited in a magnetic field, a frequency domain spectrum with a peak corresponding to the frequency of the nucleus can be scanned. The frequency, or chemical shift, is reported in parts per million (ppm), and the amplitude of the peak corresponds to the number of nuclei present in the sample (Figure 2). Both can be used to determine the concentration of a molecule in the sample (Maguire, 2014; Keun and Athersuch, 2022). ^{1}H-NMR spectroscopy is used in the majority of NMR based studies. Protons (^{1}H) are present in every metabolite and exhibit the greatest NMR signal sensitivity (Emwas et al., 2019). NMR spectra are usually recorded in water and therefore require solvent suppression (Zheng and Price, 2010; Giraudeau et al., 2015). Compared to MS, the NMR method is robust and reproducible, requires minimal sample preparation, sample measurement is rapid and robust, hence highly replicable; at the same time, it is non-destructive, no chemical derivatization is required, and all types of metabolites can be measured simultaneously and automatically (Table 1). However, the analytical sensitivity is low (10 to 100 times lower than MS), the spectra are complex and computationally intensive to deconvolute, and the NMR spectrometer requires a significant amount of physical space compared to MS. NMR detects molecules at concentrations greater than 1 μM, while MS can detect molecules at concentrations greater than 10 nM (Emwas et al., 2019; Wishart, 2019).
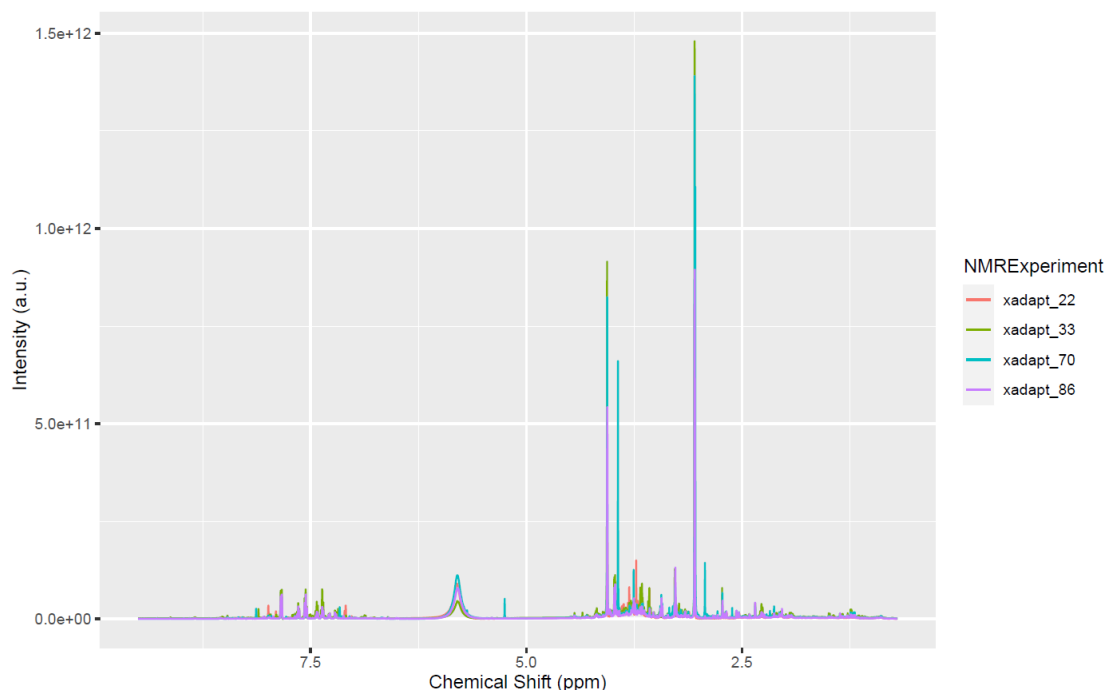
Figure 2: Representative spectra obtained in the X-Adapt study.

Slika 2: Spektri pridobljeni v okviru projekta X-Adapt.

Table 1: Representation of differences between nuclear magnetic resonance (NMR) and mass spectrometry (MS) (Wishart, 2019).

Preglednica 1: Primerjava razlik med jedrsko magnetno resonanco (NMR) in masno spektrometrijo (MS) (Wishart, 2019).

NMR	MS
Non-destructive (sample)	*Destructive (sample)*
Robust instrumentation	*Frail instrumentation*
Instrument downtime minimal	*Instrument downtime frequent*
Excellent reproducibility	*Moderate reproducibility*
Simple sample preparation	*Complex sample preparation*
No requirement for chromatography	*Requirement for chromatography*
No need for chemical derivatization	*Frequent need for chemical derivatization*
Spectra are predictable	*Spectra not very predictable*
Allows precise structure determination	*Allows partial structure determination*
Inherently quantitative	*Not inherently quantitative*
Easily automated workflow	*Difficult-to-automate workflow*
Poor to moderate sensitivity (μM)	**Excellent sensitivity (nM)**
Modest metabolite coverage	**Extensive metabolite coverage**
Very expensive instrumentation	**Moderately expensive instrumentation**
Large instrument footprint	**Small instrument footprint**
Requires cryogens	**No required cryogens**
Expensive to maintain	**Moderately expensive to maintain**
Small spectral databases	**Large spectral databases**
Few software resources	**Many software resources**

Thus far, ^{1}H-NMR has been used to investigate the modulation of metabolites on cellular stress (Lindon et al., 2003), breast cancer markers (Bro et al., 2015), acute pancreatitis (Dumas et al., 2014), the influence of the metabolome on health and disease (Marin et al., 2015), biomarkers for Crohn's disease and ulcerative colitis (Bjerrum et al., 2015), obesity (Zhang et al., 2015), and coronary heart disease and stroke (Holmes et al., 2008a; Murovec et al., 2018; Sket et al., 2018; Vignoli et al., 2019). However, it is very difficult to distinguish between microbial and human metabolites. The metabolism of all parts of the holobiont (human cells and microbial cells) is highly dynamic and variable. For this reason, some authors have used the term "dark matter" of metabolomics, which (in short) means that some metabolites have already been described, but orders of magnitude higher numbers of other metabolites remain unknown. With each study, the data increase which will aid in illuminating the metabolomic dark matter. Modern statistical approaches and data integration combined with ongoing 'omics research and methods development will accelerate the reduction of dark matter and reveal new insights, which will lead to the use of metabolomics methods in general diagnostics (da Silva et al., 2015; Peisl et al., 2018).

1.1.2 Microbial metagenomics

Classical microbiological approaches such as cultivation methods are notoriously incomplete, tedious, irreproducible, and ineffective as only 1% of microorganisms (out of 1500 species) can be easily cultured. For the study of the entire gut microbiome system (microbiome, host, and environment), top-down approaches have taken the leap forward with the introduction of 'omics technologies (Stres and Kronegger, 2019; Lin et al., 2021). For decades, amplicon sequencing was the most commonly used method in microbiome research. The most commonly used gold standard for amplicon sequencing was the sequencing of the gene for 16S rRNA. Variable regions of 16S rRNA are used to determine taxonomic profiles of the microbiota. The major weakness of this method is that we can only determine which taxa are present in the sample, and this can be effectively accomplished only down to the genus level, while species or strain resolution cannot be achieved. In addition, the functional potential of such a community remains obscured as only predictions of functional genes, metabolic pathways involved in the community of interest can be accomplished utilizing different tools such as Picrust2 tool (Langille et al., 2013b; D'Amore et al., 2016; Sinha et al., 2017; Fricker et al., 2019; Douglas et al., 2020). Recently, pipelines for automated analysis of amplicon sequences have been developed for more standardized and efficient analysis on high-performance computing clusters (HPC) (Murovec et al., 2020). The amplicon-sequencing approach is fast, simple, and requires low-cost sample preparation and analysis. However, it is not possible to distinguish living, dead or active microbes. The amplification method can lead to biases (selection of primers for PCR reaction), requires negative control, and functional information is limited (Knight et al., 2018).

Metagenomics, in contrast, uses whole genome shotgun sequencing to fragment and sequence the entire DNA pool of the microbiome in the sample, rather than just one gene (e.g., gene for 16S rRNA) as in amplicon sequencing. The data, quality control, and the information derived from this method are orders of magnitude more comprehensive and enable the recovery of information about phages, viruses, bacteria, archaea, fungi, protozoa, and human DNA. With this approach there is no need for gene prediction based on 16S rRNA as functional genes are determined by comparison to complex gene-family databases with concomitant contamination recognition and removal. With the development of novel quality control tools (KneadData), microbiome taxonomy can be deciphered at the species level with MetaPhlAn3 next to the functional genes recovered from the sample (HUMAnN3) (Brown et al., 2013; Nayfach and Pollard, 2016; Garud et al., 2019; Beghini et al., 2021). Due to the tens of thousands up to millions of variables obtained, the analysis of such datamatrices becomes computationally intensive and requires the utilization of HPC clusters. The metagenomics approach also allows us to use the latest method in microbial genomics: de novo metagenome assembly. Metagenomics can reveal microbial taxonomic and phylogenetic identity, require no PCR amplification, and enable identification of previously known and new species (MAGs) next to new gene families. However, the metagenomics wet-lab and HPC operations are currently still very costly (Knight et al., 2018).

Microbiome analyses are currently focusing on the use of metagenomics due to its wealth of data and reproducible analyses. The microbiome taxonomic description includes the representatives of the community (microbiota - bacteria, archaea, protists, fungi), while also providing the information on so-called "theatre of activity" (Figure 3). Therefore, the information provided includes not only taxonomic descriptions but also molecules produced by these taxonomic units (Whipps et al., 1988; Berg et al., 2020). This approach is becoming increasingly important as the estimates of the number of unique microbial genes per single unique human gene are becoming inherently higher over time, ranging from 50 (Qin et al., 2010) to more than 500 (as of 2022). For this reason, a holistic approach to the study of this system is required. Taking into account the considerable complexity, it becomes increasingly more evident that the disruption of the human microbiome and its activities is significantly associated to the development of various diseases, which in turn depend on the environment and lifestyle of the host (e.g., human). Various environmental factors can affect the gut microbiota: diet, medications, cultural habits, physical activity, transit time, gender, local environment, etc., to variable extent over time and space. (Schmidt et al., 2018; Deutsch and Stres, 2021).
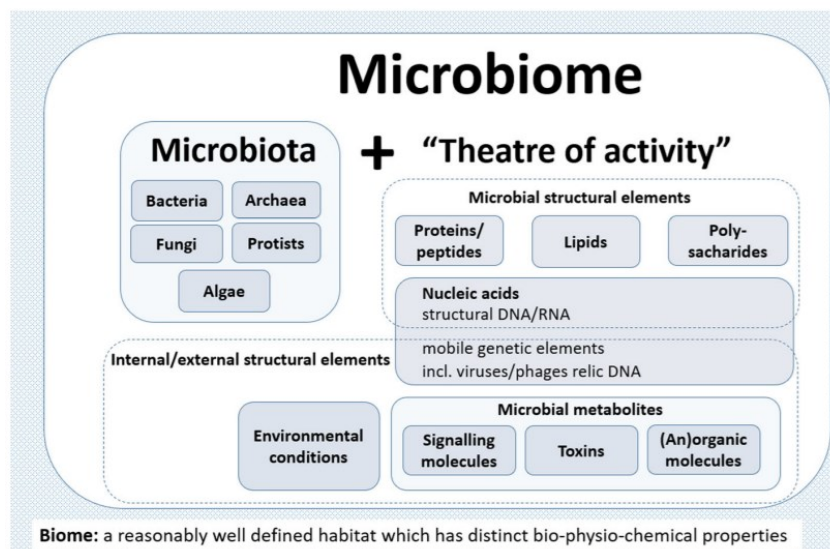
Figure 3:   Differences between "microbiota" and "microbiome" terms (Berg et al., 2020).

Slika 3:   Razlike med izrazi "mikrobiota" in "mikrobiom" (Berg in sod., 2020).

Metagenomics has become a powerful tool for understanding host-microbiome relationships and enables linking biomarkers (genera, species, functional genes) to noncommunicable diseases such as inflammatory bowel disease (Frank et al., 2007), liver cirrhosis (Qin et al., 2014), diabetes (Giongo et al., 2011; Qin et al., 2012), cardiovascular (Wang et al., 2011b) and Parkinson's disease (Scheperjans et al., 2015), colorectal cancer (Kostic et al., 2012), rheumatoid arthritis (Scher et al., 2013), obesity, metabolic syndrome and others. Therefore, metagenomics has become the currently most important approach to study the genetic potential of microbial populations in the intestinal tract. Figure 4 shows the importance of microbiome influence on our future health span (Wilkinson et al., 2021).
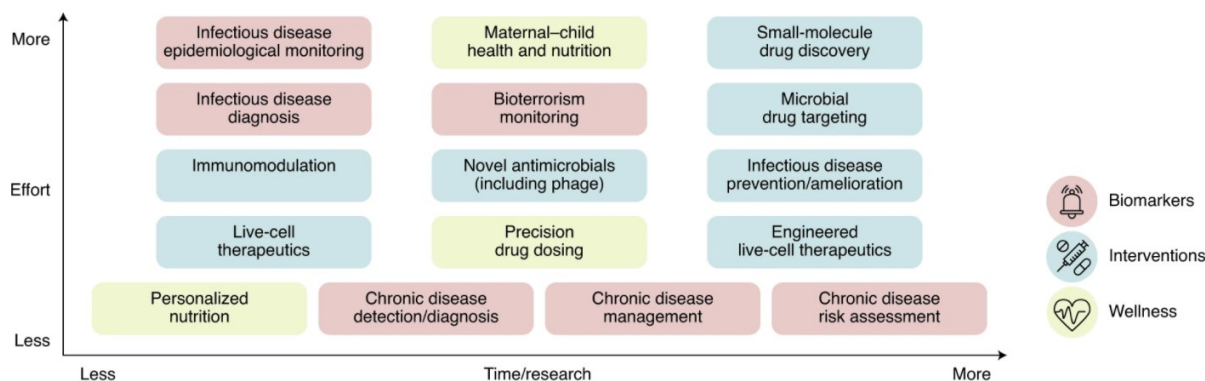
Figure 4: Increase importance of microbiome research (Wilkinson et al., 2021).
Research of microbiome is becoming more and more important from the angle of improved wellness, timely interventions and search for biomarkers.

Slika 4: Naraščajoča pomembnost raziskovanja microbiome (Wilkinson et al., 2021).
Raziskovanje mikrobioma postaja vedno bolj pomembno iz perspektive izboljšanja naših življenj, medicinskih intervencij in iskanja biomarkerjev.

1.1.3   Analysis of data

Both methods (metagenomics and metabolomics) require various steps of quality control and data processing from the sequences or spectra obtained to the final conclusions. After quality checking the sequences with programmes such as FastQC, fastp (Chen et al., 2018) or KneadData, the next step is to obtain actionable sequence data. We can determine which taxa are present in the sample (MetaPhlAn), identify strains (StrainPhlAn), and determine which functional genes are present in the sample (HUMAaN3) or even predict metabolites (MelonnPan) (Segata et al., 2012; Beghini et al., 2021; Mallick et al., 2019). To facilitate the use of such pipelines, workflows were developed by various groups, such as bioBakery or MetaBakery (in preparation by our group). These workflows simplify the use of programs for nature scientists, as they do not need to be installed separately and are already prepared as pipelines that work on HPC cluster as a Singularity images (Kurtzer et al., 2017) or Docker containers. All of these tools generate various matrices for visualisations, statistics, modelling, and machine-learning approaches (Costea et al., 2017; Quince et al., 2017; Knight et al., 2018; Moreno-Indias et al., 2021). These pipelines result in matrices of variables describing the samples.

As a second option, de novo metagenome assembly represents a second option for metagenomics data analysis and results in the assembly of novel draft genomes that may represent new and not yet described species (Yang et al., 2021). Slightly different steps are required. After quality control, the read sequences have to be assembled. There are different assemblers, including metaSPAdes (Nurk et al., 2017), megahit (Li et al., 2015) or IDBA-UD (Peng et al., 2012). Assembled sequences are binned in the next steps using binning tools, such as BinSanity (Graham et al., 2017), CONCOCT (Alneberg et al., 2014), MetaBat, MaxBin, and DAStool (Wu et al., 2016; Sieber et al., 2018).

Assembled metagenomes can be scored for quality (% completeness and % contamination) using CheckM (Parks et al., 2015) according to the MIMAG standard (> 90% complete and < 5% contamination) (Bowers et al., 2017). All MAGs obtained can be used for annotation with Prokka (Seemann, 2014) in GeneBank format or analysed with Roary (Page et al., 2015) as pan and core genomes. ezTree (Wu, 2018) can be used to extract protein-coding single-copy orthologous marker genes with functional annotation and to build maximum likelihood trees from amino acid sequences. High-throughput analysis of average nucleotide identity (ANI) of MAGs is used in FastANI (Jain et al., 2018). All the above programs are available as a single pipeline for MAGs development in MAGO and prepared for HPC computing as a Singularity image or Docker container (Murovec et al., 2020). The JSpeciesWS taxonomic threshold web service measures the probability of whether genomes belong to the same species or not based on their complete or tentative nucleotide sequence (Richter et al., 2016). For more in-depth analyses, the recently developed Genome Taxonomy Database can be utilized.

Metabolome profiling is usually performed using either targeted or untargeted methods. Targeted metabolomics studies (metabolic profiling) focus on the accurate identification and quantification of a defined group of metabolites in biological samples. Untargeted studies (metabolic fingerprinting) focus on measuring and comparing as many signals as possible in a sample set, followed by the assignment of these signals to metabolites IDs using metabolomics databases. NMR measurements generate spectra that must be processed (Bingol, 2018; Klein, 2021). The untargeted approach does not require prior knowledge of the metabolites in the sample, so its analysis can be more complex and difficult (Klein, 2021). NMR spectra can be referenced with an internal chemical shift standard, such as DSS or TSP, which are the most commonly used standards in the NMR community (Emwas, 2015; Dona et al., 2016; Emwas et al., 2016; Emwas et al., 2018). In the pre-processing step, spectra must be phased and baseline corrected. With phasing, the absorptive character and symmetry of all NMR peaks are maximized (Wishart, 2008). Baseline correction is a processing step that removes all artefacts caused by electronic distortion or incomplete digital sampling, ultimately resulting in a completely flat part of the spectra in signal-free regions (Emwas et al., 2018). Several elements of the spectra need to be removed as they represent artefacts originating from protons in water (4.5-4.9 ppm) and urea (5.5-6.1 ppm). Software for targeted approaches, such as Chenomx NMR Suite, Amix, and AssureNMR, match the obtained spectra with reference spectra (in Human Metabolome Database (HMDB) (Wishart et al., 2007a; Wishart et al., 2013; Wishart et al., 2018; Wishart et al., 2022)) to calculate the concentrations of identified metabolites in the sample (Klein, 2021). Untargeted approaches can be divided into two groups of spectra processing. The peak-picking approach requires clearly visible peaks and generates a feature list for the spectral positions of the successfully detected peaks. This approach is not able to identify low intensity signals or signals with distorted line shapes. AlpsNMR (Madrid-Gambin et al., 2020), rDolphin (Cañueto et al., 2018), or speaq 2 (Beirnaert et al., 2018) are tools that use the peak-picking approach. The other approach is spectral binning, which can be used to identify signals that are missed by the peak-picking approach. The data from binning contain a large number of features from spectral regions. However, they also contain signals from spectral noise, which can reduce the statistical power of the data analysis in the next step (Klein,

2021). Data from untargeted approaches often contain negative values. These negative values must be replaced by an affine transformation of the negative values, which is implemented in the R package mcrbin (Klein, 2021). The resulting bins or matrices of concentrations and metabolites are processed for statistical analysis (Ebbels et al., 2013; Barnes et al., 2016).

Metabolomics and metagenomics generate different data matrices with a large number of variables (taxa, functional genes, enzymatic reactions, metabolic pathways, metabolites) that are variably associated with different additional datamatrices describing the environmental factors, such as diet or patient metadata (health status, body mass index (BMI), age, etc.). These data matrices require more modern statistical approaches that use multivariate statistics (Figure 5). The high dimensionality of 'omics data can range from 300+ metabolites in NMR metabolomics to several thousand and millions of variables from microbiomes (taxa, functional genes, enzymatic reactions, metabolic pathways, predicted metabolites) and require data reduction methods (Argmann et al., 2016; Barnes et al., 2016) and nonparametric statistical methods (NPMANOVA) (Legendre and Legendre, 2012; Anderson and Walsh, 2013). Normalization must be used to remove variation between samples and make them comparable to each other (Emwas et al., 2018). To find the best normalization approach, the web tool NOREVA was developed to compare 20 different normalization methods (Yang et al., 2020). Scaling and transformation should also be applied to reduce the stronger influence of analysing features that are present in larger quantities compared to others, which means that this approach helps to distribute the data more normally (Ebbels et al., 2013; Emwas et al., 2018). There are two different approaches to data analysis. First, the unsupervised methods that do not require prior knowledge, such as principal component analysis (PCA) or hierarchical cluster analysis (HCA), utilize loadings plot created within PCA analysis to see which feature discriminates target groups of interest (Barnes et al., 2016). Second, the supervised methods assume that a known structure of patterns exists and use rules to predict new data. Supervised methods include partial least squares regression discriminant analysis (PLSDA) (Wold et al., 2001; Trygg and Wold, 2002), regression, and classification. The Variable Importance in Projection (VIP) score can be used to see which feature contributed the most to discrimination (Barnes et al., 2016). Supervised methods are very powerful and require validation methods to confirm the true relationship between different groups (Ebbels et al., 2011). Unsupervised methods may miss an interesting correlation, while supervised methods are more likely to produce false positives (Maguire, 2014). Web servers were developed to facilitate the use of these methods, such as MicrobiomeAnalyst (Dhariwal et al., 2017; Chong et al., 2020), MetaboAnalyst (Chong et al., 2018; Chong et al., 2019; Pang et al., 2021) or OmicsAnalyst (Zhou et al., 2021).
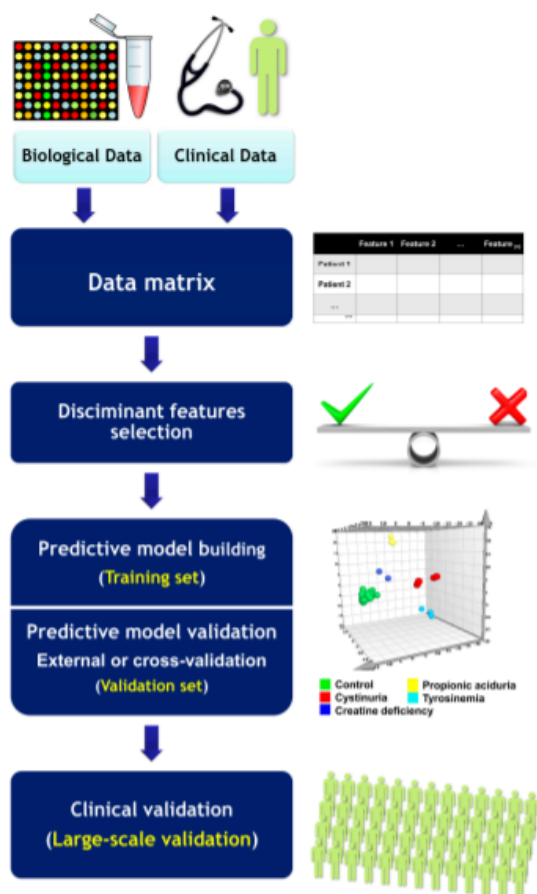
Figure 5:   Representation of data analysis in multi-omics research (Tebani et al., 2016).

Slika 5:   Prikaz analize podatkov pridobljenih z multi-omskimi metodami (Tebani in sod., 2016).

Multivariate statistical methods provide results whose feature can successfully distinguish between different groups. The next step in modern data science is the machine learning (ML) approach, which creates models that can be used in the future to diagnose, treat, and predict the health status of individuals. ML methods rely on algorithms that describe the relationship between variables (Sidey-Gibbons and Sidey-Gibbons, 2019). ML models, such as Support Vector Machines, K Nearest Neighbours, Naïve Bayes, Random Forest, and others, can be used for this purpose (Cristianini and Shawe-Taylor, 2000; Shen et al., 2003; Susnow and Dixon, 2003; Bender et al., 2007; Deo, 2015; Ekins et al., 2019). There is no clear boundary to distinguish statistical from ML methods. In short, the main goal of the statistical approach is to draw conclusions and inferences about populations based on measured data. The primary goal of ML methods, in contrast, is to make predictions. The main steps for ML are (i) importing and preparing the data set, (ii) training the ML model, (iii) testing the ML model (validating the model), (iv) evaluating the sensitivity, specificity, and accuracy of the model, (v) plotting the area under the curve and the receiver operating characteristics curve, and (vi) applying new data to the trained model (Sidey-Gibbons and Sidey-Gibbons, 2019). Regularization techniques must be used to ensure the correctness of the model. The regularization or penalty

parameter controls the complexity of the model (controls the number of features included in the prediction). Each model must be subjected to cross-validation, which means that the data must be split into a training set (for training the model) and a validation set (for validating the model). Model validation compares the predictive performance of the selected model if its performance from training is similar (Teschendorff, 2019). In recent years, AutoML platforms have been created for building ML models without human intervention JADBIO (Tsamardinos et al., 2020; Tsamardinos et al., 2022), AutoWEKA (Thornton et al., 2013; Kotthoff et al., 2017), AutoSklearn (Feurer et al., 2021), GoogleAutoML, RapidMiner) (Mustafa and Rahimi Azghadi, 2021). AutoML has already been applied in various fields of human healthcare, such as diabetes diagnosis, Alzheimer's disease, electronic medical record analysis, and medical imaging (Borkowski et al., 2019; Karaglani et al., 2020; Tsamardinos et al., 2020; Waring et al., 2020; Mustafa and Rahimi Azghadi, 2021). AutoML automates the main processes of ML, from data preparation to feature extraction and selection, algorithm selection, hyperparameter optimization and evaluation (Feurer et al., 2015; Kotthoff et al., 2017; Hutter et al., 2019; Mustafa and Rahimi Azghadi, 2021). However, experienced data scientists are still required to professionally evaluate the results obtained with AutoML (Mustafa and Rahimi Azghadi, 2021).

For proper interpretation of the obtained results, the right data integration process should be used. Identifying a combination of distinguishing characteristics satisfies biological assumptions that cannot be satisfied by univariate methods. Therefore, the combination of different statistical methods (univariate, multivariate, machine learning) provides the key to answer complex biological questions. The mixOmics-R package (Rohart et al., 2017; Singh et al., 2019) is dedicated to the multivariate analysis of biological datasets with a particular focus on data exploration, dimensionality reduction and visualisation, thus providing a systems biology approach, a wide range of methods that statistically integrate multiple datasets simultaneously to explore relationships between heterogeneous 'omics datasets to identify molecular signatures. mixOmics supports the inclusion of different types of biological data and their analysis beyond the scope of 'omics, as long as they are expressed as continuous values.

The other important issue should also be discussed. Batch effects are an important part of the natural sciences. Different processing, different samples can lead to spurious findings and obscure the true signals due to differences in experiments and methods. Biological studies depend on many different factors. This can lead to confounding factors that are unavoidable and come from biological, technical, and computational sources (Ma et al., 2019; Wang and Lê Cao, 2020). Batch effects are an obstacle to comparing the results of different studies. Traditional meta-analysis techniques for combining p-values from independent studies, such as Fisher's method, are effective but statistically conservative. If batch effects can be corrected, statistical tests can be performed on data pooled across studies, increasing the sensitivity for detecting differences between treatment groups. Removing or accounting for batch effects requires computational and analytical multivariate methods (Wang et al., 2019), such as ConQuR (Ling et al., 2021) or ComBat (Gibbons et al., 2018). Most of the above-mentioned methods were use in different projects of this work (Figure 6).
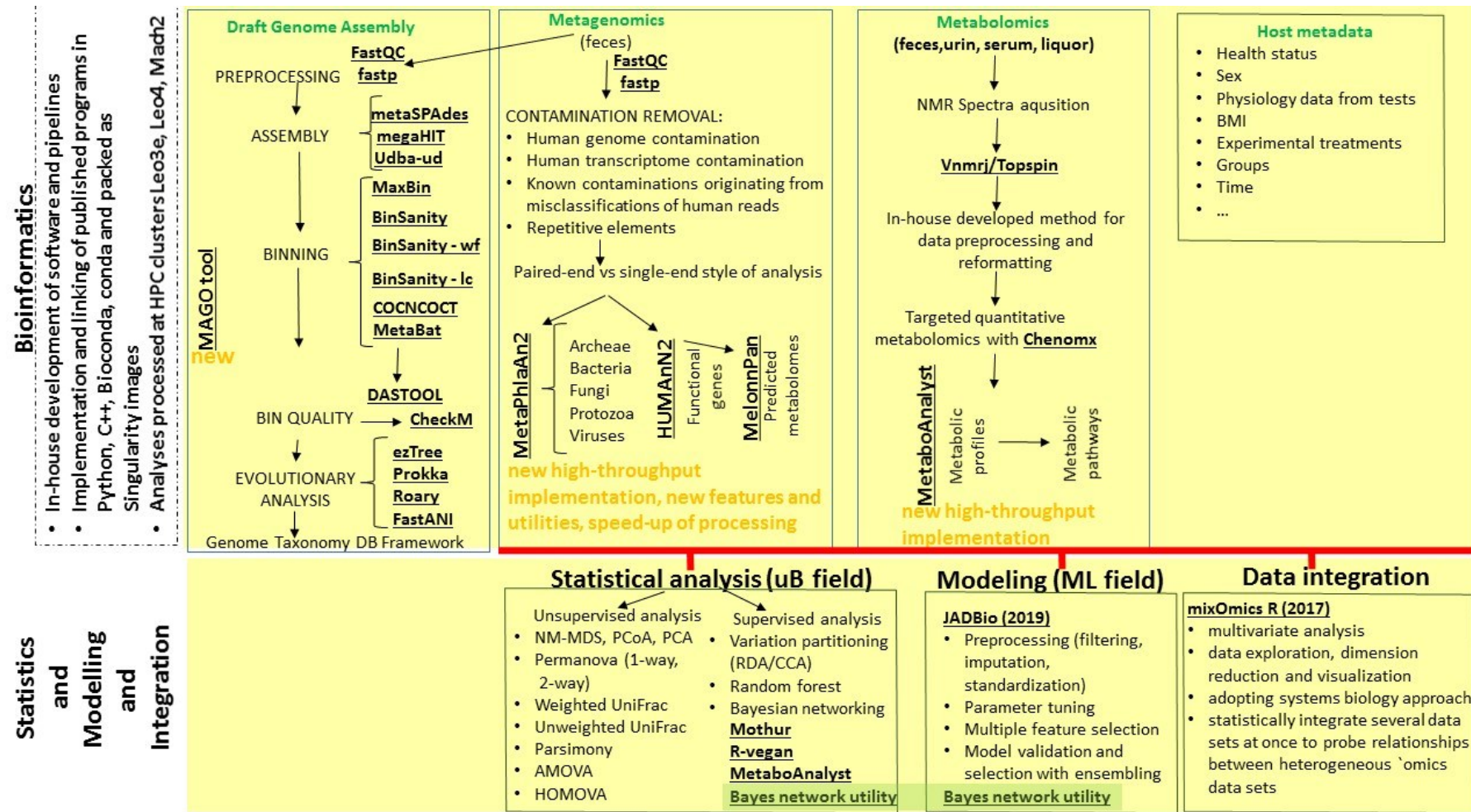
Figure 6:   Data analysis methods use in current work.

Slika 6:   Metode analize podatkov zajete v tem delu.

1.2   INACTIVITY

Physical inactivity associated with the modern sedentary lifestyle (Figure 7) is becoming a global problem and is ranked as the fourth largest behavioural risk for mortality worldwide (Kohl et al., 2012; Kelly et al., 2020a). Every adult should engage in at least 75 minutes of vigorous physical activity per week or 150 minutes of moderate physical activity per week (Sallis et al., 2016), engage in muscle training twice per week, and try to spend as little time as possible in a sedentary position (Kelly et al., 2020a). Regular physical activity reduces the risk of obesity, some cancers, diabetes, coronary heart disease, stroke, dementia, etc. (Booth et al., 2012). Several observational, short-term and long-term intervention studies have used different metabolomics methods to monitor changes in physiological levels due to inactivity as they change following different exercise regimes (Kelly et al., 2020a). Varying levels of physical activity are associated with quantifiable changes in the metabolic profile of individuals. Possible changes may be observed in metabolism of fatty acid, cholesterol and carnitine, lipolysis, the tricarboxylic acid (TCA) cycle, glycolysis, and insulin sensitivity (Kelly et al., 2020a). Metabolic syndrome has a number of risk factors associated with the development of type 2 diabetes mellitus and atherosclerotic cardiovascular disease. Biogenic amines, such as trimethylamine N-oxide (TMAO), choline and L-carnitine (all found in red meat), and branched-chain amino acids may increase the likelihood of metabolic syndrome. In contrast, histidine and lysine correlate with a lower likelihood of metabolic syndrome. Moreover, there is a plethora of molecules, and more research is needed to understand their role in the development of metabolic syndrome (Lent-Schochet et al., 2019). Inactivity also causes hypoxic conditions leading to redox imbalance, which may also be observed on metabolic levels. Differentially expressed levels of creatine, hypoxanthine, acetylcarnitine, and taurine were reported, due to hypoxic conditions (Crass and Lombardini, 1977; Franconi et al., 1985; Malcangio et al., 1989; Aureli et al., 1994; Michalk et al., 1997; Amano et al., 2003; Chen et al., 2009; Scafidi et al., 2010; Powers et al., 2011; Chen et al., 2013; Turner et al., 2015; Scheer et al., 2016; Lee et al., 2017; Sibomana et al., 2021; Wilken et al., 2022). Inactivity also leads to muscle loading (alteration I muscle protein synthesis) and heart failure (Rittweger et al., 2016), both leading to systemic hypoxemia and elevated levels of reactive oxygen substances (ROS). For the majority of people, life can be improved with moderate activity. Physical exercise is one of the main stimuli in restoring prooxidant to antioxidant balance in chronic disease patients (Vincent at el., 2007). Sitting less and moving more (low-to-moderate exercise) or a staircase approach with an increase in activity can prevent the development of metabolic syndrome (Debevec et al., 2017; Dunstan et al., 2021). However, this approach must be considered a never-ending story, which means that physical activity must continue even if health has improved (Figure 8). However, there are also people who face various health problems from birth (premature born infants, patients with spinal muscular disease) and whose possibilities of being physically active are limited, such as in spinal muscular atrophy, which will be discussed later in this work.
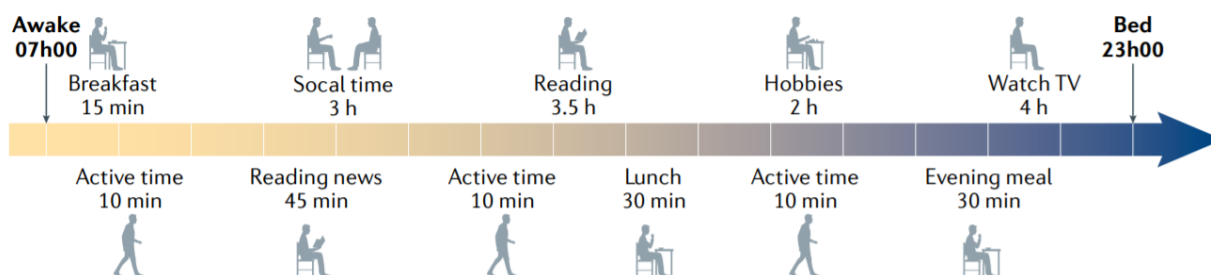
Figure 7:   Lifestyle of modern humans (Dunstan et al., 2021).

Slika 7:   Življenjski stil modernega človeka (Dunstan in sod., 2021).
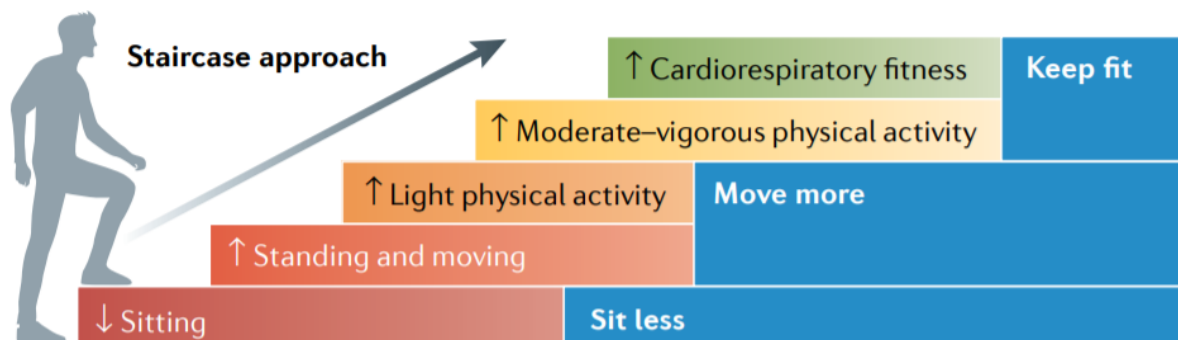
Figure 8.   Staircase approach for increased activity level (Dunstan et al., 2021).
Staircase approach should be used to increase level of activity with small steps towards reduced probabilities of reducing non-communicable disease.

Slika 8.   Postopno povečevanje aktivnosti (Dunstan in sod., 2021).
Postopno povečevanje aktivnosti z majhnimi koraki zmanjša verjetnost kroničnih bolezni.

1.3   PURPOSE OF THE RESEARCH

As stated above, three biomedically relevant reduced-exercise models form the backbone of our current work (Figure 9). The overall aim of our research was to determine the physiological responses at metabolic level to different levels of physical (in)activity and health status (PreTerm, SMA, X-Adapt) in relation to personal characteristics of participants.

To achieve this goal, it was necessary to prepare the entire infrastructure of bioinformatics analytical pathways for pre-processing of molecular data and the correct statistical, modelling, and integration approaches for data processing and interpretation.

Within each project, additional explorations were made based once different 'omics layers are made available in our measurements (Figure 8):

i)   PreTerm ((taxonomy (Bacteria, Archaea, Fungi, Protozoa, Viruses) + functional genes + metabolic pathways + predicted metabolomes + MAGs assembly)) coupled to ^{1}H-NMR metabolomics (metabolites + metabolic pathways);

ii)   SMA (^{1}H-NMR metabolomics (metabolites + metabolic pathways));

iii)   X-Adapt (^{1}H-NMR metabolomics (metabolites + metabolic pathways))

iv)   Healthy baseline (^{1}H-NMR metabolomics)

Samples	feces							urine		serum		liquor	
Omics technoloy	Metagenomics					Metabolomics							
Omics layer	Taxonomy (B,A,F,P,V)	Functional genes	metabolic pathways	predicted metabolites	MAGs	1H-NMR metabolites	metabolic pathways	1H-NMR metabolites	metabolic pathways	1H-NMR metabolites	metabolic pathways	1H-NMR metabolites	metabolic pathways
PreTerm	✓	✓	✓	✓	✓	✓	✓	✓	✓				
SMA								✓	✓				
X-adapt								✓	✓				
Healthy								✓	✓				

Figure 9:   Graphical presentation of collected samples and 'omics layers.
Graphical presentation of collected samples and 'omics layers in PreTerm, SMA and X-Adapt projects. In addition, a healthy urine database was also collected, and it represents healthy baseline for Slovenian NMR urinary database.

Slika 9:   Grafična predstavitev pobranih vzorcev in 'omskih nivojev
Grafična predstavitev pobranih vzorcev in 'omskih nivojev v projektih PreTerm, SMA in X-Adapt. Dodatno so bili še pobrani vzorci zdravih, ki predstavljajo bazno linijo za Slovensko NMR podatkovno bazo.

1.4   HYPOTHESES

1.4.1   PreTerm related

The PreTerm-related hypotheses are discussed in the published paper presented in chapter 2.1.7 and additionally in chapters 2.2.1 and 3.1.6.

H0: No significant difference exists between preterm and term groups of participants at the levels of faecal or urine metabolomes or faecal metagenomes.

H1: There are significant differences between preterm and term groups of participants in faecal and urine metabolomes that can be linked to their physical performance in experiments and physiological data at exercise and rest.

H2: There are significant differences at the level of metagenomics makeup of both groups, giving rise to identification of specific metabolic pathways differing between the two groups and their gut environment characteristics.

H3: Term and preterm gut samples contain specific MAGs associated with differences in gut environmental conditions between the two groups.

1.4.2 SMA related

The SMA-related hypothesis is discussed in published paper presented in chapter 2.1.3 and additionally in chapter 3.1.4.

H0: There are no significant differences in metabolomes before and after treatment.

H1: There are significant differences in urine (systemic) and liquor (local) metabolomes before and after treatment with gene therapy, enabling identification of characteristic metabolic pathways discerning the two groups.

1.4.3 Hypotheses of merged dataset

Hypotheses of the merged dataset are discussed in chapters 2.2.2 and 3.1.7.

H0: There is no significant difference between metabolomes of prematurely born, born on time, before SMA treatment, and after SMA treatment groups.

H1: There are significant differences in urine metabolomes that enable identification of biomarker pools and metabolic pathways delineating various groups under investigation.

2 SCIENTIFIC WORKS

2.1 PUBLISHED SCIENTIFIC WORKS

2.1.1 Computational framework for high-quality production and large-scale evolutionary analysis of metagenome assembled genomes

Murovec B., **Deutsch L.**, Stres B. 2019. Computational framework for high-quality production and large-scale evolutionary analysis of metagenome assembled genomes. Molecular Biology and Evolution, 37, 2: 593-598

Abstract

Microbial species play important roles in different environments and the production of high-quality genomes from metagenome data sets represents a major obstacle to understanding their ecological and evolutionary dynamics. Metagenome-Assembled Genomes Orchestra (MAGO) is a computational framework that integrates and simplifies metagenome assembly, binning, bin improvement, bin quality (completeness and contamination), bin annotation, and evolutionary placement of bins via detailed maximum-likelihood phylogeny based on multiple marker genes using different amino acid substitution models, next to average nucleotide identity analysis of genomes for delineation of species boundaries and operational taxonomic units. MAGO offers streamlined execution of the entire metagenomics pipeline, error checking, computational resource distribution and compatibility of data formats, governed by user-tailored pipeline processing. MAGO is an open-source-software package released in three different ways, as a singularity image and a Docker container for HPC purposes as well as for running MAGO on a commodity hardware, and a virtual machine for gaining a full access to MAGO underlying structure and source code. MAGO is open to suggestions for extensions and is amenable for use in both research and teaching of genomics and molecular evolution of genomes assembled from small single-cell projects or large-scale and complex environmental metagenomes.

For my personal contributions as a doctoral student and author of this thesis, please refer to Table 2 (page 142).

Computational Framework for High-Quality Production and Large-Scale Evolutionary Analysis of Metagenome Assembled Genomes

Boštjan Murovec,[1] Leon Deutsch,[2] and Blaz Stres [ID] *,[2,3,4,5,6]

[1]Laboratory for Artificial Sight and Automation, Faculty of Electrical Engineering, University of Ljubljana, Ljubljana, Slovenia

[2]Department of Animal Science, Biotechnical Faculty, University of Ljubljana, Ljubljana, Slovenia

[3]Center for Clinical Neurophysiology, Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia

[4]Institute of Sanitary Engineering, Faculty of Civil and Geodetic Engineering, University of Ljubljana, Ljubljana, Slovenia

[5]Department for Automation, Biocybernetics and Robotics, Jozef Štefan Institute, Ljubljana, Slovenia

[6]Department of Microbiology, Institute of Microbiology, University of Innsbruck, Innsbruck, Austria

*Corresponding authors: E-mails: blaz.stres@bf.uni-lj.si; blaz.stres@uibk.ac.at

Associate editor: Michael Rosenberg

Abstract

Microbial species play important roles in different environments and the production of high-quality genomes from metagenome data sets represents a major obstacle to understanding their ecological and evolutionary dynamics. Metagenome-Assembled Genomes Orchestra (MAGO) is a computational framework that integrates and simplifies metagenome assembly, binning, bin improvement, bin quality (completeness and contamination), bin annotation, and evolutionary placement of bins via detailed maximum-likelihood phylogeny based on multiple marker genes using different amino acid substitution models, next to average nucleotide identity analysis of genomes for delineation of species boundaries and operational taxonomic units. MAGO offers streamlined execution of the entire metagenomics pipeline, error checking, computational resource distribution and compatibility of data formats, governed by user-tailored pipeline processing. MAGO is an open-source-software package released in three different ways, as a singularity image and a Docker container for HPC purposes as well as for running MAGO on a commodity hardware, and a virtual machine for gaining a full access to MAGO underlying structure and source code. MAGO is open to suggestions for extensions and is amenable for use in both research and teaching of genomics and molecular evolution of genomes assembled from small single-cell projects or large-scale and complex environmental metagenomes.

Key words: metagenomics, evolutionary analyses, microbial draft genomes, species boundaries, FastANI, genome assembly and binning.

Microbial species play important roles in different environments characterized by a span of organismal complexities. The shotgun sequencing coupled to metagenomic analyses are used to study microbial communities in these environments. The analysis and biological interpretation of sequence information derived from complex communities or single-amplified cell communities represented as metagenome or whole-genome sequencing data sets, respectively, is challenging and crucially depends on sophisticated computational resources and analyses. These include various pieces of software and steps (e.g., read assembly, binning, annotation, bin evaluation) next to program-specific settings, file format conversions and decision points that require and consume substantial time, computational resources and may introduce unintended bias (Sczyrba et al. 2017). Obtaining genomes from metagenomes is an emerging approach with the potential for large-scale recovery of high-quality near-complete genomes amenable for analyses of their evolutionary divergence, evolutionary dynamics, and abundance in original samples (Meyer et al. 2018).

Advances in computational tools have improved our ability to address relevant evolutionary questions. However, computational costs for hundreds of samples are measured in tenths of thousands of CPU hours. The development of highly successful tools such as FastQC (Andrews 2010), fastp (Chen et al. 2018), IDBA-UD (Peng et al. 2012), megaHIT (Li et al. 2015), metaSPAdes (Nurk et al. 2017), maxBin (Wu et al. 2016), MetaBAT (Kang et al. 2015), CONCOCT (Alneberg et al. 2014), BinSanity (Graham et al. 2017), Dereplication-Aggregation Scoring Tool (Sieber et al. 2018), CheckM (Parks et al. 2015), ezTree (Wu, 2018), and lessons learned through the Critical Assessment of Metagenomic Information (CAMI; Sczyrba et al. 2017; Meyer et al. 2018; Fritz et al. 2019) enabled the field of molecular evolution of Bacteria and Archaea domains to progress from being a descriptive to an experimental endeavor, providing insight into

MBE
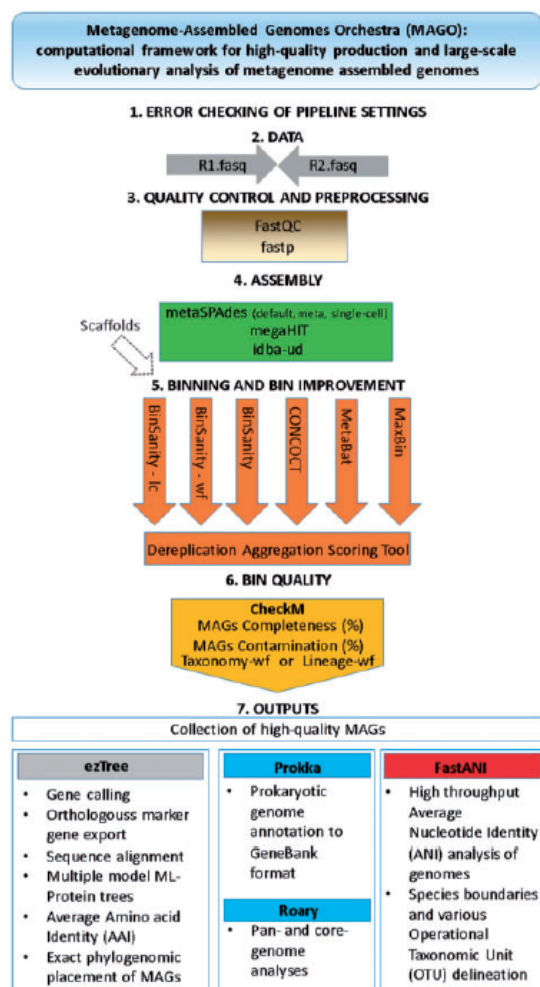
Fig. 1. A schematic representation of steps integrated within MAGO starting from the input of raw sequencing data to MAGs, bin quality checking and the production of a collection of high-quality MAGs. These are further utilized in analysis of evolutionary relationships to produce maximum-likelihood (ML) phylogenomic placement, MAGs annotation, and core/pan genome calculations next to determination of species boundaries and operational taxonomic units at genomic level. The outputs are easily integrated into recently developed tools (e.g., MEGA-X, Kumar et al. 2018; GTDB-Tk, Parks et al. 2018; MAGpy, Stewart et al. 2019).

evolutionary wealth of novel metagenome-assembled genomes (MAGs), novel microbial lineages uncovered from the environment, hence substantially revising and expanding the tree of life (Parks et al. 2017; Parks et al. 2018) and evolutionary dynamic in complex environments and medicine (Lin and Kussell, 2019; Garud et al. 2019). Although the tools are widely used, a number of limitations (supplementary table S1, Supplementary Material online) and their dispersed and boutique nature is limiting their integration and presents an obstacle to their reproducible use within community, their further adoption alongside the ubiquitous increases in sequencing volumes, study complexity (Jain et al. 2018),

emerging standards (Sczyrba et al. 2017; Bowers et al. 2017), and technology upgrades (e.g., Nanopores).

To date, no uniform piece of software exists that would integrate efficiently, scalable and reproducibly all the steps linking the raw outputs from the sequencing platform (i.e., sequence data sets) over the steps of sequence quality trimming, assembly, binning, bin improvement, bin quality control, bin annotation, to evolutionary and phylogenomic placement of bins based on multiple orthologous marker genes on protein level, provide core- and pan-genome analyses and species boundary delineation through fast average nucleotide identity (ANI) of resulting draft genomes. The field-wide analysis standards are emerging due to the ongoing efforts (Sczyrba et al. 2017; Meyer et al. 2018; Fritz et al. 2019); however, the lack of reproducible framework makes it difficult to embrace these standards, perform meta-analyses of existing data (Schloss et al. 2009; Parks et al. 2017) or simply remap and extend past analyses (Parks et al. 2018; Jain et al. 2018) to evolutionary dynamics (Garud et al. 2019).

A single software platform, Metagenome Assembled Genomes Orchestra (MAGO) (fig. 1; supplementary table S1, Supplementary Material online) was developed to fill this gap and to overcome the limitations (supplementary table S2, Supplementary Material online) by integrating an ensemble of previously developed tools, streamlining their performance and deliver compatibility of data formats, together with additional features for error checking, effective computational resource use, governed by user-tailored pipeline processing (as specified by a textual configuration file). MAGO currently makes use of the three most effective assemblers and six binners put forward by CAMISIM (Fritz et al. 2019) and AMBER (Meyer et al. 2018) studies, respectively. The resulting bins are further improved by additional (the seventh) binner, Dereplication-Aggregation Scoring Tool (Sieber et al. 2018) and evaluated by CheckM according to their quality (% completeness and % contamination; Parks et al. 2015) in line with MIMAG standard (Bowers et al. 2017). CheckM utilizes a broader set of orthologous protein marker genes specific to the position of each MAG within a reference genome tree and information about collocation of these genes, based on amino acid identity between marker genes. Finally, the produced collection of high quality MAGs can be used to extract protein-coding single-copy orthologous marker genes using functional annotation and build maximum likelihood trees from amino acid sequences with different amino acid substitution models within MAGO using ezTree (Wu, 2018). The resulting alignment file can be exported to build user specific trees in existing high-end software (e.g., MEGA, Kumar et al. 2018). To annotate and calculate core- and pan-genomes MAGO integrates Prokka (Seemann, 2014) and Roary (Page et al. 2015) and makes outputs (fasta, gbk) available for additional downstream analyses of genome rearrangements (e.g., Mauve, Darling et al. 2010). FastANI (Jain et al. 2018) is utilized for high-throughput ANI analysis of MAGs that is used to define species boundaries and Operational Taxonomic Unit (OTU) delineation at various thresholds of ANI. All outputs are readily made available in structured directories for additional

Metagenome-Assembled Genomes Orchestra · doi:10.1093/molbev/msz237 **MBE**
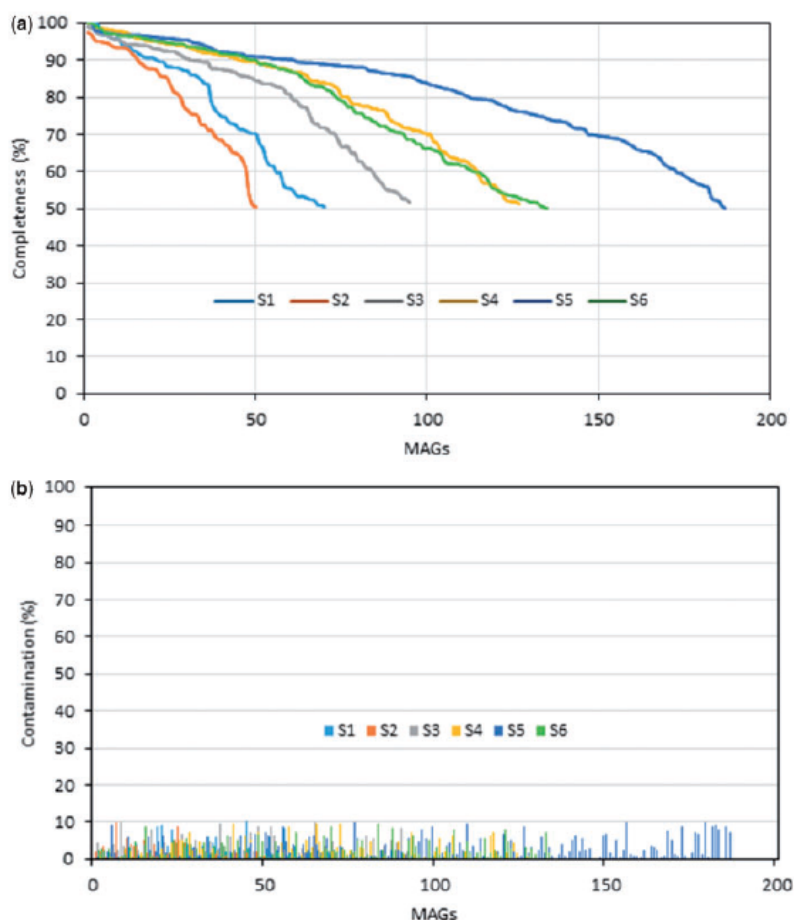
Fig. 2. Overview of the basic quality metrics of MAGs reconstructed from the moose rumen microbiome collection (samples S1–6) (supplementary table S3, Supplementary Material online; Svartström et al. 2017): (*A*) completeness (>50%); (*B*) contamination (<10%).

inspection and inclusion in other types of analyses tools (e.g., MEGA-X, Kumar et al. 2018; GTDB-Tk, Parks et al. 2018; MAGpy, Stewart et al. 2019). In total, MAGO consists of a number ($n = 53$) of externally developed pieces of software (supplementary table S1, Supplementary Material online) and >9,000 lines of Python code integrated into seamless workflow to perform error checking of pipeline configuration and to prevent suboptimal utilization of computational resources.

To overcome the constraints of web-based implementations of existing software and the known software limitations described above (supplementary table S2, Supplementary Material online) MAGO was made available as a singularity image (https://www.sylabs.io/singularity/; last accessed September 04, 2019) and a Docker container (https://www.docker.com; last accessed September 04, 2019) for high performance computing (HPC) purposes, and also as a VirtualBox (https://www.virtualbox.org/; last accessed September 04, 2019) virtual machine (as outlined in supplementary materials and methods, Supplementary Material online). By making MAGO an open-source-software package under the Commons Creative Attribution CC-BY License (https://

creativecommons.org/licenses/; last accessed September 04, 2019) the software is free and open to modifications by other researchers. It is available for download at the project website (http://mago.fe.uni-lj.si; last accessed October 28, 2019). The accompanying preprepared example pipelines and test data set document necessary information about the use of MAGO, enhance reproducibility as the entire pipeline settings can now easily be shared as a single textual pipeline file between researchers, and results reproduced independently (supplementary figs. S1 and S2, Supplementary Material online).

The abilities of MAGO are attested by the quality of the underlying pieces of software (supplementary table S1, Supplementary Material online) and their respective publications. Increasingly complex model data sets spanning CAMI (Sczyrba et al. 2017) and EBI (https://www.ebi.ac.uk/ena/data/view/PRJEB8286; last accessed September 04, 2019) were used in benchmarking MAGO (supplementary table S3, Supplementary Material online; results not shown). The Genome Assembly Gold-standard Evaluations (GAGE) and single-cell amplified genome project (Salzberg et al. 2012; Kogawa et al. 2018) were used for realistic pure culture

595

MBE
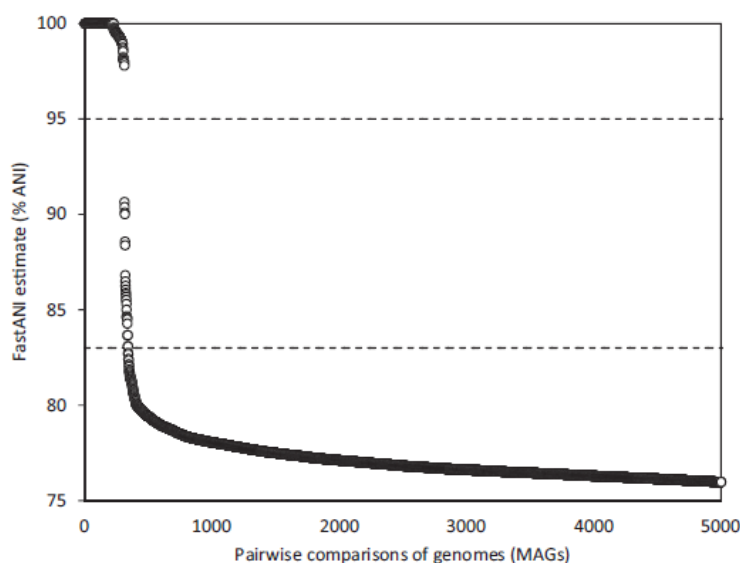
Fig. 3. Genetic discontinuity observed in the wild moose rumen MAGs shown for the first 5,000 pairwise genome comparisons (supplementary table S3, Supplementary Material online). Values of FastANI estimates in the ANI range of 75–100% are shown. The 95% and 83% ANI thresholds of FastANI estimates serve to delineate comparisons belonging to the same species (>95% intraspecies ANI) or different species (<83% interspecies ANI).

data analyses (supplementary table S3, Supplementary Material online; supplementary figs. S3–S7, Supplementary Material online). Finally, a number of real case metagenomics data sets ($n = 106$; $s = 0.4$ TB; supplementary table S3, Supplementary Material online) were analyzed: 1) the moose rumen microbiome (Svartström et al. 2017; figs. 2 and 3), and 2) longitudinal American pre/term delivery microbiomes (Goltsman et al. 2018; supplementary figs. S4–S9, Supplementary Material online).

Unless otherwise stated, in analyses of 280 GB data set of the moose rumen microbiome collection (supplementary table S3, Supplementary Material online; Svartström et al. 2017) all parameters used were the default for each subroutine. After initial sequence quality control (FastQC, fastp), each sample was assembled (MEGAHIT) and binned individually (MaxBin, metaBAT, and Concoct), aggregated and dereplicated (Dereplication-Aggregation Scoring Tool). CheckM was used to assess the quality of resulting MAGs (% completeness; % contamination). Single-sample binning produced a total of 3,012 bins. The distribution of the produced MAGs into high- and medium-quality MAGs was based on the criteria defined by the minimum information about a metagenome-assembled genome (MIMAG) standards (Bowers et al. 2017) (high: >90% completeness and <5% contamination, presence of 5S, 16S, and 23S rRNA genes, and at least 18 tRNAs; medium: ≥50% completeness and <10% contamination). Given that few of the MAGs with >90% completeness and <5% contamination in general pass the MIMAG thresholds regarding the presence of rRNA and tRNA genes due to known issues relating to the difficulties in assembly of rRNA regions, the MAGs of high quality are described as "near complete" in general (Bowers et al. 2017). Medium quality

bins ($n = 670$) represented $22.2 \pm 3.4\%$ of all bins, whereas 75%, 80% complete bins (10% contamination) (Stewart et al. 2019) next to near complete bins represented $14.7 \pm 3.4\%$ ($n = 443$), $12.9 \pm 2.9\%$ ($n = 389$), and $6.5 \pm 1.2\%$ (197) of all recovered MAGs, respectively. In general, MAGO enabled to recover 13 MAGs (80% complete; 10% contamination; dereplicated) per each 10 GB of input sequence data.

The resulting MAGs obtained in this study were first used to explore the existence of genetic discontinuity among the microbial species as observed in large collections of complete genomes from unrelated studies (Jain et al. 2018). The bimodal distribution, with the vast majority (99.8%) of the total genome comparisons showing either > 95% intraspecies ANI or <83% interspecies ANI values, was observed also for the pairwise comparisons of MAGs recovered in this study (fig. 3). It is highly likely that the discontinuity represents a true biological signature, confirming the existence of sequence-discrete populations in natural environments. Although the exact biological mechanisms giving rise to this phenomenon were not explored in this study, the existence of genetic discontinuity in various environments provides opportunity to reconsider its potential origins: 1) decreased recombination frequency below 95% ANI; 2) dispersal limitations in habitats; 3) reduced diversity due to ongoing competition; 4) stochastic events over long periods of time, and provides opportunity to extend analyses from Bacterial and Archaeal domain toward plasmids (Nurk et al. 2017) and viruses (Sutton et al. 2019) for which MAGO can be adopted. In addition, the reconstructed MAGs were compared with a large and heterogeneous collection of characterized prokaryotic genomes ($n = 91, 761$; Jain et al. 2018). The majority of MAGs recovered in this study exhibited ANI < 83% (i.e., interspecies ANI

values) with genomes in the collection. According to the species demarcation cut-off of ~95% ANI the MAGs recovered from actively fermenting wild moose rumen represent potentially new species amenable for detailed genomic analyses.

MAGO efficiently alleviates the metagenome data analysis bottleneck and provides an important and straightforward-to-implement step toward making the future large-scale evolutionary analyses of MAGs efficient, flexible, scalable and reproducible, enforcing the MIMAG standard. Its outputs are easily integrated into downstream pipelines such as The Genome Taxonomy Database (GTDB) to establish a standardized microbial taxonomy based on genome phylogeny (http://gtdb.ecogenomic.org/; last accessed September 04, 2019). MAGO is open to suggestions for extensions and is amenable for use in both research and teaching of genomics and molecular evolution of genomes assembled from small single-cell projects or large-scale and complex environmental metagenomes.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

References

Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ, Andersson AF, Quince C. 2014. Binning metagenomics contigs by coverage and composition. *Nat Methods.* 11(11):1144–1146.

Andrews A. 2010. FastQC: a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/; last accessed September 04, 2019.

Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Eloe-Fadrosh EA, et al. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol.* 35(8):725–731.

Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34(17):i884–i890.

Darling ACE, Mau BT, Perna NT. 2010. Progressive mauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5(6):e11147.

Fritz A, Hofmann P, Majda S, Dahms E, Dröge J, Fiedler J, Lesker TR, Belmann P, DeMaere MZ, Darling AE, et al. 2019. CAMISIM: simulating metagenomes and microbial communities. *Microbiome* 7(1):17. DOI:10.1186/s40168-019-0633-0636.

Garud NR, Good BH, Hallatschek O, Pollard KS. 2019. Evolutionary dynamics of bacteria in the gut microbiome within and across hosts. *PLoS Biol.* 17(1):e3000102.

Goltsman DSA, Sun CL, Proctor DM, Digiulio DB, Robaczewska A, Thomas BC, Shaw GM, Stevenson DK, Holmes SP, Banfield JF, et al. 2018. Metagenomic analysis with strain-level resolution reveals fine-scale variation in the human pregnancy microbiome. *Genome Res.* 28(10):1467–1480.

Graham ED, Heidelberg JF, Tully BJ. 2017. BinSanity: unsupervised clustering of environmental microbial assemblies using coverage and affinity propagation. *PeerJ* 5:e3035.

Jaffe AL, Castelle CJ, Dupont CL, Banfield JF. 2019. Lateral gene transfer shapes the distribution of RuBisCO among Candidate Phyla Radiation bacteria and DPANN archaea. *Mol Biol Evol.* 36(3):435–446.

Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun.* 9(1):5114.

Kang DD, Froula J, Egan R, Wang Z. 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from microbial communities. *PeerJ* 3:e1165.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol.* 35(6):1547–1549.

Kogawa M, Hosokawa M, Nishikawa Y, Mori K, Takeyama H. 2018. Obtaining high-quality draft genomes from uncultured microbes by cleaning and co-assembly of single-cell amplified genomes. *Sci Rep.* 8(1):2059.

Li D, Liu CM, Luo R, Sadakane K, Lam TW. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31(10):1674–1676.

Lin M, Kussell E. 2019. Inferring bacterial recombination rates from large-scale sequencing datasets. *Nat Methods.* 16(2):199–204.

Meyer F, Hofmann P, Belmann P, Garrido-Oter R, Fritz A, Sczyrba A, McHardy AC. 2018. AMBER: assessment of Metagenome BinnERs. *Giga Sci.* 7:1–8.

Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile metagenomics assembler. *Genome Res.* 27(5):824–834.

Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, Fookes M, Falush D, Keane JA, Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31(22):3691–3693.

Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25(7):1043–1055.

MBE

Parks DH, Rinke C, Chuvochina M, Chaumeil PA, Woodcroft BJ, Evans PN, Hugenholtz P, Tyson W. 2017. Recovery of nearly 8, 000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol.* 2(11):1533–1542.

Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil PA, Hugenholtz P. 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol.* 36(10):996–1004.

Peng Y, Leung HCM, Yiu SM, Chin F. 2012. IDBA-UD: a de novo assembler for single-cell and metagenomics sequencing data with highly uneven depth. *Bioinformatics* 28(11):1420–1428.

Salzberg SL, Phillippy AM, Zimin A, Puiu D, Magoc T, Koren S, Treangen TJ, Schatz MC, Delcher AL, Roberts M, et al. 2012. GAGE: a critical evaluation of genome assemblies and assembly algorithms. *Genome Res.* 22(3):557–567.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, et al. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol.* 75(23):7537–7541.

Sczyrba A, Hofmann P, Belmann P, Koslicki D, Janssen S, Dröge J, Gregor I, Majda S, Fiedler J, Dahms E, et al. 2017. Critical Assessment of Metagenome Interpretation – a benchmark of metagenomics software. *Nat Methods.* 14(11):1063–1071.

Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30(14):2068–2069.

Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, Banfield JF. 2018. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat Microbiol.* 3(7):836–843.

Stewart RD, Auffret MD, Snelling TJ, Roehe R, Watson M. 2019. MAGpy: a reproducible pipeline for the downstream analysis of metagenome-assembled genomes (MAGs). *Bioinformatics* 35(12):2150–2152.

Stewart RD, Auffret MD, Warr A, Walker AW, Roehe R, Watson M. 2019. Compendium of 4,941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. *Nat Biotechnol.* 37(8):953–961.

Sutton TDS, Clooney AG, Ryan FJ, Ross RP, Hill C. 2019. Choice of assembly software has a critical impact on virome characterisation. *Microbiome* 7:12.

Svartström O, Alneberg J, Terrapon N, Lombard V, de Bruijn I, Malmsten J, Dalin A-M, Muller EEL, Shah P, Wilmes P, et al. 2017. Ninety-nine de novo assembled genomes from the moose (*Alces alces*) rumen microbiome provide new insights into microbial plant biomass degradation. *ISME J.* 11(11):2538–2551.

Wu YW, Simmons BA, Singer S. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomics datasets. *Bioinformatics* 32(4):605–607.

Wu YW. 2018. ezTree: an automated pipeline for identifying phylogenetic marker genes and inferring evolutionary relationships among uncultivated prokaryotic draft genomes. *BMC Genomics* 19(S1):921.

2.1.2 General unified microbiome profiling pipeline (GUMPP) for large scale, streamlined and reproducible analysis of bacterial 16S rRNA data to predicted microbial metagenomes, enzymatic reactions and metabolic pathways

Abstract

General Unified Microbiome Profiling Pipeline (GUMPP) was developed for large scale, streamlined and reproducible analysis of bacterial 16S rRNA data and prediction of microbial metagenomes, enzymatic reactions and metabolic pathways from amplicon data. GUMPP workflow introduces reproducible data analyses at each of the three levels of resolution (genus; operational taxonomic units (OTUs); amplicon sequence variants (ASVs)). The ability to support reproducible analyses enables production of datasets that ultimately identify the biochemical pathways characteristic of disease pathology. These datasets coupled to biostatistics and mathematical approaches of machine learning can play a significant role in extraction of truly significant and meaningful information from a wide set of 16S rRNA datasets. The adoption of GUMPP in the gut-microbiota related research enables focusing on the generation of novel biomarkers that can lead to the development of mechanistic hypotheses applicable to the development of novel therapies in personalized medicine.

For my personal contributions as a doctoral student and author of this thesis, please refer to Table 2 (page 142).

metabolites

MDPI

Article

General Unified Microbiome Profiling Pipeline (GUMPP) for Large Scale, Streamlined and Reproducible Analysis of Bacterial 16S rRNA Data to Predicted Microbial Metagenomes, Enzymatic Reactions and Metabolic Pathways

Boštjan Murovec [1], Leon Deutsch [2] and Blaž Stres [2,3,4,5,*]

1   Faculty of Electrical Engineering, University of Ljubljana, Tržaška 25, SI-1000 Ljubljana, Slovenia; bostjan.murovec@fe.uni-lj.si
2   Biotechnical Faculty, University of Ljubljana, Jamnikarjeva 101, SI-1000 Ljubljana, Slovenia; leon.deutsch@bf.uni-lj.si
3   Faculty of Civil and Geodetic Engineering, University of Ljubljana, Jamova 2, SI-1000 Ljubljana, Slovenia
4   Department of Automation, Jožef Stefan Institute, Biocybernetics and Robotics, Jamova 39, SI-1000 Ljubljana, Slovenia
5   Department of Microbiology, University of Innsbruck, Technikerstrasse 25d, A-6020 Innsbruck, Austria
*   Correspondence: blaz.stres@bf.uni-lj.si; Tel.: +386-41-567-633

Abstract: General Unified Microbiome Profiling Pipeline (GUMPP) was developed for large scale, streamlined and reproducible analysis of bacterial 16S rRNA data and prediction of microbial metagenomes, enzymatic reactions and metabolic pathways from amplicon data. GUMPP workflow introduces reproducible data analyses at each of the three levels of resolution (genus; operational taxonomic units (OTUs); amplicon sequence variants (ASVs)). The ability to support reproducible analyses enables production of datasets that ultimately identify the biochemical pathways characteristic of disease pathology. These datasets coupled to biostatistics and mathematical approaches of machine learning can play a significant role in extraction of truly significant and meaningful information from a wide set of 16S rRNA datasets. The adoption of GUMPP in the gut-microbiota related research enables focusing on the generation of novel biomarkers that can lead to the development of mechanistic hypotheses applicable to the development of novel therapies in personalized medicine.

Keywords: 16S rRNA; amplicon; Mothur; PICRUSt 2; Piphillin; genus; OTU; ASV; predicted metagenomes; predicted enzymatic reactions; predicted metabolic pathways; reproducible analyses; human microbiome; gut; intestine; mice

1. Introduction

The gut microbiota is composed of a huge number of different bacteria, archaea, fungi and protozoa, next to viruses and various mobile elements [1,2]. All these microbes interact with the host, environmental stimuli and each other, thus producing an enormous diversity of chemical compounds that play a key role in host development, wellbeing and aging [3–7]. The advent of large scale microbiome studies generates analytical opportunities to understand how these communities operate and respond to their complex environmental stimuli [8]. Although knowledge of taxonomy and functional genes of microorganisms are both important, functional genes are more directly related to enzymatic reactions and metabolic pathways. It is increasingly recognized that the microbiome influences the host health state and disease progression. For instance, disease progression can range from mild gastrointestinal symptoms to inflammatory bowel disease and colorectal and liver cancer [9]. In addition, a range of diseases have been implicated in metabolic imbalances, ranging from metabolic syndrome and obesity to autoimmune diseases, psychological disorders and infections [9].

Amplicon sequencing of 16S rRNA has served as the key approach of the last decade for the understanding microbial community structure, dynamics and how organisms might influence or be influenced by environmental conditions [10]. Extensive sequencing of bacterial communities is generating large collections of datasets available through public repositories such as European Bioinformatics Institute (https://www.ebi.ac.uk/ accessed on 30 April 2021), CuratedMetagenomicsData [11], or individual studies [12]. These data have so far been described on the level of 16S rRNA taxonomy utilizing either (i) genus [12], (ii) 97–98.5% 16S rRNA identity operational taxonomic units (OTU) [13] or (iii) amplicon sequence variants (ASV) [14,15]. However, the processing and analyses of such datasets are highly diverse due to the high number of published and benchmarked pieces of software [16–20] and reports that lack significant technical details despite the Human Microbiome Project outlines and introduction of standard operating procedures [21–23].

In addition, this wealth of 16S rRNA data gives access to an untapped pool of information beyond the 16S rRNA taxonomy (genus, OTU, ASV), such as predicted functional genes, enzymatic reactions and metabolic pathways (Figure S1). The tools such as MicrobiomeAnalyst [24,25], PICRUSt [26], PICRUSt2 [27], Tax4Fun [28]; Tax4Fun2 [29] and Piphillin [30,31] link 16S rRNA sequence information to representative genome sequences and approximate metagenomics functional gene content relevant for the interpretation of the studied human disease phenomena and clinical metadata [32]. As a number of unexplored and large datasets encompassing thousands of samples and corresponding metadata are made available in repositories (e.g., [12,33] the analyses (genus, OTU, ASV) and improved predicted metagenomic, enzymatic and metabolic pathway datasets have the potential to unravel important taxonomic, functional, biochemical and metabolic findings (Figure S1).

However, in order to accomplish such intensive large scale data analyses effective workflows are required. These workflows should ideally (i) integrate various pieces software, (ii) streamline input and output formats, (iii) accommodate large datasets, (iv) maintain portability between benchtop PC and high performance computing clusters (HPC), (v) enable flexible (customizable) but also reproducible analyses (setting documentation) that can be (vi) shared with and utilized by other interested researchers.

In this study, we introduce a workflow (Figure 1) that integrates Human Microbiome Project tested procedures for amplicon sequence analysis with one of the most popular programs Mothur [34], and PICRUSt2 [27] for prediction of metagenomic functional genes, enzymatic reactions and metabolic pathways. In addition, the workflow presented here generates also formatted inputs for Piphillin [30,31], another popular sister program for metagenomic predictions. The benchmarking of the integrated programs such as Mothur, PICRUSt2, Piphillin and other comparable sister programs were already reported before in numerous studies [16–20,23,27,30,31]. The inbuilt Human Microbiome Project standard operating procedures can be tailored according to user analytical preferences and sequencing details. The whole workflow is delivered as portable all-inclusive container (Singularity [35]; https://sylabs.io accessed on 14 April 2021) amenable for teaching or/and research purposes, using personal computer or HPC. Depending on the size of data and complexity of analyses (genus-, OTU-, ASV- levels), the GUMPP workflow enables maximum utilization of information present in the original 16S rRNA amplicon datasets by producing additional three data types approaching multiomics view of the microbiome: metagenomics functional genes, enzymatic reactions and metabolic pathways. All four data types can serve as inputs for machine learning to unravel novel mechanistic insight into human disease development in relation to microbiome characteristics. To showcase the efficient analyses and utilization of computing resources two datasets describing human ($n = 307$) and mice gut ($n = 365$) were used for demonstration purposes.

GUMPP overview
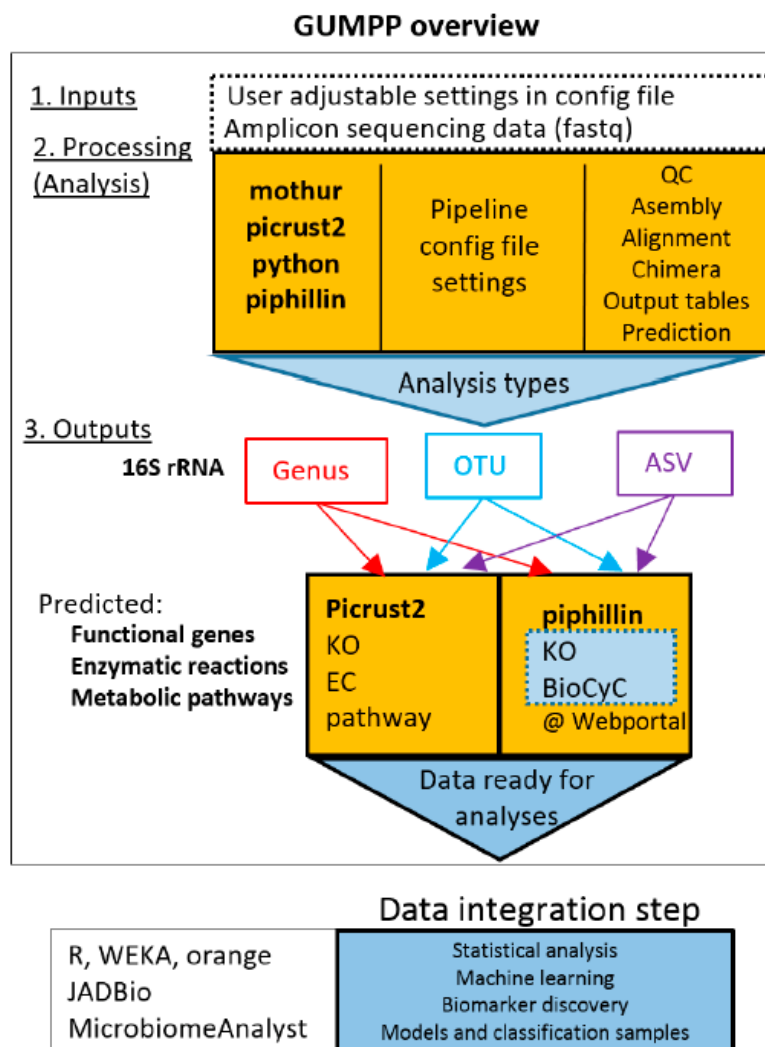
Figure 1. Schematic representation of the General Universal Microbiome Profiling Pipeline (GUMPP). The integral part consists of Mothur, PICRUSt2 and Piphillin outputs. Paired-end or single-end fastq sequence are used as input for mothur processing. The resulting biom and fasta files serve as an input for PICRUSt2. The data can be analyzed at genus-, OTU- and ASV- levels. QC–sequence quality control; OTU-Operational Taxonomic Units (generally 97% identity of 16S rRNA); ASV–Amplicon Sequence Variants (unique sequence variants). KO–KEGG Orthologs (Kyoto Encyclopedia of Genes and Genomes); EC-Enzyme Commission number; BioCyc-BioCyc collection of Pathway/Genome Databases. For each level, four output tables are generated (Please see Figures S1 and S2 for additional information). The resulting data can be analyzed in the data integration step using a variety of distinct machine learning approaches.

2. Results and Discussion

2.1. Design of GUMPP Workflow

GUMPP (http://gumpp.fe.uni-lj.si, accessed on 24 May 2021) is a freely available skeleton application for executing Mothur [34] using paired-end fastq files and executing the PICRUSt2 analyses next to producing also Piphillin [30,31] web-server input files (Figure 1). A single GUMPP run can process an arbitrary number of input files. Inputs are preprocessed by an integrated Mothur (V1.44.1) script in conjunction with Silva database (version 138), and creates biom and fasta representative sequence files as input for PICRUSt2

and outputs necessary for Piphillin [30,31]. The workflow was designed to support three levels of analysis differing in the increased extent of utilized information and fairness in data treatment: genus-, OTU- and ASV- levels (Figures S1 and S2). Users may freely replace the built in scripts and databases with their own. Customization of the built-in script (http://gumpp.fe.uni-lj.si) is also possible by template parameters.

The primary design goal of the GUMPP application was to deliver efficient analyses and utilization of computing resources. The application relies on recently developed Singularity container technology ([35]; https://sylabs.io accessed on 14 April 2021) making the pipeline straightforward to use as all its ingredients are fully integrated, preinstalled and preconfigured in a ready-made Singularity image. These consist of the Mothur and PICRUSt2 programs, the needed Mothur scripts, two Silva taxonomy databases (V138 and V138 seed), a few supporting utilities written in C++, as well as a skeleton framework consisting of slightly less than 11,000 lines of Python code which orchestrates the execution of individual pieces and takes care of executing programs and building their command lines. The actual parameters under which the workflow is executed are at the control of the user (Figure 2, ESM Figures R1–R3).

```
bostjan@Carnott:~$ singularity run /home/bostjan/gumpp_v1.simg /home/bostjan/gumpp_example_script.txt


GUMPP: General Unified Microbiome Profiling Pipeline:   V1.0   2021-Apr-07

Developers: Blaz Stres, blaz.stres@fgg.uni-lj.si
            Bostjan Murovec, bostjan.murovec@fe.uni-lj.si



-------------------------------------------------------------------
Initializing application
-------------------------------------------------------------------

Current date and time: 2021-05-04__15-10-25.


Determining number of processors: 64


Determining amount of system memory ...
    total:     527 GB
    free:       23 GB
    available: 520 GB



-------------------------------------------------------------------
Applying initial configuration parameters
-------------------------------------------------------------------

Configuration file:
    /home/bostjan/gumpp_example_script.txt

Input directory:
    /home/bostjan/Mothur_MiSeq

Output directory:
    /home/bostjan/Mothur_MiSeq_out_2021_April_ASV

History of workflow executions will be preserved.


Number of threads is not specified in the config file.
Applying number of processors from the operating system: 64


Determining imposed memory limit ...
    Memory consumption is not limited.
```

Figure 2. An example of the program startup and the initial checkups done by the Python code.

Aside from reproducible execution of the workflow and the control of algorithm settings, GUMPP offers some additional benefits. First, results of Mothur preprocessing may optionally be stored in a specially crafted storage area, where each result is associated with its full context (hash of input files, Mothur script and its values of template parameters, or other relevant information). This enables efficient workflow re-executions with different Mothur and PICRUSt2 parameters. When GUMPP detects that upon its re-execution only PICRUSt2 parameters are changed, it instantly recycles the previously obtained Mothur results. This opens up a possibility of efficient experimenting with changed PICRUSt2 parameters to observe their impact on end results. In addition, Mothur processing is split into a common and an analysis specific part. If only analysis type or its related parameters are changed, the previously computed common results are again recycled instantly, which is a significant time saver, since the common part consists of e.g., sequence alignment to a taxonomy database. The system also enables crash recovery: in the case of GUMPP interruption during e.g., PICRUSt2 step (operating system crash, power outage, abort due to administrative policies on High-Performance Computer (HPC), upon restart only the PICRUSt2 step is re-executed. Crash recovery is completely automatic and transparent. A user need not to specify any directives to inform GUMPP that execution is being repeated.

The system is suitable for autonomous execution on domestic hardware as well as on HPC facilities. All command-line parameters and intermediate file formats are handled automatically by the system, enabling the experienced users to prescribe their own parameters for PICRUSt2 or for template Mothur script parameters in order to finetune the workflow execution.

In order to aid in documenting analyses and inspection of execution, GUMPP stores an accurate verbatim copy of its screen output as a part of end report. Also, the actual command lines, standard output streams, standard error streams and exit codes of individual programs are stored on a disk in a hierarchical way for easy navigation, inspection and debugging. Analysis setup relies on configuration files, where a complete workflow configuration is prescribed and hence also documented. GUMPP presented in this study thus builds on the highly popular and tested programs that were benchmarked in numerous past studies as reported before [16–20,27,30,31].

2.2. Reanalysis and Extension of Mice Gut Microbiome Data Using GUMPP: The Choice of Level of Analysis (Genus, OTU, ASV) Is far from Arbitrary

Mice data analysis using GUMPP enabled us to explore a technical question of how user reports on different taxonomic levels (genus; OTU; ASV) affected the exact relationships between underlying samples when studied utilizing the four data types (16S rRNA; functional genes; enzyme reactions; metabolic pathways). The results of Mantel test between taxonomic levels (Figure 3) show that the correlations between 16S rRNA vs. KO, 16S rRNA vs. EC and 16S rRNA vs. pathways decreased from 0.90, 0.91 and 0.90 at genus level, to 0.75, 0.75 and 0.76 at OTU level, and to 0.61, 0.61 and 0.66 at ASV level, respectively (n = 9999 permutations, $p < 0.0002$). The fact that ASV type of analysis resulted in lower correlations between datatypes is in line with past observations that there is little congruency between rather variable taxonomic descriptions of microbial communities and their corresponding even more diverse metagenomic functional gene makeup [36].

A between level analysis for each data type separately (Figure 4) illustrates the relationships between data of the same type, obtained using a different taxonomic level of analysis (Genus, OTU or ASV). The correlations > 0.88, describing the relationships between samples were retained only for distance matrices from genus and OTU levels of analyses and were also reproduced in all four data types (Figure 4). On the other hand, the initially high correlation between OTU and ASV at 16S rRNA level dropped below 0.55 for KO, EC and Pathway datasets, reflecting the increased number of categories (genus = 148, OUT = 1328, ASV = 13,244) and their different numerical abundance [11]. These results illustrate how the user selected levels of taxonomic assignment of the sequence data can affect the relationships between samples. Switching from genus level to utilizing ASV level of analysis does not only represent a way to maximize information content of the

underlying 16S rRNA sequences [30], but it also represents a distorting transformation of the information due to the many predominantly biological limitations of such analyses: (i) differences in 16S rRNA gene copy numbers range from 1 to 15 in bacteria and 1 to 5 in Archaea [37], hence a frequently recovered sequence may represent a high copy number taxon of lesser abundance, or a low copy number taxon of higher abundance. This 16S copy number of the organism that contributed the sequence is estimated and data adjusted accordingly by utilizing PICRUSt2 [27] in GUMPP; (ii) intragenomic heterogeneity of 16S rRNA operons can be as large as 20.4%. Genus level classification encompasses rather divergent sequences of that specific genus into one category. On the other hand, single nucleotide polymorphism present within e.g., 10 copies of 16S rRNA operon within one organism represent distinct ASVs. In comparison to genus level analysis 16S rRNA variants of one organism are split to several ASV categories inflating ASV estimates of microbial taxonomic diversity and of functional diversity of underlying metagenomes [38–40]; (iii) In contrast, almost identical 16S rRNA copies and hence the lack of differences found within some genera do not enable stratification of species and strains present within, falsely deflating the number of present ASVs [10,38–42]; (iv) different hypervariable regions of 16S rRNA utilized in amplicon sequencing can result in additional distortion of signal relative to each other [43] hence compromising direct comparison of the results between studies utilizing distinct primers.
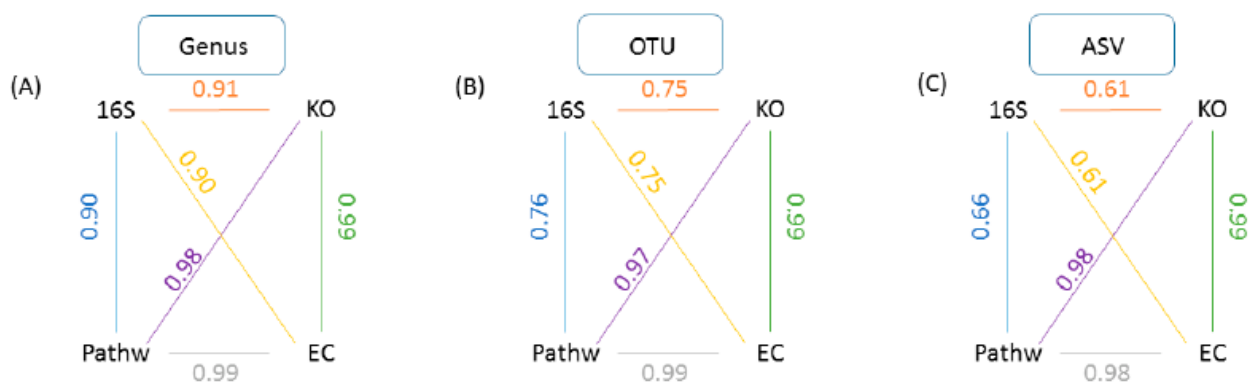
Figure 3. A within level analysis for all derived data types. A schematic representation of GUMPP generated data types analyzed at each of the three levels of 16S rRNA analysis (**A**) genus, (**B**) OTU, (**C**) ASV for the same sequence dataset and extended further to respective predicted functional genes (KO), enzymatic reactions (EC) and metabolic pathways (Pathw). Numbers designate the Mantel test correlation coefficients between various pairs of data types: (i) 16S and functional genes (KO)(orange), (ii) 16S and enzymatic reactions (EC) (yellow), (iii) 16S and metabolic pathways (Pathw) (blue), (iv) pathw and EC (pur-ple), (v) pathw and EC (grey), (vi) KO and EC (green). All analyses were performed with 9999 permutations and were statistically significant ($p = 0.0001$).

These cautionary notes listed above are intended to raise the awareness of the biological caveats of the genus, OUT and ASV levels of analyses for users. From this integrative view of biological influences the genus level analysis fits a more reserved type of analysis with arguably lower resolution, but congruent with an existing microbial taxonomy system in comparison to the ASV level of analysis, whereas OTU represents a compromise [14,44]. By utilizing ASV some genera expand into species and strains that have sufficient diversity within the 16S rRNA and contribute to ASVs, while other genera that contain species and strains with identical 16S rRNA in the region analyzed do not [14,44]. This biological distinction between genus, OTU or ASV levels of analysis has potentially large implications for the information forwarded to subsequent data types (functional genes, enzymatic reactions, metabolic pathays) irrespective of program utilized (PICRUSt, Tax4Fun, Piphillin or GUMPP).

Recent research highlights the risk of splitting a single bacterial genome into separate clusters when ASVs are used to analyze 16S rRNA gene sequence data. Although there is also a risk of clustering ASVs from different species into the same OTU when using broad distance thresholds, those risks are of less concern than artificially splitting a genome into separate ASVs and OTUs [14,44]. Based on the results presented here (Figures 3 and 4), the choice of level of analysis (genus, OTU, ASV) is far from arbitrary and may lead researchers to draw different biological conclusions. The work presented in this study highlights the utility of GUMPP that enables researchers to analyze the data at all three levels at the same time, generates functional gene, enzymatic reactions and metabolic pathways datasets for downstream machine learning exploration in relation to human diseases [44].
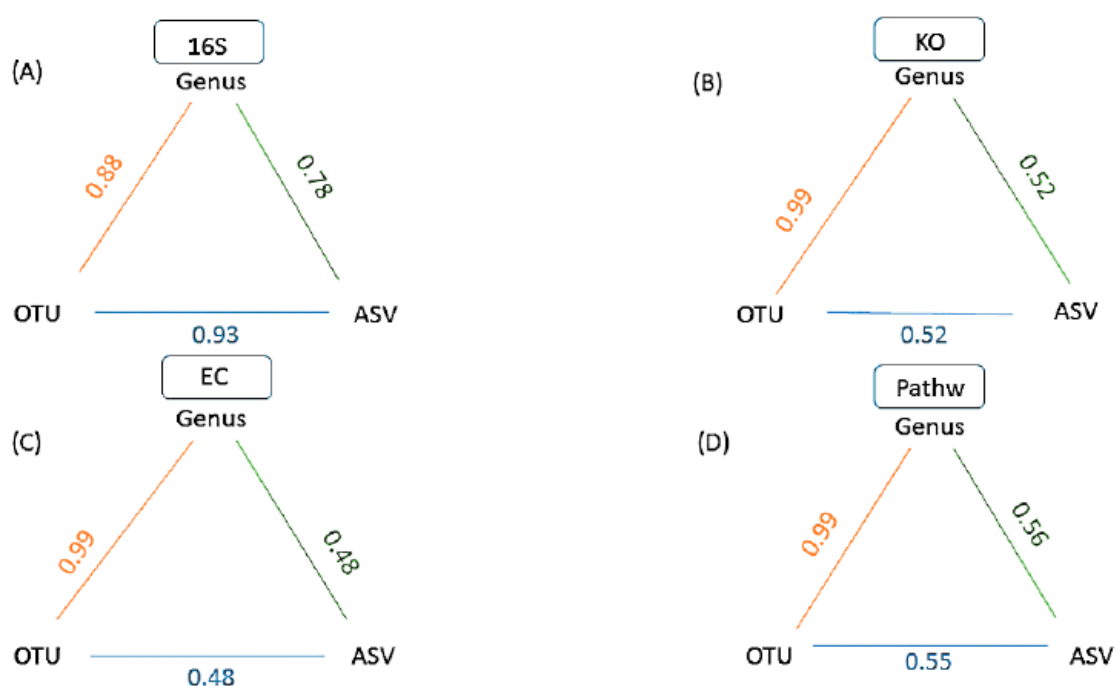
Figure 4. A schematic representation of GUMPP results showing a between level correlations for each data type: (**A**)16S rRNA (16S), (**B**) functional genes (KO), (**C**) enzymatic reactions (EC) and (**D**) metabolic pathways (Pathw). Numbers designate the Mantel test correlation coefficients between various pairs of levels for the same data type: (i) Genus and OTU (orange), (ii) OTU and ASV (blue), (iii) Genus and ASV (green). All analyses were performed with 9999 permutations and were statistically significant ($p = 0.0001$).

2.3. Reanalysis and Extension of Human Gut Microbiome Data Using GUMPP

In this study a reanalysis of published human gut data ($n = 307$) [45] was conducted utilizing GUMPP at the levels of 16S rRNA, predicted metagenomes, enzymatic reactions and metabolic pathways. Differences between the gastrointestinal patients ($n = 121$) from a single ward and 186 healthy volunteers were explored. This effectively enabled us to reproduce previously reported findings [45] utilizing GUMPP. Analyses were extended to three additional data types: predicted functional genes, enzymatic reactions and metabolic pathways. First, as reported before in the original study [45], gut microbial community description was not sufficient to differentiate the subjects based on their underlying five broad medical diagnoses: (i) ulcerative colitis; (ii) Crohn's disease, (iii) tumor (pancreatic, gastric or liver cancer), (iv) infection (pneumonia, cholangitis, hepatitis, gastritis or pancreatitis) and (v) other (cirrhosis or peptic ulcers, unidentifiable abdominal pain) [45]. The three mixed clusters independent of the underlying medical diagnosis were also reproduced (Figure S3), showing the robustness of GUMPP analysis. Second, by calculating the

statistical power for each medical diagnosis a much larger number of samples (within each medical diagnosis) would be needed ($n > 1000$) to be able to build classification models for each diagnosis (Table S2). Third, the PCA representation confirmed the existence of a core microbiome in healthy individuals as described in the original study [45]. Human gut microbiome in patients was disturbed and significantly altered relative to the healthy microbiome (Figure S3).

Extending the original 16S rRNA analysis by GUMPP derived datasets (functional genes (KO), enzymatic reactions (EC), metabolic pathways (pathway)) enabled us to explore the differences between the gastrointestinal patients and healthy volunteers utilizing machine learning. This coupling between GUMPP produced datasets and machine learning enabled us to generate, train and validate four separate models for classification of samples (Figure S3; Supplementary Electronic Material) using JADBIO AutoML approach [46,47]. In short, at all four data levels, logistic ridge regression with penalty hyperparameter lambda = 0.1 was selected as the best interpretable model with AUC metrics of 0.937 (16S rRNA), 0.949 (KO), 0.954 (EC), and 0.947 (pathway) (Figure S3). For the best microbial feature selection, LASSO algorithm was selected for the most differentiating pathways, and Test-Budgeted Statistically Equivalent Signature (SES) algorithm was selected for the search of the most differentiating 16S rRNA, KO and EC between groups of patients and healthy individuals. Models based on KO and EC data performed better than those based on 16S rRNA and pathway data (Figure S4).

The optimization of model selection allowed us to reliably identify microbial features (taxa, functional genes, enzymatic reactions, metabolic pathways) from datasets analyzed and produced by GUMPP (Figure 1, Figures S2 and S3) that discriminated between gut microbiomes of gastrointestinal patients and healthy volunteers: 25 taxonomy level 16S rRNA OTUs, four KOs, 12 ECs and 15 pathways (Table S1). As the complete in-depth biological description of these results is beyond the scope of this study, the major differences between the healthy in diseased groups at the level of metabolic pathways are reported (Figure 5). The following findings are highlighted as proof of concept of GUMPP extended data analysis: lactocepin (EC:3.4.21.96; K01361) was identified in this study as one of the most important features at the level of functional genes and enzymatic reactions distinguishing healthy from IBD, UC and CD. High lactocepin in healthy cohort is involved in the selective degradation of pro-inflammatory chemokines, leading to reduced cell infiltration and reduced inflammation in IBD models [48,49]. Further, Cu+-exporting ATPase were also found to be significantly increased in healthy, hence acting at the level of enzymatic reactions in metabolism [50]. In contrast, the elevated values of the P-type Mg^{2+} transporter observed in gastrointestinal patients were previously shown to be important for increased virulence in *Escherichia coli* and *Salmonella thyphimurium* [51]. Similarly, higher activity of enzyme maltose-6'-phosphate glucosidase were identified in the maltose degradation pathway of *Enterococcus faecalis* leading to increased virulence of this pathogen [52]. Another important enzyme NADH oxidase that exerts the main protection against oxidative stress in the human gut was low in the healthy group [53]. Thiazole component of thiamine diphosphate biosynthesis pathway I and thiamine phosphate synthase were identified as important for separation between healthy and diseased individuals [54–56]. One of the distinguishing features was also the peptidoglycan biosynthesis pathway IV, previously described in *Ruminococcus gnavus*, which is abundant in the intestines of patients with Crohn's disease [57]. Bifibacterium shunt was identified as another pathway that has been previously shown to be important in providing positive health benefits to their host with its metabolic activities [58].

These results illustrate the insight supported by GUMPP into the potential differences in the gut microbiomes, functional genes, enzymatic reactions and metabolic pathways between the diffuse group of gastrointestinal patients (five medical diagnoses) and healthy cohort coupled to machine learning.
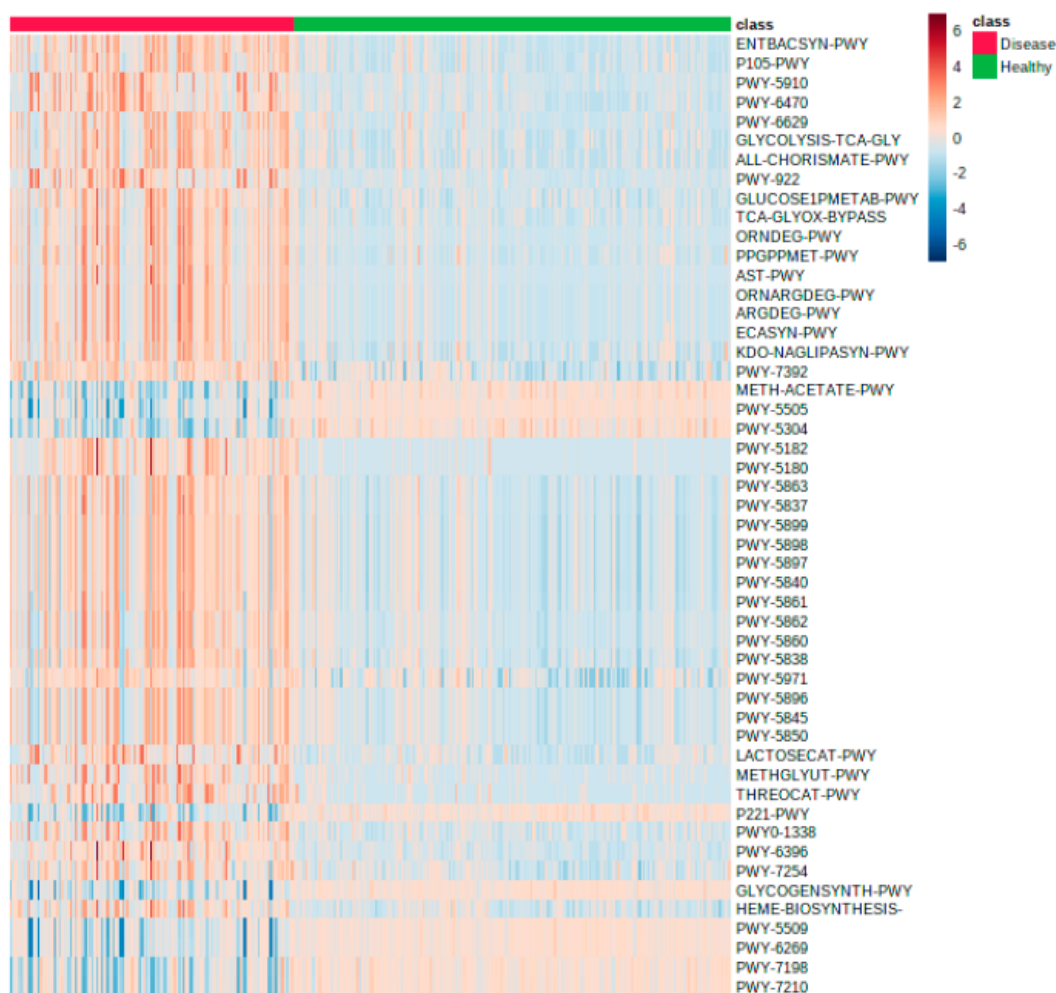
Figure 5. Heatmap showing the differences between the gastrointestinal patients (*n* = 121; red) from a single ward compared to 186 healthy volunteers (green) utilizing metabolic pathway information produced by GUMPP workflow from 16S rRNA data published before [45,59]. The first 50 most informative pathways are shown.

3. Materials and Methods

3.1. GUMPP Implementation

GUMPP utilization is described in user manual, Electronic Supplementary Materials, Config file, all available as part of this publication at http://gumpp.fe.uni-lj.si. Analyses running GUMPP were executed on a Dual Xeon system with 32 CPU cores (64 hyper-threads), 512 GB of RAM and 6 TB SATA disk. The runtime depends on the data size, sequencing depth and type of analysis (genus-, OTU-, ASV- level). For instance, human gut microbiome data analysis consisted of 307 samples, that each contained independent forward (R1) and reverse (R2) files. In total, it took <10 h, <50 h and <60 h runtime to finalize genus-, OTU- and ASV- levels of analyses, respectively. Similarly, runtime of analyzing less deeply sequenced mice dataset (*n* = 365 paired-end samples) took <4, <16 and <18 h to finalize genus-, OTU- and ASV- levels of analyses, respectively. Portability and HPC performance of the GUMPP generated in this study was confirmed on Leo3e (https://www.uibk.ac.at/zid/systeme/hpc-systeme/leo3e/ accessed on 30 April 2021) and Leo4 (https://www.uibk.ac.at/zid/systeme/hpc-systeme/leo4/ accessed on 30 April 2021) HPC infrastructure of the University of Innsbruck as described recently [60].

3.2. Sequence Data Collections

The workflow was tested using two large collections of data sets arising from human [45,59] and mice experiments ([7]; https://mothur.org/ accessed on 30 April 2021). In short, a multi-disease hospitalized cohort included various gastroenterological pathologies: ulcerative colitis, Crohn's disease, tumor, infection, cirrhosis and peptic ulcer, unidentifiable abdominal pain. Gastrointestinal patients ($n = 121$) from a single ward were compared to 186 healthy volunteers [45] in order to fine-map the gut microbiota dysbiosis, using the bacterial (V3 V4) amplicon sequencing. In total, 6.6 million pairs of sequences were analyzed with an average coverage of 35,484 pairs of sequence reads from the 16S rRNA gene.

The mice dataset explored the separation between daily murine fecal samples ($n = 360$) obtained from C57BL/6 male and female mice at 0 to 9 (early) and 141 to 150 (late) days after weaning [7]. In total, 4.3 million pairs of sequence reads from the 16S rRNA gene with an average coverage of 9913 pairs of V4V5 reads per sample [22] were analyzed. During the first 150 days post weaning mice were allowed ad libitum feed with no specific influence in order to monitor whether the rapid change in weight at 10 days post weaning (obesity) affected the stability microbiome compared to the microbiome observed between days 140 and 150.

3.3. Statistical Analyses and Machine Learning

The two 16S rRNA sequence data collections were analyzed using GUMPP and according to three layers of information, namely genus, 97% OTU and ASV, and the additional three data types were calculated using PICRUSt2 integrated in GUMPP: predicted metagenomes; enzyme reactions; metabolic pathways. Piphillin-ready outputs for clinical exploration were calculated alongside, formatted and prepared. The underlying settings used in these analyses are part of the GUMPP configuration file and can be utilized and shared among researchers for reproducibility and ease of additional calculations. The resulting genus level data analysis of human gut microbiomes (four data matrices (16S rRNA; metagenomes; enzyme reactions; metabolic pathways) were subjected to machine learning in JADBIO [47] (version 1.1.164) for identification of microbial, genetic, enzymatic and pathway variables responsible for separation of the healthy and patient groups.

JADBIO [47] provides high-quality predictive models for diagnostics using state-of-the-art statistical and machine learning methods. Personal analytical biases and methodological statistical errors were eliminated from the analysis by autonomous exploration of several settings in modeling steps, exploring wide analytical space and producing convincing discovered features to discriminate between patients and healthy individuals. The JADBIO approach was adopted for modeling because of number of reason: First, automated parameter and algorithm selection without human inference enables testing and coverage of a wide machine learning algorithm-settings space. Second, JADBIO includes several algorithms for feature selection and modeling (linear regression, SVM, decision trees, random forest and Gaussian kernel SVMs) and all possible options with different parameters are tested during the process. Third, the obtained models were trained with different configurations of sub-data of the original dataset (all results are cross-validated with recently developed Bootstrap Bias Corrected CV (BBC-CV) [61]). Fourth, analyses were run on data with biomedical characteristics (sparse matrices, nonnormal distributions). Algorithm, hyperparameter and space selection protocols (AHPS) in JADBIO were used for selecting the most appropriate algorithm for preprocessing and transformation of a given dataset, for feature selection and modeling. The output of AHPS step was then evaluated through the configuration evaluation protocol in order to find the optimal model configuration for a given dataset [46,47]. JADBIO 1.1.164 was used with extensive tuning effort and 6 CPU cores in modelling various dataset selections. All four datasets were split 70 to 30 according to machine learning protocols. The training set (70% of the data in the dataset) was used to build the best interpretable models and the rest of the data (30%) was used for performance validations at all four levels of data analysis (16S rRNA genus level (424 features), KO (6126 features), EC (1887 features), pathways (365 features)). The

area under the curve (AUC) metric was used to evaluate model performance. In total, the analytical space of algorithms and their corresponding settings was explored and 5960 of models and their individual settings were tested for genus and 11,920 for functional gene, enzymatic reactions and metabolic pathways, before the optimal configuration for the most informative model were obtained.

In addition to this, statistical power analysis of human microbiome data was performed [45,59] on all four data levels: 16S rRNA, KO, EC and pathways, between patients with different diseases and healthy individuals and according to presence/absence of the disease. Data was cube root normalized and mean centered. False discovery rate set to 0.1 was used in MetaboAnalyst module prepared for data analysis of population and metabolic studies [62].

All models created in analyses of the human gastrointestinal dataset can also be run on the local machine and are provided as part of the supplementary data (for local model execution, see the instructions in the electronic supplementary materials).

Mice data (n = 365) were processed and analyzed as described above in order to explore the differences between the four data types (16S rRNA; metagenomes; enzyme reactions; metabolic pathways) in terms of consistency of intersample relationships between the three layers of information routinely utilized in studies (genus; OTU; ASV). The intersample relationships were assessed by Mantel tests ($p < 0.0002$) utilizing (i) Pearson and (ii) Spearman correlation between data matrices (Bray-Curtis distance measure) and permutations (n = 9999) in either vegan-R [63] and/or PAST software (version 2.17c) [64]. The Mantel test tests the correlation between two distance matrices. It is non-parametric test and computes the significance of the correlation through permutations of the rows and columns of the input distance matrices.

4. Conclusions

By including the user preferences of genus, OTU or ASV type of analyses, GUMPP is the first workflow that introduces traceability and portability of all its parameters used in analyses. The workflow integrates and orchestrates end to end the inputs and outputs of the highly cited programs Mothur, PICRUSt2 and Pipihillin, controlled by Python code, delivered as portable Singularity image and accompanied by customizable configuration files. The whole GUMPP workflow can be executed for teaching or/and research purposes using personal computer or HPC. The ability to support reproducible analyses enables production of datasets that match multiomics layers of information, such as metagenomics, metaproteomics and metabolomics that ultimately identify the biochemical pathways characteristic of certain pathology [8]. These datasets coupled to biostatistics and mathematical approaches of machine learning can play significant role in extraction of truly significant and meaningful information from wide array of previously unexplored datasets (e.g., [45,59]) in relation to (i) a number of diseases (metabolic [65] or neurodegenerative [66] diseases), (ii) medical interventions, manipulations of bacteria-gut-brain axis [67] or (iii) treatment strategies for complex diseases [68]. The adoption of GUMPP in the gut-microbiota related research enables focusing on the identification of novel biomarkers that can lead to the development of mechanistic hypotheses applicable to the development of novel therapies in personalized medicine [2,9].

References

1. Stres, B.; Kronegger, L. Shift in the paradigm towards next-generation microbiology. *FEMS Microbiol. Lett.* **2019**, *366*. [CrossRef]
2. Vernocchi, P.; Del Chierico, F.; Putignani, L. Gut Microbiota Profiling: Metabolomics Based Approach to Unravel Compounds Affecting Human Health. *Front. Microbiol.* **2016**, *7*. [CrossRef]
3. Wu, J.; Wang, K.; Wang, X.; Pang, Y.; Jiang, C. The role of the gut microbiome and its metabolites in metabolic diseases. *Protein Cell* **2020**. [CrossRef]
4. Visconti, A.; Le Roy, C.I.; Rosa, F.; Rossi, N.; Martin, T.C.; Mohney, R.P.; Li, W.; de Rinaldis, E.; Bell, J.T.; Venter, J.C.; et al. Interplay between the human gut microbiome and host metabolism. *Nat. Commun.* **2019**, *10*, 4505. [CrossRef]
5. Lee-Sarwar, K.A.; Lasky-Su, J.; Kelly, R.S.; Litonjua, A.A.; Weiss, S.T. Metabolome-Microbiome Crosstalk and Human Disease. *Metabolites* **2020**, *10*, 181. [CrossRef] [PubMed]
6. Kappel, B.A.; De Angelis, L.; Heiser, M.; Ballanti, M.; Stoehr, R.; Goettsch, C.; Mavilio, M.; Artati, A.; Paoluzi, O.A.; Adamski, J.; et al. Cross-omics analysis revealed gut microbiome-related metabolic pathways underlying atherosclerosis development after antibiotics treatment. *Mol. Metab.* **2020**, *36*. [CrossRef] [PubMed]
7. Wilmanski, T.; Rappaport, N.; Earls, J.C.; Magis, A.T.; Manor, O.; Lovejoy, J.; Omenn, G.S.; Hood, L.; Gibbons, S.M.; Price, N.D. Blood metabolome predicts gut microbiome alpha-diversity in humans. *Nat. Biotechnol.* **2019**, *37*, 1217–1228. [CrossRef] [PubMed]
8. Jiang, D.; Armour, C.R.; Hu, C.; Mei, M.; Tian, C.; Sharpton, T.J.; Jiang, Y. Microbiome Multi-Omics Network Analysis: Statistical Considerations, Limitations, and Opportunities. *Front. Genet.* **2019**, *10*. [CrossRef]
9. Wang, Q.; Wang, K.; Wu, W.; Giannoulatou, E.; Ho, J.W.K.; Li, L. Host and microbiome multi-omics integration: Applications and methodologies. *Biophys. Rev.* **2019**, *11*. [CrossRef]
10. Poretsky, R.; Rodriguez-R, L.M.; Luo, C.; Tsementzi, D.; Konstantinidis, K.T. Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. *PLoS ONE* **2014**, *9*. [CrossRef]
11. Pasolli, E.; Schiffer, L.; Manghi, P.; Renson, A.; Obenchain, V.; Truong, D.T.; Beghini, F.; Malik, F.; Ramos, M.; Dowd, J.B.; et al. Accessible, curated metagenomic data through ExperimentHub. *Nat. Methods* **2017**, *14*. [CrossRef]
12. Rühlemann, M.C.; Hermes, B.M.; Bang, C.; Doms, S.; Moitinho-Silva, L.; Thingholm, L.B.; Frost, F.; Degenhardt, F.; Wittig, M.; Kässens, J.; et al. Genome-wide association study in 8,956 German individuals identifies influence of ABO histo-blood groups on gut microbiome. *Nat. Genet.* **2021**, *53*. [CrossRef] [PubMed]
13. Mysara, M.; Vandamme, P.; Props, R.; Kerckhof, F.M.; Leys, N.; Boon, N.; Raes, J.; Monsieurs, P. Reconciliation between operational taxonomic units and species boundaries. *FEMS Microbiol. Ecol.* **2017**, *93*. [CrossRef] [PubMed]
14. Schloss, P.D. Amplicon sequence variants artificially split bacterial genomes into separate clusters. *bioRxiv* **2021**. [CrossRef]
15. Callahan, B.J.; McMurdie, P.J.; Holmes, S.P. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J.* **2017**, *11*. [CrossRef]
16. Nilakanta, H.; Drews, K.L.; Firrell, S.; Foulkes, M.A.; Jablonski, K.A. A review of software for analyzing molecular sequences. *BMC Res. Notes* **2014**, *7*, 1–9. [CrossRef]

17. Pollock, J.; Glendinning, L.; Wisedchanwet, T.; Watson, M. The Madness of Microbiome: Attempting To Find Consensus "Best Practice" for 16S Microbiome Studies. *Appl. Environ. Microbiol.* **2018**, *84*. [CrossRef]
18. Schloss, P.D. Reintroducing mothur: 10 Years Later. *Appl. Environ. Microbiol.* **2020**, *86*, e02343-19. [CrossRef]
19. López-García, A.; Pineda-Quiroga, C.; Atxaerandio, R.; Pérez, A.; Hernández, I.; García-Rodríguez, A.; González-Recio, O. Comparison of Mothur and QIIME for the Analysis of Rumen Microbiota Composition Based on 16S rRNA Amplicon Sequences. *Front. Microbiol.* **2018**, *9*. [CrossRef]
20. Winand, R.; Bogaerts, B.; Hoffman, S.; Lefevre, L.; Delvoye, M.; Braekel, J.V.; Fu, Q.; Roosens, N.H.; Keersmaecker, S.C.; Vanneste, K. Targeting the 16S rrna gene for bacterial identification in complex mixed samples: Comparative evaluation of second (illumina) and third (oxford nanopore technologies) generation sequencing technologies. *Int. J. Mol. Sci.* **2019**, *21*, 298. [CrossRef]
21. Turnbaugh, P.J.; Ley, R.E.; Hamady, M.; Fraser-Liggett, C.M.; Knight, R.; Gordon, J.I. The human microbiome project. *Nature* **2007**, *449*. [CrossRef] [PubMed]
22. Kozich, J.J.; Westcott, S.L.; Baxter, N.T.; Highlander, S.K.; Schloss, P.D. Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform. *Appl. Environ. Microbiol.* **2013**. [CrossRef] [PubMed]
23. Prodan, A.; Tremaroli, V.; Brolin, H.; Zwinderman, A.H.; Nieuwdorp, M.; Levin, E. Comparing bioinformatic pipelines for microbial 16S rRNA amplicon sequencing. *PLoS ONE* **2020**, *15*. [CrossRef]
24. Dhariwal, A.; Chong, J.; Habib, S.; King, I.L.; Agellon, L.B.; Xia, J. MicrobiomeAnalyst: A web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. *Nucleic Acids Res.* **2017**, *45*. [CrossRef] [PubMed]
25. Chong, J.; Liu, P.; Zhou, G.; Xia, J. Using MicrobiomeAnalyst for comprehensive statistical, functional, and meta-analysis of microbiome data. *Nat. Protoc.* **2020**, *15*. [CrossRef] [PubMed]
26. Langille, M.G.; Zaneveld, J.; Caporaso, J.G.; McDonald, D.; Knights, D.; Reyes, J.A.; Clemente, J.C.; Burkepile, D.E.; Vega Thurber, R.L.; Knight, R.; et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* **2013**, *31*. [CrossRef] [PubMed]
27. Douglas, G.M.; Maffei, V.J.; Zaneveld, J.R.; Yurgel, S.N.; Brown, J.R.; Taylor, C.M.; Huttenhower, C.; Langille, M.G.I. PICRUSt2 for prediction of metagenome functions. *Nat. Biotechnol.* **2020**, *38*. [CrossRef]
28. Aßhauer, K.P.; Wemheuer, B.; Daniel, R.; Meinicke, P. Tax4Fun: Predicting functional profiles from metagenomic 16S rRNA data. *Bioinformatics* **2015**, *31*. [CrossRef]
29. Wemheuer, F.; Taylor, J.A.; Daniel, R.; Johnston, E.; Meinicke, P.; Thomas, T.; Wemheuer, B. Tax4Fun2: Prediction of habitat-specific functional profiles and functional redundancy based on 16S rRNA gene sequences. *Environ. Microb.* **2020**, *15*. [CrossRef]
30. Narayan, N.R.; Weinmaier, T.; Laserna-Mendieta, E.J.; Claesson, M.J.; Shanahan, F.; Dabbagh, K.; Iwai, S.; DeSantis, T.Z. Piphillin predicts metagenomic composition and dynamics from DADA2-corrected 16S rDNA sequences. *BMC Genom.* **2020**, *21*. [CrossRef]
31. Iwai, S.; Weinmaier, T.; Schmidt, B.L.; Albertson, D.G.; Poloso, N.J.; Dabbagh, K.; DeSantis, T.Z. Piphillin: Improved Prediction of Metagenomic Content by Direct Inference from Human Microbiomes. *PLoS ONE* **2016**, *11*. [CrossRef] [PubMed]
32. Sun, S.; Jones, R.B.; Fodor, A.A. Inference-based accuracy of metagenome prediction tools varies across sample types and functional categories. *Microbiome* **2020**, *8*. [CrossRef] [PubMed]
33. Salosensaari, A.; Laitinen, V.; Havulinna, A.S.; Meric, G.; Cheng, S.; Perola, M.; Valsta, L.; Alfthan, G.; Inouye, M.; Watrous, J.D.; et al. Taxonomic signatures of cause-specific mortality risk in human gut microbiome. *Nat. Commun.* **2021**, *12*, 2671. [CrossRef]
34. Schloss, P.D.; Westcott, S.L.; Ryabin, T.; Hall, J.R.; Hartmann, M.; Hollister, E.B.; Lesniewski, R.A.; Oakley, B.B.; Parks, D.H.; Robinson, C.J.; et al. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **2009**, *75*, 7537–7541. [CrossRef] [PubMed]
35. Kurtzer, G.M.; Sochat, V.; Bauer, M.W. Singularity: Scientific containers for mobility of compute. *PLoS ONE* **2017**, *12*. [CrossRef]
36. Turnbaugh, P.J.; Hamady, M.; Yatsunenko, T.; Cantarel, B.L.; Duncan, A.; Ley, R.E.; Sogin, M.L.; Jones, W.J.; Roe, B.A.; Affourtit, J.P.; et al. A core gut microbiome in obese and lean twins. *Nature* **2009**, *457*. [CrossRef]
37. Stoddard, S.F.; Smith, B.J.; Hein, R.; Roller, B.R.; Schmidt, T.M. rrnDB: Improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. *Nucleic Acids Res.* **2015**, *43*. [CrossRef]
38. Větrovský, T.; Baldrian, P. The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PLoS ONE* **2013**, *8*. [CrossRef] [PubMed]
39. Nguyen, N.P.; Warnow, T.; Pop, M.; White, B. A perspective on 16S rRNA operational taxonomic unit clustering using sequence similarity. *NPJ Biofilms Microb.* **2016**, *2*, 16004. [CrossRef]
40. Pei, A.Y.; Oberdorf, W.E.; Nossa, C.W.; Agarwal, A.; Chokshi, P.; Gerz, E.A.; Jin, Z.; Lee, P.; Yang, L.; Poles, M.; et al. Diversity of 16S rRNA genes within individual prokaryotic genomes. *Appl. Environ. Microbiol.* **2010**, *76*. [CrossRef]
41. Sun, D.L.; Jiang, X.; Wu, Q.L.; Zhou, N.Y. Intragenomic heterogeneity of 16S rRNA genes causes overestimation of prokaryotic diversity. *Appl. Environ. Microbiol.* **2013**, *79*. [CrossRef] [PubMed]
42. Huse, S.M.; Dethlefsen, L.; Huber, J.A.; Mark Welch, D.; Relman, D.A.; Sogin, M.L. Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. *PLoS Genet.* **2008**, *4*. [CrossRef]
43. Soriano-Lerma, A.; Pérez-Carrasco, V.; Sánchez-Marañón, M.; Ortiz-González, M.; Sánchez-Martín, V.; Gijón, J.; Navarro-Mari, J.M.; García-Salcedo, J.A.; Soriano, M. Influence of 16S rRNA target region on the outcome of microbiome studies in soil and saliva samples. *Sci. Rep.* **2020**, *10*. [CrossRef]

44. Joos, L.; Beirinckx, S.; Haegeman, A.; Debode, J.; Vandecasteele, B.; Baeyen, S.; Goormachtig, S.; Clement, L.; De Tender, C. Daring to be differential: Metabarcoding analysis of soil and plant-related microbial communities using amplicon sequence variants and operational taxonomical units. *BMC Genom.* **2020**, *21*. [CrossRef] [PubMed]

45. Mahnic, A.; Breskvar, M.; Dzeroski, S.; Skok, P.; Pintar, S.; Rupnik, M. Distinct Types of Gut Microbiota Dysbiosis in Hospitalized Gastroenterological Patients Are Disease Non-related and Characterized With the Predominance of Either Enterobacteriaceae or Enterococcus. *Front. Microbiol.* **2020**, *11*. [CrossRef]

46. Mustafa, A.; Rahimi Azghadi, M. Automated Machine Learning for Healthcare and Clinical Notes Analysis. *Computers* **2021**, *10*, 24. [CrossRef]

47. Tsamardinos, I.; Charonyktakis, P.; Lakiotaki, K.; Borboudakis, G.; Zenklusen, J.C.; Juhl, H.; Chatzaki, E.; Lagani, V. Just Add Data: Automated Predictive Modeling and BioSignature Discovery. *bioRxiv* **2020**. [CrossRef]

48. Hörmannsperger, G.; von Schillde, M.A.; Haller, D. Lactocepin as a protective microbial structure in the context of IBD. *Gut Microbes* **2013**, *4*. [CrossRef]

49. von Schillde, M.A.; Hörmannsperger, G.; Weiher, M.; Alpert, C.A.; Hahne, H.; Bäuerl, C.; van Huynegem, K.; Steidler, L.; Hrncir, T.; Pérez-Martínez, G.; et al. Lactocepin secreted by Lactobacillus exerts anti-inflammatory effects by selectively degrading proinflammatory chemokines. *Cell Host Microbe* **2012**, *11*. [CrossRef]

50. Osman, D.; Patterson, C.J.; Bailey, K.; Fisher, K.; Robinson, N.J.; Rigby, S.E.; Cavet, J.S. The copper supply pathway to a Salmonella Cu,Zn-superoxide dismutase (SodCII) involves P(1B)-type ATPase copper efflux and periplasmic CueP. *Mol. Microbiol.* **2013**, *87*. [CrossRef]

51. Subramani, S.; Perdreau-Dahl, H.; Morth, J.P. The magnesium transporter A is activated by cardiolipin and is highly sensitive to free magnesium in vitro. *eLife* **2016**, *5*. [CrossRef] [PubMed]

52. Joyet, P.; Mokhtari, A.; Riboulet-Bisson, E.; Blancato, V.S.; Esparíz, M.; Magni, C.; Hartke, A.; Deutscher, J.; Sauvageot, N. Enzymes Required for Maltodextrin Catabolism in Enterococcus faecalis Exhibit Novel Activities. *Appl. Environ. Microbiol.* **2017**, *83*. [CrossRef] [PubMed]

53. Yan, M.; Yin, W.; Fang, X.; Guo, J.; Shi, H. Characteristics of a water-forming NADH oxidase from Methanobrevibacter smithii, an archaeon in the human gut. *Biosci. Rep.* **2016**, *36*. [CrossRef]

54. Yoshii, K.; Hosomi, K.; Sawane, K.; Kunisawa, J. Metabolism of Dietary and Microbial Vitamin B Family in the Regulation of Host Immunity. *Front. Nutr.* **2019**, *6*. [CrossRef]

55. LeBlanc, J.G.; Milani, C.; de Giori, G.S.; Sesma, F.; van Sinderen, D.; Ventura, M. Bacteria as vitamin suppliers to their host: A gut microbiota perspective. *Curr. Opin. Biotechnol.* **2013**, *24*. [CrossRef] [PubMed]

56. Rodionov, D.A.; Arzamasov, A.A.; Khoroshkin, M.S.; Iablokov, S.N.; Leyn, S.A.; Peterson, S.N.; Novichkov, P.S.; Osterman, A.L. Micronutrient Requirements and Sharing Capabilities of the Human Gut Microbiome. *Front. Microbiol.* **2019**, *10*. [CrossRef]

57. Henke, M.T.; Kenny, D.J.; Cassilly, C.D.; Vlamakis, H.; Xavier, R.J.; Clardy, J. Ruminococcus gnavus, a member of the human gut microbiome associated with Crohn's disease, produces an inflammatory polysaccharide. *Proc. Natl. Acad. Sci. USA* **2019**, *116*. [CrossRef]

58. O'Callaghan, A.; van Sinderen, D. Bifidobacteria and Their Role as Members of the Human Gut Microbiota. *Front. Microbiol.* **2016**, *7*. [CrossRef]

59. Mahnic, A.; Rupnik, M. Different host factors are associated with patterns in bacterial and fungal gut microbiota in Slovenian healthy cohort. *PLoS ONE* **2018**, *13*. [CrossRef]

60. Murovec, B.; Deutsch, L.; Stres, B. Computational Framework for High-Quality Production and Large-Scale Evolutionary Analysis of Metagenome Assembled Genomes. *Mol. Biol. Evol.* **2020**, *37*. [CrossRef]

61. Tsamardinos, I.; Rakhshani, A.; Lagani, V. Performance-Estimation Properties of Cross-Validation-Based Protocols with Simultaneous Hyper-Parameter Optimization | SpringerLink. In *Artificial Intelligence: Methods and Applications*; Likas, A., Blekas, K., Kalles, D., Eds.; Springer International Publishing: Cham, Switzerland, 2014.

62. Chong, J.; Soufan, O.; Li, C.; Caraus, I.; Li, S.; Bourque, G.; Wishart, D.S.; Xia, J. MetaboAnalyst 4.0: Towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res.* **2018**, *46*. [CrossRef] [PubMed]

63. Dixon, P. VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* **2003**, *14*, 927–930. [CrossRef]

64. Hammer, O.; Harper, D.A.T.; Ryan, P.D. PAST: Paleontological statistics software package for education and data analysis. *Palaeontol. Electron.* **2001**, *1*, 9.

65. Proffitt, C.; Bidkhori, G.; Moyes, D.; Shoaie, S. Disease, Drugs and Dysbiosis: Understanding Microbial Signatures in Metabolic Disease and Medical Interventions. *Microorganisms* **2020**, *8*, 1381. [CrossRef] [PubMed]

66. Rosario, D.; Boren, J.; Uhlen, M.; Proctor, G.; Aarsland, D.; Mardinoglu, A.; Shoaie, S. Systems Biology Approaches to Understand the Host-Microbiome Interactions in Neurodegenerative Diseases. *Front. Neurosci.* **2020**, *14*. [CrossRef] [PubMed]

67. Sarkar, A.; Lehto, S.M.; Harty, S.; Dinan, T.G.; Cryan, J.F.; Burnet, P.W.J. Psychobiotics and the Manipulation of Bacteria-Gut-Brain Signals. *Trends Neurosci.* **2016**, *39*. [CrossRef]

68. Vijay, A.; Valdes, A.M. The Metabolomic Signatures of Weight Change. *Metabolites* **2019**, *9*, 67. [CrossRef]

2.1.3 Spinal muscular atrophy after nusinersen therapy: improved physiology in pediatric patients with no significant change in urine, serum, and liquor ^{1}H-NMR metabolomes in comparison to an age-matched, healthy cohort

Deutsch L., Osredkar D., Plavec J., Stres B. 2021. Spinal muscular atrophy after nusinersen therapy: improved physiology in pediatric patients with no significant change in urine, serum, and liquor ^{1}H-NMR metabolomes in comparison to an age-matched, healthy cohort. Metabolites, 11, 4: 206, doi. https://doi.org/10.3390/metabo11040206, 15 p.

Abstract

Spinal muscular atrophy (SMA) is a genetically heterogeneous group of rare neuromuscular diseases and was until recently the most common genetic cause of death in children. The effects of 2-month nusinersen therapy on urine, serum, and liquor ^{1}H-NMR metabolomes in SMA males and females were not explored yet, especially not in comparison to the urine ^{1}H-NMR metabolomes of matching male and female cohorts. In this prospective, single-centered study, urine, serum, and liquor samples were collected from 25 male and female pediatric patients with SMA before and after 2 months of nusinersen therapy and urine samples from a matching healthy cohort (n = 125). Nusinersen intrathecal application was the first therapy for the treatment of SMA by the Food and Drug Administration (FDA) and the European Medicines Agency (EMA). Metabolomes were analyzed using targeted metabolomics utilizing 600 MHz ^{1}H-NMR, parametric and nonparametric multivariate statistical analyses, machine learning, and modeling. Medical assessment before and after nusinersen therapy showed significant improvements of movement, posture, and strength according to various medical tests. No significant differences were found in metabolomes before and after nusinersen therapy in urine, serum, and liquor samples using an ensemble of statistical and machine-learning approaches. In comparison to a healthy cohort, ^{1}H-NMR metabolomes of SMA patients contained a reduced number and concentration of urine metabolites and differed significantly between males and females as well. Significantly larger data scatter was observed for SMA patients in comparison to matched healthy controls. Machine learning confirmed urinary creatinine as the most significant, distinguishing SMA patients from the healthy cohort. The positive effects of nusinersen therapy clearly preceded or took place devoid of significant rearrangements in the ^{1}H-NMR metabolomic makeup of serum, urine, and liquor. Urine creatinine was successful at distinguishing SMA patients from the matched healthy cohort, which is a simple systemic novelty linking creatinine and SMA to the physiology of inactivity and diabetes, and it facilitates the monitoring of SMA disease in pediatric patients through non-invasive urine collection.

For my personal contributions as a doctoral student and author of this thesis, please refer to Table 2 (page 142). The hypothesis from section 1.4.2 from this work were discussed in this paper.

metabolites

MDPI

Article

Spinal Muscular Atrophy after Nusinersen Therapy: Improved Physiology in Pediatric Patients with No Significant Change in Urine, Serum, and Liquor 1H-NMR Metabolomes in Comparison to an Age-Matched, Healthy Cohort

Leon Deutsch [1], Damjan Osredkar [2,3], Janez Plavec [4] and Blaž Stres [1,5,6,7,*]

1. Department of Animal Science, Biotechnical Faculty, University of Ljubljana, SI-1000 Ljubljana, Slovenia; leon.deutsch@bf.uni-lj.si
2. Department of Pediatric Neurology, University Children's Hospital, University Medical Centre Ljubljana, SI-1000 Ljubljana, Slovenia; damjan.osredkar@kclj.si
3. Faculty of Medicine, University of Ljubljana, SI-1000 Ljubljana, Slovenia
4. National Institute of Chemistry, NMR Center, SI-1000 Ljubljana, Slovenia; janez.plavec@ki.si
5. Department of Automation, Biocybernetics and Robotics, Jožef Stefan Institute, SI-1000 Ljubljana, Slovenia
6. Faculty of Civil and Geodetic Engineering, Institute of Sanitary Engineering, University of Ljubljana, SI-1000 Ljubljana, Slovenia
7. Department of Microbiology, University of Innsbruck, A-6020 Innsbruck, Austria
* Correspondence: blaz.stres@bf.uni-lj.si; Tel.: +386-41567633

Abstract: Spinal muscular atrophy (SMA) is a genetically heterogeneous group of rare neuromuscular diseases and was until recently the most common genetic cause of death in children. The effects of 2-month nusinersen therapy on urine, serum, and liquor 1H-NMR metabolomes in SMA males and females were not explored yet, especially not in comparison to the urine 1H-NMR metabolomes of matching male and female cohorts. In this prospective, single-centered study, urine, serum, and liquor samples were collected from 25 male and female pediatric patients with SMA before and after 2 months of nusinersen therapy and urine samples from a matching healthy cohort (*n* = 125). Nusinersen intrathecal application was the first therapy for the treatment of SMA by the Food and Drug Administration (FDA) and the European Medicines Agency (EMA). Metabolomes were analyzed using targeted metabolomics utilizing 600 MHz 1H-NMR, parametric and nonparametric multivariate statistical analyses, machine learning, and modeling. Medical assessment before and after nusinersen therapy showed significant improvements of movement, posture, and strength according to various medical tests. No significant differences were found in metabolomes before and after nusinersen therapy in urine, serum, and liquor samples using an ensemble of statistical and machine learning approaches. In comparison to a healthy cohort, 1H-NMR metabolomes of SMA patients contained a reduced number and concentration of urine metabolites and differed significantly between males and females as well. Significantly larger data scatter was observed for SMA patients in comparison to matched healthy controls. Machine learning confirmed urinary creatinine as the most significant, distinguishing SMA patients from the healthy cohort. The positive effects of nusinersen therapy clearly preceded or took place devoid of significant rearrangements in the 1H-NMR metabolomic makeup of serum, urine, and liquor. Urine creatinine was successful at distinguishing SMA patients from the matched healthy cohort, which is a simple systemic novelty linking creatinine and SMA to the physiology of inactivity and diabetes, and it facilitates the monitoring of SMA disease in pediatric patients through non-invasive urine collection.

Keywords: spinal musular atrophy; nusinersen; 1H-NMR metabolomics; males; females; serum; liquor; urine; healthy control cohort

1. Introduction

The 5q spinal muscular atrophy (SMA) is a rare neuromuscular disorder, which leads to progressive atrophy and weakening of skeletal muscles due to the progressive loss of motor neurons [1]. It is characterized by degeneration of the motor neurons in the anterior horn of the spinal cord, resulting in atrophy and weakness of the voluntary muscles of the limbs and trunk. With the incidence of about one in 11,000 live births, it was, until the development of disease-modifying drugs, the most common genetic cause of child deaths [2]. The gene for the survival motor neuron (SMN) was localized in 1990 by two different groups [3,4] on chromosome 5q13 and in 1996 was identified as the disease-causing gene. The SMN protein is produced by two genes, the telomeric SMN1 gene, which is deleted or interrupted in patients with SMA, and the centromeric SMN2 gene, which differs from SMN1 by five nucleotides and is present in several copies (Figure S1). SMN1 produces full-length transcripts, while SMN2 in 90% produces transcripts without exon 7, because of the C to T mutation that produces an exon-splicing suppressor sequence [5,6]. In SMA patients, SMN2 is present in at least one copy and is usually truncated because of C to T substitution (c.840C $\rightarrow$ T), since in 90–95% of the cases, the exon is spliced out of the produced SMN. The remaining 10% produce the SMN protein in full length, which indicates a low SMN expression [7]. SMN is a 38 kDa protein that is expressed in all somatic tissues and is located in the nucleus, in the cytoplasm, and in the axons of motor neurons. Its role is not yet fully understood, but the phenotype of the SMA depends largely on the number of SMN2 gene copies present [8].

However, SMA is a motor neuron disease that also affects the skeletal muscle, heart, kidney, liver, pancreas, spleen, bone, connective tissues, and immune systems [9].

Novel therapies for the treatment of SMA have emerged that modify the natural course of the disease by modifying the expression or replacing mutated genes involved in the development of SMA. Such treatment options are currently nusinersen [10] (an antisense oligonucleotide that modifies mRNA splicing) (Figure S1), onasemnogene abeparvovec [11] (gene replacement therapy), and risdiplam [12] (a small molecule that modifies mRNA splicing). Nusinersen was the first drug approved for SMA. It was approved by the US Food and Drug Administration (FDA) in December 2016 and by the European Medicines Agency (EMA) in June 2017 [13,14]. Nusinersen (Figure S1) is an antisense oligonucleotide that promotes the inclusion of exon 7 into mRNA transcripts of SMN2. It binds to an intronic splice site in intron 7 of SMN2 and inhibits the action of other splicing factors, thereby promoting exon 7 incorporation into the mRNA. This leads to the production of a fully functional SMN protein. Antisense oligonucleotides do not cross the blood–brain barrier, which means that they must be administered intrathecally [15–17]. The nusinersen drug improved motor function and increased the amplitude of muscle action potential of the ulnar and peroneal nerve. An autopsy analysis showed the uptake of nusinersen into motor neurons throughout the spinal cord and into neurons in the brainstem [18]. Increased motor function was manifested as an increased ability to sit or walk independently [19–21], increased bite force [22], and increased hand strength [23]. Although nusinersen is administered intrathecally, which requires a lot of expertise, it has been shown that such application is well tolerated and safe [24]. Nusinersen is available in many countries for most types of SMA patients, depending mostly on inclusion criteria and financing defined by the country of residence. However, there is an enormous need for real-world evidence of nusinersen efficacy, for better understanding of the variability of effect and side effects in a broader cohort of SMA patients [25].

In the last two decades, new approaches have been developed in the natural sciences. These include various 'omics techniques: from genomics, which is an important method for establishing the SMA diagnosis, to metatranscriptomics, proteomics, and metabolomics to monitor and investigate disease [7]. A systemic approach to therapy during early development is most likely to maximize the positive clinical outcome. Metabolomics is a useful method to evaluate the metabolites that we can identify in different biological samples, leading to the end of a cascade of biological processes, hence helping us to under-

stand the molecular phenotype and the underlying metabolic mechanisms [26]. Nuclear magnetic resonance spectroscopy (NMR) is one of the most widely used approaches due to its minimal sample preparation, non-destructive measurement method, quantitative aspect, and high reproducibility [27–29]. 1H-NMR has been used to study the modulation of metabolites to cellular stress [30], breast cancer markers [31], acute pancreatitis [32], influence of metabolome on health and disease [33], biomarkers for Crohn's disease and ulcerative colitis [34], obesity [35], coronary heart disease and stroke [36], next to physiological deconditioning through inactivity and hypoxia [37]. In line with these reports, serum creatinine was only recently identified as a potential biomarker for monitoring the SMA progression of denervation with decreasing levels reflecting the severity of the disease [38]. However, it is currently unknown whether (i) creatinine concentration showed improvement upon early medical intervention with nusinersen therapy, (ii) whether these changes are detectable in liquor, serum, or urine samples using 1H-NMR metabolomics approaches, and (iii) in relation to age- and sex- matched healthy male and female cohorts, next to (iv) whether additional biomarkers can be identified using urine samples relative to healthy cohorts.

In this study, the effects of nusinersen on the 1H-NMR metabolomes of three bodily fluids urine, liquor, or serum were explored in male and female pediatric patients with SMA (Figure S2). Matched healthy male and female cohorts (Table S1) were used to explore urine as the more ubiquitous and accessible matrix for SMA detection and monitoring. Ensemble multivariate statistical approaches (nonparametric and parametric) coupled to machine learning were used to interrogate the metabolomics data in order to establish the significance of differences between SMA before and after treatment and a healthy cohort in order to build the respective sample classification model based on the most important and validated urine biomarkers for the first time.

2. Results and Discussion

2.1. Comparison of 1H-NMR Metabolomes of Urine, Serum, and Liquor Samples before and after Nusinersen Intervention: Positive Effects of Nusinersen Therapy Clearly Preceded or Took Place Devoid of Significant Rearrangements in the Metabolomic Makeup of Serum, Urine, or Liquor

SMA samples of urine, serum, and liquor were collected before and after the 4th application of nusinersen therapy and processed as described below. Medical checkup before the first and after the 4th treatment showed significant improvements at the level of better movement, easier writing and sitting or standing, and feeling more strength next to easier finger extension, which were all measured according to The Children's Hospital of Philadelphia Infant Test of Neuromuscular Disorders (CHOP INTEND) [39], Hammersmith Functional Motor Scale (HFMS), or Expanded Hammersmith Functional Motor Scale (HFMSE) [40] scales or Motor Function Measurement (MFM) [41] tests (Table 1). Twenty patients showed improvement in moving, alongside 8 in head control, 7 in eating, 3 in breathing, 5 in wheelchair control, 10 in tiredness, 4 in hygiene, 5 in mood, 7 in speech, 2 in sleep, and one in excretion after the 4th application of nusinersen.

In contrast, the nonparametric tests (npMANOVA; $p > 0.05$; FDR corrected) used in our analyses showed no significant differences in metabolites in samples collected before and after the 4th application of nusinersen, irrespective of the sample matrix (urine, serum, liquor), sex, or data transformation and normalization procedures. These findings were further corroborated by additional parametric analyses using analyses as implemented within Metaboanalyst (PLSDA, random forest) (Figure 1) and extensive modelling using Just Add Data Bio (JADBIO), all showing no significant difference between the metabolomes collected before and after 4th application of nusinersen. Lastly, the exploration of statistical power also supported the same conclusion. The differences between the metabolic profiles of samples collected before and after the 4th application of nusinersen were so small that at least a two orders of magnitude larger sample size per group (amounting to thousands of samples) would be needed to detect significant differences between the two SMA groups at the level of metabolites and their concentrations obtained by 1H-NMR.

Table 1. Clinical data obtained from spinal muscular atrophy (SMA) patients. Due to the small cohort size, patients' data were anonymized.

Patient	Sex	SMA Type	SMN Copies	Age at 1st App	Weight at 1st App	Height at 1st App	Summary Score at 1st App	Ambulatory at 1st App	Ambulatory at 7th App	Ambulatory Change	Other
1	1	2	3	13.5	57	152	7.5	0	0	No	More strength in the legs while laying on the back
2	1	2	4	5.8	16	103	30	0			More strength
3	2	3	3	4.3	12.4	92	87.5	1	1	No	More strength
4	2	2	3	11.7	42	134	7.5	0	0	No	She stands easier when supported when going to the toilet
5	2	2	3	1.7	10	78	45.3	0	0	No	Stronger
6	1	3	4	7.6	37	136	100	1	1	No	Walks easier
7	2	2	4	12.3	22.5	135	54.7	0	0	No	Writes easier
8	2	2	3	8.6	22.5	126	15	0	0	No	Sits easier and better torso control
9	2	2	4	18.8	50	143	15	0	0	No	Talks easier and moves upper and lower limbs easier
10	2	3	3	13.8	44	160	100	1	1	No	Muscle pain after long walk
11	2	2	3	11.3	13	125	15.6	0	0	No	Stronger voice
12	1	2	3	11.6	33	144	7.8	0	0	No	More easily extends fingers
13	1	2	3	15.4	19	150	6.3	0	0	No	No changes observed
14	2	3	3	5.2	16.5	105	52.5	0	0	No	Easier movement
15	2	2	4	2.3	10.2	82	55	0	0	No	More strength
16	2	1	4	1.3	12.3	74.5	31.3	0	0	No	
17	1	2	3	6.4	13	114	17.2	0	0	No	Better movement
18	2	3	3	18.6	44	154	100	1			
19	2	2	3	1	7.6	75	59.4	0	0	No	Movement better
20	1	2	3	9.8	39.5	146	20.3	0	0	No	More strength by physiotherapists
21	1	3	4	14.2	58	174	97.5	1	1	No	Can walk further
22	2	2	4	5.9	14	110	30	0	0	No	Better movement
23	2	2	3	3.3	24	102	47.5	0	0	No	No difference
24	1	3	3	13	41	152	85	0	0	No	Better movement
25	1			9.1	30.4	138.5					
26	1			1.41	10						
27	1			3.92	45.2	152					
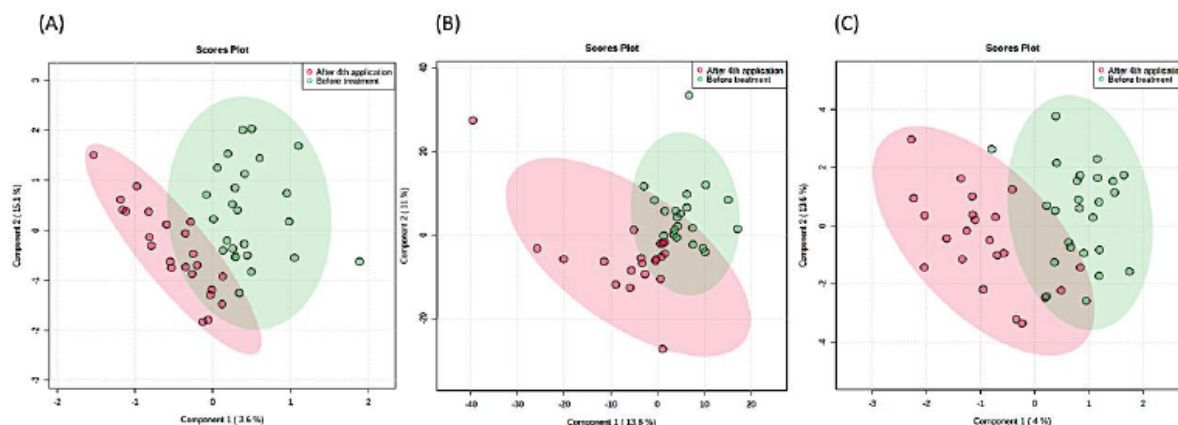
Figure 1. A Partial least square discriminant analysis (PLSDA) of 1H-NMR metabolomes of SMA patients based on (**A**) urine, (**B**) serum, and (**C**) liquor before (red) and after (green) the 4th application of nusinersen. Ellipses designate 95% confidence intervals for each group. The differences are not significant (ensemble statistical approach (npMANOVA, MetaboAnalyst, JADBio)).

Taken together, these results show that the positive effects of nusinersen therapy (Table 1) preceded or took place devoid of significant changes in the actual metabolomics makeup of all three matrices: serum, urine, or liquor (Figure 1). Therefore, the existence of subtle differences could be explored only using approaches utilizing detection thresholds beyond the routine 1H-NMR approaches, such as mass spectrometry methods with higher sensitivity and the possible identification of metabolites present at nM concentrations. However, the relative ease of sample preparation, the ability to quantify metabolite levels, the high level of experimental reproducibility, and the inherently nondestructive nature of NMR spectroscopy support the selection of 1H-NMR as the preferred platform for clinical metabolomic studies [42].

The congruency between the results obtained using various bodily fluids (urine, serum, liquor) showed that the lack of differences in the metabolite profile due to nusinersen therapy were reproducible on all three matrices and hence indeed remarkably small. Our approach also introduced urine as a more straightforward sampling approach to monitor the overall physical status in SMA patients relative to a healthy population compared to more complex serum or liquor samples utilized so far.

2.2. Search for Additional SMA Biomarkers Utilizing Routine 1H-NMR in Urine: Sex Differences, Lower Overall Metabolite Concentrations and Diversity, and Creatinine Content

The collection of urine samples from SMA patients and also from the healthy cohort enabled us to compare the two groups in search of biomarkers for the delineation of SMA and healthy controls next to the detailed targeted 1H-NMR metabolite biomarker search. The ensemble statistical approaches (npMANOVA, MetaboAnalyst; JADBIO) applied to the data matrices (all analyzed metabolites *x* all samples) clearly identified the existence of significant differences between male and female metabolism on one side next to differences between the healthy cohort and SMA on the other.

First, npMANOVA showed the importance of sex (F = 54.9; $p = 0.0001$) and SMA status before or after treatment (F = 20.7; $p = 0.0001$) as significant, while their interaction or SMA status itself (pre vs. post) were not, irrespective of the three different approaches to data preparation and transformation (composition (%), Box–Cox, Log(x + 1)).

Second, the same distinction between the groups of metabolites detectable in urine samples was obtained using PLSDA and randomForest classification as implemented in MetaboAnalyst approach (Figure 2, Figure 3, Figures S7 and S8), clearly showing that the differences between male and female physiology were significant at the level of overall urine data encompassing SMA and healthy cohort ($p < 0.05$). Using the same data sizes

and matched composition, the differences between males and females were still observable in the healthy cohort (Figure 3) ($p < 0.05$), while they were less pronounced in the SMA dataset due to the larger scatter observed in SMA groups of males and females ($p < 0.85$).
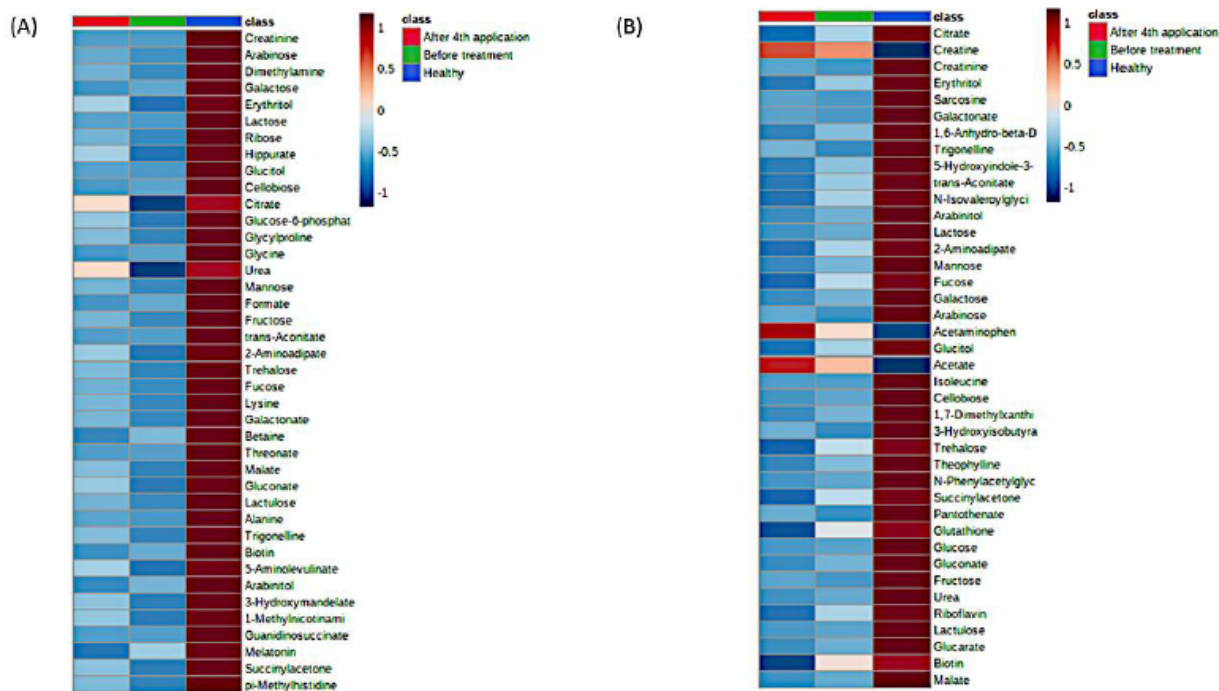
Figure 2. MetaboAnalyst heatmaps of urine 1H-NMR metabolomes representing existing differences between SMA cohort (before and after) and healthy cohort, for males (**A**) and females (**B**). The differences on metabolic level between the before-treatment (green) and after 4th application (red) groups were not significant due to the large variability in SMA samples. The most differentiating metabolites were selected by the PLSDA variable importance in projection (VIP) score, where decreasing metabolites were presented with negative values (blue color) and increasing metabolites were presented with positive values (red color). Individual data are presented in the Supplementary Material (Figures S7 and S8).

We further explored differences in measured concentrations and numbers of different metabolites between SMA and the healthy cohort. The cumulative concentration of metabolites found in the urine of SMA patients corresponded to 58% of those found in the healthy cohort, which was a significant reduction ($p < 0.05$). The higher concentrations of metabolites observed in males relative to females were not significant due to the twice as large scatter observed in SMA groups of males and females ($p < 0.05$). These results show the general picture of a significantly lower overall concentration of urine metabolites in SMA patients relative to the healthy cohort. In addition, the presence/absence pattern of metabolites, i.e., the number of detected metabolites was also significantly higher ($p < 0.001$) in the healthy cohort relative to SMA; however, it was not significantly different between SMA male and female patients, again, due to the large scatter observed in SMA groups of males and females, relative to that observed in matching healthy cohorts of males and females (Table S2). In comparison, the healthy male cohort exhibited a significantly larger number of metabolties ($n = 178.2 \pm 10.8$) in comparison to the healthy female cohort ($n = 154.9 \pm 24$) used in this study, providing future guidance for introducing female participants to such experiments. This is the first report describing the existence of such differences in the context of SMA; hence, the underlying mechanisms for the observed significant differences in metabolic makeup between SMA and healthy next to males and females warrant further analyses in the future.
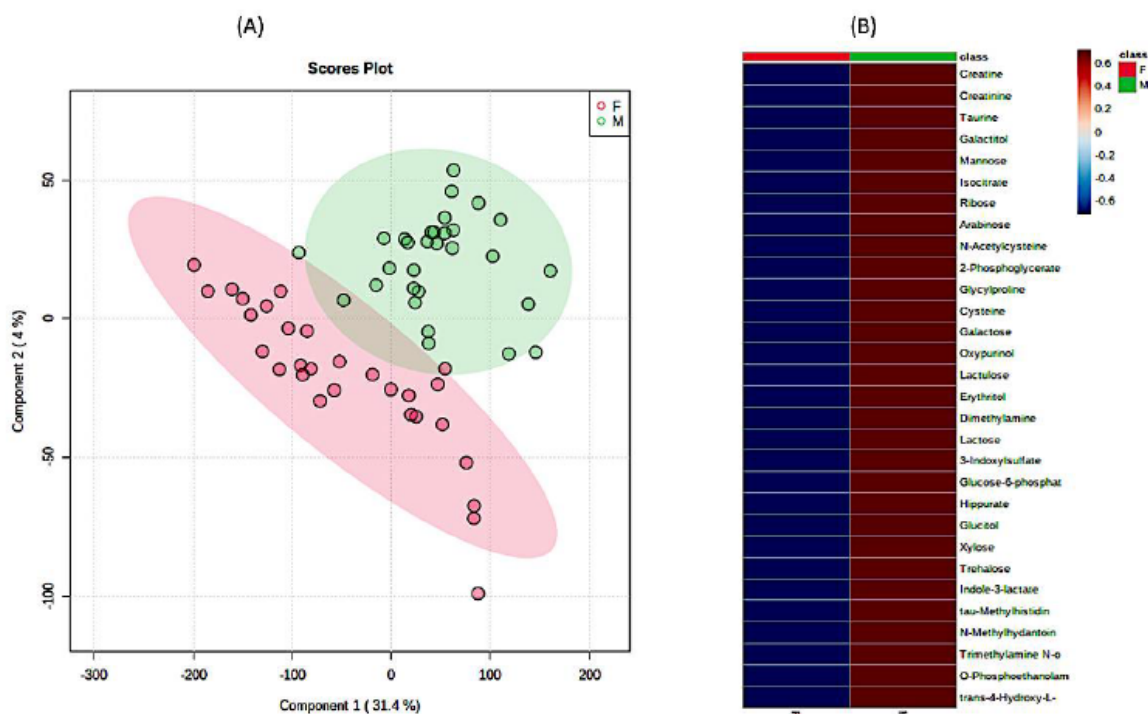
Figure 3. The significant differences in urine 1H-NMR metabolomic profiles between females and males in matched healthy cohorts. (**A**) PLSDA analysis and (**B**) heatmap analysis representing existing differences in the first 30 most important metabolites. (F—females (red), M—males (green)). The most differentiating metabolites were selected by the PLSDA VIP score, where decreasing metabolites had negative values (blue color) and increasing metabolites had positive values (red color).

Some tentative parallels exist with research conducted in the fields of exercise and inactivity. In the context of exercise medicine, analyses of exercise showed that runners experienced a profound systemic shift in blood metabolites related to energy production (especially from the lipid super pathway) and that following the 3-day exercise period, significant 2-fold or higher increases in 75 metabolites persisted for longer than a day [43]. A recent review on exercise metabolic changes revealed that in total, at least 196 metabolites changed their concentration significantly within 24 h after exercise in at least studies, signifying the importance of daily exercise bouts for the maintenance of metabolic diversity and concentration [44]. On the other hand, in the context of body deconditioning due to physical inactivity, similar findings as observed here for the SMA cohort were reported in controlled bed-rest studies. Three-week inactivity resulted in a 30% reduction in the number of statistically significantly connected metabolites, a 2.5 times reduction in the number of interactions, and diminished metabolic diversity within the human body. Conversely, the short-term complete inactivity exhibited also rather similar physiological changes (Figure S3) such as insulin resistance, bone and muscle resorption, constipation, changes in lipid metabolism, and progressively negative interactions with microbiome [29,45,46] that are all listed as part of the SMA makeup as well.

Third, extensive statistical analyses using MetaboAnalyst (Figure S4) and machine learning JADBIO modeling (Figure 4) were adopted to explore the importance of metabolites measured in urine samples. In total, 60,340 models were trained based on the complete urine metabolomics dataset comprising the SMA before, SMA after, and Healthy participants. PLSDA based on all three groups showed that the only significant difference existed again between SMA on one side and the healthy cohort on the other (Figure 4A). The most interpretable model was identified as the ridge logistic regression with the penalty hyperpa-

rameter lambda equal to 0.1 with an area under the curve (AUC) value of 0.958 (Table S3). In addition to AUC, all other threshold metrics were also statistically significantly different from baseline. Data were preprocessed with constant removal and standardized. Features were selected based on Test-Budgeted Statistically Equivalent Signature (SES) algorithm with hyperparameters: maxK = 2, alpha = 0.01, and budget = 3 *nvars. The performance of the model when only creatinine from urine is used is 97.378% (with 95% CI from 94.043% to 100%). Out of all metabolites, creatinine measured in urine samples was the most significant metabolite uniformly separating the healthy group from the SMA group (Figure 4B and Figure S4), whereas no additional metabolic features could be identified to separate SMA before from SMA after, in analogy with the other tests described in this study (Figure S5). We used the trained model on the test part of our data (30% of our total dataset) and achieved the validation performance with an AUC of 0.970. Our work shows that decreasing creatinine concentrations in urine can be used as an additional easy way to measure differentiations between SMA patients and healthy groups (Figure 5 and Figure S9) in analogy with creatinine in serum, which was only recently put forward as a potential biomarker for monitoring the SMA progression of denervation with decreasing levels, reflecting the severity of the disease [38]. However, in response to the open question put forward of whether creatinine responded to molecular therapies [38], the results presented in this study showed that the levels of creatinine did not change significantly in response to the application of nusinersen therapy. Inclusion and analysis of the urinary 1H-NMR metabolomics data following extended nusinersen therapy is projected to further answer this question in the future.
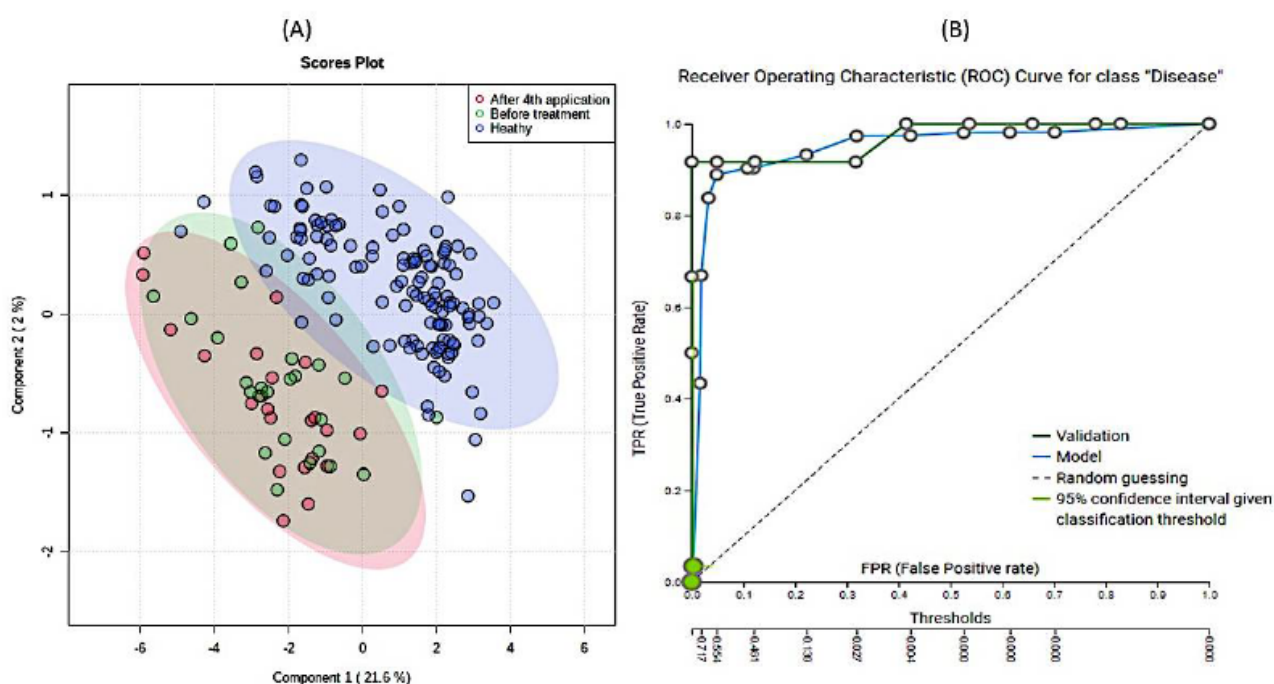
Figure 4. Results of partial least square discriminant analysis (PLSDA) (**A**) and (**B**) Receiver Operating Characteristics curve of modeling of the data. A PLSDA-based ordination of 1H-NMR urine metabolomes: healthy (blue), before (red), and after (green) 4th application of nusinersen. Ellipses designate 95% confidence intervals for each group (**A**). (**B**) Receiver Operating Characteristic (ROC) curve for SMA patients obtained with model. The Just Add Data Bio (JADBIO) model is available as part of the electronic Supplementary Material for the classification of novel 1H-NMR metabolomic data.
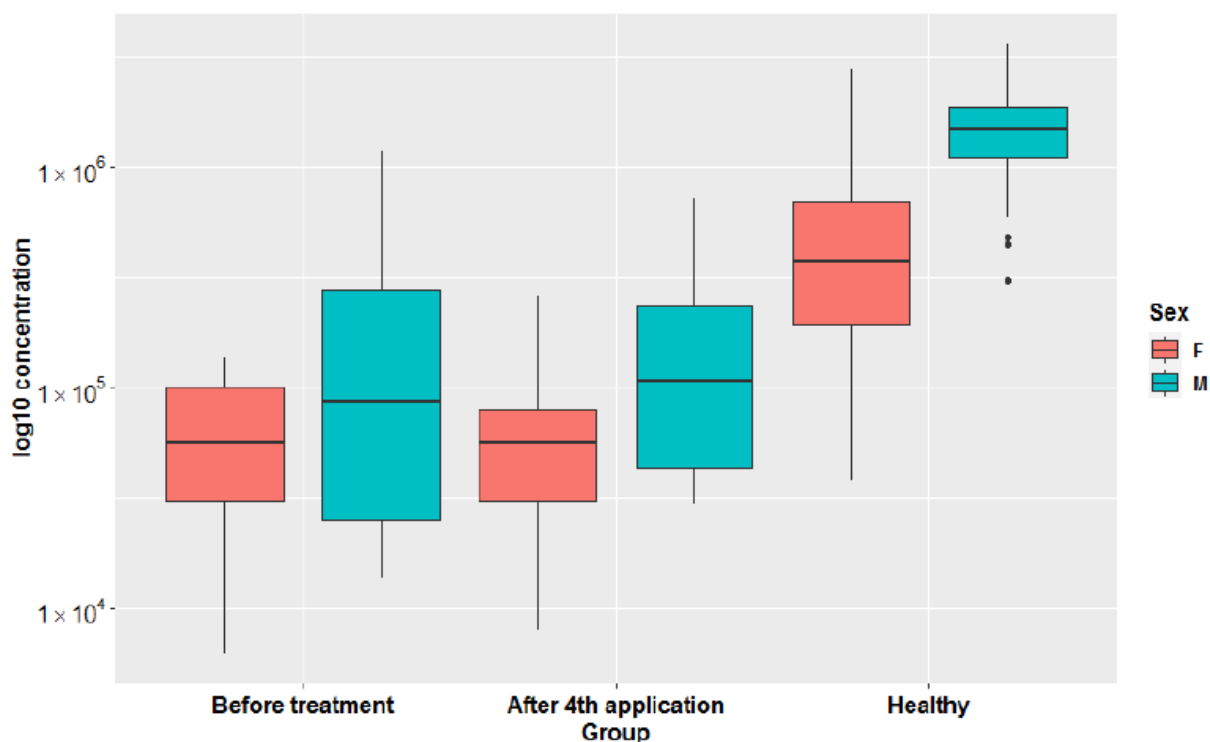
Figure 5. Box-plots representing log10 transformed creatinine concentrations in all three groups (before treatment, after 4th application, healthy) in females and males separately. Original concentrations are presented in Figure S9.

Spinal muscular atrophy patient data suggest that spinal muscular atrophy is a disease affecting neurons, which has diverse consequences for multiple tissues (skeletal muscle, heart, kidney, liver, pancreas, spleen, bone, connective tissues, intestinal tract, and immune systems) [9]. Space-exploration studies adopting bed-rest, e.g., our PlanHab experiment [29,37,45,46], represent a controlled environment for elucidation of the effects physical inactivity as such. Valuable insight was obtained into body deconditioning including insulin resistance, bone and muscle resorption, constipation, mood changes such as depression, negative changes in lipid metabolism, and inflammatory interactions with the microbiome and cardiovascular hypertension (Figure S3) [29,45,46], next to modifications in bacterial metabolism and mucosal turnover in the gut, contributing to the transfer of inflammatory compounds into the bloodstream [47].

These similarities point to a complex interplay and joint effects of physical inactivity and the congenital SMA disease of the patients. We highlight the fact that these are all distinct medical conditions which, despite their different etiologies, share certain characteristics of their metabolic phenotype and clinical characteristics that deserve further exploration.

Decreased creatinine concentrations were observed in urine samples of bed-rest immobilized healthy male participants of the planetary habitat exploration studies [37] (Figure S6). The reintroduction of exercise effectively alleviated and completely reversed the negative effects observed in the PlanHab project [29,37,45,46]. In another study, bed-rest immobilized participants that received vibrational therapy [48] showed numerous benefits relative to controls and could be considered as a step in the physical activation of SMA patients after nusinersen therapy, similar to the prevention and treatment of many diseases, including diabetes and obesity in the future.

51

3. Materials and Methods

3.1. Patients and Healthy Volunteers

This was a single-center study. Biological samples were collected from all patients, clinically diagnosed, and genetically confirmed with SMA. Patients were younger than 19 years and were treated with nusinersen at the Department of Child, Adolescent and Developmental Neurology at the University Children's Hospital Ljubljana, Slovenia.

In March 2017, nusinersen was available through the early access program. Five children received the application before was approved by the European Medicines Agency (EMA). After the approval, The Health Insurance Institute of Slovenia offered the treatment to all children and eligible young adults. Those who decided on treatment were enrolled in the study between March 2017 and by June 2020 and received four consecutive applications of nusinersen.

The study and all experimental protocols were approved by the National Medical Ethics Committee of Republic of Slovenia (0120-305/2018/6 and 0120-305/2018/11). All participants and/or their legal guardians signed the informed consent. The study was registered at ClinicalTrials.gov (accessed on 29 January 2021) under the identifier NCT04587492.

In total (Table S1), 48 samples of urine (25 before treatment and 23 after the 4th application of nusinersen), 41 samples of serum (21 before and 22 after), and 46 samples of liquor (24 before and 22 after) were collected by medical professionals at the University Children's Hospital Ljubljana, Slovenia. The SMA cohort consisted of 15 female patients (age: 8.8 ± 5.5 years; height: 131 ± 21 cm; weight: 26 ± 15 kg) and 10 male patients (age: 9.3 ± 5.1 years; height: 131 ± 20 cm; weight: 28 ± 16 kg). The matching healthy female and male cohort consisted 48 female volunteers (age: 9.4 ± 3.5 years; height: 136 ± 20 cm; weight: 32 ± 13 kg) and 77 healthy male volunteers (age: 9.6 ± 4.3 years; height: 142 ± 24 cm; weight: 36 ± 16 kg). Daily urine samples were collected for three consecutive days to capture daily variation in routines and dietary habits. All samples (SMA and the healthy group) were included in a newly established Slovenian Urine NMR database (manuscript in preparation).

3.2. Evaluation

We analyzed the number of SMN2 copies in all genetically confirmed SMA patients. Evaluation of patients were performed before the initiation of treatment, before the 5th application (after 6 months). Neurological, pulmological, gastroenterological, endocrinological, and psychological exams were performed by pediatric specialists before the start of treatment. The physical capabilities were performed by standardized test by a physiotherapist: The Children's Hospital of Philadelphia Infant Test of Neuromuscular Disorders (CHOP INTEND) [39], Hammersmith Functional Motor Scale (HFMS), or Expanded Hammersmith Functional Motor Scale (HFMSE) [40] scales or Motor Function Measurement (MFM) [41] test. Testing was performed by a physiotherapist trained to use the tests, depending on the age and capabilities of the patient, before the treatment and at all follow-up examinations. The same test was used for all time-points in each particular patient.

3.3. Treatment with Nusinersen

All patients were treated with intrathecal (IT) application of nusinersen. The application was performed under controlled hospital environment. The majority of applications were performed without sedation or under sedation with midazolam or with a combination of midazolam and ketamine. The IT application in very anxious patients was performed under general anesthesia. All patients were recommended to be well hydrated for at least one day before application.

The IT injection of nusinersen was performed on days 0, 14, 30, and 60 in all patients in a standard dose of 5 mL (12 mg/mL). In smaller children, the dose was appropriately reduced after the application; all children were advised to lie prone for 2 h to reduce the risk of post lumbar puncture (LP) symptoms. To reduce the risk of post lumbar puncture

symptoms, all patients were advised to lie prone for at least 2 h after IT application and were monitored for potential side effects for the entire duration of the study.

After the first application, patients were dismissed from the hospital one day after the application and on the same day in the evening for the following applications.

3.4. Sample Collection

For all patients treated with nusinersen who received the medicine intrathecally, samples of urine, blood serum, and cerebrospinal fluid (CSF) were collected before the initiation of treatment, and after the 4th (2 months) application of nusinersen. All obtained samples were frozen at -20 °C for further analysis. Replicate stability analyses were performed as described before [29,37] (Figure S2).

3.5. Metabolome Analysis Using Proton Nuclear Magnetic Resonance (^{1}H-NMR)

All collected samples were centrifuged (1.5 mL) at $10,000 \times g$ for 30 min to remove fine particles. Then, 400 μL of supernatant were mixed with 200 μL ^{1}H-NMR buffer as described before [49] and stored at -25 °C until analysis. Serum samples were filtered using 3 kDa colons (Amicom Ultra 3 kDa (Merck Millipore, Burlington, MA, USA)) for the additional removal of large molecules [50]. Before analysis, samples were thawed at room temperature and transferred into a 5 mm NMR tube. TSP was used as an internal standard for quantification, as described before [49]

An Agilent Technologies DD2 600 MHz spectrometer equipped with a 5 mm HCN Cold probe was used for the acquisition of NMR spectra at 25 °C. The ^{1}H NMR spectra of the samples were recorded with a spectral width of 9.0 kHz, relaxation delay of 2.0 s, 32 scans, and 32 K data points. A double-pulsed field gradient spin echo (DPFGSE) pulse sequence was used for water suppression. Total correlated spectrum (TOCSY) was measured with ^{1}H spectral widths of 7.0 kHz, 4096 complex points, a relaxation delay of 1.5 s, 32 transients, and 144 time increments. An exponential and cosine-squared function were used for apodization. Zeros were filled before Fourier transform. VNMRJ (Agilent/Varian) software was used for processing urine and liquor NMR spectra.

Serum spectra were acquired on the same spectrometer equipped with a 24-sample automation system processed with Topspin v. 4.0.9 software (Bruker, Billerica, MA, USA).

Metabolites were identified with the support of the Chenomx Compound Library extended to the Human Metabolome Database [51,52], giving access to chemical shift profiles of 674 compounds used in analyses. The number of database-derived chemical shift profiles of metabolites used in analyses was further decreased by the procedures described below.

3.6. Bioinformatics and Statistical Analysis

The resulting spectra were consequently analyzed using targeted quantitative metabolomics using Chenomx NMR Suite version 8.6 (Chenomx, Inc., Edmonton, AB, Canada). For the latter, all spectra were randomly ordered for spectral fitting using the ChenomX profiler. An ensemble approach to data analysis was utilized, employing three different approaches to asymmetric sparse matrix data analysis, establishing significant differences between tested groups as follows: nonparametric MANOVA (PERMANOVA) [53], MetaboAnalyst [54,55], and JADBIO [56].

First, for npMANOVA, each compound concentration obtained was analyzed as described before in three different ways [29,37]: (i) by dividing the measured concentration by the concentration of all metabolites in that sample; (ii) Box–Cox; or (iii) $\log(x + 1)$ transformed. The significance of metabolic differences between various groups of samples was tested using ANOSIM, NP-MANOVA, and expressed as an overlap in non-metric multidimensional scaling (nm-MDS) trait space (using Euclidean distance measures). The stress function was used to select the dimensionality reduction, whereas Shepard's plots were used to describe the correspondence between the target and obtained ranks [57].

Benjamini–Hochberg significance correction for multiple comparisons was used as described before [45,46].

Second, for MetaboAnalyst, log or cube root transformation in connection to Mean or Pareto scaling was utilized as implemented in MetaboAnalyst [54,55] followed by supervised classification using partial least squares discriminant analysis (PLSDA) method, random forest (RF), and pathway enrichment analysis. PLSDA results were cross-validated with a caret package implemented in MeatboAnalyst. The most important metabolites identified by PLSDA were determined according to variable importance in projection (VIP). The randomForest package implemented in MetaboAnalyst was used for supervised classification between different groups of interest. The most important features defined by RF were ranked by a mean decrease in classification accuracy. Hierarchical clustering was performed according to the VIP scores to obtain a heat map representing differences in metabolic profiles between samples and groups. Euclidean distance, Pearson's correlation, and Spearman's correlation were used as similarity measures and Ward's linkage was used as a clustering algorithm. Statistical power for the identification of significant differences before and after treatment was also calculated using MetaboAnalyst Statistical Power module. MetaboAnalyst and ggplot2 were used for graph generating.

KEGG human pathway libraries were used for metabolic pathway and enrichment analysis. For topological analysis, the globaltest analysis method and relative betweenness centrality were used. Significant pathways were determined using the raw p-value, Holm–Bonferroni p-adjusted value, and adjusted p-value using the False Discovery Rate. The impact of pathways was calculated using pathway topology analysis.

Metabolite Set Enrichment (MSEA) was used to identify biologically significant patterns between quantitative metabolome data from different groups. HMDB compound names were used to link to the KEGG database. Enrichment analysis was performed using the globaltest package implemented in MetaboAnalyst. The enrichment ratio was calculated by dividing observed hits and expected hits.

Last, Just Add Data Bio (JADBIO), a web-based auto machine learning platform for analyzing potential biomarkers [56], was used. The JADBIO platform was designed for predictive modeling and to provide high-quality predictive models for diagnostics using state-of-the-art statistical and machine learning methods. Personal analytic biases and methodological statistical errors were eliminated from the analysis by the autonomous exploration of various settings in modeling steps producing more convincing discovered features to discriminate between SMA and the healthy group. JADBIO 1.1.164 with extensive tuning effort and 6 CPU was used to model various dataset selections next to the overall 336 metabolites observed in urine samples in all groups (healthy versus SMA group) by splitting the total urine metabolite data into a training set and a test set in a 70:30 ratio. The training set was used for model training and the test set was used for model evaluation.

The resulting model can be obtained as part of Supplementary Material (ESM2) and run with java executor for the classification of novel urine samples based on 1H-NMR metabolomes in further exploration.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/metabo11040206/s1, Figure S1: Healthy individuals have active protein SMN1, Figure S2: Sample collection and data analysis scheme, Figure S3: Heatmap plot showing the relationship between parameters describing human physiology, psychology, and intestinal environment that differed significantly at the end of the PlanHab experiment, Figure S4: Feature importance plots (MetaboAnalyst PLSDA analysis), Figure S5: Progressive feature inclusion plot, Figure S6: Heatmap of the 50 most important urine metabolites of the PlanHab metabolomes. Figure S7: Heatmap representing 40 the most discriminative metabolic features for female SMA patients before and after treatment and matched healthy cohort, Figure S8: Heatmap representing the 40 most discriminative metabolic features for male SMA patients before and after treatment and matched healthy cohort, Figure S9: Box-plots representing creatinine concentrations in all three groups (Before treatment, After 4th application, Healthy) in females and males separately. Table S1: Descriptive statistics of

analyzed cohorts, Table S2: The number of detected metabolites in healthy cohort relative to SMA. Table S3: Performance metrics, Instructions for running a model on a local machine, model.zip (model, java executor and test data included).

References

1. Melki, J. Advances in Spinal Muscular Atrophy Research. In *Spinal Muscular Atrophy–Disease Mechanisms and Therapy*; Sumner, C.J., Paushkin, S., Ko, C., Eds.; Academic Press: London, UK, 2017; pp. xxiii–xxiv.
2. Sugarman, E.A.; Nagan, N.; Zhu, H.; Akmaev, V.R.; Zhou, Z.; Rohlfs, E.M.; Flynn, K.; Hendrickson, B.C.; Scholl, T.; Sirko-Osadsa, D.A.; et al. Pan-ethnic carrier screening and prenatal diagnosis for spinal muscular atrophy: Clinical laboratory analysis of >72,400 specimens. *Eur. J. Hum. Genet.* **2012**, *20*, 27–32. [CrossRef] [PubMed]
3. Gilliam, T.C.; Brzustowicz, L.M.; Castilla, L.H.; Lehner, T.; Penchaszadeh, G.K.; Daniels, R.J.; Byth, B.C.; Knowles, J.; Hislop, J.E.; Shapira, Y. Genetic homogeneity between acute and chronic forms of spinal muscular atrophy. *Nature* **1990**, *345*. [CrossRef] [PubMed]
4. Melki, J.; Abdelhak, S.; Sheth, P.; Bachelot, M.F.; Burlet, P.; Marcadet, A.; Aicardi, J.; Barois, A.; Carriere, J.P.; Fardeau, M. Gene for chronic proximal spinal muscular atrophies maps to chromosome 5q. *Nature* **1990**, *344*. [CrossRef]
5. Lorson, C.L.; Androphy, E.J. An exonic enhancer is required for inclusion of an essential exon in the SMA-determining gene SMN. *Hum. Mol. Genet.* **2000**, *9*. [CrossRef] [PubMed]
6. Lefebvre, S.; Bürglen, L.; Reboullet, S.; Clermont, O.; Burlet, P.; Viollet, L.; Benichou, B.; Cruaud, C.; Millasseau, P.; Zeviani, M. Identification and characterization of a spinal muscular atrophy-determining gene. *Cell* **1995**, *80*. [CrossRef]
7. Smeriglio, P.; Langard, P.; Querin, G.; Biferi, M.G. The Identification of novel biomarkers is required to improve adult SMA patient stratification, diagnosis and treatment. *J. Pers. Med.* **2020**, *10*, 75. [CrossRef] [PubMed]
8. Lunn, M.R.; Wang, C.H. Spinal muscular atrophy. *Lancet* **2008**, *371*. [CrossRef]
9. Yeo, C.J.J.; Darras, B.T. Overturning the paradigm of spinal muscular atrophy as just a motor neuron disease. *Pediatr. Neurol.* **2020**, *109*, 12–19. [CrossRef]
10. Chiriboga, C.A. Nusinersen for the treatment of spinal muscular atrophy. *Expert Rev. Neurother.* **2017**, *17*. [CrossRef]
11. Al-Zaidy, S.A.; Kolb, S.J.; Lowes, L.; Alfano, L.N.; Shell, R.; Church, K.R.; Nagendran, S.; Sproule, D.M.; Feltner, D.E.; Wells, C.; et al. AVXS-101 (onasemnogene abeparvovec) for SMA1: Comparative study with a prospective natural history cohort. *J. Neuromuscul. Dis.* **2019**, *6*. [CrossRef]

12. Poirier, A.; Weetall, M.; Heinig, K.; Bucheli, F.; Schoenlein, K.; Alsenz, J.; Bassett, S.; Ullah, M.; Senn, C.; Ratni, H.; et al. Risdiplam distributes and increases SMN protein in both the central nervous system and peripheral organs. *Pharmacol. Res. Perspect.* **2018**, *6*. [CrossRef]

13. Ramdas, S.; Servais, L. New treatments in spinal muscular atrophy: An overview of currently available data. *Expert Opin. Pharm.* **2020**, *21*. [CrossRef] [PubMed]

14. Corey, D.R. Nusinersen, an antisense oligonucleotide drug for spinal muscular atrophy. *Nat. Neurosci.* **2017**, *20*. [CrossRef] [PubMed]

15. Singh, N.K.; Singh, N.N.; Androphy, E.J.; Singh, R.N. Splicing of a critical exon of human survival motor neuron is regulated by a unique silencer element located in the last intron. *Mol. Cell. Biol.* **2006**, *26*. [CrossRef]

16. Hua, Y.; Sahashi, K.; Hung, G.; Rigo, F.; Passini, M.A.; Bennett, C.F.; Krainer, A.R. Antisense correction of SMN2 splicing in the CNS rescues necrosis in a type III SMA mouse model. *Genes Dev.* **2010**, *24*. [CrossRef]

17. Rigo, F.; Hua, Y.; Krainer, A.R.; Bennett, C.F. Antisense-based therapy for the treatment of spinal muscular atrophy. *J. Cell Biol.* **2012**, *199*. [CrossRef]

18. Finkel, R.S.; Chiriboga, C.A.; Vajsar, J.; Day, J.W.; Montes, J.; De Vivo, D.C.; Yamashita, M.; Rigo, F.; Hung, G.; Schneider, E.; et al. Treatment of infantile-onset spinal muscular atrophy with nusinersen: A phase 2, open-label, dose-escalation study. *Lancet* **2016**, *388*. [CrossRef]

19. Chiriboga, C.A.; Swoboda, K.J.; Darras, B.T.; Iannaccone, S.T.; Montes, J.; De Vivo, D.C.; Norris, D.A.; Bennett, C.F.; Bishop, K.M. Results from a phase 1 study of nusinersen (ISIS-SMN(Rx)) in children with spinal muscular atrophy. *Neurology* **2016**, *86*. [CrossRef] [PubMed]

20. De Vivo, D.C.; Bertini, E.; Swoboda, K.J.; Hwu, W.L.; Crawford, T.O.; Finkel, R.S.; Kirschner, J.; Kuntz, N.L.; Parsons, J.A.; Ryan, M.M.; et al. Nusinersen initiated in infants during the presymptomatic stage of spinal muscular atrophy: Interim efficacy and safety results from the Phase 2 NURTURE study. *Neuromuscul. Disord.* **2019**, *29*. [CrossRef] [PubMed]

21. Osredkar, D.; Jilková, M.; Butenko, T.; Loboda, T.; Golli, T.; Fuchsová, P.; Rohlenová, M.; Haberlova, J. Children and young adults with spinal muscular atrophy treated with nusinersen. *Eur. J. Paediatr. Neurol.* **2021**, *30*, 1–8. [CrossRef]

22. Kruse, T.; Heller, R.; Wirth, B.; Glöggler, J.; Wurster, C.D.; Ludolph, A.C.; Braumann, B. Maximum bite force in patients with spinal muscular atrophy during the first year of nusinersen therapy—A pilot study. *Acta Myol.* **2020**, *39*. [CrossRef]

23. De Wel, B.; Goosens, V.; Sobota, A.; Van Camp, E.; Geukens, E.; Van Kerschaver, G.; Jagut, M.; Claes, K.; Claeys, K.G. Nusinersen treatment significantly improves hand grip strength, hand motor function and MRC sum scores in adult patients with spinal muscular atrophy types 3 and 4. *J. Neurol.* **2020**. [CrossRef] [PubMed]

24. Stolte, B.; Totzeck, A.; Kizina, K.; Bolz, S.; Pietruck, L.; Mönninghoff, C.; Guberina, N.; Oldenburg, D.; Forsting, M.; Kleinschnitz, C.; et al. Feasibility and safety of intrathecal treatment with nusinersen in adult patients with spinal muscular atrophy. *Ther. Adv. Neurol. Disord.* **2018**, *11*. [CrossRef]

25. Michelson, D.; Ciafaloni, E.; Ashwal, S.; Lewis, E.; Narayanaswami, P.; Oskoui, M.; Armstrong, M.J. Evidence in focus: Nusinersen use in spinal muscular atrophy: Report of the guideline development, dissemination, and implementation subcommittee of the American academy of neurology. *Neurology* **2018**, *91*. [CrossRef]

26. Peisl, B.Y.L.; Schymanski, E.L.; Wilmes, P. Dark matter in host—Microbiome metabolomics: Tackling the unknowns—A review. *Anal. Chim. Acta* **2018**, *1037*, 13–27. [CrossRef] [PubMed]

27. Smolinska, A.; Blanchet, L.; Buydens, L.M.; Wijmenga, S.S. NMR and pattern recognition methods in metabolomics: From data acquisition to biomarker discovery: A review. *Anal. Chim. Acta* **2012**, *750*, 82–97. [CrossRef]

28. Murovec, B.; Makuc, D.; Kolbl Repinc, S.; Prevorsek, Z.; Zavec, D.; Sket, R.; Pecnik, K.; Plavec, J.; Stres, B. ^{1}H NMR metabolomics of microbial metabolites in the four MW agricultural biogas plant reactors: A case study of inhibition mirroring the acute rumen acidosis symptoms. *J. Environ. Manag.* **2018**, *222*, 428–435. [CrossRef]

29. Sket, R.; Debevec, T.; Kublik, S.; Schloter, M.; Schoeller, A.; Murovec, B.; Mikus, K.V.; Makuc, D.; Pecnik, K.; Plavec, J.; et al. Intestinal metagenomes and metabolomes in healthy young males: Inactivity and hypoxia generated negative physiological symptoms precede microbial dysbiosis. *Front. Physiol.* **2018**, *9*. [CrossRef]

30. Lindon, J.C.; Holmes, E.; Nicholson, J.K. So what's the deal with metabonomics? *Anal. Chem.* **2003**, *75*, 384a–391a. [CrossRef]

31. Bro, R.; Kamstrup-Nielsen, M.H.; Engelsen, S.B.; Savorani, F.; Rasmussen, M.A.; Hansen, L.; Olsen, A.; Tjonneland, A.; Dragsted, L.O. Forecasting individual breast cancer risk using plasma metabolomics and biocontours. *Metabolomics* **2015**, *11*, 1376–1380. [CrossRef]

32. Dumas, M.E.; Kinross, J.; Nicholson, J.K. Metabolic phenotyping and systems biology approaches to understanding metabolic syndrome and fatty liver disease. *Gastroenterology* **2014**, *146*, 46–62. [CrossRef]

33. Marin, L.; Miguelez, E.M.; Villar, C.J.; Lombo, F. Bioavailability of dietary polyphenols and gut microbiota metabolism: Antimicrobial properties. *Biomed. Res. Int.* **2015**, *2015*, 905215. [CrossRef]

34. Bjerrum, J.T.; Wang, Y.; Hao, F.; Coskun, M.; Ludwig, C.; Gunther, U.; Nielsen, O.H. Metabonomics of human fecal extracts characterize ulcerative colitis, Crohn's disease and healthy individuals. *Metabolomics* **2015**, *11*, 122–133. [CrossRef] [PubMed]

35. Zhang, C.; Yin, A.; Li, H.; Wang, R.; Wu, G.; Shen, J.; Zhang, M.; Wang, L.; Hou, Y.; Ouyang, H.; et al. Dietary modulation of gut microbiota contributes to alleviation of both genetic and simple obesity in children. *EBioMedicine* **2015**, *2*, 968–984. [CrossRef]

36. Holmes, E.; Loo, R.L.; Stamler, J.; Bictash, M.; Yap, I.K.; Chan, Q.; Ebbels, T.; De Iorio, M.; Brown, I.J.; Veselkov, K.A.; et al. Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature* **2008**, *453*, 396–400. [CrossRef]

37. Šket, R.; Deutsch, L.; Prevoršek, Z.; Mekjavić, I.B.; Plavec, J.; Rittweger, J.; Debevec, T.; Eiken, O.; Stres, B. Systems view of deconditioning during spaceflight simulation in the planhab project: The departure of urine 1 H-NMR metabolomes from healthy state in young males subjected to bedrest inactivity and hypoxia. *Front. Physiol.* **2020**, *11*. [CrossRef]

38. Alves, C.R.R.; Zhang, R.; Johnstone, A.J.; Garner, R.; Nwe, P.H.; Siranosian, J.J.; Swoboda, K.J. Serum creatinine is a biomarker of progressive denervation in spinal muscular atrophy. *Neurology* **2020**, *94*. [CrossRef] [PubMed]

39. Glanzman, A.M.; Mazzone, E.; Main, M.; Pelliccioni, M.; Wood, J.; Swoboda, K.J.; Scott, C.; Pane, M.; Messina, S.; Bertini, E.; et al. The children's hospital of philadelphia infant test of neuromuscular disorders (CHOP INTEND): Test development and reliability. *Neuromuscul. Disord.* **2010**, *20*. [CrossRef]

40. Pera, M.C.; Coratti, G.; Forcina, N.; Mazzone, E.S.; Scoto, M.; Montes, J.; Pasternak, A.; Mayhew, A.; Messina, S.; Sframeli, M.; et al. Content validity and clinical meaningfulness of the HFMSE in spinal muscular atrophy. *BMC Neurol.* **2017**, *17*. [CrossRef] [PubMed]

41. Bérard, C.; Payan, C.; Hodgkinson, I.; Fermanian, J. A motor function measure for neuromuscular diseases. Construction and validation study. *Neuromuscul. Disord.* **2005**, *15*. [CrossRef]

42. Emwas, A.H.; Roy, R.; McKay, R.T.; Tenori, L.; Saccenti, E.; Gowda, G.A.N.; Raftery, D.; Alahmari, F.; Jaremko, L.; Jaremko, M.; et al. NMR spectroscopy for metabolomics research. *Metabolites* **2019**, *9*, 123. [CrossRef]

43. Nieman, D.C.; Shanely, R.A.; Gillitt, N.D.; Pappan, K.L.; Lila, M.A. Serum metabolic signatures induced by a three-day intensified exercise period persist after 14 h of recovery in runners. *J. Proteome Res.* **2013**, *12*. [CrossRef] [PubMed]

44. Schranner, D.; Kastenmüller, G.; Schönfelder, M.; Römisch-Margl, W.; Wackerhage, H. Metabolite concentration changes in humans after a bout of exercise: A systematic review of exercise metabolomics studies. *Sports Med. Open* **2020**, *6*. [CrossRef] [PubMed]

45. Sket, R.; Treichel, N.; Debevec, T.; Eiken, O.; Mekjavic, I.; Schloter, M.; Vital, M.; Chandler, J.; Tiedje, J.M.; Murovec, B.; et al. Hypoxia and inactivity related physiological changes (constipation, inflammation) are not reflected at the level of gut metabolites and butyrate producing microbial community: The PlanHab study. *Front. Physiol.* **2017**, *8*. [CrossRef]

46. Sket, R.; Treichel, N.; Kublik, S.; Debevec, T.; Eiken, O.; Mekjavic, I.; Schloter, M.; Vital, M.; Chandler, J.; Tiedje, J.M.; et al. Hypoxia and inactivity related physiological changes precede or take place in absence of significant rearrangements in bacterial community structure: The PlanHab randomized trial pilot study. *PLoS ONE* **2017**, *12*. [CrossRef]

47. Roager, H.M.; Hansen, L.B.S.; Bahl, M.I.; Frandsen, H.L.; Carvalho, V.; Gobel, R.J.; Dalgaard, M.D.; Plichta, D.R.; Sparholt, M.H.; Vestergaard, H.; et al. Colonic transit time is related to bacterial metabolism and mucosal turnover in the gut. *Nat. Microbiol.* **2016**, *1*. [CrossRef]

48. Hoff, P.; Belavý, D.L.; Huscher, D.; Lang, A.; Hahne, M.; Kuhlmey, A.K.; Maschmeyer, P.; Armbrecht, G.; Fitzner, R.; Perschel, F.H.; et al. Effects of 60-day bed rest with and without exercise on cellular and humoral immunological parameters. *Cell. Mol. Immunol.* **2015**, *12*. [CrossRef]

49. Beckonert, O.; Keun, H.C.; Ebbels, T.M.; Bundy, J.; Holmes, E.; Lindon, J.C.; Nicholson, J.K. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat. Protoc.* **2007**, *2*, 2692–2703. [CrossRef]

50. Ravanbakhsh, S.; Liu, P.; Bjorndahl, T.C.; Mandal, R.; Grant, J.R.; Wilson, M.; Eisner, R.; Sinelnikov, I.; Hu, X.; Luchinat, C.; et al. Accurate, fully-automated NMR spectral profiling for metabolomics. *PLoS ONE* **2015**, *10*. [CrossRef]

51. Markley, J.L.; Brüschweiler, R.; Edison, A.S.; Eghbalnia, H.R.; Powers, R.; Raftery, D.; Wishart, D.S. The future of NMR-based metabolomics. *Curr. Opin. Biotechnol.* **2017**, *43*. [CrossRef]

52. Wishart, D.S.; Feunang, Y.D.; Marcu, A.; Guo, A.C.; Liang, K.; Vazquez-Fresno, R.; Sajed, T.; Johnson, D.; Li, C.; Karu, N.; et al. HMDB 4.0: The human metabolome database for 2018. *Nucleic Acids Res.* **2018**, *46*, D608–D617. [CrossRef]

53. Legendre, P.; Legendre, L.F.J. *Numerical Ecology*, 3rd ed.; Elsevier: Amsterdam, The Netherlands, 2012; Volume 24, p. 1006.

54. Chong, J.; Soufan, O.; Li, C.; Caraus, I.; Li, S.; Bourque, G.; Wishart, D.S.; Xia, J. MetaboAnalyst 4.0: Towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res.* **2018**, *46*. [CrossRef]

55. Chong, J.; Wishart, D.S.; Xia, J. Using metaboanalyst 4.0 for comprehensive and integrative metabolomics data analysis. *Curr. Protoc. Bioinform.* **2019**, *68*, e86. [CrossRef] [PubMed]

56. Tsamardinos, I.; Charonyktakis, P.; Lakiotaki, K.; Borboudakis, G.; Zenklusen, J.C.; Juhl, H.; Chatzaki, E.; Lagani, V. Just add data: Automated predictive modeling and biosignature discovery. *bioRxiv* **2020**. [CrossRef]

57. Murovec, B.; Kolbl, S.; Stres, B. Methane yield database: Online infrastructure and bioresource for methane yield data and related metadata. *Bioresour. Technol.* **2015**, *189*, 217–223. [CrossRef] [PubMed]

2.1.4 The importance of objective stool classification in fecal ^{1}H-NMR metabolomics: exponential increase in stool crosslinking is mirrored in systemic inflammation and associated to fecal acetate and methionine

Deutsch L., Stres B. 2021. The importance of objective stool classification in fecal ^{1}H-NMR metabolomics: exponential increase in stool crosslinking is mirrored in systemic inflammation and associated to fecal acetate and methionine. Metabolites, 11, 3: 172, doi. https://doi.org/10.3390/metabo11030172, 16 p.

Abstract

Past studies strongly connected stool consistency-as measured by Bristol Stool Scale (BSS)-with microbial gene richness and intestinal inflammation, colonic transit time and metabolome characteristics that are of clinical relevance in numerous gastro intestinal conditions. While retention time, defecation rate, BSS but not water activity have been shown to account for BSS-associated inflammatory effects, the potential correlation with the strength of a gel in the context of intestinal forces, abrasion, mucus imprinting, fecal pore clogging remains unexplored as a shaping factor for intestinal inflammation and has yet to be determined. Our study introduced a minimal pressure approach (MP) by probe indentation as measure of stool material crosslinking in fecal samples. Results reported here were obtained from 170 samples collected in two independent projects, including males and females, covering a wide span of moisture contents and BSS. MP values increased exponentially with increasing consistency (i.e., lower BSS) and enabled stratification of samples exhibiting mixed BSS classes. A trade-off between lowest MP and highest dry matter content delineated the span of intermediate healthy density of gel crosslinks. The crossectional transects identified fecal surface layers with exceptionally high MP and of <5 mm thickness followed by internal structures with an order of magnitude lower MP, characteristic of healthy stool consistency. The MP and BSS values reported in this study were coupled to reanalysis of the PlanHab data and fecal ^{1}H-NMR metabolomes reported before. The exponential association between stool consistency and MP determined in this study was mirrored in the elevated intestinal and also systemic inflammation and other detrimental physiological deconditioning effects observed in the PlanHab participants reported before. The MP approach described in this study can be used to better understand fecal hardness and its relationships to human health as it provides a simple, fine scale and objective stool classification approach for the characterization of the exact sampling locations in future microbiome and metabolome studies.

For my personal contributions as a doctoral student and author of this thesis, please refer to Table 2 (page 142).

metabolites

MDPI

Article

The Importance of Objective Stool Classification in Fecal 1H-NMR Metabolomics: Exponential Increase in Stool Crosslinking Is Mirrored in Systemic Inflammation and Associated to Fecal Acetate and Methionine

Leon Deutsch [1] and Blaz Stres [1,2,3,4,*]

1. Biotechnical Faculty, University of Ljubljana, Jamnikarjeva 101, SI-1000 Ljubljana, Slovenia; leon.deutsch@bf.uni-lj.si
2. Faculty of Civil and Geodetic Engineering, University of Ljubljana, Jamova 2, SI-1000 Ljubljana, Slovenia
3. Department of Automation, Jožef Stefan Institute, Biocybernetics and Robotics, Jamova 39, SI-1000 Ljubljana, Slovenia
4. Department of Microbiology, University of Innsbruck, Technikerstrasse 25d, A-6020 Innsbruck, Austria
* Correspondence: blaz.stres@bf.uni-lj.si; Tel.: +386-41-567-633

check for updates

Abstract: Past studies strongly connected stool consistency—as measured by Bristol Stool Scale (BSS)—with microbial gene richness and intestinal inflammation, colonic transit time and metabolome characteristics that are of clinical relevance in numerous gastro intestinal conditions. While retention time, defecation rate, BSS but not water activity have been shown to account for BSS-associated inflammatory effects, the potential correlation with the strength of a gel in the context of intestinal forces, abrasion, mucus imprinting, fecal pore clogging remains unexplored as a shaping factor for intestinal inflammation and has yet to be determined. Our study introduced a minimal pressure approach (MP) by probe indentation as measure of stool material crosslinking in fecal samples. Results reported here were obtained from 170 samples collected in two independent projects, including males and females, covering a wide span of moisture contents and BSS. MP values increased exponentially with increasing consistency (i.e., lower BSS) and enabled stratification of samples exhibiting mixed BSS classes. A trade-off between lowest MP and highest dry matter content delineated the span of intermediate healthy density of gel crosslinks. The crossectional transects identified fecal surface layers with exceptionally high MP and of <5 mm thickness followed by internal structures with an order of magnitude lower MP, characteristic of healthy stool consistency. The MP and BSS values reported in this study were coupled to reanalysis of the PlanHab data and fecal 1H-NMR metabolomes reported before. The exponential association between stool consistency and MP determined in this study was mirrored in the elevated intestinal and also systemic inflammation and other detrimental physiological deconditioning effects observed in the PlanHab participants reported before. The MP approach described in this study can be used to better understand fecal hardness and its relationships to human health as it provides a simple, fine scale and objective stool classification approach for the characterization of the exact sampling locations in future microbiome and metabolome studies.

Keywords: feces; stool; classification; Bristol stool scale; BSS; JADBIO; gel network; intestinal abrasion

1. Introduction

The contractile patterns of small intestine propagate toward the colon and are caused by the enteric nervous system (the interstitial cells of Cajal) that generate slow waves of smooth muscle contraction and contribute to the transit rates along the intestine [1–3]. The number and the length of peristaltic waves, i.e., circular constrictions propagating aborally, determine chime transport that decreases along the gut in the same proportion as the volume of luminal content declines by absorption of nutrients and water [4,5]. Repeated contractions are essential for maintenance of a steady-state bacterial population as mixing

of chime helps to overcome flow, and controlled contractions by the colon strongly influence microbiota density and composition. Consequently, flow and mixing play a major role in shaping the microbial metabolic reactions and interactions with the host [6]. In addition, with increasing filling of distal intestinal segments (i.e., constipation), the motility of the proximal intestine is inhibited, the number of peristaltic waves decreases while the number of stationary segmenting contractions increases [1–3]. In addition to peristaltic waves, stationary or segmenting contractions are pushing, mixing and separating chime into segments. These contractions are isolated at a single site without spatio-temporal pattern and are responsible for inflicting intestinal abrasions during prolonged and increasing constipation. The derived intestinal abrasions are highly amenable for further microbial colonization at locations with a thin (e.g., small intestine) or modified mucus layer (e.g., abrasions, imprinting, reduced mucus thickness, increased porosity and modified mucus glycosylation pattern; [7]). Local pH values in the lumen differentially affect the growth of different bacteria and drive changes in microbiota composition. The key factors influencing the delicate regulation of colonic pH are epithelial water absorption, nutrient inflow, and luminal buffering capacity [5].

Stool consistency was strongly associated with the gut microbiota richness and composition, enterotypes, increased local inflammation, lipopolysaccharides and bacterial growth rates [8,9]. Stool consistency is generally assessed using the widely adopted Bristol Stool Scale (BSS) [10,11]. In addition, long colonic transit time corresponded to lower BSS score, higher microbial richness and a shift in colonic metabolism from carbohydrate fermentation towards protein catabolism [12], interlinking thus the status of systemic levels of metabolites with microbiome and BSS. This central role of BSS as a major important confounding factor affecting microbial physiology, significant microbial community rearrangements and interactions with host showcases the importance of the local conditions for transformation of microbial physiology to a causative phenotype [3,12–15]. Strong connection of stool consistency, i.e., BSS, with microbial gene richness and intestinal inflammation were reported recently for 61 severely obese subjects [16]. Other recent communications [8,9,17,18] focused on BSS stool consistency in (healthy) male and female subjects due to its established correlation with colonic transit time, inflammation, microbiome and metabolome characteristics that are of clinical relevance in numerous gastrointestinal conditions. We recently reported on strong association of BSS, retention time and defecation frequency with increased systemic inflammation, insulin resistance, cardiovascular deconditioning, depression, increased levels of the genus *Bacteroides* and their virulence genes in healthy males after prolonged physical inactivity and hypoxia (PlanHab project) [3,13,14].

The BSS was also shown to demonstrate substantial validity and reliability in general, although difficulties arose around clinical decision points (BSS Types 2/3, 4/5) [19]. Although BSS can be easily evaluated by participants themselves a substantial intra- and inter-rater variance were observed [20,21] due to sensation of straining during defecation. Consequently, self-rating was susceptible to subjective bias despite its effectiveness in clinical use. BSS offered a reliable surrogate measure of stool consistency only when rated by well-trained expert [22]. The non-uniform or mixed makeup of fecal samples, exhibiting two or three or more different BSS classes was identified as another large source of variability in BSS values, preventing the exact classification of mixed form samples and allowing classification errors when ascribing one BSS class to such mixed form samples. These observations point to the need for improved validity and reliability through modifications to the BSS [19].

While retention time, defecation rate, BSS but not water activity have been shown to account for BSS-associated inflammatory effects [3,8,9,13,14,16–18] the potential correlation with the strength of a gel [23] in the context of intestinal abrasion forces remains unexplored. In response to external force (stress) complex materials either maintain rigidity, deform semipermanently (viscoelastic materials) or permanently (plastic materials). In the semi-solid materials (i.e., pastes) such as fecal matter minimal pressure (MP; force per unit area) required to induce permanent deformation is proportional to the density of crosslinks with

stiffer gels having a higher density of crosslinks [23]. Its importance as a shaping factor for intestinal abrasions and inflammation is currently unclear and has yet to be determined.

In this study, BSS and classical parameters (total solids, water content) [24,25] were recorded for a heterogenous collection of samples and mapped to MP measured in longitudinal and lateral transects of fecal samples. Random samples from prospective male-female study were collected and upgraded with healthy young male samples collected within the controlled four-week bed-rest space-exploration project PlanHab [26]. Gut environment was explored from the perspective of ecosystem development [15] in order to elucidate the relationship between the progressive increase in stool consistency (MP, dry matter) and the progressive intestinal and systemic inflammation observed over the course of the PlanHab project. The MP values reported in this study were coupled to reanalysis of the PlanHab data and fecal 1H-NMR metabolomes reported before [3,13,14] to explore the association between stool consistency (MP), individual signatures in 1H-NMR metabolites and the elevated intestinal, systemic inflammation and other detrimental physiological deconditioning effects observed in the PlanHab participants reported before [3,13,14,26–43].

2. Results and Discussion

2.1. Exploration of the Tripartite Relationship between BSS, Dry Matter Content and Novel MP Values

Fecal samples were collected from the prospective male-female random study that served to provide the backbone observations on the tripartite relationship between BSS, dry matter content and novel MP values. This relationship was further amended by superposition of samples collected within the PlanHab project ([3,13,14,26–43]; Table S1; Figures S1 and S2) that was designed to capture systemic body deconditioning parameters in response to three-week controlled bed-rest inactivity and hypoxia, a simulation of the space exploration environment. In summary, the decision of the host to reduce physical activity to three-week 24/7 bed-rest resulted in significant increase in insulin resistance, muscle resorption, bone demineralization and other numerous adaptations, hence the complexity of human body physiological responses were collated from the PlanHab literature and summarized in Table S1. The following observations became apparent from the exploration of the tripartite relationship between BSS, dry matter content and MP values:

First, exponential increase in MP [23] values was observed with decreasing BSS (Figure 1A) in 78 fresh stool samples exhibiting a wide array of fecal consistencies (43 males; 35 females) (Spearman $r = -0.86$, $p < 0.0001$). No significant difference could be detected between male and female samples in this study ($p > 0.05$). Samples within the same BSS class contained highly heterogeneous MP values, giving rise to five times larger variability in observed MP values within the three lowest BSS categories (1–3) (Figure 1A). The same observed relationships between MP and dry matter content were observed also for the PlanHab project samples ($n = 92$) [3,13,14]. The exponential function (Figure 1A) has little meaning in describing the relationship between MP and BSS beyond establishing the existence of nonlinear relationship in stool hardness (MP) and BSS. These results show that fecal samples exhibit continuous (MP) rather than discrete (BSS) characteristics, hence overcoming the discrete BSS boundaries.
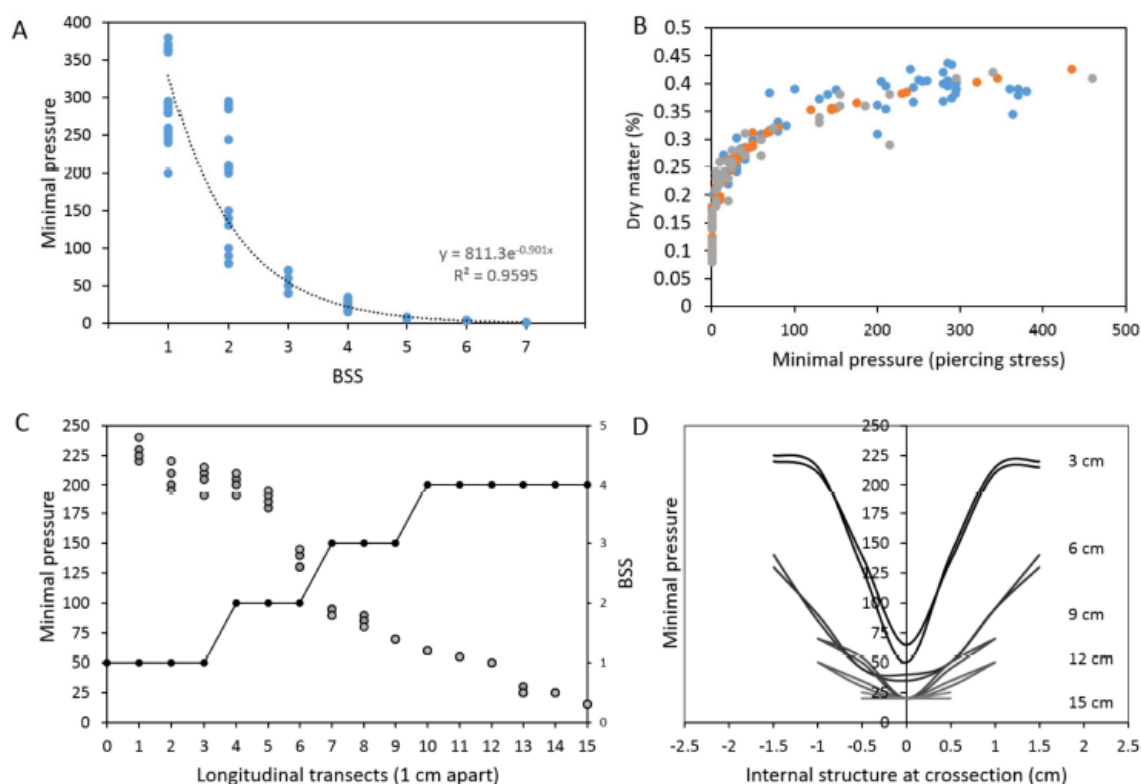
Figure 1. (**A**) Large heterogeneity in stool surface minimal pressure (MP (g/3.14 mm^2)) was identified within the same BSS class. Note the nonlinear increase in MP. (**B**) Healthy BSS values (3–4) were concentrated around the trade-off between the lowest MP and highest dry matter content. The intermediate density of crosslinks is beneficial for maintenance of human health as based on (•) 78 sample collection and our PlanHab project data (n = 96; [3,13,14]) (•) and (◦)). (**C**) An example of stool containing mixed BSS classes (•) and more than an order of magnitude difference in surface minimal pressure (◦) along the longitudinal transect of fecal specimen scanned from all 4 sides (n = 4). (**D**) The crossectional (lateral) transects of the same stool example as shown in (**C**) and their respective internal MP values at locations 3, 6, 9, 12 and 15 cm from the fecal tip. A decrease in surface MP and relatively small change in internal MP values can be seen along fecal specimen. Please consult Figures S4 and S5 for more details.

Second, the overall relationship between MP and dry matter content (Figure 1B) for all samples corresponded to rather uniform asymptotic function (y = 0.1343x$^{0.196}$, R^2 = 0.89). A surprisingly nonlinear, continuous and complex relationship between the two, and the rather disturbing increase in MP at almost negligible increase in observed dry matter content was identified (Figure 1B). Such dependence in fact confirms that the increasing molecular weight of the polymer chain connecting the crosslinks [23] generated the higher density of crosslinks. These results also suggests that at high MP the more rigid gel network, residing for a prolonged time in intestinal environment during constipation, could hardly be remodeled by intestinal muscles (peristaltic waves and segmenting contractions) without any abrasion being inflicted to the soft intestinal mucus and tissues. The correlation between DM and BSS (Figure S3) further confirmed the linear relationship between the two (R^2 = 0.92) in this study, including the large overlaps between DM in various BSS classes also observed before [3,24]. In this sense (Figure 1A,B) MP was more informative for fine scale and conclusive stratification of stool samples than DM.

Third, the intersection of lowest MP values and highest dry matter content (i.e., the apparent breaking point) corresponded to BSS values designated as healthy (BSS 3–4) (Figure 1B), despite the lower correlation between BSS and dry matter content (Spearman r = −0.76, p < 0.0001) in comparison to correlation between MP and dry matter above. This

clearly implicates intermediate MP values as highly important for maintenance of microbial activities, mixing and hence maintenance of human health. Beneficial characteristics at such density of gel crosslinks exist at intermediate MP to support non-inflammatory interactions between the microbiome and the host. The further increase in MP at roughly the same dry matter content was highly associated with progressive increase in gut and systemic inflammation reported before [3,13,14] for the same samples in the PlanHab project, showing detrimental effects of excessive crosslinking (Figure 1B).

Four, in this study, the longitudinal mapping of surface MP was performed over the entire length of stool sample to illustrate the difficulty of assigning uniform BSS to complex fecal samples (Figure 1C; Figure S4). These results show that more than an order of magnitude difference in MP can be detected along the length of a single fecal sample (Figure S1) and testifies that MP approach described in this study was able to discern fine grained internal, local differences (four transects over the same sample) that could not have been observed through the sole use of BSS for the whole fecal sample.

Five, lateral transects (Figure 1D) were inspected at 3 cm equidistance over the stool longitudinal transect and showed the existence of resistant surface layer followed by much softer internal structures exhibiting an order of magnitude lower MP values, that were characteristic of healthy, inflammation-free stool consistency (Figure S5).

From measurements obtained in this study it is conceivable that fecal matter is apparently easily mixed by intestinal tract contractions without accompanying negative symptoms characteristic of physical abrasions up to MP < 75. In this study, MP < 75 roughly corresponded to the clinically relevant boundary between BSS 2 and 3, i.e., the boundary between the BSS constipated and BSS normal values, that was so far hard to discern from visual inspection of fecal surface [19]. In contrast, MP > 75 already corresponded to prolonged intestinal transit and signs of constipation [3,13,14] at rather comparable dry matter content. In addition, internal differences in structure could be observed, such as progressively harder surface (e.g., MP~300) with the core MP values only twofold higher than in healthy makeup (Figure 1D). On the other hand, MP < 30 corresponded to the category of loose/watery fecal samples in analogy with the degree of stool characteristics described recently (hard/lumpy; normal; loose/watery) [22]. These two MP boundaries (MP < 30; MP > 75; Figure 1B) in fact correspond well to the clinically relevant boundaries for BSS > 4 and BSS < 3, respectively, that have been hard to determine unequivocally by visual inspection of specimen and BSS assignment only. In addition, fecal samples showing characteristics of BSS3 and 4 were shown to likely comprise multiple stool forms mixed together (Figure 1C), increasing thus BSS assignment errors when categorizing such mixed-forms [22]. In this sense, MP approach introduced in this study resolved this problem (Figure 1B) by introducing the continuous scale.

Taken together, in this study we established an exponential relationship between BSS and MP on one side and a complex saturation curve –like relationship between dry matter and MP. The PlanHab samples with decreased BSS values (Figure S1) were reported to be associated with a number of detrimental physiological and psychological characteristics next to intestinal inflammation (Figure S2; Table S1; [3,13,14,26,27,36]). Consequently, the increased stool resistance to remodeling as measured in this study by MP in the PlanHab samples was apparently related to intestinal abrasions, diet associated mucus imprinting and surface pore clogging on stool. These parameters were all shown to exert selective pressure on gut microbiome, its gene expression and metabolic activities, generating thus metabolic makeup associated with observed local and systemic inflammation [3,8,13,14,16], mirrored also in the urinary metabolomes of the same PlanHab project [26]. This clearly showed that parameters other than BSS accurately described clinically relevant fecal hardness.

The relationship observed in this study showed the generalizability and the potential of the MP approach to unequivocally characterize sampling sites in mixed samples (mixed BSS) for future metagenomic and metabolomic studies including mapping of the biochemical nature of the intestinal environment. Within-sample variation is still unre-

solved problem due to the observed variation between sampling sites in samples with inconsistent structure [44,45]. MP approach resolved this problem with exact characterization of the sampling microlocation. Its small surface area (d = 2 mm; p = 3.14 mm^2) enabled measurements to be uniformly repeated on many different locations over the same fecal sample to provide a multitude of measurements and hence a microscale estimate for particular location or consistency transect of fecal matter, or to map the longitudinal or lateral BSS and MP characteristics of a fecal specimen, hence linking the compactness of material to exact sample location for stratified analyses of metabolomes, host physiology, immune responses or microbiome.

Our results were obtained from samples collected in two independent projects, including males and females, covering a wide span of moisture contents and stool consistencies determined as described before [3,8,9,13,14,16–18]. Our data do provide independent evidence for the exponential association between stool consistency and MP as measure of crosslinking (Figure 1) corresponding to the intestinal and also systemic inflammation observed before [3,13,14] (Figures S1 and S2). In addition, simple two dimensional classification by dry matter and MP enables fast mapping of fecal samples for comparisons with unprecedented resolution, surpassing that of user dependent BSS values that have been also criticized for lack of consistency between studies [20].

In this study we presented an approach of minimal pressure (MP) required to induce permanent deformation (piercing) that is proportional to the density of crosslinks with stiffer gels having a higher density of crosslinks. The value of this novel approach comes from the relationship between the stress (compressive loading; forcer per unit area) and strain (deformation) that a particular material exhibits in general. Entangled polymers such as fecal matter are variably characterized by a mixture of physical entanglements between polymer chains and also chemical crosslinks that give the material gel-like properties and can span brittle-ductile material behavior [23]. Different approaches to measurement of material characteristics exist such as penetrometer [46], viscosimeter [47,48] and texture analyzer [25,49,50] and are utilized depending on the necessary sample pretreatment (flattening, homogenization, mixing, averaging, subsampling), complexity of the apparatus (e.g., TA.XTExpressC), measurement approaches (viscosity, stickiness, hardness), and the necessity to record the fine-scale 3D structure of the non-homogenous material specimen. In this sense the MP concept presented in this study and terminology of minimal pressure (MP) represent a significant extension to the existing approaches analyzing stool consistency as site-specific measurements of stool characteristics important for identifying unique metabolome and microbiome signatures are enabled, linking them to exact sampling locations. Further, MP approach does not require any pretreatment (e.g., homogenization, packing, flattening) but enables 3D mapping of the sampling locations relevant for biogeography of intestinal environment. MP approach described in this study provides clinical benefits from being able to more precisely classify BSS group 1, 2 and 3, to better delineate clinically relevant boundaries of consistencies (2/3 BSS; 3/4 BSS), i.e., delineating the central optimal span in MP relevant for the medical delineation between classes.

In addition, the MP approach operates on unmodified fecal sample that can be stored at 4 °C and rewarmed, enables fine-scale longitudinal, crossectional site-specific measurements before actual subsampling for various chemical and molecular analyses, giving rise to descriptions of fecal sample locations relevant for 3D biogeography. Finally, the MP approach does not require sample pretreatment but supports simple direct measurement devoid of complex or expensive apparatus, and hence represents a cheap and operator independent, reproducible and objective alternative that can be utilized globally.

2.2. The PlanHab Project Metabolite Signatures Characteristic of High MP

MP values recorded for the PlanHab samples were used to show that large variability in stool consistency was hidden within BSS values (Figure 2) recorded for participants [3,13,14,26]. First, these data show the high inter-individual heterogeneity despite the fact that the PlanHab project experiment was conducted under strictly controlled

conditions, including diet, immobilization, oxygen level, hydration, circadian rhythm, 24/7 medical surveillance. The three-week experiment under different conditions resulted in progressive body deconditioning (Table S1) next to constipation and increased intestinal and systemic inflammatory responses (Figures S1 and S2). The MP data for the same samples illustrate the profound increase in MP values of progressively constipated participants (Figure 2) that matched with the deconditioning and inflammation markers recorded before (Figure S2). The dose dependent increase in MP in the most affected PlanHab experimental variants illustrated the importance of the MP for the detection of modified intestinal conditions, characterized by five to six times higher MP, previously linked to the access of pathogens and endotoxins to the epithelium through physical mucus compaction [51] and abrasions due to long-term residence and regular intestinal muscle contractions [3,13,14].
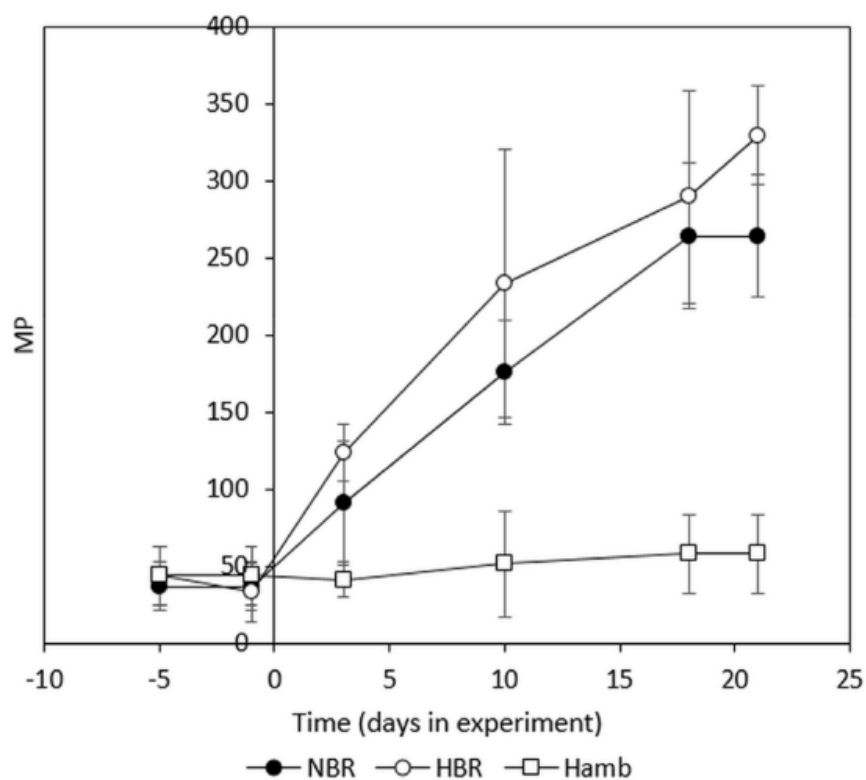
Figure 2. Presentation of MP values for the PlanHab samples in relation to the experimental variants described with BSS (Figure S6). Error bars designate standard deviation. NBR-normoxic bedrest, HBR-hypoxic bedrest, HAmb-hypoxic ambulatory variants of the PlanHab experiment.

Second, to further explore the utility of MP strategy in metabolomic analyses of intestinal tracts, MP was utilized in reanalysis of our previously published fecal 1H-NMR metabolomes from the same samples obtained from the 4-week PlanHab project [14] (Table S1). The corresponding 1H-NMR data were reanalyzed using the latest ChenomX 8.6 software and linked to MP data collected for the same samples. Principal coordinate analysis of the PlanHab 1H-NMR metabolomes showed the existence of rather unique and highly individualized metabolic signatures over the course of the PlanHab experiment (Figure S6). Essentially, each sample received unique MP value, showcasing the much finer and continuous resolution for the locations from which the samples were collected. Consequently, much larger numbers of samples would need to be collected to model MP relative to complexity of 1H-NMR metabolomes. Power analysis estimates

suggested two orders of magnitude larger sample size in the range of 10.000 would be required as observed in recent metagenomic studies [52,53].

Third, when the same 1H-NMR data were utilized on more coarse scale to explore the more recent classification [22] with the loose/watery, normal/healthy and hard/lumpy classification [22], the two MP boundaries identified in this study (MP < 30; MP > 75; Figure 1B) matched the clinically relevant boundaries for BSS < 3 and BSS > 4, giving rise to three rather broad categories of MP classes: MP1 < 30; 30 < MP2 < 75; MP3 > 75. These classes further match with the loose/watery normal/healthy and hard/lumpy classification [22] and were used in machine learning and modeling in search of significant differences in metabolomics. Analyses utilizing BSS assignments were described before [14]. Ridge logistic regression model with penalty hyperparameter 1.0 as the predictive algorithm was selected as the best interpretable model (AUC = 0.783). Constant Removal and standardization and LASSO feature selection (Penalty = 1.0, Lambda = 1.558e-01) were used in this context. The output of selected features with LASSO regression was used for prediction with ridge logistic regression model. Both were chosen with automatic machine learning process as the best option form 168952 trained models. Acetate and methionine were selected as reference signatures from the 174 analyzed metabolites. According to Individual Conditional Expectation (ICE) plots, increased concentrations of acetate increased the probability of such a sample being classified in the low minimum pressure group, while increased concentrations of methionine increased the likelihood of such a sample being classified as the medium to high MP group. The 95% CI of the model performance achievable by using acetate alone ranged from 94.1 to 100%. The addition of methionine to the analysis, the model performance was effectively close to 100%. The best interpretable model was validated on the test data, selecting acetate and methionine as predictive features, and with AUC = 0.833, the validation of the model was successful. According to the receiver operating characteristic curves (ROC), the model showed better performance in classifying in MP1 and MP3 groups (Figure 3A,C), with lower or higher MP values. On the other hand, the performance was lowest in MP2 group (Figure 3B). This further showcases the large inter-individual differences and variable responses to the same diet in the PlanHab participants. Power analysis showed that two orders of magnitude larger samples would be needed to effectively build submodels for the metabolic diversity of the apparently healthy gut metabolomes.

Taken together, metabolomes belonging to MP1 and MP3 groups outside the central span of MP2 values possessed sufficient information for acceptable sample classification. Surprisingly, the makeup of the intermediate MP2 group did not exhibit any characteristic signatures (Figure 2B). A wide array of metabolic makeups mirroring characteristic of inter-individual differences in physiological makeup of healthy microbiomes in feces were reported before [12]. From this it follows that lower and upper extremes (i.e., MP1 or MP3) contained features distributed characteristically enabling their separation. JADBIO machine learning and modelling [54] showed that out of all measured metabolites, the two most important features responsible for separation of the three groups were acetate, a short chain fatty acid, produced by microbes, and methionine, a compound involved in regulation of metabolic processes, the innate immune system, digestive functioning in mammals next to their lipid metabolism, activation of endogenous antioxidant enzymes (methionine sulfoxide reductase A), and the biosynthesis of glutathione to counteract oxidative stress [55]. Finally, methionine restriction was shown to decrease DNA damage and carcinogenic processes, averting arterial, neuropsychiatric, and neurodegenerative disease [55]. Fecal acetate was shown to be inversely related to acetate absorption from the human colon, and high circulating acetate concentrations were negatively correlated to insulin sensitivity [56]. In our past analyses of the PlanHab urine 1H-NMR metabolomes [26] elevated acetate concentrations were observed in most constipated participants, exhibiting insulin resistance, modified fat oxidation, bone demineralization, muscle deconditioning and depression (Table S1; Figure S1; Figure S2) [27,34,39,40]. The fact that the PlanHab data was derived from a medically prescreened cohort receiving defined and synchronized

diet, characteristic of Western Diet [3,27,34] in tightly controlled environment of the 4-week PlanHab study (levels of exercise, circadian rhythm, medical care, oxygen pressure) enabled us to study the existence of significant differences between fecal metabolomes and exact fecal makeup at the sampling location (MP), rather than average BSS assignment (Figure S1) described before [3]. Therefore the observed differences in the acetate and methionine levels in fecal samples stem from the conserved differences in their uptake as a result of responses to inactivity, coupled with Western type of the diet utilized in the PlanHab project.
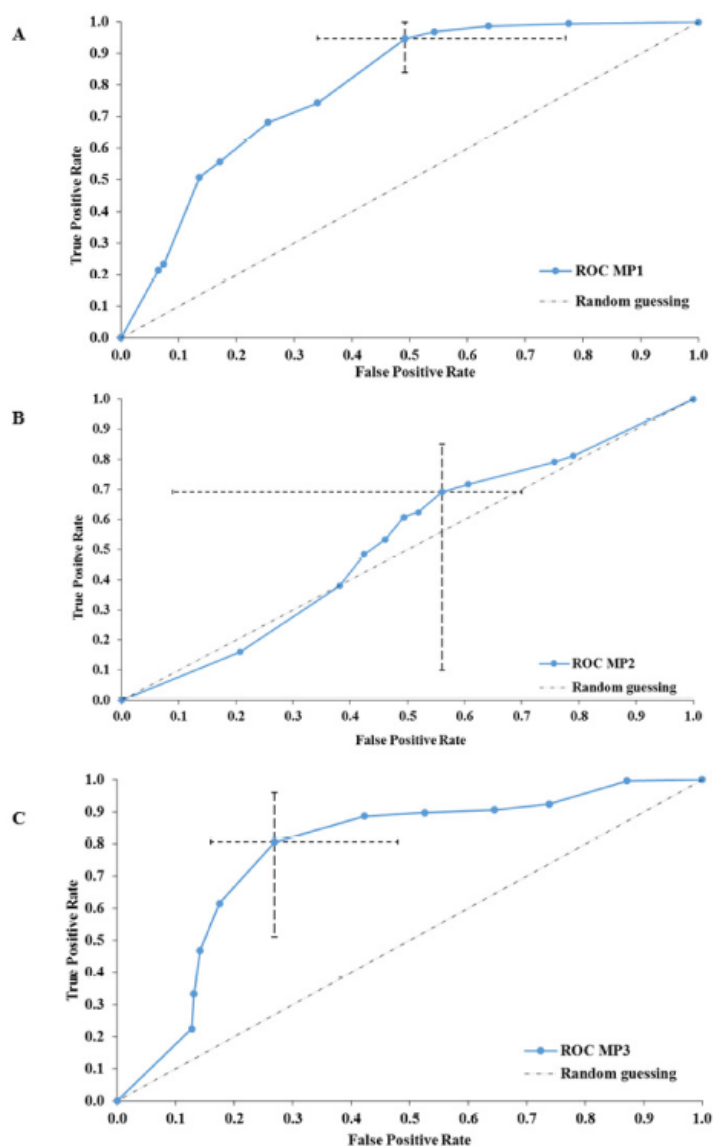
Figure 3. An overview of the model performance in three different groups between fecal metabolomes in delineation with three different groups of MP (MP1 < 30 (**A**), 30 < MP2 < 75 (**B**) and MP3 > 75 (**C**)). Please note the rather uncharacteristic and hard to classify makeup of metabolomes observed in the intermediate, healthy, group (**B**). Horizontal and vertical dashed lines represent 95% confidence interval for false positive and true positive rates, respectively. The black dashed line represents model performance in case of random guessing; blue line represents training mean performance of the model.

67

The inclusion of the newly described parameters such as MP, acetate and methionine extends our previous findings on the significant parameters associated with detrimental effects of inactivity on human body [3,13,14,26] and its mechanisms: the decision of the host to reduce physical activity gave rise to increased fecal electrical conductivity that lead to decreased BSS (constipation) and increased bile acids (BA) levels, while at the same time reduced indole concentrations at retained high electrical conductivity resulted in higher intestinal inflammation (EDN) levels (Figure S2). The relationships observed in Bayesian modelling [14] identified that nonlinear responses take place over the network as small changes in indole levels exhibited unexpectedly large effects on BA content. Taken together, this clearly shows that introduction of nonlinear MP as an extension to BSS and observing an increasing MP in fecal samples describing fecal surface deformability, represents the link between physical abrasions due to intestinal muscle contractions, microbial indole and acetate production and intestinal inflammation marker EDN reported before [3,13,14,26].

3. Materials and Methods

3.1. Fecal Sample Collection and Analysis

Samples utilized in analyses in this study were collected within the prospective study (n = 78) and the PlanHab project (n = 96), anonymized by coding following collection and subsequently characterized for various physical and chemical characteristics as described before [3,13,14,26]. All participants gave written informed consent after receiving detailed information regarding the study protocol and all experimental procedures. Ethics Committee permission was obtained from the National Ethics Committee of the Republic of Slovenia and is held by Jožef Stefan Institute.

The participants of the prospective study represented a one-time sampling cohort spanning 78 fresh stool samples exhibiting a wide array of fecal consistencies of otherwise healthy nonsmoking 43 males and 35 females. The participants of the prospective study were represented by healthy, non-obese (BMI < 30 kg/m^2) males and females (22–42 years). Baseline male and female characteristics were as following; age (28 $\pm$ 5 years and 29 $\pm$4 years); body mass (76 $\pm$ 10.5 kg and 61$\pm$ 8 kg), BMI (23 $\pm$ 5 kg/m^2 and 23$\pm$ 3 kg/m^2), respectively.

Further, the PlanHab healthy male participants were characterized by numerous clinically relevant measurements to assert absence of disease with a state of physical, mental, and social welfare. Their baseline characteristics were as following; age (27 $\pm$ 6 years); body mass (76.7 $\pm$ 11.8 kg), BMI (23.7 $\pm$ 3 kg/m^2) [27,34]. The participants underwent 5 days of baseline data collection during which participants were ambulant, 21 intervention days and 5–14 days of medical follow-up. The PlanHab project participants (n = 11) were provided with an individually tailored, standardized, and controlled diet throughout the intervention as described before [3,27,34]. Energy requirements were assessed with the Harris-Benedict method, and correction factors of 1.4 and 1.2 were used to account for activity levels in the ambulatory phases and the bed rest phases, respectively. In addition to a controlled intake of fat (30%) and protein (1.2 g per kg body mass), sodium intake was set to 3500 mg per day. Participants were supplemented with 1000 IU vitamin D3 per day. Fluid intake was ad libitum, but participants were encouraged to drink at least 28.5 mL per kg per day. Importantly, menu plans were cycled in the same way for each participant across the three experimental conditions, adjusting the quantity according to activity factors above. The collected fecal samples (Figure S1) thus represent a longitudinal transect where intestinal tracts developed progressive constipation and a number of systemic physiological deconditioning symptoms (Figure S2) [3,13,14,26]. In total, 96 samples were collected over the course of the PlanHab experiment for 11 participants. Their data relevant for this study (BSS, inflammation (eosinophile derived neurotoxin, systemic inflammation) are presented in Figures S1 and S2.

The BSS score was assigned immediately after the collection of specimen as before [3,10,11,13,14].

Water content of the fresh sample was determined by collecting samples (~70 g) into pre-weighted 200 mL collection jars and sample mass determined by second weighing of

the jar. Water content of a sample was determined by drying at 60 °C for 48 h. The sample was cooled in a laboratory desiccator and weighed again. The water content of a sample was calculated as the difference in final and initial mass of the sample, divided by the initial mass [57]. The dry matter content of the fresh sample was calculated by subtracting water content from 1.

3.2. Measurements of Minimal Pressure

Here, we measure minimal pressure of 170 fecal samples with different BSS consistencies using a defined flat-cut stainless steel probe (d = 2 mm; s = 3.14 mm^2) with attachable adjustable weights (1–400 g) at constant temperature (25 °C). Probe indentation [23] by gravity that is independent of probe velocity tests primarily the response of the gel network and hence the measured hardness shows a strong dependence on the increasing molecular weight of the polymer chain connecting the crosslinks [23] generating the higher density of crosslinks [23].

In practice the approach is used to determine the minimum weight per unit area needed to pierce through the surface of fecal specimen. The utility was developed in a way that additional weights were attached to the stainless steel rod until the weight was sufficient to pierce through the material. Due to its small surface area the MP measurements were uniformly repeated on many different locations over each specimen. This enabled us to provide a multitude of measurements and hence a microscale estimate for particular location of fecal specimen. Longitudinal transects were probed 0.5 cm apart, then the specimen was rotated for 90° and measured again. Lateral transects were obtained after the specimen was cut at locations of 3 cm, 6 cm, 9 cm, 12 cm and 15 cm. Readings were recorded for two perpendicular transects within each specimen.

The MP measurements at 37 °C were in general 10% lower in comparison to those determined at 24 °C. In addition, MP measurements taken within the 24 and 48 h stored at 4 °C and measured at 24 °C after 60 min reheating were not significantly different.

Principal coordinate analysis was conducted on Box-Cox transformed metabolomics data and Benjamini-Hochberg significance correction for multiple comparisons was used in non-parametric npMANOVA as described before [3,13].

3.3. Intestinal Metabolome Analysis Using Proton Nuclear Magnetic Resonance (1H-NMR)

The 1H-NMR intestinal metabolomic data published before [14] were reanalyzed using a novel version of ChenomX 8.6 and analyzed utilizing machine learning approach (Just Add Data Bio (JADBIO, version 1.1.182)) [54].

In essence, the reanalyzed data were obtained as described before [14]: fecal samples (200 mg of dry matter) were resuspended in 800 µL of NMR phosphate buffer and centrifuged at 10,000 g for 30 min at 4 °C to remove fine particles. Samples were filtered through 0.22 µm HPLC compatible filters (Millipore, Germany), 400 µL aliquots were mixed with 200 uL 1H-NMR buffer as described before [58] and stored at −25 °C until analysis. Phosphate buffer (pH 7.4) was prepared by weighing 1.443 g Na_2HPO_4, 0.263 g NaH_2PO_4, 2 mM TSP, and 1 mM NaN_3 into 50 mL volumetric flask. Ten milliliter of D_2O was added and filled up to 50 mL with Milli-Q water. Before analysis, samples were thawed at room temperature, centrifuged at 12,000 g for 5 min at 4 °C. In total, 550 µL of each sample was transferred into 5 mm NMR tube.

1H-NMR spectra were acquired on an Agilent Technologies DD2 600 MHz NMR spectrometer equipped with 5 mm HCN Cold probe. The 2D experiments were measured on Agilent Technologies (Varian) VNMRS 800 MHz NMR spectrometer equipped with 5 mm HCN Cold probe. All experiments were measured at 25 °C. 1H-NMR spectra of the samples were recorded with spectral width of 9.0 kHz, relaxation delay 2.0 s, 32 scans and 32 K data points. Water signal was suppressed using Double-pulsed field gradient spin echo (DPFGSE) pulse sequence. Heteronuclear single quantum coherence spectrum (HSQC) was acquired for 1H and 13C dimensions and total correlated spectrum (TOCSY) was measured with 1H spectral widths of 7.0 kHz, relaxation delay 1.5 s, 160 number of transient and

128 time increments. For apodization of acquired spectra, we used exponential and a cosine-squared functions. NMR spectra were processed using VMRJ (Agilent/Varian) and Sparky (UCSF) software and MestReNova.

The resulting spectra were consequently analyzed using targeted quantitative metabolomics using Chenomx NMR Suite version 8.6 (2020; Chenomx, Canada). All spectra were randomly ordered for spectral fitting using ChenomX profiler. Metabolites analyzed in this study were identified using the support of Chenomx Compound Library extended by Human Metabolome Data Base [59].

Automated Machine Learning was used to identify most important metabolic features separating the groups of fecal matter plasticity. Samples were divided into three separate groups indicating low ((MP1 < 30), medium (30 < MP2 < 75), and high (MP3 > 75) MP, and 174 metabolic features were analyzed for possible differentiation. JADBIO version 1.1.182 was used for model generation. Data were split in a 70:30 ratio for model training (70%) and model validation (30%). An extensive tuning effort with six CPU cores was used to compute the most interpretable classification model, which was selected based on the area under the curve (AUC) metric among 168952 trained models. Different algorithms with different combinations of tuned parameters were used for feature selection (LASSO regression and test-budgeted statistically equivalent signature) and for prediction (ridge logistic regression, support vector machines, classification random forest and classification trees). Metabolic data were preprocessed with constant removal and standardized. LASSO feature selection (penalty = 1.0, lambda = 1.558e-01) was used for metabolite feature selection. The output of feature selection was used for obtaining the best interpretable model using the predictive ridge logistic regression algorithm (with hyper-parameter penalty equal to 1.0).

The machine learning process described in this study was adopted for several reasons: (i) automation in parameter and algorithm selection results in reduced bias and human interference; (ii) the approach includes several different ML algorithms (linear regression, SVM, decision tree, random forest and Gaussian kernel SVMs and automatically choose the most interpretable model based on AUC metric; (iii) the resulting models were trained with different configurations on different sub-samples of the original dataset (cross-validation); (iv) focus on relevant humanly interpretable models. Consequently, algorithm, hyperparameter and space selection (AHPS) as implemented in JADBIO was used for selecting the most suitable algorithm for preprocessing and transformation of a given dataset, its feature selection and modeling. The output of AHPS step was analyzed by configuration evaluation protocol (CEP) in order to find the optimal configuration reported in this study [54,60].

To evaluate model classification, a receiver-operating characteristic curve (ROC curve) was constructed for all three groups, plotting the true-positive rate (sensitivity) against the false-positive rate (1-specificity). Individual conditional expectation (ICE) plots revealed the nature of the contribution of each metabolite feature to the model.

4. Conclusions

In this study, a minimal pressure (MP) approach utilizing probe indentation of intact fecal samples was introduced as a measure of stool consistency. MP values recorded over a spectrum of moisture contents increased exponentially relative to BSS and enabled stratification of samples exhibiting mixed BSS classes. A trade-off between lowest MP and highest dry matter content delineated the span of intermediate healthy density of gel crosslinks. The crossectional transects identified fecal surface layers with exceptionally high MP suggestive of mucus imprinting overlying internal fecal structures with an order of magnitude lower MP characteristic of healthy stool consistency. The exponential association between stool consistency and MP determined in this study was mirrored in the elevated intestinal and systemic inflammation next to other detrimental physiological deconditioning effects observed in the PlanHab participants reported before. High inter-individual differences in fecal 1H-NMR metabolomes derived from a wide spectrum of MP showed the importance of the exact sampling location in future microbiome and metabolome studies. In conclusion,

we believe that the MP approach described in this study it can be used to better understand fecal hardness and its relationships to human health as it provides a simple, fine scale and objective stool classification approach.

Supplementary Materials: The following are available online at https://www.mdpi.com/2218-1989/11/3/172/s1, Figure S1: Changes in Bristol stool scale values (A) and retention time (as time between particular defecations) (B) during run-in (week 1) and subsequent 3-week experimental phase of the PlanHab project, Figure S2: Heatmap plot showing the relationship between parameters describing human physiology, psychology, and intestinal environment that differed significantly at the end of the PlanHab experiment, Figure S3: A scatter plot depicting the relationship between dry matter (% DM) and BSS class assignments, Figure S4: Schematic representation of BSS and MP along the longitudinal transect of fecal specimen, Figure S5: A schematic representation of crossectional (internal) MP values for fecal bolus (upper) and along the longitudinal transect of fecal specimen of mixed type, for which a uniform BSS classification is ambiguous, Figure S6: Principal Coordinate Analysis of fecal 1H-NMR metabolomes sampled over the course of the 4-week PlanHab project, Table S1: A compilation of the PlanHab project publications containing relevant sources of information.

Author Contributions: Conception and design of the study (B.S.), data collection (L.D., B.S.), data preparation and analysis (L.D., B.S.), writing and critical revision of the manuscript (L.D., B.S.). All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the National Committee for Medical Ethics at the Ministry of Health of the Republic of Slovenia (Ethical approved number is 88/04/12).

Informed Consent Statement: All subjects gave written informed consent to publish this paper.

Data Availability Statement: BSS, minimal pressure, dry matter content data next to 1H-NMR fecal metabolomes of the PlanHab project are made available upon request to the corresponding author. Model with test data for clustering samples according to metabolite profiles are included in the supplementary material. Instructions for running a model on local machine are included in the electronic supplementary material.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ehrlein, H.J.; Schemann, M. *Gastrointestinal Motility*; Technische Universität München: Munich, Germany, 2005.
2. Johnson, L.R.; Ghishan, F.K.; Kaunitz, J.D.; Merchant, J.L.; Said, H.M.; Wood, J.D. *Physiology of the Gastrointestinal Tract*; Johnson, L.R., Kaunitz, J.D., Said, H.M., Ghishan, F.K., Merchant, J.L., Wood, J.D., Eds.; Academic Press: London, UK, 2012; p. 2308.
3. Sket, R.; Treichel, N.; Debevec, T.; Eiken, O.; Mekjavic, I.; Schloter, M.; Vital, M.; Chandler, J.; Tiedje, J.M.; Murovec, B.; et al. Hypoxia and Inactivity Related Physiological Changes (Constipation, Inflammation) Are Not Reflected at the Level of Gut Metabolites and Butyrate Producing Microbial Community: The PlanHab Study. *Front. Physiol.* **2017**, *8*. [CrossRef] [PubMed]
4. Cremer, J.; Segota, I.; Yang, C.Y.; Arnoldini, M.; Sauls, J.T.; Zhang, Z.; Gutierrez, E.; Groisman, A.; Hwa, T. Effect of flow and peristaltic mixing on bacterial growth in a gut-like channel. *Proc. Natl. Acad. Sci. USA* **2016**, *113*. [CrossRef]

5. Cremer, J.; Arnoldini, M.; Hwa, T. Effect of water flow and chemical environment on microbiota growth and composition in the human colon. *Proc. Natl. Acad. Sci. USA* **2017**, *114*. [CrossRef] [PubMed]

6. Arnoldini, M.; Cremer, J.; Hwa, T. Bacterial growth, flow, and mixing shape human gut microbiota density and composition. *Gut Microbes* **2018**, *9*. [CrossRef]

7. Glover, L.E.; Lee, J.S.; Colgan, S.P. Oxygen metabolism and barrier regulation in the intestinal mucosa. *J. Clin. Investig.* **2016**, *126*. [CrossRef]

8. Vandeputte, D.; Falony, G.; Vieira-Silva, S.; Tito, R.Y.; Joossens, M.; Raes, J. Stool consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial growth rates. *Gut* **2016**, *65*, 57–62. [CrossRef] [PubMed]

9. Tigchelaar, E.F.; Bonder, M.J.; Jankipersadsing, A.; Fu, J.; Wijmenga, C.; Zhernakova, A. Gut microbiota composition associated with stool consistency. *Gut* **2016**, *65*, 540–542. [CrossRef]

10. Heaton, K.W.; Radvan, J.; Cripps, H.; Mountford, R.A.; Braddon, F.E.M.; Hughes, A.O. Defecation frequency and timing, and stool form in the general population: A prospective study. *Gut* **1992**, *33*, 818–824. [CrossRef]

11. Lewis, S.J.; Heaton, K.W. Stool form scale as a useful guide to intestinal transit time. *Scand. J. Gastroenterol.* **1997**, *32*, 920–924. [CrossRef]

12. Roager, H.M.; Hansen, L.B.S.; Bahl, M.I.; Frandsen, H.L.; Carvalho, V.; Gobel, R.J.; Dalgaard, M.D.; Plichta, D.R.; Sparholt, M.H.; Vestergaard, H.; et al. Colonic transit time is related to bacterial metabolism and mucosal turnover in the gut. *Nat. Microbiol.* **2016**, *1*. [CrossRef] [PubMed]

13. Sket, R.; Treichel, N.; Kublik, S.; Debevec, T.; Eiken, O.; Mekjavic, I.; Schloter, M.; Vital, M.; Chandler, J.; Tiedje, J.M.; et al. Hypoxia and inactivity related physiological changes precede or take place in absence of significant rearrangements in bacterial community structure: The PlanHab randomized trial pilot study. *PLoS ONE* **2017**, *12*. [CrossRef]

14. Sket, R.; Debevec, T.; Kublik, S.; Schloter, M.; Schoeller, A.; Murovec, B.; Mikus, K.V.; Makuc, D.; Pecnik, K.; Plavec, J.; et al. Intestinal Metagenomes and Metabolomes in Healthy Young Males: Inactivity and Hypoxia Generated Negative Physiological Symptoms Precede Microbial Dysbiosis. *Front. Physiol.* **2018**, *9*. [CrossRef] [PubMed]

15. Falony, G.; Vieira-Silva, S.; Raes, J. Richness and ecosystem development across faecal snapshots of the gut microbiota. *Nat. Microbiol.* **2018**, *3*, 526–528. [CrossRef]

16. Aron-Wisnewsky, J.; Prifti, E.; Belda, E.; Ichou, F.; Kayser, B.D.; Dao, M.C.; Verger, E.O.; Hedjazi, L.; Bouillot, J.L.; Chevallier, J.M.; et al. Major microbiota dysbiosis in severe obesity: Fate after bariatric surgery. *Gut* **2019**, *68*, 70–82. [CrossRef]

17. Hadizadeh, F.; Walter, S.; Belheouane, M.; Bonfiglio, F.; Heinsen, F.A.; Andreasson, A.; Agreus, L.; Engstrand, L.; Baines, J.F.; Rafter, J.; et al. Stool frequency is associated with gut microbiota composition. *Gut* **2017**, *66*, 559–560. [CrossRef] [PubMed]

18. Vandeputte, D.; Kathagen, G.; D'Hoe, K.; Vieira-Silva, S.; Valles-Colomer, M.; Sabino, J.; Wang, J.; Tito, R.Y.; De Commer, L.; Darzi, Y.; et al. Quantitative microbiome profiling links gut community variation to microbial load. *Nature* **2017**, *551*, 507–511. [CrossRef]

19. Blake, M.R.; Raker, J.M.; Whelan, K. Validity and reliability of the Bristol Stool Form Scale in healthy adults and patients with diarrhoea-predominant irritable bowel syndrome. *Aliment. Pharmacol. Ther.* **2016**, *44*. [CrossRef] [PubMed]

20. Chumpitazi, B.P.; Self, M.M.; Czyzewski, D.I.; Cejka, S.; Swank, P.R.; Shulman, R.J. Bristol Stool Form Scale reliability and agreement decreases when determining Rome III stool form designations. *Neurogastroenterol. Motil. Off. J. Eur. Gastrointest. Motil. Soc.* **2016**, *28*. [CrossRef]

21. Derrien, M.; van Passel, M.W.; van de Bovenkamp, J.H.; Schipper, R.G.; de Vos, W.M.; Dekker, J. Mucin-bacterial interactions in the human oral cavity and digestive tract. *Gut Microbes* **2010**, *1*. [CrossRef] [PubMed]

22. Matsuda, K.; Akiyama, T.; Tsujibe, S.; Oki, K.; Gawad, A.; Fujimoto, J. Direct measurement of stool consistency by texture analyzer and calculation of reference value in Belgian general population. *Sci. Rep.* **2021**, *11*, 2400. [CrossRef] [PubMed]

23. Grillet, A.M.; Wyatt, N.B.; Gloe, L.M. Polymer Gel Rheology and Adhesion. In *Rheology*; De Vicente, J., Ed.; IntechOpen: Rijeka, Croatia, 2012; pp. 59–80.

24. de Loubens, C.; Dubreuil, A.; Lentle, R.G.; Magnin, A.; El Kissi, N.; Faucheron, J.L. Rheology of human faeces and pathophysiology of defaecation. *Tech. Coloproctology* **2020**, *24*, 323–329. [CrossRef] [PubMed]

25. Aichbichler, B.W.; Wenzl, H.H.; Santa Ana, C.A.; Porter, J.L.; Schiller, L.R.; Fordtran, J.S. A comparison of stool characteristics from normal and constipated people. *Dig. Dis. Sci.* **1998**, *43*, 2353–2362. [CrossRef] [PubMed]

26. Šket, R.; Deutsch, L.; Prevoršek, Z.; Mekjavić, I.B.; Plavec, J.; Rittweger, J.; Debevec, T.; Eiken, O.; Stres, B. Systems View of Deconditioning During Spaceflight Simulation in the PlanHab Project: The Departure of Urine 1 H-NMR Metabolomes from Healthy State in Young Males Subjected to Bedrest Inactivity and Hypoxia. *Front. Physiol.* **2020**, *11*. [CrossRef]

27. Debevec, T.; Bali, T.C.; Simpson, E.J.; Macdonald, I.A.; Eiken, O.; Mekjavic, I.B. Separate and combined effects of 21-day bed rest and hypoxic confinement on body composition. *Eur. J. Appl. Physiol.* **2014**, *114*, 2411–2425. [CrossRef] [PubMed]

28. Debevec, T.; Simpson, E.J.; Mekjavic, I.B.; Eiken, O.; Macdonald, I.A. Effects of prolonged hypoxia and bed rest on appetite and appetite-related hormones. *Appetite* **2016**, *107*, 28–37. [CrossRef]

29. Keramidas, M.E.; Kolegard, R.; Mekjavic, I.B.; Eiken, O. PlanHab: Hypoxia exaggerates the bed-rest-induced reduction in peak oxygen uptake during upright cycle ergometry. *Am. J. Physiol Heart Circ. Physiol.* **2016**, *311*, H453–H464. [CrossRef] [PubMed]

30. Louwies, T.; Jaki Mekjavic, P.; Cox, B.; Eiken, O.; Mekjavic, I.B.; Kounalakis, S.; De Boever, P. Separate and Combined Effects of Hypoxia and Horizontal Bed Rest on Retinal Blood Vessel Diameters. *Invest. Ophthalmol. Vis. Sci.* **2016**, *57*, 4927–4932. [CrossRef] [PubMed]

31. Rittweger, J.; Debevec, T.; Frings-Meuthen, P.; Lau, P.; Mittag, U.; Ganse, B.; Ferstl, P.G.; Simpson, E.J.; Macdonald, I.A.; Eiken, O.; et al. On the combined effects of normobaric hypoxia and bed rest upon bone and mineral metabolism: Results from the PlanHab study. *Bone* **2016**, *91*, 130–138. [CrossRef] [PubMed]

32. Rullman, E.; Fernandez-Gonzalo, R.; Mekjavic, I.B.; Gustafsson, T.; Eiken, O. MEF2 as upstream regulator of the transcriptome signature in human skeletal muscle during unloading. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **2018**, *315*, R799–R809. [CrossRef] [PubMed]

33. Rullman, E.; Mekjavic, I.B.; Fischer, H.; Eiken, O. PlanHab (Planetary Habitat Simulation): The combined and separate effects of 21 days bed rest and hypoxic confinement on human skeletal muscle miRNA expression. *Physiol. Rep.* **2016**, *4*. [CrossRef] [PubMed]

34. Simpson, E.J.; Debevec, T.; Eiken, O.; Mekjavic, I.; Macdonald, I.A. PlanHab: The combined and separate effects of 16 days of bed rest and normobaric hypoxic confinement on circulating lipids and indices of insulin sensitivity in healthy men. *J. Appl. Physiol.* **2016**, *120*, 947–955. [CrossRef]

35. Morrison, S.A.; Mirnik, D.; Korsic, S.; Eiken, O.; Mekjavic, I.B.; Dolenc-Groselj, L. Bed Rest and Hypoxic Exposure Affect Sleep Architecture and Breathing Stability. *Front. Physiol.* **2017**, *8*. [CrossRef] [PubMed]

36. Strewe, C.; Zeller, R.; Feuerecker, M.; Hoerl, M.; Kumprej, I.; Crispin, A.; Johannes, B.; Debevec, T.; Mekjavic, I.; Schelling, G.; et al. PlanHab study: Assessment of psycho-neuroendocrine function in male subjects during 21 d of normobaric hypoxia and bed rest. *Stress* **2017**, *20*, 131–139. [CrossRef] [PubMed]

37. Debevec, T.; Ganse, B.; Mittag, U.; Eiken, O.; Mekjavic, I.B.; Rittweger, J. Hypoxia Aggravates Inactivity-Related Muscle Wasting. *Front. Physiol.* **2018**, *9*, 494. [CrossRef]

38. Salvadego, D.; Keramidas, M.E.; Kolegard, R.; Brocca, L.; Lazzer, S.; Mavelli, I.; Rittweger, J.; Eiken, O.; Mekjavic, I.B.; Grassi, B. PlanHab(*): Hypoxia does not worsen the impairment of skeletal muscle oxidative function induced by bed rest alone. *J. Physiol.* **2018**, *596*, 3341–3355. [CrossRef] [PubMed]

39. Stavrou, N.A.M.; Debevec, T.; Eiken, O.; Mekjavic, I.B. Hypoxia Exacerbates Negative Emotional State during Inactivity: The Effect of 21 Days Hypoxic Bed Rest and Confinement. *Front. Physiol.* **2018**, *9*, 26. [CrossRef]

40. Stavrou, N.A.M.; Debevec, T.; Eiken, O.; Mekjavic, I.B. Hypoxia Worsens Affective Responses and Feeling of Fatigue during Prolonged Bed Rest. *Front. Psychol.* **2018**, *9*, 362. [CrossRef]

41. Strewe, C.; Zeller, R.; Feuerecker, M.; Hoerl, M.; Matzel, S.; Kumprej, I.; Crispin, A.; Johannes, B.; Debevec, T.; Mekjavic, I.B.; et al. PlanHab Study: Consequences of combined normobaric hypoxia and bed rest on adenosine kinetics. *Sci. Rep.* **2018**, *8*, 1762. [CrossRef] [PubMed]

42. Sarabon, N.; Mekjavic, I.B.; Eiken, O.; Babic, J. The Effect of Bed Rest and Hypoxic Environment on Postural Balance and Trunk Automatic (Re)Actions in Young Healthy Males. *Front. Physiol.* **2018**, *9*, 27. [CrossRef] [PubMed]

43. Ciuha, U.; Kounalakis, S.; McDonnell, A.C.; Mekjavic, I.B. Seasonal variation of temperature regulation: Do thermoregulatory responses "spring" forward and "fall" back? *Int. J. Biometeorol.* **2020**, 1–11. [CrossRef]

44. Gorzelak, M.A.; Gill, S.K.; Tasnim, N.; Ahmadi-Vand, Z.; Jay, M.; Gibson, D.L. Methods for Improving Human Gut Microbiome Data by Reducing Variability through Sample Processing and Storage of Stool. *PLoS ONE* **2015**, *10*. [CrossRef] [PubMed]

45. Wesolowska-Andersen, A.; Bahl, M.I.; Carvalho, V.; Kristiansen, K.; Sicheritz-Pontén, T.; Gupta, R.; Licht, T.R. Choice of bacterial DNA extraction method from fecal material influences community structure as evaluated by metagenomic analysis. *Microbiome* **2014**, *2*. [CrossRef] [PubMed]

46. Nakaji, S.; Fukuda, S.; Iwane, S.; Murakami, H.; Tamura, K.; Munakata, A.; Sugawara, K. New method for the determination of fecal consistency and its optimal value in the general population. *J. Gastroenterol. Hepatol.* **2002**, *17*. [CrossRef]

47. Wenzl, H.H.; Fine, K.D.; Schiller, L.R.; Fordtran, J.S. Determinants of decreased fecal consistency in patients with diarrhea. *Gastroenterology* **1995**, *108*. [CrossRef]

48. Eherer, A.J.; Santa Ana, C.A.; Porter, J.; Fordtran, J.S. Effect of psyllium, calcium polycarbophil, and wheat bran on secretory diarrhea induced by phenolphthalein. *Gastroenterology* **1993**, *104*. [CrossRef]

49. Gao, M.; Feng, L.; Jiang, T. Browning inhibition and quality preservation of button mushroom (Agaricus bisporus) by essential oils fumigation treatment. *Food Chem.* **2014**, *149*. [CrossRef] [PubMed]

50. de Lavergne, M.D.; van Delft, M.; van de Velde, F.; van Boekel, M.; Stieger, M. Dynamic texture perception and oral processing of semi-solid food gels: Part 1: Comparison between QDA, progressive profiling and TDS. *Food Hydrocoll.* **2015**, *43*, 207–217. [CrossRef]

51. Datta, S.S.; Preska Steinberg, A.; Ismagilov, R.F. Polymers in the gut compress the colonic mucus hydrogel. *Proc. Natl. Acad. Sci. USA* **2016**, *113*. [CrossRef]

52. Gacesa, R.; Kurilshikov, A.; Vich Vila, A.; Sinha, T.; Klaassen, M.A.Y.; Bolte, L.A.; Andreu-Sánchez, S.; Chen, L.; Collij, V.; Hu, S.; et al. The Dutch Microbiome Project defines factors that shape the healthy gut microbiome. *BioRxiv* **2020**. [CrossRef]

53. Salosensaari, A.; Laitinen, V.; Havulinna, A.; Meric, G.; Cheng, S.; Perola, M.; Valsta, L.; Alfthan, G.; Inouye, M.; Watrous, J.D.; et al. Taxonomic Signatures of Long-Term Mortality Risk in Human Gut Microbiota. *MedRxiv* **2020**. [CrossRef]

54. Tsamardinos, I.; Charonyktakis, P.; Lakiotaki, K.; Borboudakis, G.; Zenklusen, J.C.; Juhl, H.; Chatzaki, E.; Lagani, V. Just Add Data: Automated Predictive Modeling and BioSignature Discovery. *BioRxiv* **2020**. [CrossRef]

55. Martínez, Y.; Li, X.; Liu, G.; Bin, P.; Yan, W.; Más, D.; Valdivié, M.; Hu, C.A.; Ren, W.; Yin, Y. The role of methionine on metabolism, oxidative stress, and diseases. *Amino Acids* **2017**, *49*. [CrossRef]

73

56. Müller, M.; Hernández, M.A.G.; Goossens, G.H.; Reijnders, D.; Holst, J.J.; Jocken, J.W.E.; van Eijk, H.; Canfora, E.E.; Blaak, E.E. Circulating but not faecal short-chain fatty acids are related to insulin sensitivity, lipolysis and GLP-1 concentrations in humans. *Sci. Rep.* **2019**, *9*. [CrossRef] [PubMed]
57. Contijoch, E.J.; Britton, G.J.; Yang, C.; Mogno, I.; Li, Z.H.; Ng, R.; Llewellyn, S.R.; Hira, S.; Johnson, C.; Rabinowitz, K.M.; et al. Gut microbiota density influences host physiology and is shaped by host and microbial factors. *Elife* **2019**, *8*. [CrossRef] [PubMed]
58. Beckonert, O.; Keun, H.C.; Ebbels, T.M.; Bundy, J.; Holmes, E.; Lindon, J.C.; Nicholson, J.K. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat. Protoc.* **2007**, *2*, 2692–2703. [CrossRef]
59. Wishart, D.S.; Jewison, T.; Guo, A.C.; Wilson, M.; Knox, C.; Liu, Y.F.; Djoumbou, Y.; Mandal, R.; Aziat, F.; Dong, E.; et al. HMDB 3.0-The Human Metabolome Database in 2013. *Nucleic Acids Res.* **2013**, *41*, D801–D807. [CrossRef]
60. Mustafa, A.; Rahimi Azghadi, M. Automated Machine Learning for Healthcare and Clinical Notes Analysis. *Computers* **2021**, *10*, 24. [CrossRef]

2.1.5 Systems view of deconditioning during spaceflight simulation in the PlanHab project: the departure of urine ^{1}H-NMR metabolomes from healthy state in young males subjected to bedrest inactivity and hypoxia

Šket R., **Deutsch L.**, Prevoršek Z., Mekjavić I.B., Plavec J., Rittweger, J., Debevec T., Eiken O., Stres B. 2020. Deutsch L., Stres B. 2021. Systems view of deconditioning during spaceflight simulation in the PlanHab Project: The departure of urine ^{1}H-NMR metabolomes from healthy state in young males subjected to bedrest inactivity and hypoxia. Frontiers in Physiology, 11: 532271, doi. https://doi.org/10.3389/fphys.2020.532271, 15 p.

Abstract

We explored the metabolic makeup of urine in prescreened healthy male participants within the PlanHab experiment. The run-in (5 day) and the following three 21-day interventions [normoxic bedrest (NBR), hypoxic bedrest (HBR), and hypoxic ambulation (HAmb)] were executed in a crossover manner within a controlled laboratory setup (medical oversight, fluid and dietary intakes, microbial bioburden, circadian rhythm, and oxygen level). The inspired O2 (FiO$_2$) fraction next to inspired O2 (PiO$_2$) partial pressure were 0.209 and 133.1 ± 0.3 mmHg for the NBR variant in contrast to 0.141 ± 0.004 and 90.0 ± 0.4 mmHg (approx. 4,000 m of simulated altitude) for HBR and HAmb interventions, respectively. ^{1}H-NMR metabolomes were processed using standard quantitative approaches. A consensus of ensemble of multivariate analyses showed that the metabolic makeup at the start of the experiment and at HAmb endpoint differed significantly from the NBR and HBR endpoints. Inactivity alone or combined with hypoxia resulted in a significant reduction of metabolic diversity and increasing number of affected metabolic pathways. Sliding window analysis (3 + 1) unraveled that metabolic changes in the NBR lagged behind those observed in the HBR. These results show that the negative effects of cessation of activity on systemic metabolism are further aggravated by additional hypoxia. The PlanHab HAmb variant that enabled ambulation, maintained vertical posture, and controlled but limited activity levels apparently prevented the development of negative physiological symptoms such as insulin resistance, low-level systemic inflammation, constipation, and depression. This indicates that exercise apparently prevented the negative spiral between the host's metabolism, intestinal environment, microbiome physiology, and proinflammatory immune activities in the host.

Systems View of Deconditioning During Spaceflight Simulation in the PlanHab Project: The Departure of Urine ^{1}H-NMR Metabolomes From Healthy State in Young Males Subjected to Bedrest Inactivity and Hypoxia

Robert Šket[1], Leon Deutsch[1], Zala Prevoršek[1], Igor B. Mekjavić[2], Janez Plavec[3], Joern Rittweger[4], Tadej Debevec[2,5], Ola Eiken[6] and Blaz Stres[1,2,7,8,9*]

[1] Department of Animal Science, Biotechnical Faculty, University of Ljubljana, Ljubljana, Slovenia, [2] Department of Automation, Biocybernetics and Robotics, Jožef Stefan Institute, Ljubljana, Slovenia, [3] National Institute of Chemistry, NMR Center, Ljubljana, Slovenia, [4] German Aerospace Center, Institute of Aerospace Medicine, Muscle and Bone Metabolism, Köln, Germany, [5] Faculty of Sports, University of Ljubljana, Ljubljana, Slovenia, [6] Department of Environmental Physiology, Swedish Aerospace Physiology Centre, KTH Royal Institute of Technology, Stockholm, Sweden, [7] Faculty of Civil and Geodetic Engineering, Institute of Sanitary Engineering, University of Ljubljana, Ljubljana, Slovenia, [8] Laboratory for Clinical Toxicology, Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia, [9] Department of Microbiology, University of Innsbruck, Innsbruck, Austria

We explored the metabolic makeup of urine in prescreened healthy male participants within the PlanHab experiment. The run-in (5 day) and the following three 21-day interventions [normoxic bedrest (NBR), hypoxic bedrest (HBR), and hypoxic ambulation (HAmb)] were executed in a crossover manner within a controlled laboratory setup (medical oversight, fluid and dietary intakes, microbial bioburden, circadian rhythm, and oxygen level). The inspired O_2 (F_iO_2) fraction next to inspired O_2 (P_iO_2) partial pressure were 0.209 and 133.1 ± 0.3 mmHg for the NBR variant in contrast to 0.141 ± 0.004 and 90.0 ± 0.4 mmHg (approx. 4,000 m of simulated altitude) for HBR and HAmb interventions, respectively. ^{1}H-NMR metabolomes were processed using standard quantitative approaches. A consensus of ensemble of multivariate analyses showed that the metabolic makeup at the start of the experiment and at HAmb endpoint differed significantly from the NBR and HBR endpoints. Inactivity alone or combined with hypoxia resulted in a significant reduction of metabolic diversity and increasing number of affected metabolic pathways. Sliding window analysis (3 + 1) unraveled that metabolic changes in the NBR lagged behind those observed in the HBR. These results show that the negative effects of cessation of activity on systemic metabolism are further aggravated by additional hypoxia. The PlanHab HAmb variant

76

that enabled ambulation, maintained vertical posture, and controlled but limited activity levels apparently prevented the development of negative physiological symptoms such as insulin resistance, low-level systemic inflammation, constipation, and depression. This indicates that exercise apparently prevented the negative spiral between the host's metabolism, intestinal environment, microbiome physiology, and proinflammatory immune activities in the host.

INTRODUCTION

Metabolomics has developed into a technology-driven discipline enabling improved data collection, analysis, and interpretation. In particular, ^{1}H-NMR spectroscopy has received significant attention since it is non-destructive, non-biased, quantitative, and at the same time requires no sample derivatization (Emwas et al., 2019), is reproducible, quantitative, and enables identification of unknown novel compounds routinely in complex biological systems, such as human body or built environments (Murovec et al., 2018; Sket et al., 2018; Emwas et al., 2019).

The PlanHab project encompasses the two faceted nature of spaceflight, where human physiological responses are coupled to microbial responses to inactivity on one side and 21-day (prolonged) confination within built environment, similar to hospital settings, on the other (Debevec et al., 2014a; Simpson et al., 2016). The combined effects of 21-day inactivity/unloading and hypoxia were investigated in a controlled manner (crossover design) using medically prescreened cohort of healthy male volunteers. The experiment was executed adopting the European Space Agency (ESA) and NASA core bedrest data collection SOP (Standardization of bedrest study conditions 1.5, August 2009) controlling a number of parameters such as atmospheric oxygen content, levels of exercise (immobilization), daily water and nutritional intake, circadian rhythm, and microbial ambiental and aerosol bioburden next to the 24/7 medical surveillance (Debevec et al., 2014a; Simpson et al., 2016). In this study, the PlanHab repertoire of exploration was extended by analyses of urine ^{1}H-NMR metabolomes during the run-in (5 day) and three consecutive experimental phases [21-day normoxic bedrest (NBR), hypoxic bedrest (HBR), and hypoxic ambulation (HAmb)] in healthy male test participants. Bedrest approach in experiments is widely adopted to simulate the effects of microgravity on various physiological systems of human body, especially for studies of bone, muscle, and the cardiovascular system by NASA, ESA, and Roscosmos (Hargens and Vico, 2016; Sundblad et al., 2016). On the other hand, physical inactivity in general has emerged as the fourth leading behavioral risk factor for worldwide mortality (Kelly et al., 2020). Risk of over 20 chronic conditions [e.g., coronary heart disease, stroke, type 2 diabetes, some cancers, obesity, mental health problems (e.g., depression), and neurological conditions (e.g., dementia)] is increased by physical inactivity making lack of exercise a global health problem (Kelly et al., 2020).

The past findings obtained within the PlanHab platform showed that a number of negative physiological symptoms related to obesity and metabolic syndrome developed in a dose-dependent manner over the course of 21-day experimental period in the HBR and NBR but were absent from the HAmb variant (Debevec et al., 2014b, 2016b; Rittweger et al., 2016; Simpson et al., 2016; Stavrou et al., 2016; Sket et al., 2017a,b; Strewe et al., 2017). In addition, the observed negative physiological symptoms faded effectively in 14, 10, and <4 days for HBR, NBR, and HAmb, respectively (Debevec et al., 2014b; Sket et al., 2017a,b). Also, many of the microbial parameters such as butyrate producing microbial community, the general bacterial and archaeal microbial communities were shown to respond to modifications in human intestinal environment but lagged behind the changes in human physiology and intestinal environment (Sket et al., 2017a,b). These findings suggested a time-dependent and complex interplay between the host physiology (including apparent constipation), immunity (inflammation), controlled diet, intestinal environment variables, and microbiome physiology in absence of exercise. The analyses of microbiome and associated environmental parameters suggested that the onset of inactivity gave rise to progressive shifts in intestinal environment boiling down to modified microbial metabolic activity and increased metabolism toward degradation of host mucus layer in bedrest variants (HBR, NBR) (Sket et al., 2017b). On the other hand, in the absence of such changes the healthy HAmb variant was coupled to the production of beneficial indole derivatives (Sket et al., 2017b). Further metagenomic analyses within the PlanHab platform (Sket et al., 2018) confirmed that inactivity and hypoxia resulted in a significant increase in the relative abundance of genus *Bacteroides* in HBR next to *Bacteroides* cell wall, capsule, virulence, defense, and mucin degradation genes [beta-galactosidase (EC3.2.1.23), α-L-fucosidase (EC3.2.1.51), Sialidase (EC3.2.1.18), and α-N-acetylglucosaminidase (EC3.2.1.50)] and genes coding for iron acquisition and metabolism proteins (Sket et al., 2018). In contrast, the corresponding microbial fecal metabolomes, intestinal chemical and metal profiles, and the diversity of bacterial, archaeal, and fungal microbial communities were not significantly affected within the timeframe using the experimental set-up of the PlanHab project (Sket et al., 2018). The fact that the genus *Bacteroides* and proteins involved in iron acquisition and metabolism, cell wall, capsule, virulence, and mucin degradation were also enriched at the end of HBR revealed that significantly increased constipation and electrical conductivity led to decreased intestinal metal availability that consequently affected the expression of codependent and coregulated genes in *Bacteroides* genomes. Data integration utilizing Bayesian

network analysis resulted in the establishment of the first hierarchical model describing the onset of inactivity-mediated deconditioning over time (Sket et al., 2018).

The PlanHab wash-out period corresponded to a reintroduction of exercise, vertical position, and posture maintenance that resulted in stepwise amelioration of the negative physiological symptoms, indicating that physical activity as such introduced changes into the crosstalk between the host physiology, microbial physiology, mucin degradation, and proinflammatory immune activities within the host (Sket et al., 2017a,b, 2018). This observation was based on the fact that the observed progressive decrease in some of the parameters (e.g., defecation frequency, intestinal indole content) and concomitant increase in other (e.g., intestinal electrical conductivity, inflammatory markers) preceded or took place in absence of significant changes at the levels of microbial taxonomy, the corresponding functional genes, intestinal metabolomes, and accompanying metal profiles (Sket et al., 2017a,b, 2018).

Metabolic signal can be divided into three categories, human, microbial, and human-microbial cometabolites (Dumas et al., 2017; Wilmanski et al., 2019) and can represent a significant portion of dissolved organic matter in blood and urine. Hence, the selection of metabolomics layer for in-depth analysis of the PlanHab project-derived urine samples thus represents a logical continuation of efforts to discern and improve our understanding of the timing and the consequences of 21-day inactivity and hypoxia on human pathophysiology.

As there is a lack of data and understanding on the progressive changes in human metabolic responses coupled to microbial metabolites in the absence of exercise, we hypothesized that reduction in physical activity (complete inactivity) would (i) result in structured and significant changes in urine metabolomes of healthy participants; (ii) enable identification of significant groupings of experimental variants; (iii) provide discriminant metabolites between observed physiological states; (iv) enable the construction of metabolic network of co-occurring metabolites; (v) provide insight into the time-dependent changes in metabolomes; and finally (vi) enlighten the significantly different metabolic pathways between the experimental variants and also relative to the healthy initial state. In addition, the systemic hypoxia due to inactivity (HBR) versus ambulation in hypoxia (HAmb) was predicted to be an additional important factor aggravating the observed physiological changes within the 21-day PlanHab execution, unraveling the difference due to retained physical activity levels, hydrostatic pressures, and posture-related muscle activity in HAmb (Debevec et al., 2014a; Miles-Chan and Dulloo, 2017; Sket et al., 2017a,b).

METHODS

Experimental Setup
Experimental setup, registration, approval, recruitment, medical prescreening, acquisition of clinical data and supervision, and hypoxic facility next to the detailed outline of the PlanHab study were prepared and conducted according to the European Space

Agency's standardization plan for bedrest studies (ESA, 2009), including sample size calculation and were extensively detailed before (Debevec et al., 2014a, 2016a; Rittweger et al., 2016; Simpson et al., 2016; Sket et al., 2017a,b, 2018; Stavrou et al., 2016; Strewe et al., 2017). In short, for this study, each healthy male, participant, characterized by numerous clinically relevant measurements to assert absence of disease with a state of physical, mental, and social welfare, underwent 5 days of baseline data collection during which participants were ambulant, 21 intervention days and 5–14 days of medical follow-up. The participants underwent the following three protocols: (1) normobaric NBR (fraction of inspired O_2 (F_iO_2) = 0.209; partial pressure of inspired O_2 (P_iO_2) = 133.1 ± 0.3 mmHg); (2) normobaric hypoxic ambulatory confinement (HAmb; F_iO_2 = 0.141 ± 0.004; P_iO_2 = 90.0 ± 0.4 mmHg; ~4,000 m simulated altitude); and (3) normobaric HBR (F_iO_2 = 0.141 ± 0.004; P_iO_2 = 90.0 ± 0.4 mmHg; ~4,000 m simulated altitude). Altogether, 11 healthy men underwent all three campaigns in randomized crossover design of PlanHab project. Subjects were enrolled by project manager and randomly allocated between campaigns using Latin square design method. Sample size was determined based on previous reports on bedrest studies to obtain sufficient predictive power ≥ 0.80 (Traon et al., 2007; Angerer et al., 2014; Debevec et al., 2014a, 2016a,b; Simpson et al., 2016; Sundblad et al., 2016). For detailed experimental protocols, please see Debevec et al. (2014a); Sket et al. (2017b). In essence, the combined effects of 21-day complete inactivity and hypoxia on healthy participants were examined within the PlanHab study utilizing 11 healthy medically prescreened participants in the crossover design under strictly controlled conditions according to ESA/NASA core bedrest data collection SOP in order to determine significant differences between samples and experimental variants relative to healthy baseline data collection.

The PlanHab Project Acquisition of Clinical, Exercise, Dietary, and Ambiental Data
Acquisition of clinical, exercise, dietary, and ambiental data were described in detail before (Debevec et al., 2014a, 2016a; Simpson et al., 2016). The in-house database (Sket et al., 2017b) containing over 13,000 entries based on all measured variables in the PlanHab experiment (i.e., clinical, inflammation, immune, human physiology, and nutrition data next to the experimental design and characteristics of the participants) was checked for consistency and updated with recent publications related to the PlanHab project (Debevec et al., 2014a, 2016b, 2018; Keramidas et al., 2016; Louwies et al., 2016; Rittweger et al., 2016; Rullman et al., 2016, 2018; Simpson et al., 2016; Morrison et al., 2017; Strewe et al., 2017; Salvadego et al., 2018; Sarabon et al., 2018; Stavrou et al., 2018a,b; Ciuha et al., 2020). The in-house database was used to identify parameters that differed significantly between the experimental variants over the course of the experiment as described before (Sket et al., 2017b, 2018). This resulted in 48 parameters describing

diet, intestinal metabolites, immune, and chemical parameters next to human physiology that were significantly different between NBR, HBR, and HAmb variants ($p < 0.05$; corrected for multiple comparisons). These served as the basis for the linking of observed body deconditioning to urine metabolites observed in this study.

Participants

After initial prescreening according to NASA and ESA guidelines for bedrest studies, the data of 11 participants that finished all three interventions were included in our analysis with the following baseline characteristics (mean ± SD): age = 27 ± 6 years; body mass = 76.7 ± 11.8 kg; stature = 179 ± 3 cm; BMI = 23.7 ± 3.0 kg m^{-2}; body fat = 21 ± 5%; maximal oxygen uptake = 44.3 ± 6.1 ml kg^{-1} min^{-1} (Debevec et al., 2014a; Šket et al., 2017a).

Sample Collection

Urine samples were collected aseptically on a daily basis in the early morning during the 5 days of run-in period and 21 days of intervention periods. In total, 523 samples were obtained, aliquoted, and frozen at $-20°C$ for further analyses.

Urine Metabolome Analysis Using Proton Nuclear Magnetic Resonance

Urine samples (600 µl) were centrifuged at 10,000 × g for 30 min at 4°C to remove fine particles. Samples were filtered through 0.22 µm HPLC-compatible filters (Millipore, Germany), 400 µl aliquots were mixed with 200 µl ^{1}H-NMR buffer as described before (Beckonert et al., 2007) and stored at $-20°C$ until analysis. Before analysis, samples were thawed at room temperature and centrifuged at 12,000 × g for 5 min at 4°C; 550 µl of each sample was transferred into 5 mm NMR tube as described before (Murovec et al., 2018).

Proton nuclear magnetic resonance (^{1}H-NMR) spectra were acquired on an Agilent Technologies DD2 600 MHz NMR spectrometer equipped with 5 mm HCN Cold probe. 2D experiments were measured on Agilent Technologies (Varian) VNMRS 800 MHz NMR spectrometer equipped with 5 mm HCN Cold probe. All experiments were measured at 25°C. ^{1}H-NMR spectra of the samples were recorded with spectral width of 9.0 kHz, relaxation delay 2.0 s, 32 scans, and 32 K data points. Water signal was suppressed using double-pulsed field gradient spin-echo (DPFGSE) pulse sequence. Heteronuclear single quantum coherence spectrum (HSQC) for ^{1}H- and ^{13}C-dimensions (2D NMR) was acquired with spectral widths of 9.0 and 40 kHz for ^{1}H- and ^{13}C-dimensions, respectively, and 1,536 complex points for ^{1}H-dimension, relaxation delay 1.5 s, 160 number of transients, and 128 time increments. Total correlated spectrum (TOCSY) was measured with ^{1}H spectral widths of 7.0 kHz, 4,096 complex points, relaxation delay 1.5 s, 32 number of transients, and 144 time increments. The ^{1}H and 2D spectra were apodized with an exponential function and a cosine-squared function, respectively, and zero filled before Fourier transform. NMR spectra were processed and analyzed using VNMRJ (Agilent/Varian) and Sparky (UCSF) software and MestReNova.

The resulting spectra were consequently analyzed in two complementary ways: (i) human expert chemometric untargeted metabolomics, including 2D spectra, and (ii) targeted quantitative metabolomics using Chenomx NMR Suite version 8.3 (Chenomx, Inc.) For the latter, all spectra were randomly ordered for spectral fitting using ChenomX profiler. Metabolites were thus identified with the support of Chenomx Compound Library extended by Human Metabolome Data Base (Wishart et al., 2009; Markley et al., 2017), giving access to chemical shift profiles of 674 compounds used in analyses. The number of database derived chemical shift profiles of metabolites used in analyses was further decreased by the procedures described below.

Bioinformatic and Statistical Analysis of Urine Metabolomes

Two different approaches to asymmetric sparse matrix data analysis were adopted (Legendre and Legendre, 2012), as each compound concentration was (i) normalized by dividing the measured concentration into the total concentration of all metabolites in that sample and (ii) by Box-Cox or log2 transformation (Šket et al., 2018). The metabolites that were present in less than 5% of the samples (i.e., < the size of the smallest experimental group of samples in analysis) were excluded from further analysis.

The significance of difference in the metabolic characteristics of various groups of samples was tested using ANOSIM, NP-MANOVA, expressed as an overlap in non-metric multidimensional scaling (nm-MDS) trait space using Gower and Euclidean distance measures, and finally the dimensionality reduction selected through stress function and inspection of Shepard's plots of correspondence between target and obtained ranks. To analyze the relationship between starting and endpoints of each variant, and also between the endpoints of particular variants, a number of established approaches were used: weighted UniFrac, uweighted UniFrac, analysis of molecular variance (AMOVA), HOMOVA, LEfSe, indicator species, and Metastats tests with 999 permutations were used as implemented in mothur (Schloss et al., 2009). Multiple-group comparisons were performed using Benjamini-Hochberg false discovery rate (FDR). Multiple test correction (Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2001), was used as described before (Šket et al., 2017a,b, 2018).

Associations between urine metabolites were calculated using non-linear Spearman correlation as implemented in mothur (Schloss et al., 2009), and significant interactions ($p < 0.005$) were used for further network analysis. Software Cytoscape (Shannon et al., 2003) was used to create interaction networks between the significantly different groups of metabolomes identified in the previous section, giving thus rise to two groups: (i) the beginning of the experiment and endpoint of HAmb on one side and (ii) the endpoints of experimental variants NBR and HBR at the other. Network characteristics were described using parameters, e.g., clustering coefficient, number of nodes and edges, and network density next to centrality measures such as betweenness and closeness (Shannon et al., 2003).

Furthermore, a complementary analysis using a completely distinct analytical approach utilizing a dedicated MetaboAnalyst tool (Xia et al., 2009) was adopted. The supervised classification using random forest method and pathway mapping were utilized where measured metabolites were compared with human metabolome database for identification of the affected metabolic pathways (Wishart et al., 2007). Pathway enrichment analysis was performed using global ANCOVA and topology analysis using relative-betweeness centrality in MetaboAnalyst (Goeman et al., 2004; Chong et al., 2019).

The final type of analysis introduced a sliding window analysis of the relationships between the recorded metabolic profiles. Metabolomes belonging to a particular day over the run-in and experimental phase were binned together using window size of 3 days and the increment step size of 1 day. For each window, the urine metabolites and their distribution between samples were used to calculate the mean values of 3 days span for all three experimental variants (HBR, NBR, HAmb). Furthermore, the metabolic windows of 3 days calculated for different experimental variants were compared with the first 3 days of baseline data collection using permutational multivariate analysis of variance (PERMANOVA) tests with 9,999 permutations to assess the significance of differences between multiple-group comparisons and elucidate the possible trends in changes of significance within each and between different windows.

RESULTS

The Extent of Body Deconditioning in the PlanHab Project

The in-house PlanHab database reported before (Šket et al., 2017b) enabled us to incorporate novel recently reported parameters within the PlanHab project (**Supplementary Table 1**) and identify 48 variables from other substudies within the PlanHab project that differed significantly between the experimental variants describing the clinical, inflammation, immune, human physiology, and nutrition characteristics of the participants (**Figure 1**). The results show clear separation between the HAmb variant and the inactive HBR and NBR variants. In addition, the variables were clearly separated into two broad response clusters with a number of variable subtypes, showing the complexity of the developed physiological and nutritional responses. The healthy levels of measured variables were retained for the major part of the measured variables in HAmb and hence constitute the least-affected phenotype, whereas those observed for HBR and NBR were classified as characteristic of insulin resistance (type 2 diabetes), low-level systemic inflammation, constipation, depression, symptoms related to metabolic syndrome, obesity, and body deconditioning due to inactivity. A number of specific changes can be observed in human physiology in response to either hypoxia or inactivity under hypoxia that are beyond the scope of this work and were already described in details within the PlanHab project publications (Šket et al., 2017b, 2018; **Supplementary Table 1**).

Variations in Measured Urine Metabolites Between the Experimental Branches

Multiple comparisons using AMOVA test indicated significant shifts in metabolites between baseline data collection and endpoints of experimental variants ($p < 0.01$). Individually tested correlations between experimental variants showed that metabolites detected in HBR and NBR campaigns differed significantly from HAmb and baseline data collections (**Figure 2A**). As the metabolites detected in baseline data collection and HAmb group were not significantly different, these two groups represented rather healthy physiological signatures, as observed before in the PlanHab literature (**Supplementary Table 1**). Multiple comparisons of the most significant metabolites according to ANOVA significance testing (**Figures 2B–C**) confirmed the joint clustering of HAmb and baseline data collections as healthy physiological signatures on one side in contrast to HBR and NBR campaigns as affected states on the other. In this respect, the joint branching of the baseline data collection of healthy participants with HAmb variant represented thus the rather healthy human physiological signatures on one side with NBR and HBR experimental variants representing severely affected participants on the other (**Figure 1**; **Supplementary Table 1**).

The metabolites most involved in separation of the two experimental branches (healthy vs. affected) listed within **Figures 2B,C** represent the classes of microbial metabolites (e.g., acetate, formate, hippurate), human-microbe cometabolites (e.g., trimethyl amine, hippurate, carnitine, acetyl carnitine, cresol, phenyl acetyl glycine), and human-derived metabolites involved in ATP synthesis (e.g., creatinine, choline, guanidinoacetate, hypoxanthine, xanthine), DNA (purine) metabolism (e.g., uric acid, xanthine, hypoxanthine), tricarboxylic acid cycle (e.g., succinate, citrate), muscle mitochondria (e.g., isoleucine), generation of reactive oxygen species (ROS; e.g., xanthine, hypoxanthine), bile acid metabolism (e.g., taurine), and others. It can be seen that numerous metabolites were associated and could be hence involved with distinct complex physiological responses detailed in **Figure 1**.

Of interest, the three collections of run-in baseline data metabolomes obtained from healthy and medically prescreened participants were not significantly different (PERMANOVA test; $p > 0.05$; FDR corrected). This shows that urinary metabolomes obtained during the run-in baseline data collection were representative of healthy normal males.

Interaction Network Analysis of Co-occurring Metabolites

Interaction network of metabolites characteristic of the healthy state showed us 177 statistically significant connected metabolites (i.e., nodes; Spearman correlation $p < 0.005$) with a total of 1,769 edges representing the co-occurrence patterns between metabolites (**Figure 3**). In contrast, the interaction network in affected participants of NBR and HBR variants showed a severe reduction of more than 30% in the number of statistically significantly connected metabolites and a 2.5 times reduced number of their interactions. This testifies that a reduction in

FIGURE 1 | Heatmap plot showing the relationship between parameters describing human physiology, psychology, and intestinal environment that differed significantly at the end of the PlanHab experiment ($n = 48$; $p < 0.05$; FDR corrected) that are now part of the new version of the in-house PlanHab database (Šket et al., 2017b) based on all measured variables within the project. The inset to the left represents the magnitude of z-normalized data.

81

FIGURE 2 | Schematic overview of the detected changes in urine metabolites. **(A)** Comparison of metabolite groups showing the start and HAmb, HBR, and NBR experimental variants using three different tests. The dotted and solid lines designate significant and non-significant differences between the groups. **(B)** Heatmap of the 50 most important urine metabolites according to AMOVA significance testing constructed using Euclidean distance measure and Ward clustering algorithm. **(C)** Graphical representation of 15 most informative metabolic features and their sample classification efficiency ranked by random forest algorithm. The insets to the right **(B,C)** represent the magnitude of z-normalized data.

physical exercise is coupled to significant reduction in metabolic diversity within human body.

Based on centrality measurements (betweenness, closeness), the most important metabolites representing the difference between healthy and affected states that were identified also using different statistical approaches (**Supplementary Table 1**) were enriched in either healthy or affected states (**Figures 2B,C**), suggesting significant shifts existed in the metabolic makeup of the human urine after introduction to inactivity within the PlanHab project and secondly very few to the project itself.

In addition, these graphical representations of metabolic co-occurrence networks clearly demonstrate the complexity of

metabolic makeup of developed metabolic states observed in the PlanHab project showing that the search for a single or a handful of biomarkers would be prohibitive and oversimplification and that a more complex approach needs to be utilized to derive important information.

Variations in Predicted Urine Metabolic Pathways

As many metabolites can be involved in different not necessarily complementary metabolic pathways, the collected metabolomics data were used to reconstruct the most important

FIGURE 3 | An overview of the complexity of the metabolic co-occurrence networks and their characteristics reported for the healthy (start; HAmb) and affected (NBR; HBR) metabolic states. The rightmost pane represents the difference between the two networks, showing the extent of lost metabolic interactions and nodes due to modifications in human physiology in response to conditions in HBR and NBR (inactivity and hypoxia).

metabolic pathways contributing to the observed differences in metabolomes. The pathways were identified based on the importance of underlying metabolites (pathway impact) and the significance of comparison between different metabolites (significance after FDR).

The metabolites involved in propanoate metabolism ($p < 0.0001$) were enriched in comparison with the start of the experiment (**Figure 4** and **Supplementary Table 2**) in all three campaigns. On the other hand, the metabolites involved in synthesis and degradation of ketone bodies with pathway impact 0.7 were enriched solely in HAmb experimental variant ($p < 0.0001$) considering FDR but not in HBR and NBR variants.

Most enriched pathway in the most affected variant of the PlanHab project, the HBR campaign, were, e.g., glycolysis or gluconeogenesis and furthermore the concentration of glucose 1-phosphate were lower at the end of HBR in comparison with the start of the experiment, whether on the other hand metabolite acetate was increased in both HBR and NBR campaigns (**Figures 3, 4**). Acetate was the main factor in HBR and NBR campaigns suggesting the enriched pyruvate metabolism.

Other significantly affected metabolic pathways enriched in HBR and NBR in comparison with the start of experiment were aminoacyl-tRNA biosynthesis, arginine and proline metabolism, beta-alanine metabolism, fructose and mannose metabolism, galactose metabolism, glycerophospholipid metabolism, methane metabolism, nitrogen metabolism, pantothenate and CoA biosynthesis, selenoamino acid metabolism, sulfur metabolism, taurine and hypotaurine metabolism, valine, leucine, and isoleucine biosynthesis (**Figure 4** and **Supplementary Table 2**).

Finally, an overview of the number of affected pathways suggested that the introduction of the participants into the PlanHab project significantly affected four metabolic pathways in HAmb in comparison with the starting metabolic makeup, whereas a five and eight times larger number of pathways were progressively affected in NBR ($n = 22$) and HBR ($n = 32$), respectively. This is in line with our observation that inactivity irrespective of hypoxia resulted in 30% reduction in the number of statistically significantly connected metabolites, a 2.5 times reduction in the number of interactions and that reduced physical exercise resulted in diminished metabolic diversity within human body.

The Sliding Window Time-Frame Analysis

Sliding window analysis enabled us to compare each bin of 3 days to the start of the experiment in order to identify the onset of significant changes in experimental variants (**Figure 5**) over

FIGURE 4 | Graphical representation of metabolic pathways according to their significance of change between experimental variants relative to the start of the experiment [significance (−log(p))] and importance of metabolites within a given pathway. y-axis: the p-Values (−log(p)) from pathway enrichment analysis using Global test for testing differentially expressed metabolites. Significantly changed urine metabolic pathways were based on KEGG database relative to the start of the experiment. x-axis: pathway impacts from the topology analysis using relative-betweeness centrality were used to estimate the importance of measured metabolites within a given metabolic pathway. Designation of changes in metabolic pathways relative to the start of the experiment: yellow, not significant changes; red, significant changes in NBR and HBR; blue, significant changes in all three variants. See **Supplementary Figure 1** and **Supplementary Table 2** for additional information. The size of the circles corresponds to pathway impact (x-axis) for simplicity.

time. The changes in metabolic makeup in both bedrest variants (HBR, NBR) deviated progressively away from the initial status until significant changes were detected by the end of the second week of the experiments. Significant changes in human urine metabolome were observed by the end of the first week in HBR, whereas the apparent delay of significant changes in NBR in comparison with HBR lasted till the day 12, and the difference can be attributed to the lower levels of oxygen in HBR. It is interesting to note that the pattern of metabolome deviation of HAmb variant from its original state actually followed an acclimation pattern. The initial effects of hypoxia were thus ameliorated in HAmb by the retained levels of exercise in this

particular variant of the PlanHab project, diurnal vertical posture maintenance activity, and hence establishment of hydrogradients within the HAmb, giving rise to overall insignificant changes in HAmb urine metabolites to the starting point during 21 days of the experiment.

DISCUSSION

The unique crossover design allowed us to include responses of the same participants to all three experimental variants, NBR, HBR, and HAmb, under the controlled nutritional,

FIGURE 5 | Analysis of significant changes in metabolic signatures over time. Sliding window analysis ($n = 3$) was adopted to elucidate the relationships between the recorded metabolic profiles. x-axis: time of metabolic signature. For each signature, metabolomes were binned together using window size of 3 days and the increment step size of plus 1 day. For each window, urine metabolites and their distribution between samples were used to calculate the mean values of 3 days span for all three experimental variants (HBR, NBR, HAmb) and compared with the first 3 days of baseline data collection. y-axis: the significance of differences between different metabolic windows over time ($p < 0.05$). Multivariate non-parametric test PERMANOVA with 9,999 permutations was used to assess the significance of differences between multiple-group comparisons and to elucidate the possible trends in changes of significance within each and also between different windows.

environmental, and experimental conditions. The same general trends of body deconditioning were recovered in this study based on ^{1}H-NMR metabolomics of urine, as described before using different sets of markers and approaches in the PlanHab subprojects (Debevec et al., 2014a, 2016b; Rittweger et al., 2016; Simpson et al., 2016; Strewe et al., 2017, 2018; Stavrou et al., 2018a,b; **Supplementary Table 1**). This shows large congruence between the various independently collected datasets within the PlanHab and the metabolomics approach used in this study. For instance, zonulin concentration in blood samples (Strewe et al., 2018) showed the same patterns as zonulin samples collected from fecal samples (Sket et al., 2017b).

In this respect, our study demonstrates that ^{1}H-NMR metabolomics coupled to standardized analytical approaches and sample preparation next to in-depth statistical analyses allows for comprehensive characterization of the physiological responses and enables the detection of subtle metabolic changes during the initial and reversible body deconditioning in response to 3-week inactivity. In comparison with HAmb, the participants involved in NBR and HBR exhibited specific and different metabolic trajectories giving rise to severely reduced metabolic diversity and hence the reduction in the number of metabolic pathways under controlled experimental and nutritional conditions

(**Figures 2, 3**). In essence, this shows a profound impact of the onset of 3-week inactivity on human physiology revealing the progressive systemic maladjustments. Finally, the Bayesian modeling in our previous work (Sket et al., 2017a,b, 2018) showed that the significant changes in human physiology in the PlanHab project preceded or took place devoid of the corresponding changes at the level of intestinal microbiome. The genus *Bacteroides* and proteins involved in iron acquisition and metabolism, cell wall, capsule, virulence, and mucin degradation were enriched solely at the end of the third week in HBR only. Apparently, constipation and electrical conductivity decreased intestinal metal availability, induced modified expression of coregulated genes in *Bacteroides* genomes (Sket et al., 2018), possibly also the zwitterionic capsular polysaccharides with anti-inflammatory properties (Neff et al., 2016).

Our findings suggest that the decision of the host to minimize physical activity under hypoxic conditions (HBR) is detectable within a few days at the level of urine metabolites using ^{1}H-NMR and by the end of the first 10 days in NBR irrespective of individual responses to food intake (Sato et al., 2018), daily composition, time of ingestion, and diurnal cycles of sleep described before (Sket et al., 2017a,b, 2018). Our results show reproducibly high flexibility of the underlying physiological

metabolic pathways in the absence of the diurnal metabolic signals from the use of skeletal muscles (Schranner et al., 2020). This is important as in the absence of the metabolic signals from the use of human skeletal muscles, the metabolomes of other body organs seem to develop primarily different metabolic changes with little similar alterations that showcase the complexity of consequences due to the lack of exercise at the organismal level (Starnes et al., 2017).

This shows that the host's metabolic and other physiological and psychological responses (Sket et al., 2017a,b, 2018; **Figure 1**) actually precede the responses of microbiome at the community structure level, but the chemical crosstalk between the two entities remains apparently responsive as based on the differences in metabolites that are known to be cometabolized by both and exchanged between the two subsystems (human and microbiome). Consequently, it is apparently the host that can be held responsible for the differences in thermodynamic niches provided to the microbes and to which microbial constituents respond. The colonic transit time was put forward as one of the most important parameters of intestinal tract related to bacterial metabolism and mucosal turnover in the gut (Roager et al., 2016), as also observed in our past studies (Sket et al., 2017a,b, 2018), and is hence a highly important factor to be considered in future metabolomics studies.

This complex crosstalk between microbiome and host's systems is influenced by innumerable environmental parameters (Rooks and Garrett, 2016), crosstalk within microbial domains (Neff et al., 2016), and human evolutionary adaptations (Murray and Montgomery, 2014). In addition, their interaction can act locally and across greater distances within the human body, with some yet undetermined temporal delays (Rooks and Garrett, 2016). However, the contribution of microbiome to metabolic conversions of exercise-induced metabolites was shown to be of significant importance (Scheiman et al., 2019) acting as natural, microbiome-encoded enzymatic processes converting muscle lactate to formate and providing it back to host. In essence, this provides support for the concept, that mammals are holobionts, dependent on microbial and host genome information for optimal performance (Rooks and Garrett, 2016; Sket et al., 2018).

The approach adopted in this study provides an opportunity to generate new hypotheses on metabolic pathway perturbation. One can indeed hypothesize that the metabolites involved in metabolic pathways identified in this study in fact act as signaling molecules [or account for lack of these (e.g., in HBR, NBR)] involved in the PlanHab symptoms as detailed in **Figure 1**: insulin resistance, low-grade inflammation, different mitochondrial function, miRNA expression in large muscles, differences in lipid oxidation, mood changes, and depression (Debevec et al., 2014a, 2016b; Rittweger et al., 2016; Simpson et al., 2016; Sket et al., 2017a,b, 2018; Strewe et al., 2017, 2018; Stavrou et al., 2018a,b; **Supplementary Table 1**). In addition to those listed above, groups of metabolites identified in this study were also associated with: (i) the chronic obstructive pulmonary disease (COPD) (Adamko et al., 2015; Ząbek et al., 2015) and included metabolites such as 3-hydroxyisovalerate, 2-hydroxyisobutyrate, creatinine, formate,

taurine, urea, choline, isoleucine, pantothenate, valine, and its degradation to beta-aminoisobutyric acid during metabolism of branched-chain amino acids suggest increased catabolism associated with COPD; (ii) cardiovascular disease as a results of associated chain of events such as tissue hypoxia (gut ischemia) due to reduced oxidative phosphorylation and energy production that lead to pulmonary hypertension, systemic inflammatory responses, and increased risk of cardiovascular disease, type 2 diabetes, depression, and osteoporosis (Jones, 2014). Phospholipids such as trimethylamine (TMA), choline, and trimethylamine-*N*-oxide (TMAO) were strongly correlated with cardiovascular disease (Senn et al., 2012); and (iii) diabetes and the metabolic syndrome where different metabolites and metabolic pathways were correlated with the onset of the disease, such as isoleucine and phenylalanine, alanine, aspartate and glutamate metabolism, glycine serine and threonine metabolism, and phenylalanine metabolism (Wang et al., 2011; Jones, 2014).

In the single study of human metabolic responses to microgravity simulated in a 45-day 6° head-down tilt bedrest (HDBR) experiment (Chen et al., 2016) utilizing ^{1}H-NMR in urine metabolomic analyses, similar changes in a limited number of biomarkers were detected (corresponding to NBR variant of our experiment), such as increased guanidinoacetate associated with enhancement of protein turnover inducing further muscle turnover, trimethylamines and taurine associated with cardiovascular diseases, and mammalian-microbial cometabolites such as acetate and hippurate, products of microbial fermentations, and dietary protein metabolism. This observation signifies congruent detection of a small number of the most informative metabolites in the two bedrest studies. However, it also shows that there is little congruency between different metabolomics studies based on the precise nature of a handful of specific metabolites to be assigned as specific biomarkers for certain disease or healthy status (Schranner et al., 2020). This is further exemplified by the incompatibilities between the methods, experimental designs, statistical approaches utilized (biomarker vs. pathways), levels of disease development, reversibility of the symptoms and conditions. However, the correspondence is markedly increased by the adoption of metabolite integration into metabolic pathways that are up- or downregulated, as shown in this study and in comparison to other studies utilizing the pathway approach where the same affected pathways have started to emerge for specific conditions (Sheedy et al., 2014; Elliott et al., 2015; Tynkkynen et al., 2019; Kelly et al., 2020).

From this it follows that no simple or single metabolic biomarker exists for delineation of particular human state (e.g., healthy vs. diseased in our experiment; trained vs. untrained; active vs. sedentary; young vs. old or any other group comparisons). In contrast, rather complex multivariate descriptions of metabolic makeup are needed to capture commonalities in human physiological states due to complex responses in human physiology, large interpersonal variability and variability over time, the fact that the same metabolites can act in different metabolic pathways and can hence act as up- or downregulated depending on the pathway.

Significant further research work will be needed to understand how the regulatory cascades of physical exercise and oxygen supply translate stimuli to various host's tissues and microbiome domains that all affect human metabolic makeup and crosstalk between the domains of holobiont. The adoption of supervised and automated analyses amenable for re-analyses once improved algorithms, databases, statistical approaches arise enable us to continuously expand and learn from the datasets at hand over time. One has to realize that long-term bedrest studies with females are significantly more challenging and hence not many studies with sufficient statistical power were reported so far to close the gap. With the concomitant methodological development, the exploration of more complex female metabolome and responses to inactivity and hypoxia can be commenced, extending our recent FemHab work on this topic (Debevec et al., 2016a). Finally, genetic and environmental parameters likely play pivotal roles and further work is needed to understand their relative contributions, how these can be managed using metabolomics as one of the most promising approaches to explore these relationships (Kelly et al., 2020).

A few limitations and concepts of this study need to be considered. First, although the sample size utilized in this study seems relatively small from the perspective of screening random populations of participants, the sample size was well within the limits of recent detailed studies adopting the bedrest format or others (David et al., 2014a,b; Thaiss et al., 2014; Chen et al., 2016). Second, the effects of supposedly limited statistical power and accompanying potential for type-II error were at least partly alleviated by the fact that the test participant population was prescreened for healthy young males according to SOP used by ESA/NASA (Thevenot et al., 2015). Third, this study was conducted according to the European Space Agency's standardization plan for bedrest studies (ESA, 2009), taking into account results of pre-experiments (Debevec et al., 2016b; Keramidas et al., 2016; Rittweger et al., 2016; Stavrou et al., 2016; Šket et al., 2017b; Strewe et al., 2017), Guidelines for Standardization of Bed Rest Studies in the Spaceflight Context (Angerer et al., 2014; Sundblad et al., 2016), and past recommendations on the sufficient sample size for measurements of the majority of routine parameters (Traon et al., 2007). Fourth, the PlanHab project was executed as crossover design experiment, hence the same participants were subjected to all experimental conditions in separate campaigns, further minimizing the overall interpersonal variability between campaigns.

In order to study metabolic deconditioning of the human body exposed to inactivity or other metabolic disorders, that may arise as a result of either acute or chronic and communicable or non-communicable diseases (**Supplementary Table 1**), the adoption of multivariate analysis of complex metabolomes in a unified framework can unravel more biologically relevant findings than search for a few biomarker metabolic or microbial species (Visconti et al., 2019; Kelly et al., 2020). In addition, ^{1}H-NMR metabolomics offers quantitative insight (Beckonert et al., 2007; Emwas et al., 2019) as it is not compositional in contrast to shotgun or amplicon metagenomics (unless deliberately transformed) (Vandeputte et al., 2017; Contijoch et al., 2019) and can be used in metabolic and computational modeling for guided decisions and health monitoring in personalized medicine approaches (Sung et al., 2016; Palumbo et al., 2018).

CONCLUSION

The PlanHab project was designed to investigate in a controlled manner the combined effects of 21-day inactivity/unloading and hypoxia on a medically prescreened cohort of healthy male volunteers in crossover design. In total, 523 urine metabolomes were analyzed and processed using standard quantitative ^{1}H-NMR approaches and ensemble of multivariate methods from three interventions: normoxic bedrest, hypoxic bedrest, and hypoxic ambulation. Results show that in contrast to hypoxic ambulation and run-in period inactivity alone or combined with hypoxia resulted in significantly reduced systemic metabolic diversity, increasing number of affected metabolic pathways, and faster metabolic deconditioning. The maintained vertical posture and controlled but limited activity in hypoxic ambulation variant prevented the development of negative physiological symptoms such as insulin resistance, low-level systemic inflammation, constipation, depression, symptoms of metabolic syndrome, and body deconditioning reported before in the PlanHab project. Metabolic and pathway diversity as a response to physical activity are apparently required to prevent the negative spiral between the host and microbiome physiology governed by intestinal environment and proinflammatory immune activities of the host. In order to study metabolic deconditioning of the human body exposed to inactivity or other metabolic disorders, the adoption of multivariate analysis of complex metabolomes in a unified framework of metabolic pathways can unravel more biologically relevant findings than a search for a few specific metabolic biomarker signatures.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by REPUBLIC OF SLOVENIA Ministry of Health National Medical Ethics Committee Štefanova 5, 1000 Ljubljana, Slovenia, http://www.kme-nmec.si/kontakt/. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

BS provided the concept for metabolome analysis and drafted the manuscript. TD and JR collected the samples. BS, RŠ, and JP designed the metabolome analyses. RŠ, BS, ZP, LD, OE, and IM conducted the research. RŠ, BS, and LD analyzed the data. RŠ

and BS provided necessary code to streamline ¹H-NMR spectra analyses and provided statistical analyses. All authors provided intellectual content at various stages of project development and manuscript preparation and approved the final version of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphys.2020.532271/full#supplementary-material

REFERENCES

Adamko, D. J., Nair, P., Mayers, I., Tsuyuki, R. T., Regush, S., and Rowe, B. H. (2015). Metabolomic profiling of asthma and chronic obstructive pulmonary disease: a pilot study differentiating diseases. *J. Allergy Clin. Immunol.* 136, 571–580.e3. doi: 10.1016/j.jaci.2015.05.022

Angerer, O., Larina, I., and Cromwell, R. (2014). *Guidelines for Standardization of Bed Rest Studies in the Spaceflight Context.* Paris: International Academy of Astronautics (IAA).

Beckonert, O., Keun, H. C., Ebbels, T. M., Bundy, J., Holmes, E., Lindon, J. C., et al. (2007). Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat. Protoc.* 2, 2692–2703. doi: 10.1038/nprot.2007.376

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Benjamini, Y., and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* 29, 1165–1188. doi: 10.1214/aos/1013699998

Chen, P., Yu, Y. B., Tan, C., Liu, H. J., Wu, F., Li, H. Y., et al. (2016). Human metabolic responses to microgravity simulated in a 45-day 6 degrees head-down tilt bed rest (HDBR) experiment. *Anal. Methods* 8, 4334–4344. doi: 10.1039/c6ay00644b

Chong, J., Wishart, D. S., and Xia, J. (2019). Using METABOANALYst 4.0 for comprehensive and integrative metabolomics data analysis. *Curr. Protoc. Bioinform.* 68:e86.

Ciuha, U., Kounalakis, S., Mcdonnell, A. C., and Mekjavic, I. B. (2020). Seasonal variation of temperature regulation: do thermoregulatory responses "spring" forward and "fall" back? *Int. J. Biometeorol.* 64, 1221–1231. doi: 10.1007/s00484-020-01898-w

Contijoch, E. J., Britton, G. J., Yang, C., Mogno, I., Li, Z. H., and Ng, R. (2019). Gut microbiota density influences host physiology and is shaped by host and microbial factors. *eLife* 8:e40553.

David, L. A., Materna, A. C., Friedman, J., Campos-Baptista, M. I., Blackburn, M. C., Perrotta, A., et al. (2014a). Host lifestyle affects human microbiota on daily timescales. *Genome Biol.* 15:R89.

David, L. A., Maurice, C. F., Carmody, R. N., Gootenberg, D. B., Button, J. E., Wolfe, B. E., et al. (2014b). Diet rapidly and reproducibly alters the human gut microbiome. *Nature* 505, 559–563. doi: 10.1038/nature12820

Debevec, T., Bali, T. C., Simpson, E. J., Macdonald, I. A., Eiken, O., and Mekjavic, I. B. (2014a). Separate and combined effects of 21-day bed rest and hypoxic confinement on body composition. *Eur. J. Appl. Physiol.* 114, 2411–2425. doi: 10.1007/s00421-014-2963-2961

Debevec, T., Mcdonnell, A. C., Macdonald, I. A., Eiken, O., and Mekjavic, I. B. (2014b). Whole body and regional body composition changes following 10-day hypoxic confinement and unloading - inactivity. *Appl. Physiol. Nutr. Metab.* 39, 386–395. doi: 10.1139/apnm-2013-0278

Debevec, T., Ganse, B., Mittag, U., Eiken, O., Mekjavic, I. B., and Rittweger, J. (2018). Hypoxia aggravates inactivity-related muscle wasting. *Front. Physiol.* 9:494. doi: 10.3389/fphys.2018.00494

Debevec, T., Pialoux, V., Ehrstrom, S., Ribon, A., Eiken, O., Mekjavic, I. B., et al. (2016a). FemHab: the effects of bed rest and hypoxia on oxidative stress in healthy women. *J. Appl. Physiol.* 120, 930–938. doi: 10.1152/japplphysiol.00919.2015

Debevec, T., Simpson, E. J., Mekjavic, I. B., Eiken, O., and Macdonald, I. A. (2016b). Effects of prolonged hypoxia and bed rest on appetite and appetite-related hormones. *Appetite* 107, 28–37. doi: 10.1016/j.appet.2016.07.005

Dumas, M. E., Rothwell, A. R., Hoyles, L., Aranias, T., Chilloux, J., Calderari, S., et al. (2017). Microbial-host co-metabolites are prodromal markers predicting phenotypic heterogeneity in behavior, obesity, and impaired glucose tolerance. *Cell Rep.* 20, 136–148. doi: 10.1016/j.celrep.2017.06.039

Elliott, P., Posma, J. M., Chan, Q., Garcia-Perez, I., Wijeyesekera, A., Bictash, M., et al. (2015). Urinary metabolic signatures of human adiposity. *Sci. Transl. Med.* 7, 262–285.

Emwas, A. H., Roy, R., Mckay, R. T., Tenori, L., Saccenti, E., Gowda, G. A. N., et al. (2019). NMR spectroscopy for metabolomics research. *Metabolites* 9:123.

ESA (2009). *Standardization of Bed Rest Study Conditions (Version 1.5) (ESTEC Contract Number 20187/06/NL/VJ)*. Paris: ESA.

Goeman, J. J., Van De Geer, S. A., De Kort, F., and Van Houwelingen, H. C. (2004). A global test for groups of genes: testing association with a clinical outcome. *Bioinformatics* 20, 93–99. doi: 10.1093/bioinformatics/btg382

Hargens, A. R., and Vico, L. (2016). Long-duration bed rest as an analog to microgravity. *J. Appl. Physiol.* 120, 891–903. doi: 10.1152/japplphysiol.00935.2015

Jones, O. A. H. (2014). *Metabolomics and Systems Biology in Human Health and Medicine*. Boston, FL: CAB International.

Kelly, R. S., Kelly, M. P., and Kelly, P. (2020). Metabolomics, physical activity, exercise and health: a review of the current evidence. *BBA Mol. Basis Dis.* 1866:165936. doi: 10.1016/j.bbadis.2020.165936

Keramidas, M. E., Kölegård, R., Mekjavic, I. B., and Eiken, O. (2016). PlanHab: hypoxia exaggerates the bed-rest-induced reduction in peak oxygen uptake during upright cycle ergometry. *Am. J. Physiol. Hear. Circ. Physiol.* 311, H453–H464. doi: 10.1152/ajpheart.00304.2016

Legendre, P., and Legendre, L. (2012). *Numerical Ecology. 3rd English edition*. Amsterdam: Elsevier Science BV.

Louwies, T., Jaki Mekjavic, P., Cox, B., Eiken, O., Mekjavic, I. B., Kounalakis, S., et al. (2016). Separate and combined effects of hypoxia and horizontal bed rest on retinal blood vessel diameters. *Invest. Ophthalmol. Vis. Sci.* 57, 4927–4932. doi: 10.1167/iovs.16-19968

Markley, J. L., Brüschweiler, R., Edison, A. S., Eghbalnia, H. R., Powers, R., Raftery, D., et al. (2017). The future of NMR-based metabolomics. *Curr. Opin. Biotechnol.* 43, 34–40. doi: 10.1016/j.copbio.2016.08.001

Miles-Chan, J. L., and Dulloo, A. G. (2017). Posture allocation revisited: breaking the sedentary threshold of energy expenditure for obesity management. *Front. Physiol.* 8:420. doi: 10.3389/fphys.2018.00420

Morrison, S. A., Mirnik, D., Korsic, S., Eiken, O., Mekjavic, I. B., and Dolenc-Groselj, L. (2017). Bed rest and hypoxic exposure affect sleep architecture and breathing stability. *Front. Physiol.* 8:410. doi: 10.3389/fphys.2017.00410

Murovec, B., Makuc, D., Kolbl Repinc, S., Prevorsek, Z., Zavec, D., Sket, R., et al. (2018). (1)H NMR metabolomics of microbial metabolites in the four MW agricultural biogas plant reactors: a case study of inhibition mirroring the acute rumen acidosis symptoms. *J. Environ. Manag.* 222, 428–435. doi: 10.1016/j.jenvman.2018.05.068

Murray, A. J., and Montgomery, H. E. (2014). How wasting is saving: weight loss at altitude might result from an evolutionary adaptation. *Bioessays* 36, 721–729. doi: 10.1002/bies.201400042

Neff, C. P., Rhodes, M. E., Arnolds, K. L., Collins, C. B., Donnelly, J., Nusbacher, N., et al. (2016). Diverse intestinal bacteria contain putative zwitterionic capsular polysaccharides with anti-inflammatory properties. *Cell Host Microb.* 20, 535–547. doi: 10.1016/j.chom.2016.09.002

Palumbo, M. C., Morettini, M., Tieri, P., Diele, F., Sacchetti, M., and Castiglione, F. (2018). Personalizing physical exercise in a computational model of fuel homeostasis. *PLoS Comput. Biol.* 14:e1006073. doi: 10.1371/journal.pcbi.1006073

Rittweger, J., Debevec, T., Frings-Meuthen, P., Lau, P., Mittag, U., Ganse, B., et al. (2016). On the combined effects of normobaric hypoxia and bed rest upon bone and mineral metabolism: results from the PlanHab study. *Bone* 91, 130–138. doi: 10.1016/j.bone.2016.07.013

Roager, H. M., Hansen, L. B. S., Bahl, M. I., Frandsen, H. L., Carvalho, V., Gobel, R. J., et al. (2016). Colonic transit time is related to bacterial metabolism and mucosal turnover in the gut. *Nat. Microbiol.* 1:16093.

Rooks, M. G., and Garrett, W. S. (2016). Gut microbiota, metabolites and host immunity. *Nat. Rev. Immunol.* 16, 341–352. doi: 10.1038/nri.2016.42

Rullman, E., Fernandez-Gonzalo, R., Mekjavic, I. B., Gustafsson, T., and Eiken, O. (2018). MEF2 as upstream regulator of the transcriptome signature in human skeletal muscle during unloading. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 315, R799–R809.

Rullman, E., Mekjavic, I. B., Fischer, H., and Eiken, O. (2016). PlanHab (Planetary Habitat Simulation): the combined and separate effects of 21 days bed rest and hypoxic confinement on human skeletal muscle miRNA expression. *Physiol. Rep.* 4:e12753. doi: 10.14814/phy2.12753

Salvadego, D., Keramidas, M. E., Kolegard, R., Brocca, L., Lazzer, S., Mavelli, I., et al. (2018). PlanHab(⁺) : hypoxia does not worsen the impairment of skeletal muscle oxidative function induced by bed rest alone. *J. Physiol.* 596, 3341–3355. doi: 10.1113/jp275605

Sarabon, N., Mekjavic, I. B., Eiken, O., and Babic, J. (2018). The effect of bed rest and hypoxic environment on postural balance and trunk automatic (Re)actions in young healthy males. *Front. Physiol.* 9:27. doi: 10.3389/fphys.2018.00027

Sato, S., Parr, E. B., Devlin, B. L., Hawley, J. A., and Sassone-Corsi, P. (2018). Human metabolomics reveal daily variations under nutritional challenges specific to serum and skeletal muscle. *Mol. Metab.* 16, 1–11. doi: 10.1016/j.molmet.2018.06.008

Scheiman, J., Luber, J. M., Chavkin, T. A., Macdonald, T., Tung, A., Pham, L. D., et al. (2019). Meta-omics analysis of elite athletes identifies a performance-enhancing microbe that functions via lactate metabolism. *Nat. Med.* 25, 1104–1109. doi: 10.1038/s41591-019-0485-4

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541. doi: 10.1128/AEM.01541-1549

Schranner, D., Kastenmuller, G., Schonfelder, M., Romisch-Margl, W., and Wackerhage, H. (2020). Metabolite concentration changes in humans after a bout of exercise: a systematic review of exercise metabolomics studies. *Sports Med. Open* 6:11.

Senn, T., Hazen, S. L., and Tang, W. H. (2012). Translating metabolomics to cardiovascular biomarkers. *Prog. Cardiovasc. Dis.* 55, 70–76. doi: 10.1016/j.pcad.2012.06.004

Shannon, P., Markiel, A., Ozier, W., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13:2498–2504. doi: 10.1101/gr.1239303.metabolite

Sheedy, J. R., Gooley, P. R., Nahid, A., Tull, D. L., Mcconville, M. J., Kukuljan, S., et al. (2014). (1)H-NMR analysis of the human urinary metabolome in response to an 18-month multi-component exercise program and calcium-vitamin-D3 supplementation in older men. *Appl. Physiol. Nutr. Metab.* 39, 1294–1304. doi: 10.1139/apnm-2014-0060

Simpson, E. J., Debevec, T., Eiken, O., Mekjavic, I. B., and Macdonald, I. A. (2016). The combined and separate effects of 16 days bed rest and normobaric hypoxic confinement on circulating lipids and indices of insulin sensitivity in healthy men. *J. Appl. Physiol.* 120, 947–955. doi: 10.1152/japplphysiol.00897.2015

Sket, R., Debevec, T., Kublik, S., Schloter, M., Schoeller, A., Murovec, B., et al. (2018). Intestinal metagenomes and metabolomes in healthy young males: inactivity and hypoxia generated negative physiological symptoms precede microbial dysbiosis. *Front. Physiol.* 9:198. doi: 10.3389/fphys.2018.00198

Sket, R., Treichel, N., Debevec, T., Mekjavic, I. B., Eiken, O., Schloter, M., et al. (2017a). Hypoxia and inactivity related physiological changes (constipation, inflammation) are not reflected at the level of gut metabolites and butyrate producing microbial community: the PlanHab study. *Front. Physiol.* 8:250. doi: 10.3389/fphys.2017.00250

Sket, R., Treichel, N., Kublik, S., Debevec, T., Eiken, O., Mekjavic, I. B., et al. (2017b). Hypoxia and inactivity related physiological changes precede or take place in absence of significant rearrangements in bacterial community structure: the PlanHab randomized pilot trial study. *PLoS One* 12:e0188556. doi: 10.1371/journal.pone.0188556

Starnes, J. W., Parry, T. L., O'neal, S. K., Bain, J. R., Muehlbauer, M. J., Honcoop, A., et al. (2017). Exercise-induced alterations in skeletal muscle, heart, liver, and serum metabolome identified by non-targeted metabolomics analysis. *Metabolites* 7:40. doi: 10.3390/metabo7030040

Stavrou, N. A., Debevec, T., Eiken, O., and Mekjavic, I. B. (2016). Hypoxia worsens affective responses and feeling of fatigue during prolonged inactivity, in Joint Life Science Meeting "Life in space for life on earth. *Paper Presented at 14th European Life Sciences Symposium & 37th Annual International Gravitational Physiology Meeting*, Toulouse.

Stavrou, N. A. M., Debevec, T., Eiken, O., and Mekjavic, I. B. (2018a). Hypoxia exacerbates negative emotional state during inactivity: the effect of 21 days hypoxic bed rest and confinement. *Front. Physiol.* 9:26. doi: 10.3389/fphys.2018.00026

Stavrou, N. A. M., Debevec, T., Eiken, O., and Mekjavic, I. B. (2018b). Hypoxia worsens affective responses and feeling of fatigue during prolonged bed rest. *Front. Psychol.* 9:362. doi: 10.3389/fphys.2018.00362

Strewe, C., Zeller, R., Feuerecker, M., Hoerl, M., Kumprej, I., Crispin, A., et al. (2017). PlanHab study: assessment of psycho-neuroendocrine function in male subjects during 21 d of normobaric hypoxia and bed rest. *Stress* 20, 131–139. doi: 10.1080/10253890.2017.1292246

Strewe, C., Zeller, R., Feuerecker, M., Hoerl, M., Matzel, S., Kumprej, I., et al. (2018). PlanHab study: consequences of combined normobaric hypoxia and bed rest on adenosine kinetics. *Sci. Rep.* 8:1762.

Sundblad, P., Orlov, O., Angerer, O., Larina, I., and Cromwell, R. (2016). Standardization of bed rest studies in the spaceflight context. *J. Appl. Physiol.* 121, 348–349. doi: 10.1152/japplphysiol.00089.2016

Sung, J., Hale, V., Merkel, A. C., Kim, P. J., and Chia, N. (2016). Metabolic modeling with big data and the gut microbiome. *Appl. Transl. Genom.* 10, 10–15. doi: 10.1016/j.atg.2016.02.001

Thaiss, C. A., Zeevi, D., Levy, M., Zilberman-Schapira, G., Suez, J., Tengeler, A. C., et al. (2014). Transkingdom control of microbiota diurnal oscillations promotes metabolic homeostasis. *Cell* 159, 514–529. doi: 10.1016/j.cell.2014.09.048

Thevenot, E. A., Roux, A., Xu, Y., Ezan, E., and Junot, C. (2015). Analysis of the human adult urinary metabolome variations with age, body mass index, and gender by implementing a comprehensive workflow for univariate and OPLS statistical analyses. *J. Proteome Res.* 14, 3322–3335. doi: 10.1021/acs.jproteome.5b00354

Traon, A. P. L., Heer, M., Narici, M. V., Rittweger, J., and Vernikos, J. (2007). From space to Earth: advances in human physiology from 20 years of bed rest studies (1986-2006). *Eur. J. Appl. Physiol.* 101, 143–194. doi: 10.1007/s00421-007-0474-z

Tynkkynen, T., Wang, Q., Ekholm, J., Anufrieva, O., Ohukainen, P., Vepsalainen, J., et al. (2019). Proof of concept for quantitative urine NMR metabolomics pipeline for large-scale epidemiology and genetics. *Int. J. Epidemiol.* 48, 978–993. doi: 10.1093/ije/dyy287

Vandeputte, D., Kathagen, G., D'hoe, K., Vieira-Silva, S., Valles-Colomer, M., Sabino, J., et al. (2017). Quantitative microbiome profiling links gut community variation to microbial load. *Nature* 551, 507–511. doi: 10.1038/nature24460

Visconti, A., Le Roy, C. I., Rosa, F., Rossi, N., Martin, T. C., Mohney, R. P., et al. (2019). Interplay between the human gut microbiome and host metabolism. *Nat. Commun.* 10:4505.

Wang, T. J., Larson, M. G., Vasan, R. S., Cheng, S., Rhee, E. P., Mccabe, E., et al. (2011). Metabolite profiles and the risk of developing diabetes. *Nat. Med.* 17, 448–453.

Wilmanski, T., Rappaport, N., Earls, J. C., Magis, A. T., Manor, O., Lovejoy, J., et al. (2019). Blood metabolome predicts gut microbiome alpha-diversity in humans. *Nat. Biotechnol.* 37, 1217–1228. doi: 10.1038/s41587-019-0233-9

Wishart, D. S., Knox, C., Guo, A. C., Eisner, R., Young, N., Gautam, B., et al. (2009). HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.* 37, 603–610. doi: 10.1093/nar/gkn810

Wishart, D. S., Tzur, D., Knox, C., Eisner, R., Guo, A. C., Young, N., et al. (2007). HMDB: the human metabolome database. *Nucleic Acids Res.* 35, 521–526. doi: 10.1093/nar/gkl923

Xia, J., Psychogios, N., Young, N., and Wishart, D. S. (2009). MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acids Res.* 37, 652–660. doi: 10.1093/nar/gkp356

Ząbek, A., Stanimirova, I., Deja, S., Barg, W., Kowal, A., Korzeniewska, A., et al. (2015). Fusion of the 1H NMR data of serum, urine and exhaled breath condensate in order to discriminate chronic obstructive pulmonary disease and obstructive sleep apnea syndrome. *Metabolomics* 11, 1563–1574. doi: 10.1007/s11306-015-0808-5

2.1.6 Exercise and interorgan communication: short-term exercise training blunts differences in consecutive daily urine ^{1}H-NMR metabolomic signatures between physically active and inactive individuals

Abstract

Physical inactivity is a worldwide health problem, an important risk for global mortality and is associated with chronic noncommunicable diseases. The aim of this study was to explore the differences in systemic urine ^{1}H-NMR metabolomes between physically active and inactive healthy young males enrolled in the X-Adapt project in response to controlled exercise (before and after the 3-day exercise testing and 10-day training protocol) in normoxic (21% O_2), normobaric (~1000 hPa) and normal-temperature (23 °C) conditions at 1 h of 50% maximal pedaling power output (W_{peak}) per day. Interrogation of the exercise database established from past X-Adapt results showed that significant multivariate differences existed in physiological traits between trained and untrained groups before and after training sessions and were mirrored in significant differences in urine pH, salinity, total dissolved solids and conductivity. Cholate, tartrate, cadaverine, lysine and N6-acetyllisine were the most important metabolites distinguishing trained and untrained groups. The relatively little effort of 1 h 50% Wpeak per day invested by the untrained effectively modified their resting urine metabolome into one indistinguishable from the trained group, which hence provides a good basis for the planning of future recommendations for health maintenance in adults, irrespective of the starting fitness value. Finally, the 3-day sessions of morning urine samples represent a good candidate biological matrix for future delineations of active and inactive lifestyles detecting differences unobservable by single-day sampling due to day-to-day variability.

For my personal contributions as a doctoral student and author of this thesis, please refer to Table 2 (page 142).

metabolites

MDPI

Article

Exercise and Interorgan Communication: Short-Term Exercise Training Blunts Differences in Consecutive Daily Urine 1H-NMR Metabolomic Signatures between Physically Active and Inactive Individuals

Leon Deutsch [1], Alexandros Sotiridis [2,3], Boštjan Murovec [4], Janez Plavec [5], Igor Mekjavic [3], Tadej Debevec [3,6] and Blaž Stres [1,3,7,*]

1 Biotechnical Faculty, Department of Animal Science, University of Ljubljana, SI-1000 Ljubljana, Slovenia; leon.deutsch@bf.uni-lj.si
2 Section of Sport Medicine and Biology of Exercise, School of Physical Education and Sport Science, National and Kapodistrian University of Athens, 17237 Athens, Greece; asotiridis@phed.uoa.gr
3 Department of Automation, Biocybernetics and Robotics, Jožef Stefan Institute, SI-1000 Ljubljana, Slovenia; igor.mekjavic@ijs.si (I.M.); tadej.debevec@fsp.uni-lj.si (T.D.)
4 Faculty of Electrical Engineering, University of Ljubljana, Jamova 2, SI-1000 Ljubljana, Slovenia; bostjan.murovec@fe.uni-lj.si
5 NMR Center, National Institute of Chemistry, SI-1000 Ljubljana, Slovenia; janez.plavec@ki.si
6 Faculty of Sports, University of Ljubljana, SI-1000 Ljubljana, Slovenia
7 Faculty of Civil and Geodetic Engineering, Institute of Sanitary Engineering, University of Ljubljana, SI-1000 Ljubljana, Slovenia
* Correspondence: blaz.stres@bf.uni-lj.si; Tel.: +386-4156-7633

check for **updates**

Abstract: Physical inactivity is a worldwide health problem, an important risk for global mortality and is associated with chronic noncommunicable diseases. The aim of this study was to explore the differences in systemic urine 1H-NMR metabolomes between physically active and inactive healthy young males enrolled in the X-Adapt project in response to controlled exercise (before and after the 3-day exercise testing and 10-day training protocol) in normoxic (21% O_2), normobaric (~1000 hPa) and normal-temperature (23 °C) conditions at 1 h of 50% maximal pedaling power output (W_{peak}) per day. Interrogation of the exercise database established from past X-Adapt results showed that significant multivariate differences existed in physiological traits between trained and untrained groups before and after training sessions and were mirrored in significant differences in urine pH, salinity, total dissolved solids and conductivity. Cholate, tartrate, cadaverine, lysine and N6-acetyllisine were the most important metabolites distinguishing trained and untrained groups. The relatively little effort of 1 h 50% W_{peak} per day invested by the untrained effectively modified their resting urine metabolome into one indistinguishable from the trained group, which hence provides a good basis for the planning of future recommendations for health maintenance in adults, irrespective of the starting fitness value. Finally, the 3-day sessions of morning urine samples represent a good candidate biological matrix for future delineations of active and inactive lifestyles detecting differences unobservable by single-day sampling due to day-to-day variability.

Keywords: exercise; trained; untrained; 1H-NMR metabolomics; human metabolome; JADBio; biomarkers

1. Introduction

Physical inactivity is a worldwide health problem ranking as the fourth most important risk for global mortality [1]. The efforts undertaken by the World Health Organization (WHO) to minimize the time spent sedentary [2] are directed at decreasing the risks for more than twenty chronic noncommunicable diseases (e.g., coronary heart disease, stroke, type 2 diabetes, obesity, metabolic syndrome, glucose insensitivity) next to mental health and neurological problems such as depression and dementia [1]. As physical activity

in the form of various types of exercise promotes wellbeing and increased quality of life, understanding the biological mechanisms through which it impacts is of central importance. Although genetics, lifestyle and environment are likely the most important parameters, their relative contributions and interactions are not well-understood.

Exercise-related stress alters the chemical steady state of the internal biochemical environment. The net result is modifications in the rate of production and consumption of various metabolites within biochemical network affecting the systemic levels of metabolites relative to the exercise intensity, muscle damage or the extent of the exercise as part of the lifelong history [3–5].

In respect to the progressively emerging picture of metabolic states characteristic of various noncommunicable diseases, a number of metabolomic studies have clearly shown that physical activity results in modifications of hundreds of metabolites associated with fatty-acid mobilization, lipolysis and metabolism, the TCA cycle, glycolysis, amino-acid metabolism, carnitine, purine and cholesterol metabolism and insulin sensitivity [1]. Based on these results, it has become obvious that while the metabolomics patterns may differ slightly between groups, it was the overall volume of exercise acting as the most important driver of the metabolomics makeup [6–9], even irrespective of hypoxia [4]. This points to the multifactorial dose–response relationship between activity (intensity, frequency, time frame (exposure measured in hours, days, weeks, years, lifelong)) and metabolomic signatures [1,4].

Metabolomics has become a technology-driven discipline focusing on improved high-throughput and large-scale data collection, analysis and interpretation. Metabolomes were characterized utilizing proton nuclear magnetic resonance (^{1}H-NMR) in minimally user-invasive biomaterial—the first morning urine. The approach of ^{1}H-NMR was utilized in this study as it is nondestructive, quantitative, cost-effective, reproducible and requires no sample derivatization [10]. Although the approach captures a modest number of metabolites ($n > 350$ in our past studies [4,5,11]) it enables identification of unknown novel compounds in complex biological matrices such as serum, saliva, urine or feces [4,10] and has been frequently utilized (>40% of studies) in observational and experimental studies next to short-term (<1 week) or long-term (>1 week) interventions [1]. The use of consecutive three-day urine ^{1}H-NMR data points was first tested recently in the form of a 3-day sliding window within the PlanHab project and showed promising results for delineation of systemic differences between groups [3–5].

The aim of this study was to explore the differences in systemic urine ^{1}H-NMR metabolomic signatures between groups of physically active and inactive individuals before and after a 10-day training protocol in normoxic (21% O_2) and normal-temperature conditions (23 °C) of the campaign number 4 within the X-Adapt: Cross-adaptation between heat and hypoxia project (Figure S1) [12]. The aim of the X-Adapt study itself was to investigate the effects of a 10-day exercise protocol on aerobic performance in young males. The X-Adapt training sessions were composed of controlled 60 min normoxic and normobaric (~1000 hPa) exercise [12] utilizing prescreened participants (graded exercise test on a cycle ergometer to determine their normoxic VO_2max and maximal power output (W_{peak}—the highest workload sustained by incremental exercise until exhaustion). In short, aerobic fitness was defined using maximal oxygen uptake (VO_2max) values (untrained VO_2max < 45 mL·kg^{-1}·min^{-1}; trained VO_2max > 55 mL·kg^{-1}·min^{-1}) [13,14]. Untrained participants were also required to not participate in organized sports, while minimal cycling and walking for commuting to work were allowed. In contrast, trained participants performed endurance-type activities (running, cycling, swimming) several times per week.

The X-Adapt urine-sample collection produced by the original project outline described before [12,15,16] (i.e.,) was augmented by including two additional urine-sampling periods, extending the project outline and resulting in the extended sample collection (Figure 1 and Figure S1). As a result of these extended urine-sampling periods there was no effect on human physiology or exercise approaches utilized in the X-Adapt project. The extended sample collection included the additional three-day baseline urine samples

before the actual start of the X-Adapt campaign 4 [12] and samples collected during the last three days of the 10-day exercise session. In total, the time span between the two sampling periods of extended sample collection contained 3 days testing, 1 day rest and 10 days exercise, amounting to almost 14 days of exercise [12]. This enabled us to capture the daily variability between the trained and untrained groups before the actual onset of the X-Adapt campaign 4 and to observe the actual systemic differences in response to the almost 14-day concerted exercise between the trained and untrained groups. In addition, this enabled us to perform additional comparisons between the various sections based on ^{1}H-NMR urine metabolomes collected uniquely over three consecutive days (Figure S1).
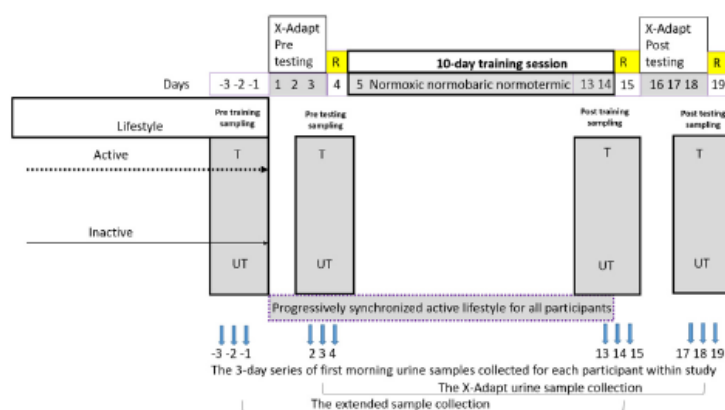
(A)
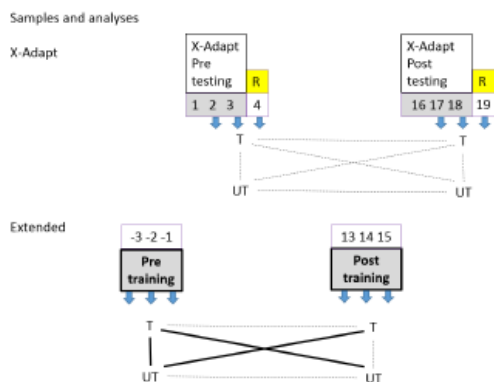
(B)

Figure 1. A schematic outline of the X-Adapt project with the two sampling collections designated below: the X-Adapt urine-sample collection and the extended sample collection (**A**). The extended sampling was conducted at days −3, −2, −1 day before the start of the campaign and at days 13, 14, 15 of the X-Adapt campaign 4. T—trained; UT—untrained group of participants. Blue arrows indicate sampling days within each of the four 3-day urine-sampling series. The X-Adapt urine-sample collection thus encompasses samples collected during the X-Adapt pretesting and post-testing periods. The extended sample collection encompasses the urine samples collected before the actual onset of the campaign (baseline) and during the last three days of training (days 13, 14, 15 of the campaign). For simplicity, the collection days are linked by hyphens to mark the compatible datasets. (**B**) A schematic representation of the X-Adapt urine-sample collection and the extended sample-collection groups with their respective analyses and comparisons delineated with lines. Solid and dashed lines designate significant and not significant differences between the groups. Analyses were conducted on overall group, sample collection and daily basis separately.

As there is a lack of data and understanding on the differences between healthy trained and untrained young males and the progressive changes in human metabolomics responses coupled to introduction of exercise, we first compiled and performed a multivariate analysis of the exercise dataset [12,15,16] and hypothesized that (i) significant differences existed in the exercise dataset between trained and untrained groups; (ii) the 2-week experimental setup would enable us to detect overall change in resting urinary metabolome 3-day sequences; (iii) significant differences existed between the trained and untrained group's urinary metabolomes despite the nonsynchronized diet of participants; (iv) the introduction of scheduled 2-week physical exercise would significantly change urine ^{1}H-NMR metabolomes in the untrained group at least; (v) discriminant metabolites could be identified between the trained and untrained groups; (vi) the extended X-Adapt experiment utilized in this study provided insight into the significantly different metabolic pathways between the trained and untrained experimental variants, signifying the importance of the training history of participants for responses in human metabolomes that were also linked to the VO$_2$max values (the maximal rate of oxygen consumption).

2. Results and Discussion

Twenty male participants 23.5 $\pm$ 2.5 years old were recruited for this study and were divided into two groups (10 participants per group) based on their physical performance (trained and untrained group). Table S3 represents their baseline characteristics. Participants in the trained group were 23 $\pm$ 2 years old, 180 $\pm$ 5 cm tall, weighed 74 $\pm$ 3 kg and had a body surface area of 1.96 $\pm$ 0.08 m^2 and body fat of 9.2 $\pm$ 2.3%. On the other side, untrained participants were 25 $\pm$ 3 years old, 179 $\pm$ 3 cm tall, weighed 85 $\pm$ 14 kg, had a body surface area of 2.05 $\pm$ 0.17 m^2 and body fat of 16.3 $\pm$ 4.9%. VO$_2$peak, W$_{peak}$ and W$_{peak}$ per kg were significantly different between the untrained and trained group. The untrained group had a lower VO$_2$peak (42 $\pm$ 5 mL/kg $\times$ x min in untrained and 58 $\pm$ 6 mL/kg $\times$ x min in trained group), lower W$_{peak}$ performance (309 $\pm$ 46 W in untrained and 364 $\pm$ 35 W in trained group) and lower W$_{peak}$ per kg (3.6 $\pm$ 0.4 W/kg in untrained and 4.9 $\pm$ 0.5 in trained group) [12,15–17].

2.1. Integrated Analysis of Exercise Data and the X-Adapt Urine-Sample Collection

In this study, exercise data reported before [17] and ^{1}H-NMR metabolomic data obtained in this study were explored. The previously reported physiological data [12,15–17] relevant for metabolomic analyses within the same 3-day series of X-Adapt pre/post-testing were analyzed. Their integrated analysis in this study showed that significant multivariate differences existed between the trained and untrained groups at pretesting (Figure 1) before the onset of the 10-day 50% W$_{peak}$ training session and after the training (PERMANOVA; F = 7.304; p(same) = 0.0001; n$_{permutations}$ = 5000). In addition, nearly significant differences (p = 0.054) existed between the pre-exercise untrained and postexercise untrained groups, suggesting a larger magnitude of changes in human exercise-related characteristics than in those leading active lifestyles.

The nonmetric multidimensional scaling (nmMDS) results also showed significant groupings separating trained from untrained (Figure 2) showing that significant differences at the level of human exercise data also remained detectable after the 10-day training period. A heatmap (Figures 3 and S2) of the measured exercise parameters shows large differences in measured parameters between trained and untrained groups, but also reflects significant interpersonal variability within each of the measured parameter. This suggests that although significant differences in the multivariate description of exercise states can be reported for the trained and untrained groups before and after the training sessions, the rate of change within the 10-day training at 50% W$_{peak}$ was significantly higher for the untrained group, as reported before [12]. This observation is further supported by detailed analyses of the exercise parameters contributing most to differences between trained and untrained groups, as VO$_2$max values in fact decreased 3.2% and increased for 9.2% in trained and untrained groups, respectively. This observation is in line with past

observations showing that the pretraining VO$_2$peak and percentage change in VO$_2$peak with training were inversely correlated, showing that the rate of adaptation is largest in less physically prepared participants [18]. In addition, the integrated exercise data reported in this study showed that trained and untrained groups responded differently, as VO$_2$max of the trained group could not be sustained by 50% W$_{peak}$ training in comparison to a further increase in the untrained group in response to 50% W$_{peak}$ training. Taken together, these results show that during the 10-day 50% W$_{peak}$ training, the trained and untrained groups were becoming more synchronized in terms of measured exercise parameters, as also suggested before [12,15–17]. A two-way PERMANOVA confirmed that participant status (trained or untrained) and 50% W$_{peak}$ training exercise (pre- or post-training) were significantly associated with the underlying multivariate exercise data ($F = 13.07$; $F = 2.57$ and p(same) = 0.0001; p(same) = 0.038), respectively), while interaction between participant status (trained or untrained) and time of training exercise (pre-or post-training) was not significant (status x exercise; $F = 0.47$; p(same) = 0.79), suggesting that the response of the two groups to the application of exercise was not uniform.
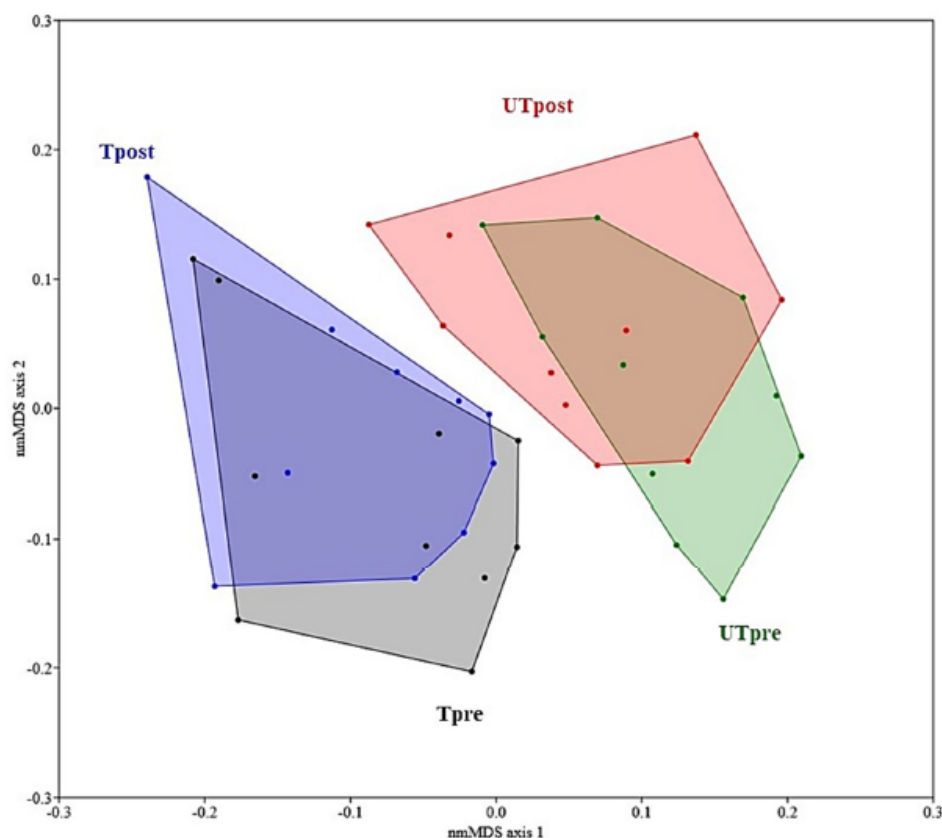
Figure 2. A nmMDS representation of physical parameters (*n* = 39) measured in X-Adapt project (UTpre (green)—untrained pre-exercise testing, UT-post (red)—untrained postexercise testing, Tpre (black)—trained pre-exercise testing, Tpost (blue)—trained postexercise testing). Stress value of nmMDS was 0.185. Please also see Figure S2 for more details.
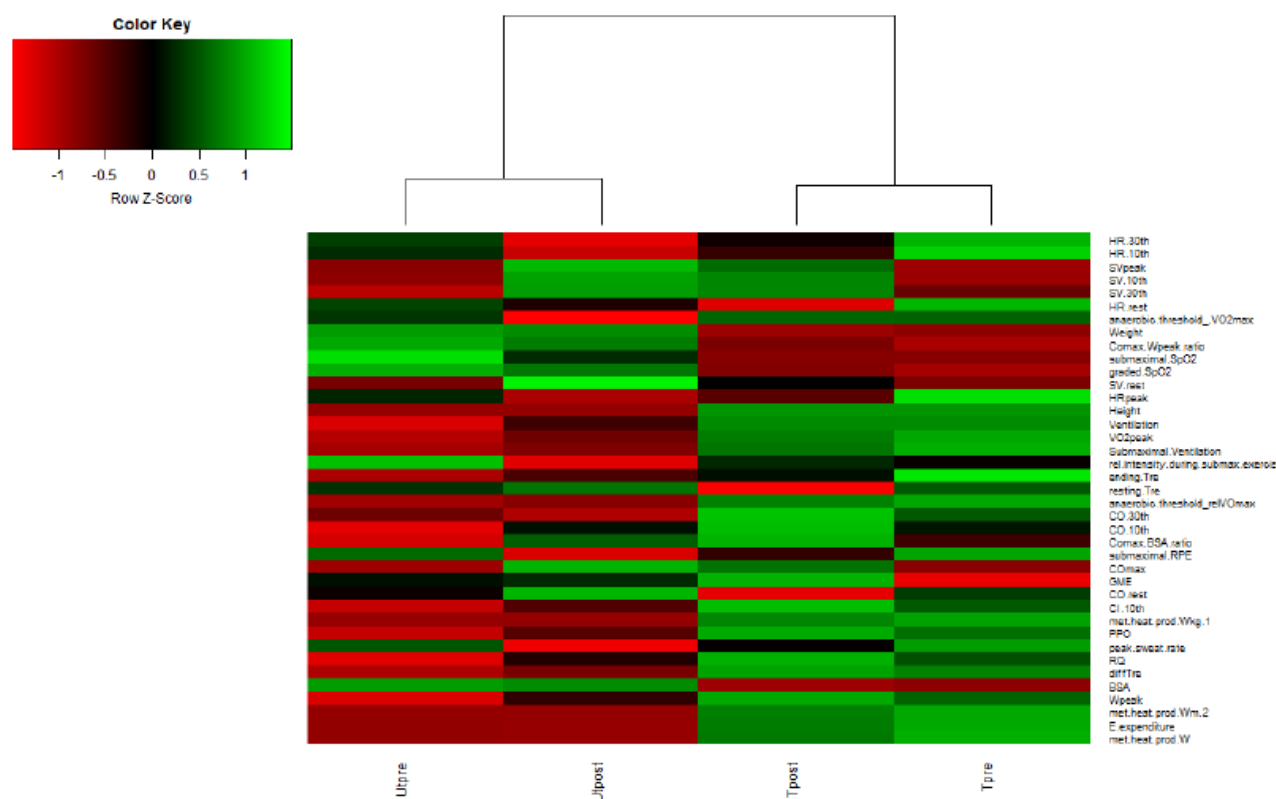
Figure 3. Heatmap representation of the underlying multivariate exercise physicochemical parameters measured in the X-Adapt project. Dendrogram clustering shows differences between the trained and untrained groups in these parameters before and after training. Higher-resolution heatmap can be found in electronic supplementary material. Abbreviations: body surface area (BSA), stroke volume (SV), hearth rate (HR) cardiac output (CO), rectal temperature (Tre), maximal power output (W_{peak}), peak power output (PPO), gross mechanical efficiency (GME), respiratory quotient (RQ), cardiac index (CI). Please see Figure S2 for the heatmap on representation per sample basis of all participants before and after exercise performance.

In contrast to results from physiological measurements, our ^{1}H-NMR analyses of the X-Adapt urine-sample collection (i.e., the 3-day morning urine samples taken within the same timeframes of X-Adapt pretraining and post-training test sessions) did not identify any significant difference between any groups (PERMANOVA; $p > 0.3$; $n_{permutations} = 5000$) (Figure 1). This is in line with past observations that metabolomes at rest (e.g., systemic morning urine samples) cannot be indicative of physical status and capacity due to their gradual return to baseline within 24 h after exercise [19]. In addition, these results point to the potentially homogenizing short-term responses in trained and untrained individuals to the standardized pre- and post-testing conducted on three consecutive days utilized in the X-Adapt study [12]. The normoxic, temperature and hypoxic tests utilized in X-Adapt were described in detail before [12,15–17]. Moreover, additional in-depth tests of statistical significance between ^{1}H-NMR metabolomes from trained and untrained groups on a day-to-day basis also did not produce significant differences (PERMANOVA; $p > 0.05$; $n_{permutations} = 5000$). These results show the lack of significant differences between the trained and untrained groups on the level of urine ^{1}H-NMR metabolomes in response to the X-Adapt pretesting and post-testing trials (Figure 1).

2.2. Differences in Urine ^{1}H-NMR Metabolomes between the Trained and Untrained Groups: The Extended Urine-Sample Collection

In order to elucidate the potentially homogenizing responses in trained and untrained individuals to the X-Adapt training regimen the extended urine-sample collection (Figure 1) was analyzed by ^{1}H-NMR. The ^{1}H-NMR fingerprints of trained and untrained groups were compared to identify the existence of internaldata-structure characteristics for the two groups of participants. The results of one-way and two-way PERMANOVA showed that significant differences existed between metabolomes of trained and untrained participants ($p < 0.01$). This was also confirmed by the two-way PERMANOVA test, showing that activity (trained/untrained) was the only parameter significantly associated with the two groups ($p = 0.0001$). Regime (pre- or post-test) and the interaction between regime and activity was insignificant ($p > 0.05$). This was also confirmed by the nonsignificant change in the number of metabolites present and the sum of their concentrations in all sampled groups (Table S1).

The Mann–Whitney test showed that significant differences in distributions between trained and untrained group existed in physical characteristics of urine such as pH, salinity, total dissolved solids (TDS) and conductivity. Salinity, conductivity and TDS were significantly higher in the untrained group than in trained, while pH was slightly more alkaline in trained (Figure S3).

Based on the nonparametric approaches described below, we used statistical methods implemented in MetaboAnalyst 5.0 [20–22]. According to the partial least-squares-discriminant analysis (PLSDA) of variable importance in the projection (VIP) scores, differences existed between the trained and untrained groups of participants at the level of cholate, tartrate, cadaverine, lysine and N6-acetyllisine (HMDB0000206) as the most distinguishing metabolites to differentiate the trained and untrained groups (Figure 4). The first three metabolites were all present at higher concentrations in the untrained group while concentrations of lysine and N6-acetylisine were higher in the trained group. Primary bile acid synthesis, glutathione metabolism, aminoacyl-tRNA biosynthesis and lysine degradation pathways were enriched in the untrained group (Figure 5).
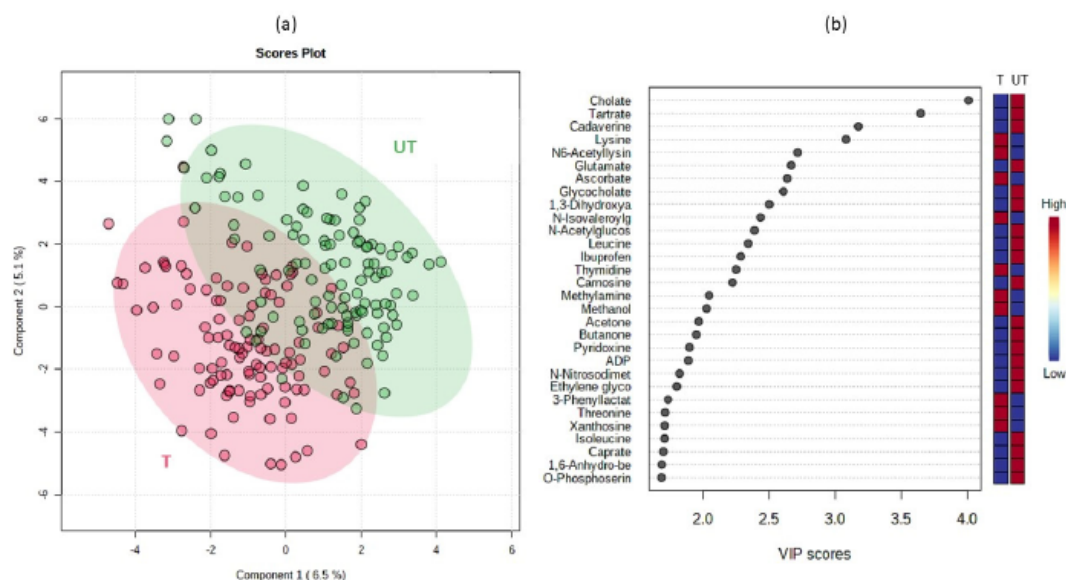
Figure 4. PLSDA ordination of metabolomics signatures present in the trained and untrained groups (**a,b**) VIP scores of the most important metabolites separating the two groups. The three-day series of urine samples of trained and untrained groups were analyzed with MetaboAnalyst. Prior to the PLSDA analysis, concentrations were transformed with Log10 normalization and scaled with Mean Centering approach. Each dot represents one sample of participant per day.
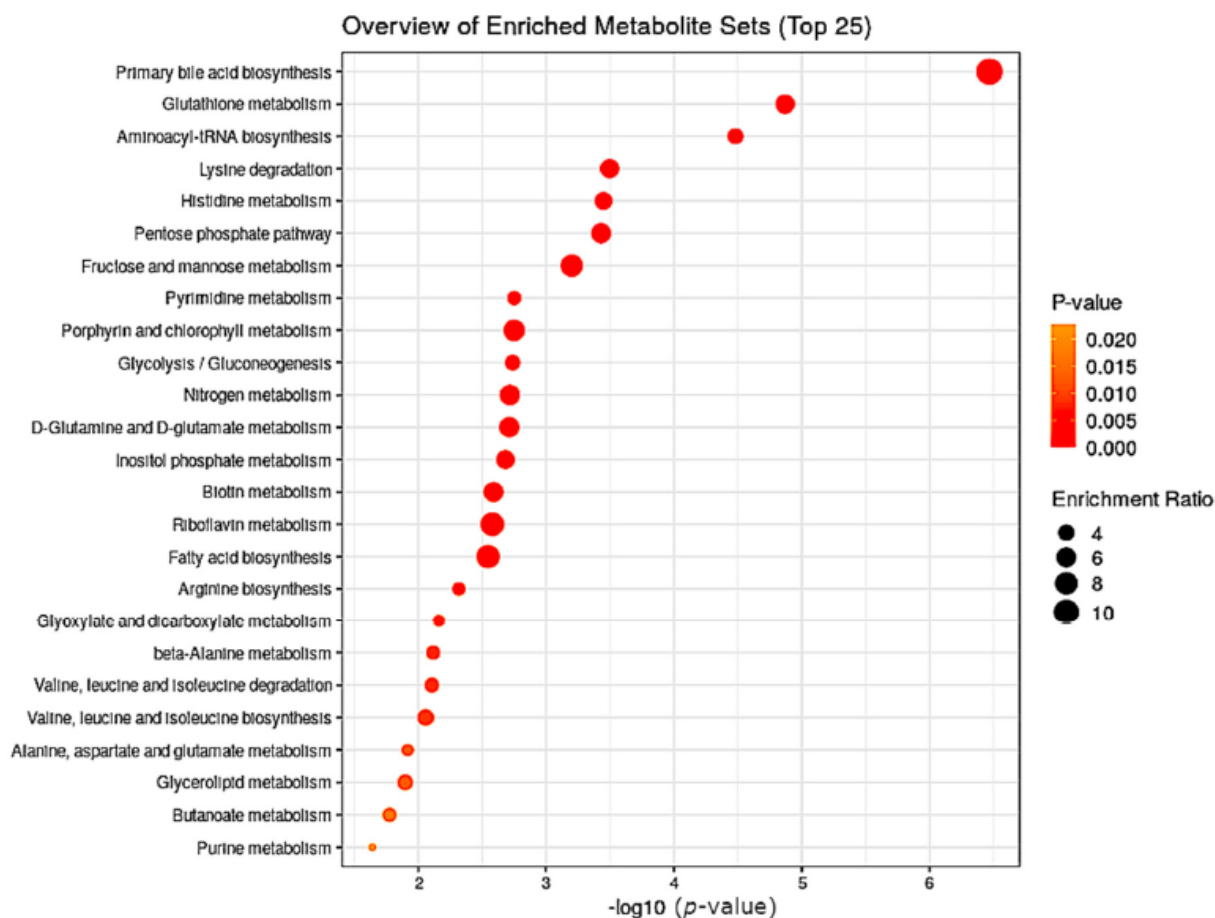
Figure 5. Primary bile-acid biosynthesis, glutathione metabolism and aminoacyl-tRNA biosynthesis were enriched in untrained group, based on increased levels of cholate (primary bile-acids biosynthesis), cadaverine (glutathione metabolism) and L-tryptophan and L-cysteine (aminoacyl-tRNA biosynthesis).

In addition to the multivariate analyses, we also performed extensive machine-learning modeling using Just Add Data Bio (JADBIO) [23] to investigate the importance of metabolites and physicochemical parameters in urine samples. A total of 181,020 models were trained using extensive tuning effort. The most interpretable model was logistic ridge regression with the penalty hyperparameter lambda of 10^{-4} and an area under the curve (AUC) value of 0.748. In addition to AUC (Figure 6a); all other thresholds were also statistically significantly different from baseline. Data were preprocessed and standardized by imputation of means and removal of constants. Features were selected based on the test-budgeted statistically equivalent signature (SES) algorithm with the following hyperparameters: maxK = 2, alpha = 0.1, and budget = 3 × nvars. PCA plot (Figure 6b) shows that differentiation based on modeled data is not complete, which means that larger groups should be formed in the future. 12 metabolites and pH were selected as the most important features for distinguishing the trained from the untrained group based on urine. Table S2 lists all the important metabolites. The major metabolite selected by JADBIO was tartrate. The power of the model obtained by using only tartrate was 73.8% (with 95% CI from 69.9% to 77.6%) (Figure S4). We applied the trained model to the test portion of our data (30% of our total dataset) and achieved validation performance with an AUC of 0.647.
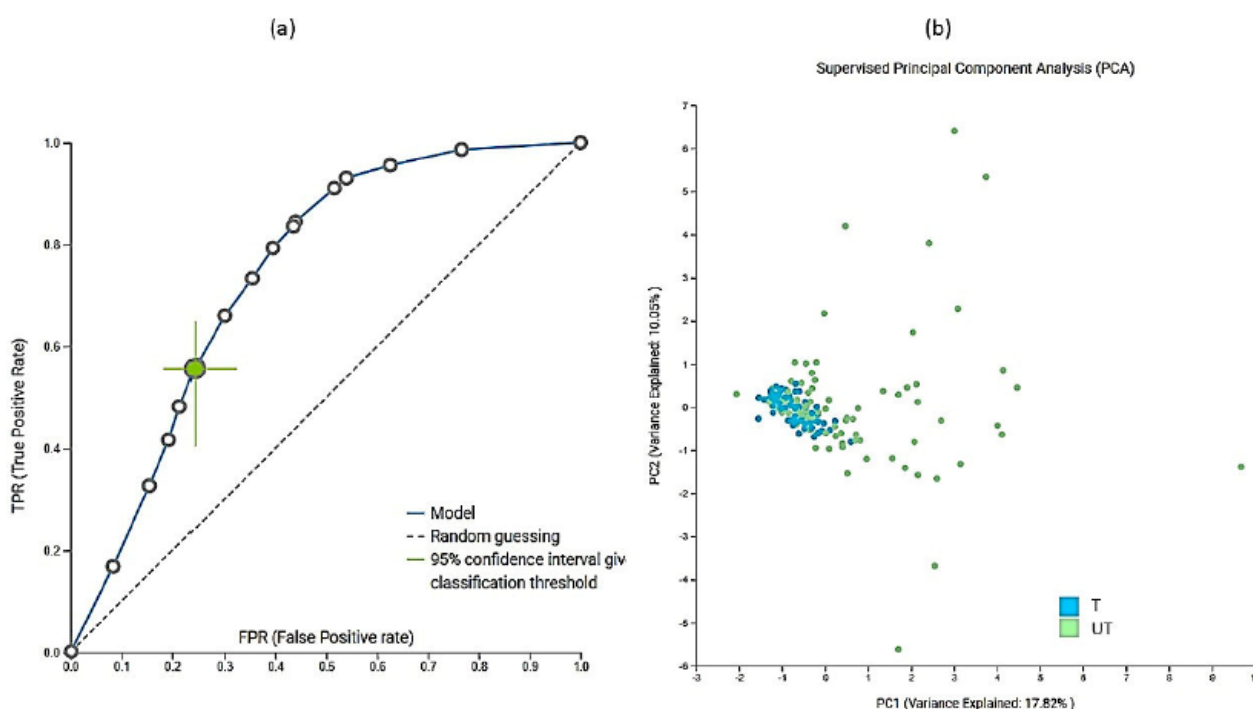
Figure 6. The receiver operator curve (ROC) (**a**) plot and PCA plot (**b**) of modeled data. The dimensionality reduction was performed within JADBio on a subset of the original dataset, keeping only the features included in the first signature. Features were standardized with statistical normalization $((x - \mu)/\sigma)$. A total of 155 samples were included in this analysis for training the model from the entire dataset. A total of 72 samples belonged to trained and 82 samples belonged to untrained group.

Some metabolites (cholate, tartrate, methanol, N-acetylglucosamine, butanone, caprate) were selected as the top 25 metabolites using the PLSDA approach in MetaboAnalyst. Both tartrate and cholate were elevated in the untrained group, which could be related to their diet. The diet of athletes is much more constant, and the diet was not standardized in the X-Adapt project. However, the decreased tartrate levels may suggest that tartrate supplementation is needed in the trained group to reduce metabolic stress, minimize muscle damage, improve hormone receptor levels, and promote recovery after resistance exercise [24,25]. L-carnitine L-tartrate supplementation increases carbohydrate oxidation rates. Endurance athletes in particular have higher carnitine uptake in skeletal muscle [24,25]. Tartrate is a nonhuman metabolite found in grapes, wine, and as an additive in foods [26]. Increased consumption of tartrate-containing foods and beverages also lowers cardiovascular risk factors such as LDL cholesterol [27,28]. Tartrate is part of glyoxylate and dicarboxylate metabolism, which was also observed in the enrichment analysis of metabolic pathways and enriched in the untrained group. Glyoxylate and dicarboxylate metabolic pathways were observed in young patients with major depressive disorder. Improving physical activity improved patients with major depressive disorder and additionally reduced other complications of cardiovascular disease [29,30]. Inactivity in the bed rest study (e.g., PlanHab) also led to the development of psychiatric problems after one week of bed rest, showing possible associations between inactivity, metabolism and mental health problems [4,5,31–44].

Cholate, on the other hand, is one of the primary bile acids that may be involved in the development of an atrophic state in myotubes [45] and in the invasion of human colon cancer cells [46], which can be observed in less active and untrained individuals. Bile acids in general have also been associated with obesity [47], higher BMI, elevated blood glucose levels [48], liver dysfunction [49] and cardiovascular health [50]. Bile acids in urine can be used for diagnostic purposes, as it has already been shown that bile acids

in urine have lower variability and higher stability than bile acids in serum [51]. Elevated cholate concentrations have also been observed in patients with gastric cancer. In our work, increased concentrations of cholate were observed in untrained individuals. A meta-analysis has previously shown that regular physical activity can prevent gastric cancer [52–54]. A single training run in amateur runners resulted in a significant decrease in circulating bile acids. Recent studies have also shown that bile-acid concentrations were higher in less fit women than in fit women [55,56].

Polyamines such as lysine and cadaverine, which were also detected in our study, have also been associated with the development of various diseases described by the common term "metabolic syndrome". It has already been shown that elevated cadaverine concentrations may correlate with intestinal disease or colon and liver cancer. Cadaverine was also elevated in the untrained group and is part of the glutathione metabolism previously described in men with type 1 diabetes [57]. Metabolic syndrome develops mainly due to inactivity or lack of exercise [58–60]. Lysine is involved in aminoacyl-tRNA biosynthesis and was increased in trained group. Aminoacyl-tRNA biosynthesis was associated with higher physical activity, a less sedentary lifestyle and high-intensity interval training [61–63]. Using metabolomes in stool and serum, the same metabolic pathway was identified as altered in endurance cross-country athletes, reflecting modifications in protein synthesis [64].

In contrast, 2-hydroxy-3-methyl- valerate was identified only with machine learning and was decreased in the trained group, confirming that it may also be involved in affecting physical function through peroxisome proliferator-activated receptor alpha (PPAR-α) activation, which is associated with microbial metabolism and insulin sensitivity [65]. PPAR-α is a hormone-receptor transcription factor involved in energy metabolism. Untrained participants in X-Adapt are less physically active and have increased levels of 2-hydroxy-3-methyl valerate, leading to possible activation of PPAR-α, as shown in functionally impaired older adults [65,66]. N6-acetyl-L-lysine is an acetylated amino acid that is increased in the trained group and plays an important role in regulating gene transcription, cell-cycle progression, apoptosis, DNA repair and cytoskeletal organization, also decreasing chances of Alzheimer's disease shown on rats. Physical activity has previously been shown to reduce the risk of age-related Alzheimer's disease [67,68] and metabolic syndrome [69].

We also observed that an increased pH increased the chance of classifying participants into a trained group. A lower urine pH was associated with chronic kidney disease [70], chronic heart failure [71] and metabolic syndrome [69,72,73].

Our analyses of the same 3-day-series data on a daily basis did not produce interpretable patterns of significant differences between the daily metabolome groups of the same 3-day sampling campaign after the correction for multiple comparisons (PERMANOVA; $p > 0.05$; $n_{permutations} = 5000$) and were not reported. This corroborates our past observation [5] on the higher resolution of 3-day series of ^{1}H-NMR metabolomes in contrast to single-day sampling.

2.3. Differences between Trained and Untrained Groups before and after Synchronizing Normoxic Training Campaign: The Extended Urine-Sample Collection

Our last analysis focused on the exploration of the extended urine-sample collection between trained and untrained (Figure 1B) to identify differences in morning urine metabolomes as a result of their original lifestyle and almost 2 weeks of 1h training at 50% W_{peak} (i.e., 3-day exercise tests, 1 day rest, 10 days 1 h training at 50% W_{peak}; Figure 1A). The results of PERMANOVA ($p(same) = 0.003$; $n_{permutations} = 5000$; Figure S5) showed that in the trained group, an active lifestyle supported significantly different metabolomic fingerprints in comparison to the untrained group (Figure S5). The differences between the trained and untrained groups were no longer significant at the end of training ($p = 0.226$), while shared metabolomics features were present within each of the groups on the relation between pre- and post-training states (horizontal lines; Figure S5) as the significant differences persisted in relation to pretrained vs. post-untrained and pre-untrained vs. post-trained (diagonal lines; Figure S5). The results of this study suggest that exercise introduced changes in

trained and untrained groups, making their endpoints not significantly different, and was accompanied by the concomitant decrease in the VO_2max values (-3.2%) in trained and increase ($+9.2\%$) in untrained groups [12,15,16,18].

When all eight groups of metabolomes (Figure 1B) were analyzed, it became apparent that the first introduction of controlled exercise at pre-exercise tests generated rather similar resting morning metabolomic urine makeup (i.e., short-term multivariate phenotype) in the two physiologically significantly different groups, while measurable changes within the exercise parameters (long-term multivariate phenotype (Figures 2 and S2) were detected much later. Consequently, the frequency of these training bouts (i.e., life-long exercise) is in fact a crucial parameter for maintaining a healthy metabolomic phenotype and VO_2max next to other exercise-related parameters. In contrast to WHO's proposed 75 min to 150 min of vigorous- to moderate-intensity training, respectively, for adults per week [1,2], our study showed that a 5 times larger exercise input was effective at bringing the urine metabolomics makeup and VO_2max values closer to the trained group, while obviously for the maintenance of an active lifestyle pursued by the trained group, much higher efforts would need to be invested. This finding is also in-line with the past observations on the difficulties in observing differences between training regimes [74], the effects of which subsided within 3 h after exercise, even in clinical populations [19]. Putting it simply, long-term exercise makes us rather similar in health, but a lack of it makes us different in disease. X-Adapt findings presented in this study on homogenizing effects of exercise are mirroring our past results from the PlanHab project on negative effects of inactivity [4,5,33,34].

To conclude, morning urine, especially as utilized in the form of 3-day sessions, has been shown to represent a good candidate biological matrix for delineation of active and inactive lifestyles in this study, detecting differences unobservable by single-day sampling. Resting morning urine metabolomes as a result of 1 h 50% W_{peak} daily activity provided a good basis for planning future recommendations for the maintenance of health in adults, irrespective of the starting fitness value. The maintenance of systemic homeostasis and the response to nutritional and environmental challenges require the coordination of multiple organs and tissues. To respond to various metabolic demands, the human body integrates and builds upon a system of interorgan communication through which one tissue can affect metabolic pathways in a distant tissue. Dysregulation of these lines of communication through lack of exercise (sedentary lifestyle) and of highly energetic diets contribute to human pathologies, including obesity, diabetes, liver disease and atherosclerosis. Increasing exercise levels in the untrained apparently has the capacity to significantly reconstitute the interorgan communication towards the levels observed in the healthy trained cohort. In addition, recent technical advances such as data-driven bioinformatics on layers of information (microbiome, proteome, metabolome) expanded our understanding of the complexity of systemic metabolic crosstalk and its underlying mechanisms [75].

3. Materials and Methods

3.1. Project Description

In this study, the fourth campaign of the X-Adapt: Cross-adaptation between heat and hypoxia–novel strategy for performance and work-ability enhancement in various environments project (ARRS research project J5-9350) was utilized as source of exercise data and urine samples for ^{1}H-NMR metabolomics analyses (Figure S1).

The main objective of the X-Adapt project was to determine the metabolic differences between trained and untrained individuals and the effects of 10 days of training on metabolism utilizing urinary metabolomics.

During the prescreening procedure, participants completed a graded exercise test on a cycle ergometer to determine their normoxic (environment with normal O_2 concentrations (e.g., 21%)) maximal-rate oxygen consumption (VO_2max) and maximal power output (W_{peak}). W_{peak} is defined as the highest workload sustained by incremental exercise until exhaustion. Aerobic fitness was defined using VO_2max values. A VO_2max of less than $45 \text{ mL·kg}^{-1}\text{·min}^{-1}$ or greater than $55 \text{ mL·kg}^{-1}\text{·min}^{-1}$ was considered a requirement for

participation in the lower-fitness (untrained) or higher-fitness (trained) group, respectively, consistent with values reported in previous studies [13,14]. To further ensure that VO_2max reflected participants' true cardiorespiratory fitness levels, untrained participants were also required to not participate in organized sports. Cycling and walking for commuting to work were allowed. Accordingly, trained participants performed endurance-type activities (running, cycling, swimming) several times per week. Participants were informed that the aim of the study was to investigate the effects of a 10-day exercise protocol on aerobic performance in young males [12,15–17].

Twenty healthy young male volunteers were recruited to participate in the study. Inclusion criteria included males between the ages of 18 and 30, nonsmokers, and unmedicated. All participants lived near the sea and had not been exposed to altitudes > 1500 m or temperatures > 30 °C for at least 1 month before the start of the study, which took place in November and December 2018. None of the participants had a history of cardiorespiratory or hematologic disease. Participants were instructed to abstain from caffeine and alcohol consumption throughout the study. They were given detailed information about the study protocol and potential risks.

The study consisted of three parts: pretraining exercise testing, a 10-day exercise training program, followed by the post-training exercise testing (Figure 1).

During the pre-exercise tests, participants completed the same maximal exercise performance test on three consecutive but separate days under thermoneutral normoxic, thermoneutral hypoxic and hot normoxic conditions as described before [12,15–17]. The order of exercise tests was randomized and counterbalanced between participants. All tests were performed at the same time of day for a given participant (± 1 h). Exercise training sessions took place in the morning hours (9:00–12:00). Participants were given a 24 h rest period before and after the 10-day training to minimize the contribution of fatigue during the exercise tests [12,15–17].

During the 10-day training session, all participants completed 60 min supervised cycling sessions daily for 10 days. Exercise was performed on a cycle ergometer (Daum, Electronic, Furth, Germany). During training, each participant pedalled at a preferred cadence (between 60 and 90 rpm), which they maintained throughout the experiment via visual and verbal feedback. Exercise intensity was relatively similar for all participants and was set at 50% of the W_{peak} calculated from the individual W_{peak} achieved during the preparatory graded normoxic exercise test. Participants were only informed of the time remaining until the end of the exercise session and were allowed to drink ad libitum during each exercise session. Heart rate and SpO_2 were measured with a finger pulse oximeter (Wristox 3100 Nonin, Plymouth, MN, USA) at 5 min intervals. Ratings of perceived exertion (RPE; 6–20) was also recorded at 5 min intervals. Ambient temperature was maintained at 24 °C. The training room was well-ventilated so that normoxic and normocapnic (normal arterial carbon dioxide pressure) conditions prevailed during training. Participants completed all exercise sessions at the same time of day. No other exercise training was allowed during the study. Sessions were supervised by at least two researchers to record exercise data and ensure that all participants maintained the desired workload at all times [12,15–17].

After the completion of 10-day training session, postexercise tests were performed. All pre- and postexercise tests were performed in a laboratory 300 m above sea level (Ljubljana, Slovenia). Trials were performed on a cycle ergometer (Daum, Electronic, Furth, Germany) and included two phases: a 30 min steady-state workout immediately followed by incremental training to exhaustion. Before (pre) and after (post) the 10-day training protocol, participants performed three trials on three consecutive days. At normal temperature and normoxic conditions (NOR), participants breathed room air (pre: partial pressure of oxygen in the inspired air (PiO_2) = 143.7 $\pm$ 0.8 mmHg, post: PiO_2 = 143.4 $\pm$ 0.7 mmHg) and exercised under thermoneutral conditions (pre: Ta = 23.2 $\pm$ 0.7 °C and relative humidity (RH) = 47.2 $\pm$ 2.2%, post: Ta = 23.2 $\pm$ 0.5 °C and RH = 46.6 $\pm$ 5.9%). In the hypoxic condition (HYP), they inspired a hypoxic gas mixture (pre: PiO_2 = 92.2 $\pm$ 1.5 mmHg, post:

$PiO_2 = 93.2 \pm 1.2$ mmHg) and exercised in thermoneutral conditions (pre: Ta = 22.8 ± 0.5 °C and RH = 51.2 ± 1.2%, post: Ta = 22.5 ± 0.6 °C and RH = 51.5 ± 1.3%). In the hot condition (HE), the participants inspired room air (pre: $PiO_2 = 142.6 \pm 1.7$ mmHg, post: $PiO_2 = 142.7 \pm 1.8$ mmHg), but exercised in a hot environment (pre: Ta = 34.1 ± 0.9 °C, RH = 48.1 ± 4.2%, post: Ta = 34.1 ± 1.1 °C and RH = 49.8 ± 3.0%) [12,15–17].

3.2. Sample Collection

Urine samples were collected in four sessions for 3 consecutive days to form 3-day series of urine samples for all participants (Figures 1 and S1): (i) 3-day baseline data of participants before the start of X-Adapt campaign, (ii) 3-day pre-exercise testing before 10-day 50% W_{peak} training, (iii) 3-day sampling of the last days of 10-day exercise; and (iv) 3-day postexercise testing after the 10-day 50% W_{peak} training. All obtained samples were frozen at −20 °C for further analysis as described before [4,5,76]. For simplicity, the X-Adapt urine-sample collection was used to denote samples collected during the X-Adapt pre-exercise and postexercise testing periods. The extended sample collection encompasses the urine samples collected before the actual onset of the campaign (baseline) and during the last three days of the 10-day 50% W_{peak} training.

3.3. NMR Metabolomics

All collected samples were centrifuged (1.5 mL) at $10,000\times g$ for 30 min to remove fine particles. Then, 600 μL of supernatant was mixed with 300 μL ^{1}H-NMR buffer as described before [77] and stored at −25 °C until analysis. Before analysis, samples were thawed at room temperature and transferred into a 5 mm NMR tube. TSP was used as an internal standard for quantification, as described before [77].

A Bruker Avance NEO 600 MHz spectrometer equipped with a 24-sample SampleCase autosampler and a 5 mm HCN Cold probe was used for the acquisition of NMR spectra at 25 °C. The 1H NMR spectra of the samples were recorded with a spectral width of 9.0 kHz, relaxation delay of 2.0 s, 32 scans and 32 K data points. A double-pulsed field gradient spin echo (DPFGSE) pulse sequence was used for water suppression. Total correlated spectrum (TOCSY) was measured with 1H spectral widths of 7.0 kHz, 4096 complex points, a relaxation delay of 1.5 s, 32 transients and 144 time increments. An exponential and cosine-squared function were used for apodization. Zeros were filled before Fourier transform. TopSpin v. 4.0.9 software (Bruker, Billerica, MA, USA) was used for processing urine NMR spectra [4,5,76,78]. AlpsNMR R package was used for the visualization of example spectra [79].

3.4. Physicochemical Parameters of Urine Samples

Urine samples were thawed at room temperature, homogenized. Additional physical chemical parameters were recorded such as pH, conductivity, total dissolved solids and salinity using Pocket pro$^+$ Multimeter 2 (Hach Company, Loveland, CO, USA).

3.5. Statistical Analysis and Machine Learning

The resulting spectra were consequently analyzed using targeted quantitative metabolomics using Chenomx NMR Suite version 8.6 (Chenomx, Inc., Edmonton, AB, Canada). For the latter, all spectra were randomly ordered for spectral fitting using the ChenomX profiler and the Human Metabolome Database (https://hmdb.ca/ (accessed on 24 April 2022)) compound names were used [80]. In this study, spectral deconvolution utilizing Chenomx and HMDB was used instead of the binning approaches with extensive normalization as described before [81,82]. An ensemble approach to data analysis was utilized, employing three different approaches to asymmetric sparse matrix data analysis, establishing significant differences between tested groups as follows: nonparametric MANOVA (PERMANOVA) [83], MetaboAnalyst [20–22], and JADBIO [23]. Heatmap of measured physiological parameters was generated using gplots R package. Data were normalized with scale function.

First, for PERMANOVA, each compound concentration obtained was analyzed as described before [4,11]. Box–Cox transformation was used. The significance of metabolic differences between various groups of samples was tested using 1-way and 2-way PERMANOVA, and expressed as an overlap in nonmetric multidimensional scaling (nm-MDS) trait space (using Euclidean distance measures). The stress function was used to select the dimensionality reduction, whereas Shepard's plots were used to describe the correspondence between the target and obtained ranks. Benjamini–Hochberg significance correction for multiple comparisons was used as described before [4,5].

Second, for MetaboAnalyst, log- or cube-root transformation in connection to Mean or Pareto scaling was utilized as implemented in MetaboAnalyst, followed by supervised classification using partial least-squares-discriminant analysis (PLSDA) method and random forest (RF). Statistical power for the identification of significant differences before and after treatment was also calculated using MetaboAnalyst Statistical Power module.

Metabolite Set Enrichment (MSEA) was used to identify biologically significant patterns between quantitative metabolome data from different groups. HMDB compound names were used to link to the KEGG database. Enrichment analysis was performed using the globaltest package implemented in MetaboAnalyst. The enrichment ratio was calculated by dividing observed hits and expected hits.

Finally, Just Add Data Bio (JADBIO), a web-based auto-machine-learning platform for analyzing potential biomarkers [23], was used. JADBIO 1.4.0 with extensive tuning effort and 6 CPU was used to model various dataset selections next to the overall 336 metabolites observed in urine samples in all groups (trained vs. untrained) by splitting the total urine metabolite data into a training set and a test set in a 70:30 ratio. The training set was used for model training and the test set was used for model evaluation.

The resulting model can be obtained as part of Supplementary Material (File S2) and run with Java executor for the classification of novel urine samples based on ^{1}H-NMR in further explorations.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Kelly, R.S.; Kelly, M.P.; Kelly, P. Metabolomics, physical activity, exercise and health: A review of the current evidence. *Biochim. Biophys. Acta. Mol. Basis Dis.* **2020**, *1866*, 165936. [CrossRef]
2. Sallis, J.F.; Bull, F.; Guthold, R.; Heath, G.W.; Inoue, S.; Kelly, P.; Oyeyemi, A.L.; Perez, L.G.; Richards, J.; Hallal, P.C.; et al. Progress in physical activity over the Olympic quadrennium. *Lancet* **2016**, *388*, 1325–1336. [CrossRef]
3. Bongiovanni, T.; Genovesi, F.; Nemmer, M.; Carling, C.; Alberti, G.; Howatson, G. Nutritional interventions for reducing the signs and symptoms of exercise-induced muscle damage and accelerate recovery in athletes: Current knowledge, practical application and future perspectives. *Eur. J. Appl. Physiol.* **2020**, *120*, 1965–1996. [CrossRef]
4. Sket, R.; Debevec, T.; Kublik, S.; Schloter, M.; Schoeller, A.; Murovec, B.; Mikus, K.V.; Makuc, D.; Pecnik, K.; Plavec, J.; et al. Intestinal Metagenomes and Metabolomes in Healthy Young Males: Inactivity and Hypoxia Generated Negative Physiological Symptoms Precede Microbial Dysbiosis. *Front. Physiol.* **2018**, *9*, 198. [CrossRef]
5. Šket, R.; Deutsch, L.; Prevoršek, Z.; Mekjavić, I.B.; Plavec, J.; Rittweger, J.; Debevec, T.; Eiken, O.; Stres, B. Systems View of Deconditioning During Spaceflight Simulation in the PlanHab Project: The Departure of Urine 1 H-NMR Metabolomes From Healthy State in Young Males Subjected to Bedrest Inactivity and Hypoxia. *Front. Physiol.* **2020**, *11*, 532271. [CrossRef]
6. Al-Khelaifi, F.; Diboun, I.; Donati, F.; Botrè, F.; Asayrafi, M.; Georgakopoulos, C.; Suhre, K.; Yousri, N.A.; Elrayess, M.A. A pilot study comparing the metabolic profiles of elite-level athletes from different sporting disciplines. *Sports Med.-Open* **2018**, *4*, 2. [CrossRef]
7. Danaher, J.; Gerber, T.; Wellard, R.M.; Stathis, C.G.; Cooke, M.B. The use of metabolomics to monitor simultaneous changes in metabolic variables following supramaximal low volume high intensity exercise. *Metabolomics* **2015**, *12*, 7. [CrossRef]
8. Zafeiridis, A.; Chatziioannou, A.C.; Sarivasiliou, H.; Kyparos, A.; Nikolaidis, M.G.; Vrabas, I.S.; Pechlivanis, A.; Zoumpoulakis, P.; Baskakis, C.; Dipla, K.; et al. Global Metabolic Stress of Isoeffort Continuous and High Intensity Interval Aerobic Exercise: A Comparative 1 H NMR Metabonomic Study. *J. Proteome Res.* **2016**, *15*, 4452–4463. [CrossRef]
9. Xiao, Q.; Moore, S.C.; Keadle, S.K.; Xiang, Y.B.; Zheng, W.; Peters, T.M.; Leitzmann, M.F.; Ji, B.T.; Sampson, J.N.; Shu, X.O.; et al. Objectively measured physical activity and plasma metabolomics in the Shanghai Physical Activity Study. *Int. J. Epidemiol.* **2016**, *45*, 1433–1444. [CrossRef]
10. Emwas, A.H.; Roy, R.; McKay, R.T.; Tenori, L.; Saccenti, E.; Gowda, G.A.N.; Raftery, D.; Alahmari, F.; Jaremko, L.; Jaremko, M.; et al. NMR Spectroscopy for Metabolomics Research. *Metabolites* **2019**, *9*, 123. [CrossRef]
11. Murovec, B.; Makuc, D.; Kolbl Repinc, S.; Prevorsek, Z.; Zavec, D.; Sket, R.; Pecnik, K.; Plavec, J.; Stres, B. ^{1}H NMR metabolomics of microbial metabolites in the four MW agricultural biogas plant reactors: A case study of inhibition mirroring the acute rumen acidosis symptoms. *J. Environ. Manag.* **2018**, *222*, 428–435. [CrossRef]
12. Sotiridis, A.; Debevec, T.; Ciuha, U.; McDonnell, A.C.; Mlinar, T.; Royal, J.T.; Mekjavic, I.B. Aerobic but not thermoregulatory gains following a 10-day moderate-intensity training protocol are fitness level dependent: A cross-adaptation perspective. *Physiol. Rep.* **2020**, *8*, e14355. [CrossRef]
13. Jay, O.; Bain, A.R.; Deren, T.M.; Sacheli, M.; Cramer, M.N. Large differences in peak oxygen uptake do not independently alter changes in core temperature and sweating during exercise. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **2011**, *301*, R832–R841. [CrossRef] [PubMed]
14. Montero, D.; Lundby, C. Refuting the myth of non-response to exercise training: 'non-responders' do respond to higher dose of training. *J. Physiol.* **2017**, *595*, 3377–3387. [CrossRef] [PubMed]
15. Sotiridis, A.; Debevec, T.; McDonnell, A.C.; Ciuha, U.; Eiken, O.; Mekjavic, I.B. Exercise cardiorespiratory and thermoregulatory responses in normoxic, hypoxic and hot environment following 10-day continuous hypoxic exposure. *J. Appl. Physiol.* **2018**, *125*, 1284–1295. [CrossRef] [PubMed]
16. Sotiridis, A.; Debevec, T.; Ciuha, U.; Eiken, O.; Mekjavic, I.B. Heat acclimation does not affect maximal aerobic power in thermoneutral normoxic or hypoxic conditions. *Exp. Physiol.* **2019**, *104*, 345–358. [CrossRef] [PubMed]
17. Sotiridis, A. Independent and Combined Effects of Heat and Hypoxic Acclimation on Exercise Performance in Humans: With Particular Reference to Cross-Adaption. Ph.D. Thesis, Jozef Stefan Institute, Ljubljana, Slovenia, 2019.

18. Armstrong, N.; Barker, A.R. Endurance training and elite young athletes. *Med. Sport Sci.* **2011**, *56*, 59–83. [CrossRef]

19. Siopi, A.; Deda, O.; Manou, V.; Kellis, S.; Kosmidis, I.; Komninou, D.; Raikos, N.; Christoulas, K.; Theodoridis, G.A.; Mougios, V. Effects of Different Exercise Modes on the Urinary Metabolic Fingerprint of Men with and without Metabolic Syndrome. *Metabolites* **2017**, *7*, 5. [CrossRef]

20. Chong, J.; Soufan, O.; Li, C.; Caraus, I.; Li, S.; Bourque, G.; Wishart, D.S.; Xia, J. MetaboAnalyst 4.0: Towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res.* **2018**, *46*, W486–W494. [CrossRef]

21. Chong, J.; Wishart, D.S.; Xia, J. Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis. *Curr. Protoc. Bioinform.* **2019**, *68*, e86. [CrossRef]

22. Pang, Z.; Chong, J.; Zhou, G.; de Lima Morais, D.A.; Chang, L.; Barrette, M.; Gauthier, C.; Jacques, P.É.; Li, S.; Xia, J. MetaboAnalyst 5.0: Narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res.* **2021**, *49*, W388–W396. [CrossRef] [PubMed]

23. Tsamardinos, I.; Charonyktakis, P.; Lakiotaki, K.; Borboudakis, G.; Zenklusen, J.C.; Juhl, H.; Chatzaki, E.; Lagani, V. Just Add Data: Automated Predictive Modeling and BioSignature Discovery. *bioRxiv* **2020**, 1–46. [CrossRef]

24. Spiering, B.A.; Kraemer, W.J.; Hatfield, D.L.; Vingren, J.L.; Fragala, M.S.; Ho, J.Y.; Thomas, G.A.; Häkkinen, K.; Volek, J.S. Effects of L-carnitine L-tartrate supplementation on muscle oxygenation responses to resistance exercise. *J. Strength Cond. Res.* **2008**, *22*, 1130–1135. [CrossRef] [PubMed]

25. Abramowicz, W.N.; Galloway, S.D. Effects of acute versus chronic L-carnitine L-tartrate supplementation on metabolic responses to steady state exercise in males and females. *Int. J. Sport Nutr. Exerc. Metab.* **2005**, *15*, 386–400. [CrossRef]

26. Vázquez-Fresno, R.; Llorach, R.; Urpi-Sarda, M.; Khymenets, O.; Bulló, M.; Corella, D.; Fitó, M.; Martínez-González, M.A.; Estruch, R.; Andres-Lacueva, C. An NMR metabolomics approach reveals a combined-biomarkers model in a wine interventional trial with validation in free-living individuals of the PREDIMED study. *Metabolomics* **2014**, *11*, 797–806. [CrossRef]

27. Garcia-Perez, I.; Posma, J.M.; Chambers, E.S.; Nicholson, J.K.; Mathers, J.C.; Beckmann, M.; Draper, J.; Holmes, E.; Frost, G. An Analytical Pipeline for Quantitative Characterization of Dietary Intake: Application To Assess Grape Intake. *J. Agric. Food Chem.* **2016**, *64*, 2423–2431. [CrossRef]

28. Domínguez-López, I.; Parilli-Moser, I.; Arancibia-Riveros, C.; Tresserra-Rimbau, A.; Martínez-González, M.A.; Ortega-Azorín, C.; Salas-Salvadó, J.; Castañer, O.; Lapetra, J.; Arós, F.; et al. Urinary Tartaric Acid, a Biomarker of Wine Intake, Correlates with Lower Total and LDL Cholesterol. *Nutrients* **2021**, *13*, 2883. [CrossRef]

29. Chen, J.J.; Xie, J.; Li, W.W.; Bai, S.J.; Wang, W.; Zheng, P.; Xie, P. Age-specific urinary metabolite signatures and functions in patients with major depressive disorder. *Aging* **2019**, *11*, 6626–6637. [CrossRef]

30. Gerber, M.; Beck, J.; Brand, S.; Cody, R.; Donath, L.; Eckert, A.; Faude, O.; Fischer, X.; Hatzinger, M.; Holsboer-Trachsler, E.; et al. The impact of lifestyle Physical Activity Counselling in IN-PATients with major depressive disorders on physical activity, cardiorespiratory fitness, depression, and cardiovascular health risk markers: Study protocol for a randomized controlled trial. *Trials* **2019**, *20*, 367. [CrossRef]

31. Debevec, T.; Bali, T.C.; Simpson, E.J.; Macdonald, I.A.; Eiken, O.; Mekjavic, I.B. Separate and combined effects of 21-day bed rest and hypoxic confinement on body composition. *Eur. J. Appl. Physiol.* **2014**, *114*, 2411–2425. [CrossRef]

32. Debevec, T.; Simpson, E.J.; Mekjavic, I.B.; Eiken, O.; Macdonald, I.A. Effects of prolonged hypoxia and bed rest on appetite and appetite-related hormones. *Appetite* **2016**, *107*, 28–37. [CrossRef] [PubMed]

33. Sket, R.; Treichel, N.; Debevec, T.; Eiken, O.; Mekjavic, I.; Schloter, M.; Vital, M.; Chandler, J.; Tiedje, J.M.; Murovec, B.; et al. Hypoxia and Inactivity Related Physiological Changes (Constipation, Inflammation) Are Not Reflected at the Level of Gut Metabolites and Butyrate Producing Microbial Community: The PlanHab Study. *Front. Physiol.* **2017**, *8*, 250. [CrossRef] [PubMed]

34. Sket, R.; Treichel, N.; Kublik, S.; Debevec, T.; Eiken, O.; Mekjavic, I.; Schloter, M.; Vital, M.; Chandler, J.; Tiedje, J.M.; et al. Hypoxia and inactivity related physiological changes precede or take place in absence of significant rearrangements in bacterial community structure: The PlanHab randomized trial pilot study. *PLoS ONE* **2017**, *12*, e0188556. [CrossRef] [PubMed]

35. Debevec, T.; Ganse, B.; Mittag, U.; Eiken, O.; Mekjavic, I.B.; Rittweger, J. Hypoxia Aggravates Inactivity-Related Muscle Wasting. *Front. Physiol.* **2018**, *9*, 494. [CrossRef]

36. Stavrou, N.A.M.; Debevec, T.; Eiken, O.; Mekjavic, I.B. Hypoxia Exacerbates Negative Emotional State during Inactivity: The Effect of 21 Days Hypoxic Bed Rest and Confinement. *Front. Physiol.* **2018**, *9*, 26. [CrossRef]

37. Stavrou, N.A.M.; Debevec, T.; Eiken, O.; Mekjavic, I.B. Hypoxia Worsens Affective Responses and Feeling of Fatigue During Prolonged Bed Rest. *Front. Psychol.* **2018**, *9*, 362. [CrossRef]

38. Strewe, C.; Zeller, R.; Feuerecker, M.; Hoerl, M.; Kumprej, I.; Crispin, A.; Johannes, B.; Debevec, T.; Mekjavic, I.; Schelling, G.; et al. PlanHab study: Assessment of psycho-neuroendocrine function in male subjects during 21 d of normobaric hypoxia and bed rest. *Stress* **2017**, *20*, 131–139. [CrossRef]

39. Strewe, C.; Zeller, R.; Feuerecker, M.; Hoerl, M.; Matzel, S.; Kumprej, I.; Crispin, A.; Johannes, B.; Debevec, T.; Mekjavic, I.B.; et al. PlanHab Study: Consequences of combined normobaric hypoxia and bed rest on adenosine kinetics. *Sci. Rep.* **2018**, *8*, 1762. [CrossRef]

40. Keramidas, M.E.; Kolegard, R.; Mekjavic, I.B.; Eiken, O. PlanHab: Hypoxia exaggerates the bed-rest-induced reduction in peak oxygen uptake during upright cycle ergometry. *Am. J. Physiol. Heart Circ. Physiol.* **2016**, *311*, H453–H464. [CrossRef]

107

41. Louwies, T.; Jaki Mekjavic, P.; Cox, B.; Eiken, O.; Mekjavic, I.B.; Kounalakis, S.; De Boever, P. Separate and Combined Effects of Hypoxia and Horizontal Bed Rest on Retinal Blood Vessel Diameters. *Investig. Ophthalmol. Vis. Sci.* **2016**, *57*, 4927–4932. [CrossRef]

42. Simpson, E.J.; Debevec, T.; Eiken, O.; Mekjavic, I.; Macdonald, I.A. PlanHab: The combined and separate effects of 16 days of bed rest and normobaric hypoxic confinement on circulating lipids and indices of insulin sensitivity in healthy men. *J. Appl. Physiol.* **2016**, *120*, 947–955. [CrossRef] [PubMed]

43. Sarabon, N.; Mekjavic, I.B.; Eiken, O.; Babic, J. The Effect of Bed Rest and Hypoxic Environment on Postural Balance and Trunk Automatic (Re)Actions in Young Healthy Males. *Front. Physiol.* **2018**, *9*, 27. [CrossRef] [PubMed]

44. Rullman, E.; Mekjavic, I.B.; Fischer, H.; Eiken, O. PlanHab (Planetary Habitat Simulation): The combined and separate effects of 21 days bed rest and hypoxic confinement on human skeletal muscle miRNA expression. *Physiol. Rep.* **2016**, *4*, e12753. [CrossRef] [PubMed]

45. Abrigo, J.; Gonzalez, F.; Aguirre, F.; Tacchi, F.; Gonzalez, A.; Meza, M.P.; Simon, F.; Cabrera, D.; Arrese, M.; Karpen, S.; et al. Cholic acid and deoxycholic acid induce skeletal muscle atrophy through a mechanism dependent on TGR5 receptor. *J. Cell. Physiol.* **2021**, *236*, 260–272. [CrossRef] [PubMed]

46. Li, S.; Ung, T.T.; Nguyen, T.T.; Sah, D.K.; Park, S.Y.; Jung, Y.D. Cholic Acid Stimulates MMP-9 in Human Colon Cancer Cells via Activation of MAPK, AP-1, and NF-κB Activity. *Int. J. Mol. Sci.* **2020**, *21*, 3420. [CrossRef]

47. Mercer, K.E.; Maurer, A.; Pack, L.M.; Ono-Moore, K.; Spray, B.J.; Campbell, C.; Chandler, C.J.; Burnett, D.; Souza, E.; Casazza, G.; et al. Exercise training and diet-induced weight loss increase markers of hepatic bile acid (BA) synthesis and reduce serum total BA concentrations in obese women. *Am. J. Physiol. Endocrinol. Metab.* **2021**, *320*, E864–E873. [CrossRef]

48. Zheng, X.; Chen, T.; Zhao, A.; Ning, Z.; Kuang, J.; Wang, S.; You, Y.; Bao, Y.; Ma, X.; Yu, H.; et al. Hyocholic acid species as novel biomarkers for metabolic disorders. *Nat. Commun.* **2021**, *12*, 1487. [CrossRef]

49. Alamoudi, J.A.; Li, W.; Gautam, N.; Olivera, M.; Meza, J.; Mukherjee, S.; Alnouti, Y. Bile acid indices as biomarkers for liver diseases I: Diagnostic markers. *World J. Hepatol.* **2021**, *13*, 433–455. [CrossRef]

50. Pushpass, R.G.; Alzoufairi, S.; Jackson, K.G.; Lovegrove, J.A. Circulating bile acids as a link between the gut microbiota and cardiovascular health: Impact of prebiotics, probiotics and polyphenol-rich foods. *Nutr. Res. Rev.* **2021**, 1–20. [CrossRef]

51. Bathena, S.P.; Thakare, R.; Gautam, N.; Mukherjee, S.; Olivera, M.; Meza, J.; Alnouti, Y. Urinary bile acids as biomarkers for liver diseases II. Signature profiles in patients. *Toxicol. Sci. Off. J. Soc. Toxicol.* **2015**, *143*, 308–318. [CrossRef]

52. Lyu, J.; Li, H.; Yin, D.; Zhao, M.; Sun, Q.; Guo, M. Analysis of eight bile acids in urine of gastric cancer patients based on covalent organic framework enrichment coupled with liquid chromatography-tandem mass spectrometry. *J. Chromatogr. A* **2021**, *1653*, 462422. [CrossRef] [PubMed]

53. Singh, S.; Edakkanambeth Varayil, J.; Devanna, S.; Murad, M.H.; Iyer, P.G. Physical activity is associated with reduced risk of gastric cancer: A systematic review and meta-analysis. *Cancer Prev. Res.* **2014**, *7*, 12–22. [CrossRef] [PubMed]

54. Abioye, A.I.; Odesanya, M.O.; Abioye, A.I.; Ibrahim, N.A. Physical activity and risk of gastric cancer: A meta-analysis of observational studies. *Br. J. Sports Med.* **2015**, *49*, 224–229. [CrossRef] [PubMed]

55. Danese, E.; Salvagno, G.L.; Tarperi, C.; Negrini, D.; Montagnana, M.; Festa, L.; Sanchis-Gomar, F.; Schena, F.; Lippi, G. Middle-distance running acutely influences the concentration and composition of serum bile acids: Potential implications for cancer risk? *Oncotarget* **2017**, *8*, 52775–52782. [CrossRef]

56. Maurer, A.; Ward, J.L.; Dean, K.; Billinger, S.A.; Lin, H.; Mercer, K.E.; Adams, S.H.; Thyfault, J.P. Divergence in aerobic capacity impacts bile acid metabolism in young women. *J. Appl. Physiol.* **2020**, *129*, 768–778. [CrossRef]

57. Darenskaya, M.A.; Chugunova, E.V.; Kolesnikov, S.I.; Grebenkina, L.A.; Semenova, N.V.; Nikitina, O.A.; Kolesnikova, L.I. Content of Carbonyl Compounds and Parameters of Glutathione Metabolism in Men with Type 1 Diabetes Mellitus at Preclinical Stages of Diabetic Nephropathy. *Bull. Exp. Biol. Med.* **2021**, *171*, 592–595. [CrossRef]

58. Liu, R.; Li, Q.; Ma, R.; Lin, X.; Xu, H.; Bi, K. Determination of polyamine metabolome in plasma and urine by ultrahigh performance liquid chromatography-tandem mass spectrometry method: Application to identify potential markers for human hepatic cancer. *Anal. Chim. Acta* **2013**, *791*, 36–45. [CrossRef]

59. Maráková, K.; Piešťanský, J.; Zelinková, Z.; Mikuš, P. Simultaneous determination of twelve biogenic amines in human urine as potential biomarkers of inflammatory bowel diseases by capillary electrophoresis–tandem mass spectrometry. *J. Pharm. Biomed. Anal.* **2020**, *186*, 113294. [CrossRef]

60. Venäläinen, M.K.; Roine, A.N.; Häkkinen, M.R.; Vepsäläinen, J.J.; Kumpulainen, P.S.; Kiviniemi, M.S.; Lehtimäki, T.; Oksala, N.K.; Rantanen, T.K. Altered Polyamine Profiles in Colorectal Cancer. *Anticancer. Res.* **2018**, *38*, 3601–3607. [CrossRef]

61. Tian, Q.; Corkum, A.E.; Moaddel, R.; Ferrucci, L. Metabolomic profiles of being physically active and less sedentary: A critical review. *Metab. Off. J. Metab. Soc.* **2021**, *17*, 68. [CrossRef]

62. Castro, A.; Duft, R.G.; Ferreira, M.L.V.; Andrade, A.L.L.; Gáspari, A.F.; Silva, L.M.; Oliveira-Nunes, S.G.; Cavaglieri, C.R.; Ghosh, S.; Bouchard, C.; et al. Association of skeletal muscle and serum metabolites with maximum power output gains in response to continuous endurance or high-intensity interval training programs: The TIMES study—A randomized controlled trial. *PLoS ONE* **2019**, *14*, e0212115. [CrossRef] [PubMed]

63. Robinson, M.M.; Dasari, S.; Konopka, A.R.; Johnson, M.L.; Manjunatha, S.; Esponda, R.R.; Carter, R.E.; Lanza, I.R.; Nair, K.S. Enhanced Protein Translation Underlies Improved Metabolic and Physical Adaptations to Different Exercise Training Modes in Young and Old Humans. *Cell Metab.* **2017**, *25*, 581–592. [CrossRef] [PubMed]

64. Tabone, M.; Bressa, C.; García-Merino, J.A.; Moreno-Pérez, D.; Van, E.C.; Castelli, F.A.; Fenaille, F.; Larrosa, M. The effect of acute moderate-intensity exercise on the serum and fecal metabolomes and the gut microbiota of cross-country endurance athletes. *Sci. Rep.* **2021**, *11*, 3558. [CrossRef] [PubMed]

65. Lustgarten, M.S.; Price, L.L.; Chalé, A.; Fielding, R.A. Metabolites related to gut bacterial metabolism, peroxisome proliferator-activated receptor-alpha activation, and insulin sensitivity are associated with physical function in functionally-limited older adults. *Aging Cell* **2014**, *13*, 918–925. [CrossRef] [PubMed]

66. Coen, P.M.; Jubrias, S.A.; Distefano, G.; Amati, F.; Mackey, D.C.; Glynn, N.W.; Manini, T.M.; Wohlgemuth, S.E.; Leeuwenburgh, C.; Cummings, S.R.; et al. Skeletal muscle mitochondrial energetics are associated with maximal aerobic capacity and walking speed in older adults. *J. Gerontol. Ser. A Biol. Sci. Med. Sci.* **2013**, *68*, 447–455. [CrossRef]

67. Liu, Y.; Liu, Z.; Wei, M.; Hu, M.; Yue, K.; Bi, R.; Zhai, S.; Pi, Z.; Song, F.; Liu, Z. Pharmacodynamic and urinary metabolomics studies on the mechanism of Schisandra polysaccharide in the treatment of Alzheimer's disease. *Food Funct.* **2019**, *10*, 432–447. [CrossRef]

68. Gronek, P.; Balko, S.; Gronek, J.; Zajac, A.; Maszczyk, A.; Celka, R.; Doberska, A.; Czarny, W.; Podstawski, R.; Clark, C.C.T.; et al. Physical Activity and Alzheimer's Disease: A Narrative Review. *Aging Dis.* **2019**, *10*, 1282–1292. [CrossRef]

69. Maalouf, N.M.; Cameron, M.A.; Moe, O.W.; Adams-Huet, B.; Sakhaee, K. Low urine pH: A novel feature of the metabolic syndrome. *Clin. J. Am. Soc. Nephrol. CJASN* **2007**, *2*, 883–888. [CrossRef]

70. Kraut, J.A.; Madias, N.E. Metabolic Acidosis of CKD: An Update. *Am. J. Kidney Dis. Off. J. Natl. Kidney Found.* **2016**, *67*, 307–317. [CrossRef]

71. Otaki, Y.; Watanabe, T.; Takahashi, H.; Hasegawa, H.; Honda, S.; Funayama, A.; Netsu, S.; Ishino, M.; Arimoto, T.; Shishido, T.; et al. Acidic urine is associated with poor prognosis in patients with chronic heart failure. *Heart Vessel.* **2013**, *28*, 735–741. [CrossRef]

72. Shimodaira, M.; Okaniwa, S.; Nakayama, T. Fasting Single-Spot Urine pH Is Associated with Metabolic Syndrome in the Japanese Population. *Med. Princ. Pract. Int. J. Kuwait Univ. Health Sci. Cent.* **2017**, *26*, 433–437. [CrossRef] [PubMed]

73. Hara, S.; Tsuji, H.; Ohmoto, Y.; Amakawa, K.; Hsieh, S.D.; Arase, Y.; Nakajima, H. High serum uric acid level and low urine pH as predictors of metabolic syndrome: A retrospective cohort study in a Japanese urban population. *Metab. Clin. Exp.* **2012**, *61*, 281–288. [CrossRef] [PubMed]

74. Kistner, S.; Rist, M.J.; Krüger, R.; Döring, M.; Schlechtweg, S.; Bub, A. High-Intensity Interval Training Decreases Resting Urinary Hypoxanthine Concentration in Young Active Men-A Metabolomic Approach. *Metabolites* **2019**, *9*, 137. [CrossRef]

75. Di Liegro, C.M.; Schiera, G.; Proia, P.; Di Liegro, I. Physical Activity and Brain Health. *Genes* **2019**, *10*, 720. [CrossRef] [PubMed]

76. Deutsch, L.; Osredkar, D.; Plavec, J.; Stres, B. Spinal Muscular Atrophy after Nusinersen Therapy: Improved Physiology in Pediatric Patients with No Significant Change in Urine, Serum, and Liquor ^{1}H-NMR Metabolomes in Comparison to an Age-Matched, Healthy Cohort. *Metabolites* **2021**, *11*, 206. [CrossRef]

77. Beckonert, O.; Keun, H.C.; Ebbels, T.M.; Bundy, J.; Holmes, E.; Lindon, J.C.; Nicholson, J.K. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat. Protoc.* **2007**, *2*, 2692–2703. [CrossRef] [PubMed]

78. Deutsch, L.; Stres, B. The Importance of Objective Stool Classification in Fecal ^{1}H-NMR Metabolomics: Exponential Increase in Stool Crosslinking Is Mirrored in Systemic Inflammation and Associated to Fecal Acetate and Methionine. *Metabolites* **2021**, *11*, 172. [CrossRef]

79. Madrid-Gambin, F.; Oller-Moreno, S.; Fernandez, L.; Bartova, S.; Giner, M.P.; Joyce, C.; Ferraro, F.; Montoliu, I.; Moco, S.; Marco, S. AlpsNMR: An R package for signal processing of fully untargeted NMR-based metabolomics. *Bioinformatics* **2020**, *36*, 2943–2945. [CrossRef]

80. Wishart, D.S.; Guo, A.; Oler, E.; Wang, F.; Anjum, A.; Peters, H.; Dizon, R.; Sayeeda, Z.; Tian, S.; Lee, B.L.; et al. HMDB 5.0: The Human Metabolome Database for 2022. *Nucleic Acids Res.* **2022**, *50*, D622–D631. [CrossRef]

81. Emwas, A.H.; Saccenti, E.; Gao, X.; McKay, R.T.; Dos Santos, V.A.P.M.; Roy, R.; Wishart, D.S. Recommended strategies for spectral processing and post-processing of 1D 1 H-NMR data of biofluids with a particular focus on urine. *Metab. Off. J. Metab. Soc.* **2018**, *14*, 31. [CrossRef]

82. Dieterle, F.; Ross, A.; Schlotterbeck, G.; Senn, H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. *Anal. Chem.* **2006**, *78*, 4281–4290. [CrossRef] [PubMed]

83. Legendre, P.; Legendre, L.F.J. *Numerical Ecology*, 3rd ed.; Elsevier: Amsterdam, The Netherlands, 2012; Volume 24, p. 1006.

2.1.7 Urine and fecal ^{1}H-NMR metabolomes differ significantly between pre-term and full-term born physically fit healthy adult males

Deutsch L., Debevec T., Millet G.P.., Osredkar D., Opara S., Šket R., Murovec B., Mramor M., Plavec J. Stres B. 2022. Urine and fecal ^{1}H-NMR metabolomes differ significantly between pre-term and full-term born physically fit healthy adult males. Metabolites, 12: X, doi. https://doi.org/10.3390/metabo12060536, 23 p.

Abstract

Preterm birth (before 37 weeks gestation) accounts for ~10% of births worldwide and remains one of the leading causes of death in children under 5 years of age. Preterm born adults have been consistently shown to be at an increased risk for chronic disorders including cardiovascular, endocrine/metabolic, respiratory, renal, neurologic, and psychiatric disorders that result in increased death risk. Oxidative stress was shown to be an important risk factor for hypertension, metabolic syndrome and lung disease (reduced pulmonary function, long-term obstructive pulmonary disease, respiratory infections, and sleep disturbances). The aim of this study was to explore the dif-ferences between preterm and full-term male participants' levels of urine and fecal proton nuclear magnetic resonance (^{1}H-NMR) metabolomes, during rest and exercise in normoxia and hypoxia and to assess general differences in human gut-microbiomes through metagenomics at the level of taxonomy, diversity, functional genes, enzymatic reactions, metabolic pathways and predicted gut metabolites. Significant differences existed between the two groups based on the analysis of ^{1}H-NMR urine and fecal metabolomes and their respective metabolic pathways, enabling the elucidation of a complex set of microbiome related metabolic biomarkers, supporting the idea of distinct host-microbiome interactions between the two groups and enabling the efficient classification of samples; however, this could not be directed to specific taxonomic characteristics.

For my personal contributions as a doctoral student and author of this thesis, please refer to Table 2 (page 142). The hypothesis from section 1.4.1 from this work were discussed in this paper.

metabolites

MDPI

Article

Urine and Fecal ^{1}H-NMR Metabolomes Differ Significantly between Pre-Term and Full-Term Born Physically Fit Healthy Adult Males

Leon Deutsch [1], Tadej Debevec [2,3], Gregoire P. Millet [4], Damjan Osredkar [5,6], Simona Opara [1], Robert Šket [7], Boštjan Murovec [8], Minca Mramor [9], Janez Plavec [10] and Blaz Stres [1,3,11,*]

1   Department of Animal Science, Biotechnical Faculty, University of Ljubljana, SI-1000 Ljubljana, Slovenia; leon.deutsch@bf.uni-lj.si (L.D.); simona.konda@gmail.com (S.O.)
2   Faculty of Sports, University of Ljubljana, SI-1000 Ljubljana, Slovenia; tadej.debevec@fsp.uni-lj.si
3   Department of Automation, Biocybernetics and Robotics, Jožef Stefan Institute, SI-1000 Ljubljana, Slovenia
4   Institute of Sport Sciences, University of Lausanne, CH-1015 Lausanne, Switzerland; gregoire.millet@unil.ch
5   Department of Pediatric Neurology, University Children's Hospital, University Medical Centre Ljubljana, SI-1000 Ljubljana, Slovenia; damjan.osredkar@kclj.si
6   Faculty of Medicine, University of Ljubljana, SI-1000 Ljubljana, Slovenia
7   Institute for Special Laboratory Diagnostics, University Children's Hospital, University Medical Centre Ljubljana, SI-1000 Ljubljana, Slovenia; robert.sket@kclj.si
8   Faculty of Electrical Engineering, University of Ljubljana, Jamova 2, SI-1000 Ljubljana, Slovenia; bostjan.murovec@fe.uni-lj.si
9   Department of Infectious Diseases, University Medical Centre Ljubljana, SI-1000 Ljubljana, Slovenia; minca.mramor@kclj.si
10   National Institute of Chemistry, NMR Center, SI-1000 Ljubljana, Slovenia; janez.plavec@ki.si
11   Institute of Sanitary Engineering, Faculty of Civil and Geodetic Engineering, University of Ljubljana, SI-1000 Ljubljana, Slovenia
*   Correspondence: blaz.stres@bf.uni-lj.si; Tel.: +386-4156-7633

check for updates

Abstract: Preterm birth (before 37 weeks gestation) accounts for ~10% of births worldwide and remains one of the leading causes of death in children under 5 years of age. Preterm born adults have been consistently shown to be at an increased risk for chronic disorders including cardiovascular, endocrine/metabolic, respiratory, renal, neurologic, and psychiatric disorders that result in increased death risk. Oxidative stress was shown to be an important risk factor for hypertension, metabolic syndrome and lung disease (reduced pulmonary function, long-term obstructive pulmonary disease, respiratory infections, and sleep disturbances). The aim of this study was to explore the differences between preterm and full-term male participants' levels of urine and fecal proton nuclear magnetic resonance (^{1}H-NMR) metabolomes, during rest and exercise in normoxia and hypoxia and to assess general differences in human gut-microbiomes through metagenomics at the level of taxonomy, diversity, functional genes, enzymatic reactions, metabolic pathways and predicted gut metabolites. Significant differences existed between the two groups based on the analysis of ^{1}H-NMR urine and fecal metabolomes and their respective metabolic pathways, enabling the elucidation of a complex set of microbiome related metabolic biomarkers, supporting the idea of distinct host-microbiome interactions between the two groups and enabling the efficient classification of samples; however, this could not be directed to specific taxonomic characteristics.

Keywords: premature birth; ^{1}H-NMR metabolomics; hypoxia; fecal metagenomics; biomarkers; activity; hypoxia

1. Introduction

Preterm birth, defined as a birth before 37 weeks gestation, accounts for approximately 10% of births worldwide. Four degrees of preterm birth are known: extreme preterm (before 28 weeks), very preterm (28–31 weeks), mild preterm (32–33 weeks) and moderate

preterm (34–36 weeks) [1]. While the mechanisms underlying preterm birth are complex, with risk factors comprising infection, cervical disease, uterine over-distention, stress and placental disorders [2,3] it remains one of the leading causes of death in children under 5 years of age [4,5]. Improved neonatology care has led to significantly increased pre-term born survival rates over the last 50 years [6]. Importantly, preterm born adults have been consistently shown to be at an increased risk for chronic disorders involving various organ systems, including cardiovascular, endocrine/metabolic, respiratory, renal, neurologic, and psychiatric disorders. These disorders either persist from infancy into adulthood or sometimes even appear in adulthood and result in a moderately (30% to 50%) increased risk of death in early to mid-adulthood in preterm compared to full-term, and even higher risks among those born at the earliest gestational ages [3]. Preterm infants were also shown to experience an imbalance between oxidants and antioxidant capacity [7]. Oxidative stress was shown to be an important risk factor for hypertension, metabolic syndrome (diabetes mellitus, dyslipidemia), lung disease (reduced pulmonary function, long-term obstructive pulmonary disease, respiratory infections, and sleep disturbances) [3].

In addition to the population level disease metrics, preterm born individuals have increased body fat mass, arterial blood pressure, and higher fasting glucose, insulin, and cholesterol levels [3,8,9]. Elevated levels of low-density lipoprotein in preterm individuals pose a greater risk of developing atherosclerosis or cardiovascular disease [10]. Preterm born individuals also experience problems with renal function due to altered nephron development [11]. Imbalance in the ratio between reactive oxygen species (ROS) and antioxidants was identified as conductive to oxidative stress [12], which was associated with increased molecular damage [13]. ROS overproduction was reported to be induced in hypoxia by the xanthine oxidase pathway, catecholamine production and increased rate of electron leakage within the mitochondria [14–18]. Taken together, past research showed that significant differences existed between preterm and full-term born adults with respect to oxidative stress induced by hypoxia, activity, or exercise [7,13,14,19,20].

Preterm birth was also shown to induce life-long pulmonary system effects and compromise ventilator control resulting in blunted hypoxic ventilatory response (HVR) in preterm infants. The PreTerm project (Slovenian Research agency (ARRS) project # J3-7536 (D), Figures S1–S3) was devised to explore whether the differences and impairments in HVR persisted with aging in physically fit young men. The differences in HVR responses between preterm born adults and their age matched full-term controls were explored during rest and exercise, in normoxia and hypoxia [21]. Hypoxia was shown to provoke a similar relative reduction in maximal aerobic power and submaximal ventilatory threshold in healthy preterm and full-term born matched controls with comparable peak oxygen consumption levels. These data suggested that exercising in normobaric hypoxia does not exert a higher ventilator and metabolic load in otherwise healthy physically fit individuals born prematurely [21]. Only recently was the post-exercise accumulation of interstitial lung water shown to be higher in adults born prematurely in hypobaric hypoxia, than in normobaric hypoxia [22].

Given the complexity of the human body and its responses to chronically elevated oxidative stress levels that may persist into adulthood and consequently contribute to the development of numerous noncommunicable diseases observed in the preterm population (diabetes, hypertension or lung disorders) [7], the systemic bodily matrices, such as urine and feces remain surprisingly unexplored by powerful high-throughput top-down analytical approaches [23]. To fill this gap, the aim of this study was to explore differences between preterm and full-term participants' urine and fecal ^{1}H-NMR metabolomes and respective enzymatic reactions, during rest and exercise, in normoxia and hypoxia. In addition, metagenomic analysis of human gut-microbiomes was conducted to assess the differences in the human gut microbiome taxonomy, diversity, functional genes, enzymatic reactions, metabolic pathways and predicted gut metabolites. In this work we hypothesized that significant differences exist between the preterm and full-term groups at the levels of

multivariate physiological initial states and their respective responses to the tests conducted (rest or exercise; normoxia or hypoxia) [24].

The analyses of large metabolomic and metagenomic datasets were hypothesized to enable the detection of the characteristic differences between the two groups at information levels not utilized before in preterm research, in addition to the efficient selection of more complex sets of biomarkers by utilizing machine learning and the exploration of vast algorithm spaces.

2. Results and Discussion

2.1. Group Characteristics in Relation to Gut Physiological Data

Thirty-seven men volunteered for this study and were divided into two groups based on the mode of delivery. Fifteen participants were born at term (full-term control) and 22 were born prematurely (preterm). Control participants were 22 ± 2 years old, weighed 76 ± 6 kg, were 180 ± 5 cm tall, had a VO_2 max of 52 ± 5 mL kg^{-1} min^{-1}, and were born at 39 ± 2 weeks. Preterm participants were 21 ± 2 years old, weighed 69 ± 7 kg, were 175 ± 7 cm tall, had a VO_2 max of 48 ± 6 mL kg^{-1} min^{-1}, and were born at 29 ± 3 weeks. Table S1 shows their baseline data [24]. Gestational age was statistically different between the two groups [24]. Twenty-one preterm and 13 full-term participants were included in the metabolomic and metagenomic part of the PreTerm study. Two full-term and one preterm participant did not collect urine and fecal samples and were excluded from this part of the PreTerm study.

Physiological exercise tests from the preterm project were already published before and showed that incremental cycling in normoxia and hypoxia resulted in increased levels of advanced oxidation protein levels (AOPP), catalase (CAT), superoxide dismutase (SOD), and nitrosative stress markers in both groups (preterm and full-term) immediately after exercise [24]. No differences were observed between normoxic and hypoxic environments. However, hypoxic exposure itself resulted in a significant increase in AOPP, and CAT and showed a trend toward an increase in the nitrosative markers control group only, but not in the preterm group. Further, in line with the above observations, the metabolic response to hypoxia may be blunted in adult preterm born adults [24]. Periodic breathing (repeated oscillations of hyperventilation followed by an apneic phase) was also different in the preterm group than in full-born adults, suggesting a possible physiological mechanism [25]. The hypoxic ventilatory response at rest was lower in preterm, but no differences in exercise were observed between the two groups [21]. Preterm born adults experienced reduced physical capacity in normoxia compared to those full-term born and have a lower hypoxic ventilatory response (HVR, ability to change ventilation in the function of blood oxygen saturation), while no such difference was observed under hypoxic conditions [21,26–31]. These reports show that preterm individuals nevertheless exhibited increased oxidative stress, antioxidant activity, and NO metabolism in acute exercise. However, under hypoxic conditions, the preterm group did not exhibit increased levels of plasma advance oxidation protein products (AOPP), catalase, and nitrosative stress markers (NOx) levels, indicating a possibly greater activation of responses resisting oxidative stress under hypoxic conditions [24,32,33].

Based on the integration of past findings obtained utilizing the same cohorts within the PreTerm project, we speculated that measurable differences existed also in the makeup of the intestinal tract characteristics. Various physiological characteristics of the gut environment were previously associated with numerous non-communicable diseases [34]. To assess these differences in the intestinal environment between preterm and full-term control, 25 additional variables were measured in human feces (Figure 1). The obtained results, surprisingly suggest that no significant differences existed in the measured intestinal parameters between the preterm and full-term participants (Permutational analysis of variance (PERMANOVA); $p > 0.05$; n permutations = 5000). In addition, no difference existed between the gut environmental characteristics before and after normoxic and hypoxic test periods (PERMANOVA; $p < 0.05$; n permutations = 5000). The use of a 3-day

sampling series enabled us to conclude that the set of 25 measured parameters reported in this study was either insufficient or not measured at an appropriate scale to detect significant differences between the intestinal tracts of the participants from the two groups. As many of the measured parameters were previously effective for detection of differences in the intestinal environment of the participants involved in the three-week bed-rest campaigns of our previous Planetary Habitat Simulation Project (PlanHab) [35–37], the results suggest a lack of long-term differences between physically fit preterm and full-term young male participants at the level of the measured intestinal parameters. The role of matching physical fitness between preterm and full-term groups for health maintenance shows that the differences in the status of intestinal tract environments were notably smaller between the active young males irrespective of the preterm birth and oxidative stress markers detected. From this, a different set of additional parameters arise (e.g., zonulin, α1-antitrypsin, eosinophile-derived neurotoxin, bile acid and derivatives, ionic strength, redox potential, mucus characteristics) focusing more intensively on the gut-feces interface and its interaction with the host. This should be used in future studies focusing on immunological, ion-selective and electrochemical characteristics next to spectral and excitation-emission analyses of dissolved organic compounds in the intestinal tract [35–39]. In contrast to our past work utilizing participants exposed to a tightly controlled environment, diet, water intake, circadian rhythm and level of exercise (in the PlanHab project [35–38]) the PreTerm project interpersonal variability in the same types of variables might have obscured differences in the intestinal parameters associated with individual lifestyle and food preferences in relation to exercise. The significant differences in physiological parameters measured in the PreTerm project were not reproduced in the measurements of the intestinal parameters in the same participants as described above.
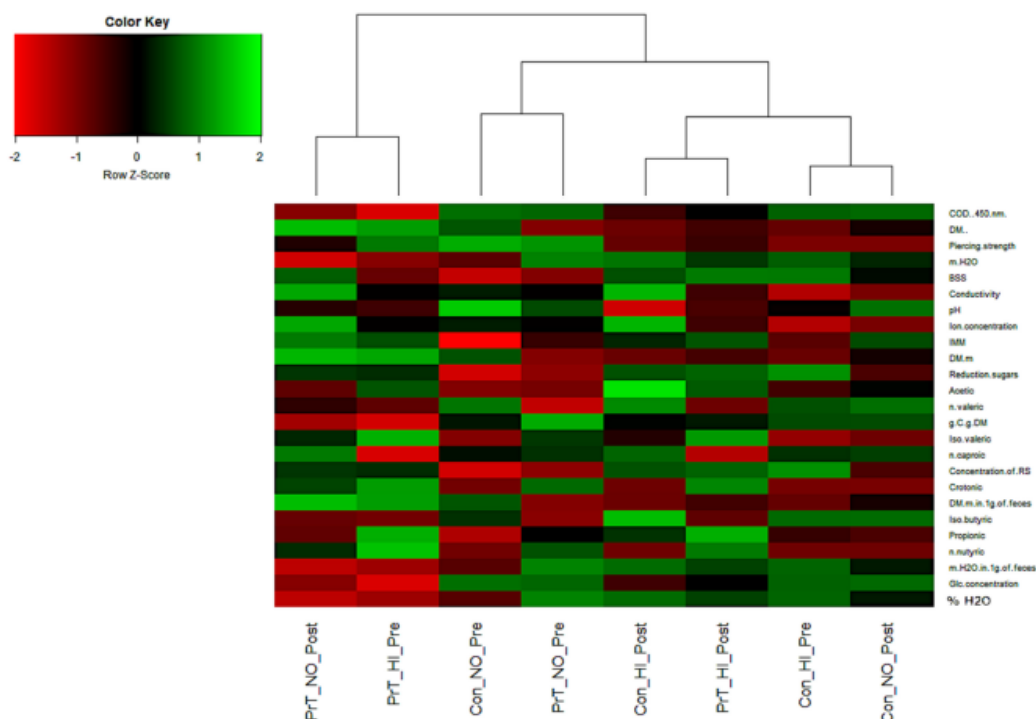
Figure 1. A heatmap summarizing the differences in the 25 measured parameters (COD—chemical oxygen demand, DM—dry matter, m—mass (g), BSS—Bristol stool scale, IMM—molecular mass index, C—carbon, RS—reduction sugars, Glc—glucose) describing the intestinal environment of the preterm and full-term groups exposed to distinct training regimes of the PreTerm project (Figures S1–S3). No significant difference was observed (PERMANOVA; $p > 0.05$; n permutations = 5000).

2.2. Multivariate Relationships in Urinary and Fecal ^{1}H-NMR Metabolomes Supported Significant Differences between Preterm and Full-Term Groups

Urine and fecal samples from the preterm and full-term groups were collected on three consecutive days before and three consecutive days after the hypoxic and normoxic tests (Figures S2–S4). In total, each participant was characterized by 12 samples (three daily consecutive samples before and after normoxic and hypoxic tests). Nonparametric analyses utilizing one-way and two-way PERMANOVA on either urine or fecal identified metabolites showed that preterm and full-term groups differed significantly ($p < 0.001$) at both metabolomic levels. In two-way PERMANOVA the training condition (hypoxic vs. normoxic) was marginally significant for urinary metabolomes ($p = 0.05$), but not for fecal metabolomes. This shows that rapid changes in human physiology take place upon the introduction of exercise, while longer bouts of exercise (weeks) would be needed to detect larger differences between physically fit preterm and full-term participants at both levels, similar to the PlanHab project [35–38]. Differences in the numbers of detected metabolites per group, test and time of sample collection, and the sum of their concentrations in all studied groups were not significantly different (Table S2). In summary, significant differences were identified in the overall makeup of urinary and fecal metabolomes between the preterm and full-term groups.

2.3. Urine ^{1}H-NMR Metabolomics

The identified differences between the preterm and full-term groups at the level of urinary ^{1}H-NMR metabolomes were explored in more detail. When comparing the urine metabolomes (ESM 2) of preterm and control participants using the PLSDA method integrated into MetaboAnalyst [40] and based on cross-validation, three components were recommended to distinguish between the two groups explaining 24% of the variation (Figure 2a,b). Acetone, tartrate, and trans-aconitate [41] were the first three of the most differentiating metabolites in urine, and all three were elevated in the control group based on VIP scores. Acetone metabolism is part of two pathways, the decarboxylation of acetoacetate that is generated during dextrose metabolism and lipolysis, or the dehydrogenation of 2-propanol. Its concentrations in exhaled breath have previously been shown to correlate strongly with acetone concentrations in the blood, as well as with other ketones and were affected by fasting, exercise, and/or disease (e.g.,) diabetes mellitus [42]. Tartrate is part of glyoxylate and dicarboxylate metabolism (Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway: ko00630 (accessed on 20 May 2022) while trans-aconitate (accompanied by creatinine) as the metabolite related to the tricarboxylic acid cycle was indicative of differences in exercise capacity [43].

D-arginine and D-ornithine metabolism, synthesis and degradation of ketone bodies (acetone), and the Warburg effect were the most enriched metabolic pathways (mostly associated with the preterm group) in the urine metabolome (Table S3, Figure 2c,d, Figures S5 and S6). Our results from the metabolome analysis were compared with the metabolomes of specific urinary disease pathways within MetaboAnalyst. Tentatively, the most interesting enriched pathways identified in the preterm group were described before in relation to systemic or tissue hypoxia (Table S4) [44,45].

The MetaboAnalyst PLSDA analysis reported decreased levels of acetone, trans-aconitate and tartrate in the preterm group, irrespective of their matching physical fitness [21,24]. To the best of our knowledge our results could be compared to a single existing study reporting significantly different sets of markers, such as citrate, hippurate, creatinine, and fumarate as crucial metabolites responsible for the differentiation of preterm adults from full-term adults [46]; however, in that study participants of matching fitness and exercise tests were not included.

115
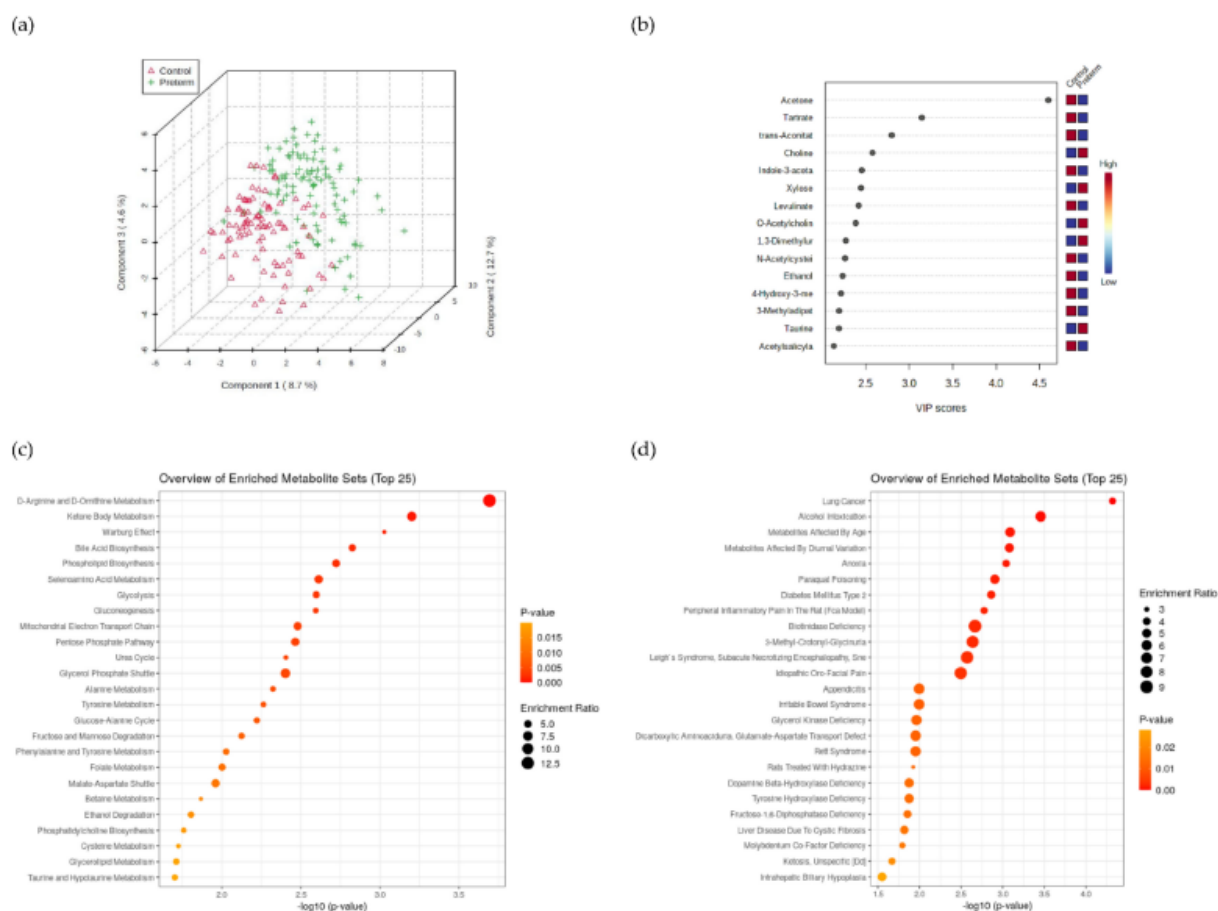
Figure 2. (a) Urine metabolomes comparing preterm and full-term born adults and the most differentiating metabolites based on PLSDA method (**b**). The most enriched pathways associated with metabolism (**c**) and diseases (**d**) based on urinary metabolomes. Enlarged (**c**,**d**) figures were added to supplementary (Figures S5 and S6).

In the present study, physically fit preterm and full-term participants were enrolled in physical exercise tests and exhibited physical performance indistinguishable between the two groups, i.e., preterm born participants' physical capacity (expressed as peak oxygen consumption) was not impaired in comparison to normal-born participants [7,21]. In contrast, other studies reported that the physical performance of preterm individuals was lower in normoxia and hypoxia [19] referring to a cohort sampled from the general population. Our study observed a lack of difference in the physical capacity related to the physical fitness of the participants in both groups, further emphasizing the importance of exercise for the maintenance of physical health, while at the same time noticing the differences in metabolic makeup of the two groups at the level of urine. These differences apparently stem from impaired autonomic function as heart rate recovery seems slower in preterm adults and could give rise to anoxia and increase their cardiovascular risk as suggested before [47,48].

2.4. ^{1}H-NMR Metabolomics of Fecal Content

To match urine sample collection, fecal samples were collected in four 3-day series as described below (Figures S2 and S3, ESM 3). In total, 12 samples were collected per person for the fecal matrix. In contrast to urine metabolomes, two components were sufficient to differentiate preterm and full-term groups in fecal metabolomes by the PLSDA

method using MetaboAnalyst [40] (Figure 3a,b). Fecal biomarkers lactate, tyrosine, and serotonin were identified as the three most efficient for differentiation between the two groups. Lactate and serotonin were significantly elevated in the preterm group while tyrosine was decreased. Metabolite set enrichment analysis (MSEA) coupled with the PLSDA reported increased lactate concentrations in the preterm group and reported that pyruvate metabolism and the Warburg effect were enriched in the preterm group. The Warburg effect was also previously associated with mitochondrial dysfunction, which also occurs in preterm infants (Table S5) [49,50]. Thyroid hormone synthesis and catecholamine biosynthesis were the first two most enriched metabolic pathways according to MSEA. In addition, an extended list of fecal metabolites analyzed with MSEA was previously correlated with fecal diseases, such as ileal Crohn's disease, and irritable bowel disease [51] as the most enriched metabolic pathways (Figure 3c,d, Figures S7 and S8, Table S6).
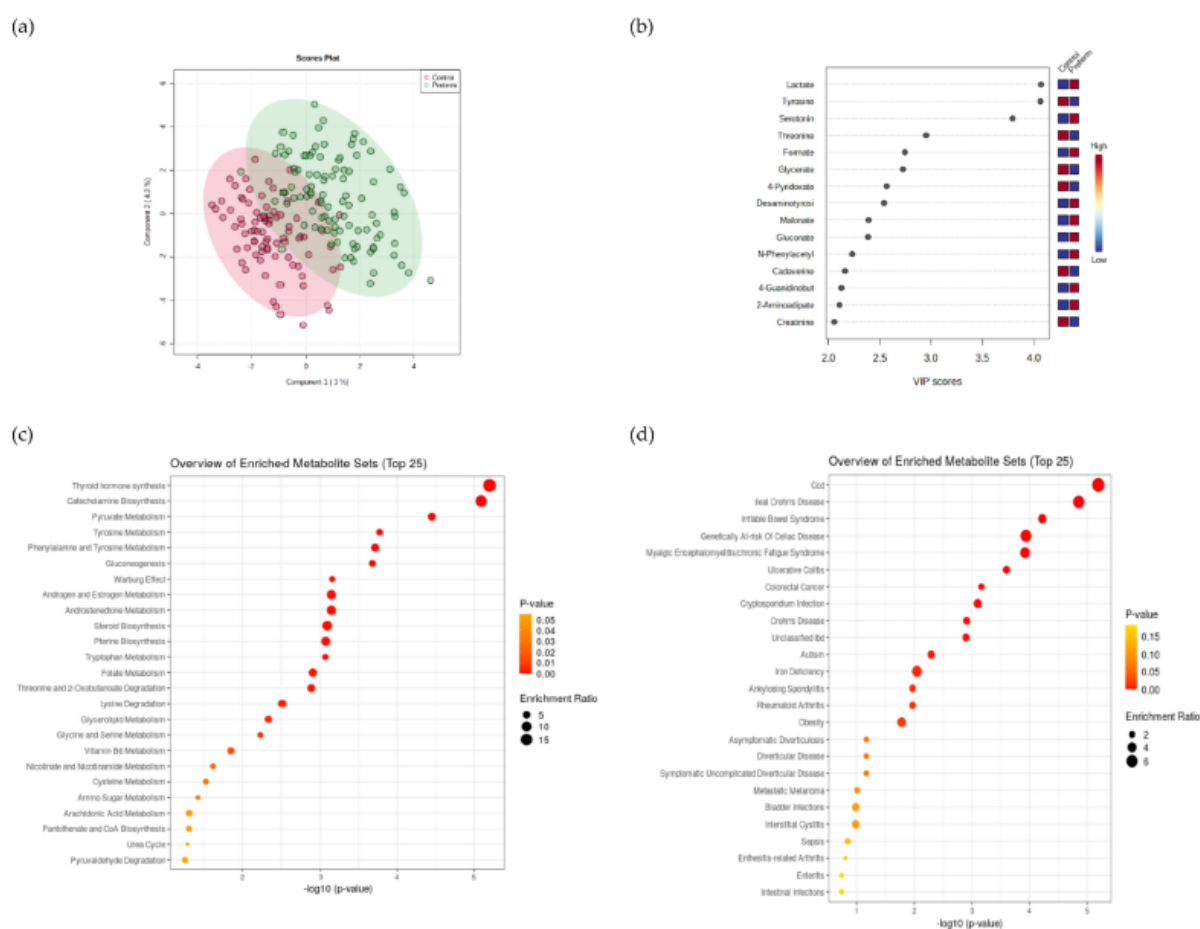
Figure 3. (**a**) Fecal metabolomes comparing preterm and full-term born adults and the most differentiating metabolites based on the PLSDA method (**b**). The most enriched pathways associated with metabolism (**c**) and diseases (**d**) based on fecal metabolomes. Enlarged (**c**,**d**) figures were added to supplementary (Figures S7 and S8).

These results constitute the first report on the significant differences in the metabolomics makeup of fecal samples between the preterm and full-term control groups, irrespective of the observed lack of differences in the 25 measured parameters of the intestinal tract (Figure 1). These results represent possibly the first evidence that systemic differences due to life-long exposure to oxidative stress actually exist and raise the question of whether these differences are linked to minute differences produced from the side of the preterm

host or from the side of the microbiome responding to these environmental signals or their mutual interaction in the form of a complex biochemical network steady state.

The rather small extent of variation (7% and 25%) between the two groups could be explained by this approach utilizing feces and urine, respectively, suggesting further and multiple sources of variation exist beyond those described in this study. We extended our interrogation of the data to provide an estimate of the cohort size that would need to be utilized in future experiments. Based on the power analysis module in Metabo-Analyst [40] at least two orders of magnitude larger cohorts amounting to a couple of thousand participants would be required in order to better discern differences at the level of fecal metabolomes. These results point to a conclusion that although fecal metabolomics makeup in physically fit young male participants was significantly different from their matched controls, these differences were independent of the normoxic or hypoxic nature of the tests (PERMANOVA; $p > 0.05$; n permutations = 5000). Apparently, characteristic long-term differences exist between the two groups at the level of fecal metabolomes, most probably linked to the fact that preterm individuals experienced increased oxidative stress, and responded with elevated antioxidant activity, and NO metabolism in the acute exercise studies reported before [24], resulting in the characteristic differences in their fecal metabolomes observed in this study.

2.5. Fecal Metagenomics: From Taxonomy, Functional Genes to Predicted Metabolomes

Significant differences in urine and fecal ^{1}H-NMR metabolomes between the preterm and full-term groups prompted us to explore whether significant differences exist at the level of the human gut microbiome. Fecal samples collected during the PreTerm project were used for shotgun sequencing. The in-house analytical pipeline utilizing bioBakery [52] was used to preprocess sequence data (Kneaddata (https://huttenhower.sph.harvard.edu/kneaddata/, accessed on 7 April 2022) and analyze the sequences at the strain level of taxonomy (MetaPhlAn3 [53]), diversity (mothur [54]), functional genes, enzymatic reactions and metabolic pathways (HUMAnN3 [53]) next to predicted metabolites (MelonnPan [55]). In total, 853 taxonomic units (kingdoms, phyla, clades, orders, families, genera, and species), 30 diversity calculators, 198,305 gene families, 183,200 enzymatic reactions, 10,974 metabolic pathways, and 80 metabolites present in the human gut microbiota were identified and analyzed. In total, 393,442 variables were considered in this search for differences between the preterm and full-term groups. Each dataset corresponding to a layer of information was analyzed separately using JADBio extensive machine learning modeling as described before [39,56,57].

2.5.1. Taxonomy and Microbial Diversity of Intestinal Tract

In contrast to the observed differences at the urine and fecal metabolomics levels described above, the taxonomic level of information did not result in significant differences between groups (PERMANOVA; $p > 0.05$; n permutations > 5000). In addition, based on the taxonomic data 181,020 JADBio models were trained using an extensive tuning effort, but no reliable biomarker or trained model could be obtained. In general, three different kingdoms were detected in all samples (archaea, bacteria, and DNA viruses). The average relative abundance of bacteria was lower in the preterm than in the full-term group (92.3% vs. 77.5%). The abundance of DNA viruses was higher in the preterm group (7.2% vs. 20%) (Figure S9a). The archaeal kingdom was least prevalent in both groups (0.5% in the control group vs. 2.5% in the preterm group). This lack of significant differences at taxonomic levels was previously attributed to large interpersonal differences between participants [58,59]; however, in this study variance within the full-term control group was at least two times larger than that observed in the preterm group and hence significantly higher ($p < 0.05$) based on the analysis of 3D coordinates after nmMDS and PCoA analysis. This points to the existence of an overarching effect shared by all preterm participants absent from the matching control full-term group.

We further point out that both groups contained matching groups of young healthy physically fit participants, in contrast to past studies exploring the differences between the preterm and general population [60]. Our results point to the fact that some microbiome-related characteristics within the preterm group were apparently shared to a larger extent within the preterm group in comparison to the full-term control group, signifying the existence of differences in the microbial makeup due to differences in the physiology of the host. The existence of an overarching effect shared by all preterm participants absent in the control group was evident from the significantly higher Shannon diversity ($p < 0.05$) (Figure S9b) in the preterm group including other diversity estimates that differed significantly between the two groups (Table S7). These two observations suggest the existence of an interplay between the increased similarity of the major preterm taxonomic categories and the diversity of a smaller highly variable list of taxa not well shared between the preterm participants in this study. It is easily envisioned that additional environmental factors shape the gut microbiome within the preterm group, making the preterm group a narrower subset of the otherwise healthy human gut.

2.5.2. Functional Genes of Human Gut Microbiome

Based on the 198,306 categories describing gene family data, in the preterm and control groups, 90,510 models were trained using an extensive tuning effort in search of biologically meaningful discriminative variables between the preterm and full-term control groups. The entire list of features was used to build and validate a trained model that achieved insignificant validation performance with an AUC and other metrics. Consequently, no significant differences could be identified between the two datasets at the level of functional gene lists.

2.5.3. Enzymatic Reactions Taking Place in Human Gut

For the aggregation of functional gene information into enzymatic reactions (Figure 4a), again 90,510 models were explored using an extensive tuning effort. The best model was ridge logistic regression with the penalty hyperparameter lambda = 1.0, with an area under the curve (AUC) value of 0.992. In addition to AUC, all other thresholds were also statistically significantly different from the baseline. Features were selected based on the Test-Budgeted Statistically Equivalent Signature (SES) algorithm with the following hyperparameters: maxK = 3, alpha = 0.1, and budget = 3 $*n_{vars}$. RXN-15378, RXN-14971, RXN-21393, and RXN-21394 were equally selected as the most important for discriminating between the preterm and control groups and were increased in the preterm group (Figure S10) ($p < 0.05$). All of the above reactions represent the enzymatic reaction succinate dehydrogenase based on the BioCyc website ([61], accessed on 23 January 2022). We used RXN-15378 to validate the trained model and obtained a validation performance with an AUC of 0.931. Succinate is a metabolite produced by both host and microbial cells and accumulates under conditions of inflammation and microbial imbalances in the intestinal tract [62]. Succinate was shown to accumulate in areas of inflammation and metabolic stress [63] and can have tissue specific but also systemic effects as a proinflammatory signaling molecule [62,64–66]. Although gut microbes represent the predominant source of succinate, it is typically rapidly consumed in the production of propionate, one of the major short chain fatty acids, by *Bacteroides* spp., *Prevotella* spp. and some members of Firmicutes [62,67]. Although the mucosal uptake of succinate as a charged molecule over the mucosal epithelia is significantly higher in the small intestine, it takes place to various extents throughout the length of the intestinal tract and requires sodium dependent transport proteins [62]. In addition to the internalized succinate provided by the microbiome, succinate also accumulates within cells under conditions of low oxygen as a metabolic signature of hypoxia, generating HIF-1α to regulate cellular responses and adapt to a low oxygen environment. At normoxia, HIF-1α is regulated by posttranslational hydroxylation and degradation by prolyl-hydroxylase activity that converts alpha-keto-glutarate to CO_2 and succinate while inactivating HIF-1α. Excess uptake of microbiome produced succinate

results in higher levels of intracellular succinate that can slow down prolyl-hydroxylase activity through product inhibition and result in an additional activation and stabilization of HIF-1α beyond its response to hypoxia itself, which can significantly augment the LPS-induced expression of proinflammatory cytokines [62,68].
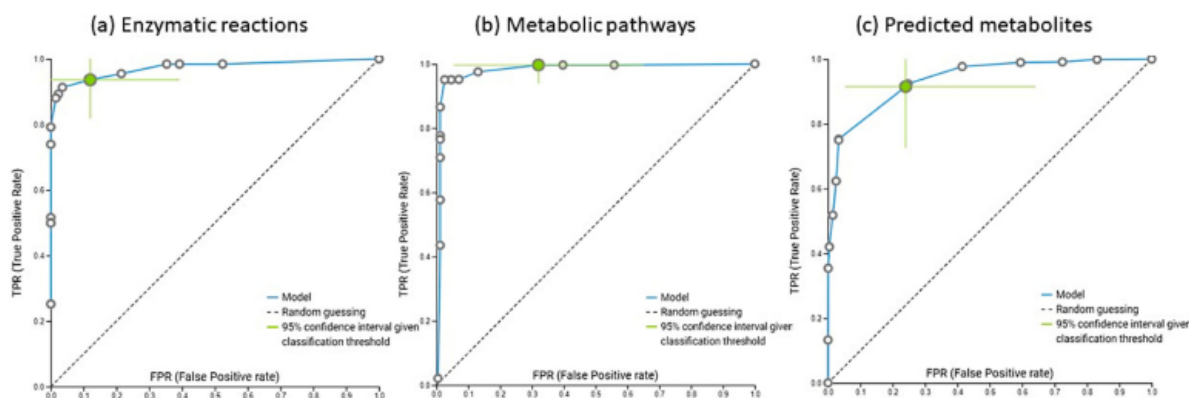
Figure 4. ROC curves of obtained models (JADBio [57]) based on the enzymatic reactions (**a**), metabolic pathways (**b**) and relaxation network predicted metabolites (**c**) produced by our in-house implementation of bioBakery3.

2.5.4. Metabolic Pathways Observed in Human Gut

Based on the metabolic pathway data (Figure 4b), 60,310 models were trained, with extensive tuning effort. The best model was ridge logistic regression with a penalty hyperparameter of 100 and an area under the curve (AUC) of 0.981. In addition to AUC, all other thresholds were also statistically significantly different from baseline. Features were selected on the basis of Lasso feature selection with a penalty = 1.5. On the basis of the MetaCyc website ([69–71]; accessed on 23 January 2022), the most important metabolic pathways were PWY-7456 (β-(1,4)-mannan degradation), PWY-7323 (superpathway of GDP-mannose-derived O-antigen building blocks biosynthesis), GLYCOLYSIS-TCA-GLYOX-BYPASS (a superpathway of glycolysis, pyruvate dehydrogenase, TCA, and glyoxylate bypass), P221-PWY (octane oxidation), and PWY-5173 (unclassified). These pathways were the most important for distinguishing the preterm group from the control group and were all increased in the preterm group (Figure S11). The entire set of selected features was used to validate the trained model and achieved a validation performance with an AUC of 1.00. The relative frequency of this response was significantly increased in the preterm group, which was also confirmed by the t-statistic ($p < 0.05$). A set of selected features was used to validate the trained model and achieved validation performance with an AUC of 1.00.

The β-(1,4)-mannan degradation (PWY-7456) belongs to *Bacteroides fragilis* in the human intestinal tract and is essential for mucosal integrity and host nutrition [72,73]. Degradation of mannan by either *Bacteroides dorei* or *Fecalibacterium prausnitzii* and *Roseburia intestinalis* promotes the growth of *Lactobacillus helveticus* and *Bifidobacterium adolescentis*, which have probiotic properties and promote the synthesis of short chain fatty acids [74] or promote the growth of commensal microbes [75,76].

The superpathway of GDP-mannose-derived O-antigen building blocks biosynthesis (PWY-7323) is involved in lipopolysaccharide (LPS) biosynthesis. Only gram-negative bacteria have LPS, and O-antigen is the part that extends the polysaccharide away from the cell surface and triggers the host cell immune response [77,78]. Gram-negative bacteria observed in preterm infants cause serious infections, such as sepsis [79] coupled with the absence of MD -2 (a protein responsible for the recognition of LPS), which leads to a higher risk of developing intestinal diseases in adults born preterm due to the impaired recognition of LPS in the past [78]. Elevated LPS levels may also contribute to inflammaging (chronic, low-grade inflammation that develops with age) [80]. This also fits our observation that

microbially produced succinate coupled with hypoxia can significantly augment LPS-induced expression of the proinflammatory cytokines [68].

GLYCOLYSIS-TCA-GLYOX-BYPASS (a superpathway of glycolysis, pyruvate dehydrogenase, TCA, and glyoxylate bypass), is a superpathway that was significantly overrepresented in the preterm group. It integrates some of the fundamental components of energy metabolism, starting with a hexose sugar and ending with CO_2 and several forms of highly reducing metabolites that can be used for adenosine triphosphate (ATP) generation. Even though acetyl-CoA is shown in this superpathway as a product of the glycolysis pathway, it is also generated by the degradation of fats and proteins and by the fermentation of many metabolites. This superpathway includes the glyoxylate cycle, which bypasses those steps in the TCA cycle that lead to a loss of CO_2, and operates in bacteria. The increased energy production in the preterm microbiome apparently coincided with the general characteristics of the preterm individuals, such as increased oxidative stress, elevated antioxidant activity, and NO metabolism in acute exercise as described above [7]. It is possible to suggest that the intestinal conditions experienced by the gut microbiome exerted additional stress on the microbial functioning as well. Further research is needed to corroborate this notion.

In line with our observation of differences in microbiome functioning, P221-PWY (octane oxidation) was shown to increase with the Westernization of the human gut and lifestyle [81,82] as volatile organic compounds, including octane, were found either in the exhaled air or feces of human subjects with diverse medical conditions associated with oxidative stress and chronic inflammation, including lung cancer [82,83] obstructive sleep apnea [84], gastrointestinal diseases [85] and NAFLD [86]. Previous studies demonstrated that several alkane-degrading bacteria were capable of using diverse compounds as a carbon source in addition to alkanes [87], which are further oxidized to fatty acids via the bacterial β-oxidation pathway (BioCyc ID: P221-PWY). The key process in octane oxidation is the alkane hydroxylase system that introduces molecular oxygen in the C1 atom of the hydrocarbons at the expense of NADH to yield primary alcohols [88] that were further linked to liver associated diseases.

The acetyl-CoA biosynthesis (PWY-5173) pathway involved in carbohydrate metabolism was also significantly increased in the preterm group. The resulting acetyl-CoA acts as a precursor in the synthesis of intestinal short chain fatty acids including butyrate and acetate [89], that are important in maintaining gut health [90]. The increased levels of acetyl-CoA biosynthesis fit nicely with the other pathways observed in this study that either contribute or consume mass flow related to this reaction. As both preterm and full-term groups were composed of healthy young physically fit males differing significantly in acetyl-CoA biosynthesis, our findings support a recent report on this pathway being one of the most variable pathways in a survey of subgroups of elite Irish athletes [91].

These overall results of the metabolic pathway analysis point to the fact that (irrespective of the heterogeneous makeup of the underlying microbiome taxonomy within the individual participant) the complex coordinated adjustments to the metabolism of the microbiome nevertheless take place and can be robustly reproduced from the integration of the sequencing information as described in this study [91–94] and can be linked to physiologically meaningful differences between groups reported before [24].

2.5.5. Predicted Water- and Lipid-Soluble Intestinal Metabolites

Our last layer of information dealt with the extended analysis of sequencing data towards water- and lipid- soluble predicted metabolites utilizing relaxation-network analysis, which has been extensively trained and validated before [52,55]. In summary, 17 metabolites (out of 81) (Table S8) predicted with MelonnPan were detected also by ^{1}H-NMR in fecal samples, showing possible interaction between two systemic metabolisms (human and microbial). None of these metabolites were chosen by machine learning. Metabolites associated with the human gut microbiota (Figure 4c) were explored utilizing JADBio and 181,020 models were trained using extensive tuning efforts. The best model was a Support Vector Machine type C-SVC with a radial basis function kernel and hyper-parameter

(cost = 10, gamma = 1.0), with an area under the curve (AUC) value of 0.976. In addition to AUC, all other thresholds were also statistically significantly different from the baseline. Feature selection was based on LASSO feature selection with a penalty = 0.25. Alpha-muricholate, putrescine, dimethyllysine, diacetylspermine, and C16 carnitine were significantly increased in the preterm group. In contrast, hydrocinnamic acid, fructose, glucose and galactose, chenodeoxycholate and deoxycholate were lower in the preterm group (Figure S12). When the trained model was applied to the test portion (30% of our total data set) validation performance with an AUC of 0.957 was obtained. In the following sections let us first review the predicted metabolites significantly increased in the preterm group.

Carnitine was increased in the preterm group and is associated with trimethylamine N-oxide (TMAO) production (Figure S12). TMAO is synthesized by the microbiota from trimethylamine (TMA), which in turn is formed from carnitine or choline. Increased choline content in the preterm group was also observed in urine metabolomics. These two molecules together (carnitine and choline), in conjunction with the microbiota, may be the most important cause of the increased likelihood of cardiovascular disease in the preterm group [8,95–100].

Putrescine and diacetylspermine are polyamines and important metabolites for the gut microbiota (Figure S12). Putrescine is synthesized by interspecies cooperation between *Escherichia coli* and *Enterococcus faeacalis* and is formed from arginine [101,102]. Elevated putrescine levels have been associated with an older gut microbiota [103], increased gut permeability, and elevated levels of inflammatory cytokines in mouse colon tissue [104]. Elevated putrescine levels led to activation of genes that regulate oxidative stress, which may lead to a parallel increased risk of developing metabolic syndrome [35–38] and irritable bowel syndrome [105].

Diacetylspermine as a polyamine metabolite was linked to cancer growth and its association with microbial biofilm formation. It is synthesized by bacterial acetylation and has been significantly upregulated in tissues with biofilms in animal models [106]. This suggests that microbial organization and biofilm formation capacity at the interface between the mucus layer and lumen might differ significantly between the preterm and full-term participants, an observation worth further exploration.

Dimethyllysine can be the end product of either host or microbial metabolism but currently little is known about its physiological roles for the host and microbiome in the Human Metabolome Database [107], ChemSpider (https://www.chemspider.com/ (accessed on 15 April 2022)) or FooDB (www.foodb.ca (accessed on 15 April 2022)) (Figure S12). A recent review of macronutrient metabolism by the human gut microbiome focusing on major fermentation byproducts and their impact on host health [108] reported that the major products of lysin were acetate, butyrate and cadaverine, hence linking this compound to the short- and long-chain fatty acid cycles associated with ulcerative colitis [109].

Alpha-muricholic acid was identified by the MelonnPan [55] relaxation network since its first use in the analyses of human samples analyzed using MelonnPan [110,111], suggesting a misclassification of rodent muricholic acid for cholic acid in humans in this approach. Nevertheless, irrespective of its MelonnPan supported assignment, it is evident that this secondary bile acid was identified at elevated levels in the preterm group, fitting into the framework of the distinct chemical makeup of the preterm gut in relation to fat metabolism and the metabolites reported in this study.

In addition to elevated metabolites identified by MelonnPan in the preterm group, the following metabolites were identified in significantly lower concentrations in the preterm group.

Deoxycholate (decreased in the preterm group) is another metabolite that interacts with microbes (Figure S12). Deoxycholate is a secondary bile acid. The human intestinal microbiota (*Bacteroides intestinalis*, *Bacteroides fragilis*, *Escherichia coli*) are involved in the production of secondary bile acids from primary bile acids, such as choline. Deoxycholate is also known to promote colon cancer. Because of the increased cholate levels in the preterm group, we would expect a greater likelihood of microbial metabolites associated with

primary bile acids, as well as increased levels of the expected metabolites in the preterm group. In contrast, deoxycholate levels were decreased in the preterm group. This could also be due to the greater urinary excretion of cholate (cholate was increased in the urine of preterm infants). Bile acids also generally induce mitochondria ROS production. Preterm infants are challenged by ROS in the first few months of life, possibly implying that the systemic response is to increased urinary excretion of bile acids in preterm infants [112–115].

Hydrocinnamic acids are a major class of phenolic acids from dietary fiber with the characteristic phenylpropanoid C6-C3 backbone that were significantly decreased in the preterm group (Figure S12). Although the polyphenol–gut microbiota interactions and their impact on human health have been known for decades, there is great inter-individual variation caused by the different individual capabilities of processing, absorbing and using these compounds effectively [116]. In light of the physiological differences between the two groups analyzed in this study, it seems plausible that differences exist also in the extent of the utilization of these polyphenols in the preterm group. In addition, lower levels of hydrocinnamic acid were observed in patients with Crohn's disease and ulcerative colitis compared to the healthy cohort. Lower levels of hydrocinnamic acid in the preterm group may lead to increased levels of circulating BCAAs, which in turn predisposes preterm born individuals to metabolic syndrome and cardiovascular disease [117–120].

The lower levels of reducing sugars fructose, glucose and galactose, in the preterm group corresponded with a greater capacity to produce short chain fatty acids (Figure S12). The metabolic reactions and predicted metabolites jointly suggest the existence of a larger metabolic flow-through of the preterm microbiome in comparison to the full-term group, pointing to significant differences in the environmental setup in the preterm gut.

In conclusion, the results presented here constitute the first report on the differences in the urine and fecal metabolomes between preterm and full-term groups of physically fit healthy young males. Clear differences were identified in the urine and fecal metabolomes next to the metabolic pathways, suggesting that systemic differences between the two groups affect the metabolism of the host as well as intestinal tract parameters and that of the underlying microbiome (Figure S13). One has to realize that studies with female participants are lacking and not many studies with sufficient statistical power were reported so far to close the gap. With the concomitant methodological developments presented in this study, the exploration of the more complex female metabolome and responses to inactivity and hypoxia can be commenced in a comparable way [35].

3. Materials and Methods

3.1. PreTerm Project: Cardio-Respiratory Responses during Hypoxic Exercise in Individuals Born Prematurely

The PreTerm project aimed to investigate the acute cardio-respiratory responses during rest and exercise in two groups of prematurely born, but otherwise healthy male adolescents and adults. In addition, this project aimed to elucidate the underlying mechanisms of the altered resting and exercise cardio-respiratory responses in prematurely born, but otherwise healthy individuals. The results from this cohort were compared to the data from control groups consisting of healthy, age and aerobic capacity-matched individuals born at full-term resulting in a unique dataset. The obtained results provide extensive basic physiological data on the development of cardiorespiratory control in individuals born prematurely, hypoxia exercise capacity and cardiorespiratory demand during hypoxic exercise in non-acclimatized individuals born prematurely [24].

Thirty-seven healthy men volunteered and gave written informed consent to participate in this study (Cardio-respiratory responses during hypoxic exercise in individuals born prematurely—ARRS research project J3-7536). All participants were free of cardiorespiratory and hematologic disease and had not been exposed to altitudes above 1500 m during the one-month period prior to the study. Twenty-two participants were born premature (gestational age $\leq$ 32 weeks; gestational weight $\leq$ 1500 g) and 15 were born full-term. The experimental protocol was approved by the National Medical Ethics Committee of Slovenia

(No. 0120-101/2016-2) and conducted in accordance with the principles of the Declaration of Helsinki. The study was also pre-registered at ClinicalTrials.gov (NCT02780908) [24].

The experimental protocol included two testing sessions in each group. On both occasions, no more than seven days apart, participants performed a graded exercise test for voluntary exhaustion. During the exercise tests, participants breathed either normoxic ambient air (fraction of inspired oxygen (FiO$_2$ = 0.209) or a humidified hypoxic air mixture (FiO$_2$ = 0.130 corresponding to a terrestrial altitude of approximately 3800 m) in a randomized, placebo-controlled manner. Indirect calorimetry, near-infrared spectroscopy and ECG measurements were performed during all tests. During both tests, participants performed a hypoxia sensitivity test to assess the hypoxic ventilatory response at rest and during exercise. In addition, selected hematological and oxidative stress markers were determined from blood samples collected before and after each hypoxia sensitivity test [24].

The two graded exercise tests were performed on an electromagnetically braked cycle ergometer (Ergo Bike Premium, Daum electronics, Fürth, Germany) under normoxic (FiO$_2$ = 0.21; PiO$_2$ = 147 mmHg) and normobaric hypoxic (FiO$_2$ = 0.13; PiO$_2$ = 91 mmHg) conditions in a randomized manner. They were blinded as to the FiO$_2$ of the gas mixture they inhaled on both occasions. Both tests were performed at the same time of day for each participant. The test protocol started at 60 W and was increased by 40 W every 2 min until exhaustion. The normoxic and hypoxic tests were performed exactly 7 days apart. During the tests, participants breathed through a face mask (Vmask, 7500 series, Hans Rudolph Inc., Shawnee, KS, USA) and oxygen uptake (VO$_2$) and ventilation (VE) were measured using a metabolic cart (Quark CPET, Cosmed, Rome, Italy). Capillary oxygen saturation (SpO$_2$) was measured using a transcutaneous finger pulse oximetry device (Nellcor, BCI 3301, Boulder, CO, USA). Fecal and urine samples were collected three consecutive days before and three consecutive days after the test under normoxic and hypoxic conditions (Figure S1) [24].

3.2. Sample Collection

Fecal and urine samples were collected three consecutive days before and three consecutive days after the test under normoxic and hypoxic conditions at the home of the participants (Figures S2 and S3). Collected samples were frozen at −20 °C immediately after collection. All participants collected 12 urine and 12 fecal samples in total. Three urine and three fecal samples were collected before and after normoxic tests, giving rise to six urine and six fecal samples per participant. The same approach was utilized for hypoxic tests, giving rise to another six urine and six fecal samples per participant. Two full-term and one preterm participant did not collect fecal and urine samples before and after the exercise test and were excluded from the metagenomic and metabolomic part of the PreTerm study.

3.3. ^{1}H-NMR Metabolomics

Samples were thawed at room temperature before preparation for NMR measurements. All collected samples were centrifuged (1.5 mL) at 10,000× g for 30 min to remove fine particles. Subsequently, 400 µL of the supernatant was mixed with 200 µL of ^{1}H-NMR buffer as previously described [121] and stored at −25 °C until analysis.

Prior to analysis, samples were thawed at room temperature and transferred to a 5 mm NMR tube. TSP was used as an internal standard for quantification, as described previously [121].

A 600 MHz Bruker Neo NMR spectrometer equipped with a 5 mm HCN Cold probe was used to record NMR spectra at 25 °C. The ^{1}H NMR spectra of the samples were recorded with a spectral width of 9.0 kHz, a relaxation delay of 2.0 s, 32 scans and 32 K data points. A double pulsed field gradient spin echo (DPFGSE) pulse sequence was used to suppress water. The total correlated spectrum (TOCSY) was measured with 1H spectral widths of 7.0 kHz, 4096 complex points, a relaxation delay of 1.5 s, 32 transients, and 144 time increments. An exponential function and a cosine squared function were used for

apodization. Zeros were filled before the Fourier transform. TopSpin (version 4.1) was used to process the NMR spectra [37–39,56,122].

Spectra Processing

NMR spectra were preprocessed with an internal script and prepared for identification with the Chenomx Compound Library, extended to the Human Metabolome Database [41,107], giving access to the chemical shift profiles of 674 compounds used in the analyses. Chemical shifts of 647 compounds were used for the identification of metabolites observed in our study. The resulting spectra were then analyzed with targeted quantitative metabolomics using Chenomx NMR Suite version 8.6 (Chenomx, Inc., Edmonton, AB, Canada). ChenomX profiler was used for randomized spectral fitting. All spectra were processed in the same way by spectral deconvolution and once metabolites were identified, urine and fecal data matrices were established, assigning 0 to a particular metabolite not detected in all samples.

3.4. Fecal Metagenomics

Fecal samples collected three days before and 1 day after normoxic and hypoxic testing were used for shotgun sequencing; 200 mg of feces were used for DNA extraction using the MagicPure Stool and Soil Genomic DNA Kit (Beijing, China) according to the manufacturer's protocol. Shotgun sequencing was performed using TruSeq Nano DNA (350) (Macrogen, Seoul, Korea).

Sequence Processing

Paired reads obtained from Macrogen were analyzed using our in-house pipeline for metagenomics sequence processing—Metabakery (in preparation). Metabakery is a re-implementation of the BioBakery [52] workflow using (https://huttenhower.sph.harvard.edu/kneaddata/, accessed on 7 April 2022) for quality control, MetaPhlAn [53] for taxonomy analysis (bacteria, archaea, fungi, protozoa, and viruses), and HUMAn3 [53] for functional genes, enzymatic reactions, and metabolic pathways determination. Additionally, the MelonnPan was used for the prediction of metabolites. Metabakery is implanted as a singularity image and prepared on high computing performance clusters. The analyses running MetaBakery were performed on a dual Xeon system with 32 CPU cores (64 hyper-threads), 512 GB RAM and 6 TB SATA hard disc at the Faculty of Electrical Engineering, University of Ljubljana.

3.5. Characterization of Fecal Samples: Bristol Stool Scale, Metabolites, pH, MWI

Fecal samples were analyzed for a number of parameters as previously described and as follows [35]: Bristol stool scale (BSS) [123], water content, pH [124], total soluble organic carbon (TSOC), short-chain fatty acids (SCFA) [125], reducing sugars (Carbohydrate determination with 4-hydroxybenzoic acid hydrazide (PAHBAH)) [126], molecular weight, and dissolved organic carbon complexity using molecular weight indices [127,128]. In addition, fecal piercing strengths, as described before, were used as previously described [39].

3.6. Statictics and Machine Learning
3.6.1. Statistics

First, the software PAST [129] was used for PERMANOVA. All obtained data matrices (NMR metabolomes—identified fecal and urinary metabolites at micromolar concentrations, microbial taxonomy, gene families, enzymatic reactions, metabolic pathways and predicted microbial metabolites) were analyzed in the same way. Each determined parameter was analyzed in three different ways as previously described [35–38,56]: (i) by dividing the measured concentration by the concentration of all metabolites in that sample; (ii) Box-Cox; or (iii) $\log(x + 1)$ transformed. The significance of the metabolic differences and microbial entities between the different sample groups were tested using ANOSIM, and NP-MANOVA, and expressed as the overlap in the non-metric multidimensional scaling

(nm-MDS) trait space (using Euclidean distance measures). The stress function was used to select the dimensionality reduction, while Shepard's plots were used to describe the correspondence between the target values and the obtained ranks. In addition, PCoA and PCA were performed on metagenomic data. Benjamini–Hochberg significance correction for multiple comparisons was used as previously described [130].

Second, for MetaboAnalyst [40], a log or cube root transformation was used in conjunction with mean or Pareto scaling as implemented in MetaboAnalyst, followed by supervised classification using the partial least squares discriminant analysis (PLSDA) method, random forest (RF), and pathway enrichment analysis. The PLSDA results were cross-validated with a caret package implemented in MeatboAnalyst. The major metabolites identified by PLSDA were determined according to the variable importance in projection (VIP). The randomForest package implemented in MetaboAnalyst was used for supervised classification between different groups of interest. The main features defined by RF were ordered according to the mean decrease in classification accuracy. Hierarchical clustering was performed according to the VIP scores to obtain a heat map representing the differences in metabolic profiles between samples and groups. Euclidean distance, Pearson's correlation and Spearman's correlation were used as similarity measures and Ward's linkage was used as a clustering algorithm. MetaboAnalyst and gplot were used to generate graphs.

KEGG libraries for human metabolic pathways were used for metabolic pathway and enrichment analysis. For topological analysis, the globaltest analysis method and relative Betweenness centrality were used. Significant pathways were determined using the raw p-value, Holm–Bonferroni p-adjusted value, and adjusted p-value using the false discovery rate. The effect of pathways was calculated using the pathway topology analysis.

Metabolite Set Enrichment (MSEA) was used to identify biologically significant patterns between quantitative metabolome data from different groups. The names of compounds in Human Metabolome Database (HMDB) were used for linkage to the KEGG database. Enrichment analysis was performed using the globaltest package implemented in MetaboAnalyst. The enrichment ratio was calculated by dividing observed hits and expected hits.

3.6.2. JADBIO Auto Machine Learning

Just Add Data Bio (JADBIO), a web-based machine learning platform for analyzing potential biomarkers [57], was used to search for biomarkers. The JADBIO platform was developed for predictive modeling and providing high-quality predictive models for diagnostics using state-of-the-art statistical and machine learning methods. Personal analytic biases and methodological statistical errors were eliminated from the analysis by autonomously exploring different settings in the modeling steps, resulting in more convincing discovered features to distinguish between different groups. JADBIO with extensive tuning effort and six CPUs was used to model different dataset choices in addition to the features observed in samples of all groups from different projects by splitting the total data into a training set and a test set in a 70:30 ratio. The training set was used to train the model and the test set was used to evaluate the model [39,56].

To assess the classification of the model, a receiver-operating characteristic curve (ROC curve) was constructed for all studied groups, plotting the true-positive rate (sensitivity) against the false-positive rate (1-specificity). Individual conditional expectation plots (ICE) showed the nature of the contribution of each feature characteristic to the model. All obtained models can be run locally using a Java executor.

Author Contributions: Conceptualization, T.D., D.O., G.P.M. and B.Š.; data collection, M.M., S.O., R.Š., L.D. and B.Š.; methodology, L.D., R.Š. and B.Š.; conceptualization for metabolomic analysis,

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the National Committee for Medical Ethics at the Ministry of Health of the Republic of Slovenia, No. 0120-101/2016-2. The study was also pre-registered at ClinicalTrials.gov (NCT02780908).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data underlying this study are available in electronic Supplementary Materials.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Moutquin, J.M. Classification and heterogeneity of preterm birth. *BJOG Int. J. Obstet. Gynaecol.* **2003**, *110* (Suppl. 20), 30–33. [CrossRef]
2. Tingleff, T.; Vikanes, Å.; Räisänen, S.; Sandvik, L.; Murzakanova, G.; Laine, K. Risk of preterm birth in relation to history of preterm birth: A population-based registry study of 213 335 women in Norway. *BJOG Int. J. Obstet. Gynaecol.* **2022**, *129*, 900–907. [CrossRef]
3. Crump, C. An overview of adult health outcomes after preterm birth. *Early Hum. Dev.* **2020**, *150*, 105187. [CrossRef]
4. Blencowe, H.; Cousens, S.; Oestergaard, M.Z.; Chou, D.; Moller, A.B.; Narwal, R.; Adler, A.; Vera Garcia, C.; Rohde, S.; Say, L.; et al. National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: A systematic analysis and implications. *Lancet* **2012**, *379*, 2162–2172. [CrossRef]
5. Liu, L.; Oza, S.; Hogan, D.; Perin, J.; Rudan, I.; Lawn, J.E.; Cousens, S.; Mathers, C.; Black, R.E. Global, regional, and national causes of child mortality in 2000-13, with projections to inform post-2015 priorities: An updated systematic analysis. *Lancet* **2015**, *385*, 430–440. [CrossRef]
6. Manley, B.J.; Doyle, L.W.; Davies, M.W.; Davis, P.G. Fifty years in neonatology. *J. Paediatr. Child Health* **2015**, *51*, 118–121. [CrossRef] [PubMed]
7. Martin, A.; Faes, C.; Debevec, T.; Rytz, C.; Millet, G.; Pialoux, V. Preterm birth and oxidative stress: Effects of acute physical exercise and hypoxia physiological responses. *Redox Biol.* **2018**, *17*, 315–322. [CrossRef]
8. Markopoulou, P.; Papanikolaou, E.; Analytis, A.; Zoumakis, E.; Siahanidou, T. Preterm Birth as a Risk Factor for Metabolic Syndrome and Cardiovascular Disease in Adult Life: A Systematic Review and Meta-Analysis. *J. Pediatrics* **2019**, *210*, 69–80. [CrossRef]
9. Kerkhof, G.F.; Breukhoven, P.E.; Leunissen, R.W.; Willemsen, R.H.; Hokken-Koelega, A.C. Does preterm birth influence cardiovascular risk in early adulthood? *J. Pediatrics* **2012**, *161*, 390–396. [CrossRef]
10. Parkinson, J.R.; Hyde, M.J.; Gale, C.; Santhakumaran, S.; Modi, N. Preterm birth and the metabolic syndrome in adult life: A systematic review and meta-analysis. *Pediatrics* **2013**, *131*, e1240–e1263. [CrossRef]
11. Gubhaju, L.; Sutherland, M.R.; Black, M.J. Preterm birth and the kidney: Implications for long-term renal health. *Reprod. Sci.* **2011**, *18*, 322–333. [CrossRef] [PubMed]

12. Lushchak, V.I. Free radicals, reactive oxygen species, oxidative stress and its classification. *Chem.-Biol. Interact.* **2014**, *224*, 164–175. [CrossRef] [PubMed]
13. Debevec, T.; Millet, G.P.; Pialoux, V. Hypoxia-Induced Oxidative Stress Modulation with Physical Activity. *Front. Physiol.* **2017**, *8*, 84. [CrossRef] [PubMed]
14. Magalhães, J.; Ascensão, A.; Viscor, G.; Soares, J.; Oliveira, J.; Marques, F.; Duarte, J. Oxidative stress in humans during and after 4 hours of hypoxia at a simulated altitude of 5500 m. *Aviat. Space Environ. Med.* **2004**, *75*, 16–22. [PubMed]
15. Pialoux, V.; Mounier, R.; Rock, E.; Mazur, A.; Schmitt, L.; Richalet, J.P.; Robach, P.; Coudert, J.; Fellmann, N. Effects of acute hypoxic exposure on prooxidant/antioxidant balance in elite endurance athletes. *Int. J. Sports Med.* **2009**, *30*, 87–93. [CrossRef] [PubMed]
16. Yuan, G.; Adhikary, G.; McCormick, A.A.; Holcroft, J.J.; Kumar, G.K.; Prabhakar, N.R. Role of oxidative stress in intermittent hypoxia-induced immediate early gene activation in rat PC12 cells. *J. Physiol.* **2004**, *557*, 773–783. [CrossRef] [PubMed]
17. Mazzeo, R.S.; Child, A.; Butterfield, G.E.; Mawson, J.T.; Zamudio, S.; Moore, L.G. Catecholamine response during 12 days of high-altitude exposure (4, 300 m) in women. *J. Appl. Physiol.* **1998**, *84*, 1151–1157. [CrossRef]
18. Kehrer, J.P.; Lund, L.G. Cellular reducing equivalents and oxidative stress. *Free Radic. Biol. Med.* **1994**, *17*, 65–75. [CrossRef]
19. Filippone, M.; Bonetto, G.; Corradi, M.; Frigo, A.C.; Baraldi, E. Evidence of unexpected oxidative stress in airways of adolescents born very pre-term. *Eur. Respir. J.* **2012**, *40*, 1253–1259. [CrossRef]
20. Powers, S.K.; Nelson, W.B.; Hudson, M.B. Exercise-induced oxidative stress in humans: Cause and consequences. *Free Radic. Biol. Med.* **2011**, *51*, 942–950. [CrossRef]
21. Debevec, T.; Pialoux, V.; Millet, G.P.; Martin, A.; Mramor, M.; Osredkar, D. Exercise Overrides Blunted Hypoxic Ventilatory Response in Prematurely Born Men. *Front. Physiol.* **2019**, *10*, 437. [CrossRef] [PubMed]
22. Debevec, T.; Poussel, M.; Osredkar, D.; Willis, S.J.; Sartori, C.; Millet, G.P. Post-exercise accumulation of interstitial lung water is greater in hypobaric than normobaric hypoxia in adults born prematurely. *Respir. Physiol. Neurobiol.* **2022**, *297*, 103828. [CrossRef] [PubMed]
23. Stres, B.; Kronegger, L. Shift in the paradigm towards next-generation microbiology. *Fems Microbiol. Lett.* **2019**, *366*, fnz159. [CrossRef]
24. Martin, A.; Millet, G.; Osredkar, D.; Mramor, M.; Faes, C.; Gouraud, E.; Debevec, T.; Pialoux, V. Effect of pre-term birth on oxidative stress responses to normoxic and hypoxic exercise. *Redox Biol.* **2020**, *32*, 101497. [CrossRef] [PubMed]
25. Lancaster, G.; Debevec, T.; Millet, G.P.; Poussel, M.; Willis, S.J.; Mramor, M.; Goričar, K.; Osredkar, D.; Dolžan, V.; Stefanovska, A. Relationship between cardiorespiratory phase coherence during hypoxia and genetic polymorphism in humans. *J. Physiol.* **2020**, *598*, 2001–2019. [CrossRef] [PubMed]
26. Vrijlandt, E.J.; Gerritsen, J.; Boezen, H.M.; Grevink, R.G.; Duiverman, E.J. Lung function and exercise capacity in young adults born prematurely. *Am. J. Respir. Crit. Care Med.* **2006**, *173*, 890–896. [CrossRef] [PubMed]
27. Svedenkrans, J.; Henckel, E.; Kowalski, J.; Norman, M.; Bohlin, K. Long-term impact of preterm birth on exercise capacity in healthy young men: A national population-based cohort study. *PLoS ONE* **2013**, *8*, e80869. [CrossRef]
28. Lovering, A.T.; Laurie, S.S.; Elliott, J.E.; Beasley, K.M.; Yang, X.; Gust, C.E.; Mangum, T.S.; Goodman, R.D.; Hawn, J.A.; Gladstone, I.M. Normal pulmonary gas exchange efficiency and absence of exercise-induced arterial hypoxemia in adults with bronchopulmonary dysplasia. *J. Appl. Physiol.* **2013**, *115*, 1050–1056. [CrossRef]
29. Clemm, H.H.; Vollsæter, M.; Røksund, O.D.; Eide, G.E.; Markestad, T.; Halvorsen, T. Exercise capacity after extremely preterm birth. Development from adolescence to adulthood. *Ann. Am. Thorac. Soc.* **2014**, *11*, 537–545. [CrossRef]
30. Farrell, E.T.; Bates, M.L.; Pegelow, D.F.; Palta, M.; Eickhoff, J.C.; O'Brien, M.J.; Eldridge, M.W. Pulmonary Gas Exchange and Exercise Capacity in Adults Born Preterm. *Ann. Am. Thorac. Soc.* **2015**, *12*, 1130–1137. [CrossRef]
31. Bates, M.L.; Farrell, E.T.; Eldridge, M.W. Abnormal ventilatory responses in adults born prematurely. *N. Engl. J. Med.* **2014**, *370*, 584–585. [CrossRef] [PubMed]
32. Kelly, M.M.; Griffith, P.B. The Influence of preterm birth beyond infancy: Umbrella review of outcomes of adolescents and adults born preterm. *J. Am. Assoc. Nurse Pract.* **2020**, *32*, 555–562. [CrossRef] [PubMed]
33. Raju, T.N.K.; Buist, A.S.; Blaisdell, C.J.; Moxey-Mims, M.; Saigal, S. Adults born preterm: A review of general health and system-specific outcomes. *Acta Paediatr.* **2017**, *106*, 1409–1437. [CrossRef]
34. Gupta, V.K.; Kim, M.; Bakshi, U.; Cunningham, K.Y.; Davis, J.M.; Lazaridis, K.N.; Nelson, H.; Chia, N.; Sung, J. A predictive index for health status using species-level gut microbiome profiling. *Nat. Commun.* **2020**, *11*, 4635. [CrossRef] [PubMed]
35. Sket, R.; Treichel, N.; Debevec, T.; Eiken, O.; Mekjavic, I.; Schloter, M.; Vital, M.; Chandler, J.; Tiedje, J.M.; Murovec, B.; et al. Hypoxia and Inactivity Related Physiological Changes (Constipation, Inflammation) Are Not Reflected at the Level of Gut Metabolites and Butyrate Producing Microbial Community: The PlanHab Study. *Front. Physiol.* **2017**, *8*, 250. [CrossRef]
36. Sket, R.; Treichel, N.; Kublik, S.; Debevec, T.; Eiken, O.; Mekjavic, I.; Schloter, M.; Vital, M.; Chandler, J.; Tiedje, J.M.; et al. Hypoxia and inactivity related physiological changes precede or take place in absence of significant rearrangements in bacterial community structure: The PlanHab randomized trial pilot study. *PLoS ONE* **2017**, *12*, e0188556. [CrossRef]
37. Sket, R.; Debevec, T.; Kublik, S.; Schloter, M.; Schoeller, A.; Murovec, B.; Mikus, K.V.; Makuc, D.; Pecnik, K.; Plavec, J.; et al. Intestinal Metagenomes and Metabolomes in Healthy Young Males: Inactivity and Hypoxia Generated Negative Physiological Symptoms Precede Microbial Dysbiosis. *Front. Physiol.* **2018**, *9*, 198. [CrossRef]

38. Šket, R.; Deutsch, L.; Prevoršek, Z.; Mekjavić, I.B.; Plavec, J.; Rittweger, J.; Debevec, T.; Eiken, O.; Stres, B. Systems View of Deconditioning During Spaceflight Simulation in the PlanHab Project: The Departure of Urine 1 H-NMR Metabolomes From Healthy State in Young Males Subjected to Bedrest Inactivity and Hypoxia. *Front. Physiol.* **2020**, *11*, 532271. [CrossRef]

39. Deutsch, L.; Stres, B. The Importance of Objective Stool Classification in Fecal 1H-NMR Metabolomics: Exponential Increase in Stool Crosslinking Is Mirrored in Systemic Inflammation and Associated to Fecal Acetate and Methionine. *Metabolites* **2021**, *11*, 172. [CrossRef]

40. Pang, Z.; Chong, J.; Zhou, G.; de Lima Morais, D.A.; Chang, L.; Barrette, M.; Gauthier, C.; Jacques, P.É.; Li, S.; Xia, J. MetaboAnalyst 5.0: Narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res.* **2021**, *49*, W388–W396. [CrossRef]

41. Wishart, D.S.; Guo, A.; Oler, E.; Wang, F.; Anjum, A.; Peters, H.; Dizon, R.; Sayeeda, Z.; Tian, S.; Lee, B.L.; et al. HMDB 5.0: The Human Metabolome Database for 2022. *Nucleic Acids Res.* **2022**, *50*, D622–D631. [CrossRef] [PubMed]

42. Arakawa, T.; Iitani, K.; Toma, K.; Mitsubayashi, K. Biosensors: Gas Sensors. In *Reference Module in Biomedical Sciences*; Elsevier: Amsterdam, The Netherlands, 2021.

43. Jahan, K.; Jungdae, P.; Sung-Soo, P.; Geum-Sook, H. Correlation analysis of human urinary metabolites related to gender and obesity using NMR-based metabolic profiling. *J. Korean Magn. Reson. Soc.* **2012**, *16*, 46–66. [CrossRef]

44. Crump, C.; Sundquist, J.; Winkleby, M.A.; Sundquist, K. Gestational age at birth and mortality from infancy into mid-adulthood: A national cohort study. *Lancet Child Adolesc. Health* **2019**, *3*, 408–417. [CrossRef]

45. Paradis, A.N.; Gay, M.S.; Wilson, C.G.; Zhang, L. Newborn hypoxia/anoxia inhibits cardiomyocyte proliferation and decreases cardiomyocyte endowment in the developing heart: Role of endothelin-1. *PLoS ONE* **2015**, *10*, e0116600. [CrossRef]

46. Perrone, S.; Negro, S.; Laschi, E.; Calderisi, M.; Giordano, M.; De Bernardo, G.; Parigi, G.; Toni, A.L.; Esposito, S.; Buonocore, G. Metabolomic Profile of Young Adults Born Preterm. *Metabolites* **2021**, *11*, 697. [CrossRef]

47. Haraldsdottir, K.; Watson, A.M.; Beshish, A.G.; Pegelow, D.F.; Palta, M.; Tetri, L.H.; Brix, M.D.; Centanni, R.M.; Goss, K.N.; Eldridge, M.W. Heart rate recovery after maximal exercise is impaired in healthy young adults born preterm. *Eur. J. Appl. Physiol.* **2019**, *119*, 857–866. [CrossRef]

48. Qiu, S.; Cai, X.; Sun, Z.; Li, L.; Zuegel, M.; Steinacker, J.M.; Schumann, U. Heart Rate Recovery and Risk of Cardiovascular Events and All-Cause Mortality: A Meta-Analysis of Prospective Cohort Studies. *J. Am. Heart Assoc.* **2017**, *6*, e005505. [CrossRef]

49. Prakash, Y.S.; Pabelick, C.M.; Sieck, G.C. Mitochondrial Dysfunction in Airway Disease. *Chest* **2017**, *152*, 618–626. [CrossRef]

50. Ten, V.S. Mitochondrial dysfunction in alveolar and white matter developmental failure in premature infants. *Pediatric Res.* **2017**, *81*, 286–292. [CrossRef]

51. Sonntag, B.; Stolze, B.; Heinecke, A.; Luegering, A.; Heidemann, J.; Lebiedz, P.; Rijcken, E.; Kiesel, L.; Domschke, W.; Kucharzik, T.; et al. Preterm birth but not mode of delivery is associated with an increased risk of developing inflammatory bowel disease later in life. *Inflamm. Bowel Dis.* **2007**, *13*, 1385–1390. [CrossRef]

52. McIver, L.J.; Abu-Ali, G.; Franzosa, E.A.; Schwager, R.; Morgan, X.C.; Waldron, L.; Segata, N.; Huttenhower, C. bioBakery: A meta'omic analysis environment. *Bioinformatics* **2018**, *34*, 1235–1237. [CrossRef] [PubMed]

53. Beghini, F.; McIver, L.J.; Blanco-Míguez, A.; Dubois, L.; Asnicar, F.; Maharjan, S.; Mailyan, A.; Manghi, P.; Scholz, M.; Thomas, A.M.; et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *eLife* **2021**, *10*, e65088. [CrossRef] [PubMed]

54. Schloss, P.D.; Westcott, S.L.; Ryabin, T.; Hall, J.R.; Hartmann, M.; Hollister, E.B.; Lesniewski, R.A.; Oakley, B.B.; Parks, D.H.; Robinson, C.J.; et al. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **2009**, *75*, 7537–7541. [CrossRef] [PubMed]

55. Mallick, H.; Franzosa, E.A.; McIver, L.J.; Banerjee, S.; Sirota-Madi, A.; Kostic, A.D.; Clish, C.B.; Vlamakis, H.; Xavier, R.J.; Huttenhower, C. Predictive metabolomic profiling of microbial communities using amplicon or metagenomic sequences. *Nat. Commun.* **2019**, *10*, 3136. [CrossRef]

56. Deutsch, L.; Osredkar, D.; Plavec, J.; Stres, B. Spinal Muscular Atrophy after Nusinersen Therapy: Improved Physiology in Pediatric Patients with No Significant Change in Urine, Serum, and Liquor 1H-NMR Metabolomes in Comparison to an Age-Matched, Healthy Cohort. *Metabolites* **2021**, *11*, 206. [CrossRef]

57. Tsamardinos, I.; Charonyktakis, P.; Lakiotaki, K.; Borboudakis, G.; Zenklusen, J.C.; Juhl, H.; Chatzaki, E.; Lagani, V. Just Add Data: Automated Predictive Modeling and BioSignature Discovery. *bioRxiv* **2020**, 1–46. [CrossRef]

58. Qin, J.; Li, R.; Raes, J.; Arumugam, M.; Burgdorf, K.S.; Manichanh, C.; Nielsen, T.; Pons, N.; Levenez, F.; Yamada, T.; et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **2010**, *464*, 59–65. [CrossRef]

59. Turnbaugh, P.J.; Ley, R.E.; Hamady, M.; Fraser-Liggett, C.M.; Knight, R.; Gordon, J.I. The human microbiome project. *Nature* **2007**, *449*, 804–810. [CrossRef]

60. Cao, B.; Stout, M.J.; Lee, I.; Mysorekar, I.U. Placental Microbiome and Its Role in Preterm Birth. *NeoReviews* **2014**, *15*, e537–e545. [CrossRef]

61. Karp, P.D.; Billington, R.; Caspi, R.; Fulcher, C.A.; Latendresse, M.; Kothari, A.; Keseler, I.M.; Krummenacker, M.; Midford, P.E.; Ong, Q.; et al. The BioCyc collection of microbial genomes and metabolic pathways. *Brief. Bioinform.* **2019**, *20*, 1085–1093. [CrossRef]

62. Connors, J.; Dawe, N.; Van Limbergen, J. The Role of Succinate in the Regulation of Intestinal Inflammation. *Nutrients* **2018**, *11*, 25. [CrossRef] [PubMed]

63. Akram, M. Citric acid cycle and role of its intermediates in metabolism. *Cell Biochem. Biophys.* **2014**, *68*, 475–478. [CrossRef]

64. Ariza, A.C.; Deen, P.M.; Robben, J.H. The succinate receptor as a novel therapeutic target for oxidative and metabolic stress-related conditions. *Front. Endocrinol.* **2012**, *3*, 22. [CrossRef] [PubMed]

65. Littlewood-Evans, A.; Sarret, S.; Apfel, V.; Loesle, P.; Dawson, J.; Zhang, J.; Muller, A.; Tigani, B.; Kneuer, R.; Patel, S.; et al. GPR91 senses extracellular succinate released from inflammatory macrophages and exacerbates rheumatoid arthritis. *J. Exp. Med.* **2016**, *213*, 1655–1662. [CrossRef]

66. Rubic, T.; Lametschwandtner, G.; Jost, S.; Hinteregger, S.; Kund, J.; Carballido-Perrig, N.; Schwärzler, C.; Junt, T.; Voshol, H.; Meingassner, J.G.; et al. Triggering the succinate receptor GPR91 on dendritic cells enhances immunity. *Nat. Immunol.* **2008**, *9*, 1261–1269. [CrossRef] [PubMed]

67. Louis, P.; Flint, H.J. Formation of propionate and butyrate by the human colonic microbiota. *Environ. Microbiol.* **2017**, *19*, 29–41. [CrossRef]

68. Tannahill, G.M.; Curtis, A.M.; Adamik, J.; Palsson-McDermott, E.M.; McGettrick, A.F.; Goel, G.; Frezza, C.; Bernard, N.J.; Kelly, B.; Foley, N.H.; et al. Succinate is an inflammatory signal that induces IL-1β through HIF-1α. *Nature* **2013**, *496*, 238–242. [CrossRef]

69. Caspi, R.; Dreher, K.; Karp, P.D. The challenge of constructing, classifying, and representing metabolic pathways. *FEMS Microbiol. Lett.* **2013**, *345*, 85–93. [CrossRef]

70. Karp, P.D.; Caspi, R. A survey of metabolic databases emphasizing the MetaCyc family. *Arch. Toxicol.* **2011**, *85*, 1015–1033. [CrossRef]

71. Caspi, R.; Billington, R.; Keseler, I.M.; Kothari, A.; Krummenacker, M.; Midford, P.E.; Ong, W.K.; Paley, S.; Subhraveti, P.; Karp, P.D. The MetaCyc database of metabolic pathways and enzymes—a 2019 update. *Nucleic Acids Res.* **2020**, *48*, D445–D453. [CrossRef]

72. Wexler, H.M. The Genus *Bacteroides*. In *The Prokaryotes*; Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E., Thompson, F., Eds.; SpringerLink: Berlin/Heidelberg, Germany, 2014; pp. 459–484.

73. Kawaguchi, K.; Senoura, T.; Ito, S.; Taira, T.; Ito, H.; Wasaki, J.; Ito, S. The mannobiose-forming exo-mannanase involved in a new mannan catabolic pathway in *Bacteroides fragilis*. *Arch. Microbiol.* **2014**, *196*, 17–23. [CrossRef] [PubMed]

74. Gao, G.; Cao, J.; Mi, L.; Feng, D.; Deng, Q.; Sun, X.; Zhang, H.; Wang, Q.; Wang, J. BdPUL12 depolymerizes β-mannan-like glycans into mannooligosaccharides and mannose, which serve as carbon sources for *Bacteroides dorei* and gut probiotics. *Int. J. Biol. Macromol.* **2021**, *187*, 664–674. [CrossRef] [PubMed]

75. Lindstad, L.J.; Lo, G.; Leivers, S.; Lu, Z.; Michalak, L.; Pereira, G.V.; Røhr, Å.K.; Martens, E.C.; McKee, L.S.; Louis, P.; et al. Human Gut Faecalibacterium prausnitzii Deploys a Highly Efficient Conserved System To Cross-Feed on β-Mannan-Derived Oligosaccharides. *mBio* **2021**, *12*, e03628-20. [CrossRef] [PubMed]

76. La Rosa, S.L.; Leth, M.L.; Michalak, L.; Hansen, M.E.; Pudlo, N.A.; Glowacki, R.; Pereira, G.; Workman, C.T.; Arntzen, M.Ø.; Pope, P.B.; et al. The human gut Firmicute *Roseburia intestinalis* is a primary degrader of dietary β-mannans. *Nat. Commun.* **2019**, *10*, 905. [CrossRef]

77. Samuel, G.; Reeves, P. Biosynthesis of O-antigens: Genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. *Carbohydr. Res.* **2003**, *338*, 2503–2519. [CrossRef]

78. Wolfs, T.G.; Derikx, J.P.; Hodin, C.M.; Vanderlocht, J.; Driessen, A.; de Bruïne, A.P.; Bevins, C.L.; Lasitschka, F.; Gassler, N.; van Gemert, W.G.; et al. Localization of the lipopolysaccharide recognition complex in the human healthy and inflamed premature and adult gut. *Inflamm. Bowel Dis.* **2010**, *16*, 68–75. [CrossRef]

79. Shah, J.; Jefferies, A.L.; Yoon, E.W.; Lee, S.K.; Shah, P.S.; Network, C.N. Risk Factors and Outcomes of Late-Onset Bacterial Sepsis in Preterm Neonates Born at. *Am. J. Perinatol.* **2015**, *32*, 675–682. [CrossRef]

80. Kim, K.A.; Jeong, J.J.; Yoo, S.Y.; Kim, D.H. Gut microbiota lipopolysaccharide accelerates inflamm-aging in mice. *BMC Microbiol.* **2016**, *16*, 9. [CrossRef]

81. Phan, J.; Nair, D.; Jain, S.; Montagne, T.; Flores, D.V.; Nguyen, A.; Dietsche, S.; Gombar, S.; Cotter, P. Alterations in Gut Microbiome Composition and Function in Irritable Bowel Syndrome and Increased Probiotic Abundance with Daily Supplementation. *mSystems* **2021**, *6*, e01215-21. [CrossRef]

82. Pirola, C.J.; Salatino, A.; Quintanilla, M.F.; Castaño, G.O.; Garaycoechea, M.; Sookoian, S. The influence of host genetics on liver microbiome composition in patients with NAFLD. *EBioMedicine* **2022**, *76*, 103858. [CrossRef]

83. Poli, D.; Carbognani, P.; Corradi, M.; Goldoni, M.; Acampa, O.; Balbi, B.; Bianchi, L.; Rusca, M.; Mutti, A. Exhaled volatile organic compounds in patients with non-small cell lung cancer: Cross sectional and nested short-term follow-up study. *Respir. Res.* **2005**, *6*, 71. [CrossRef] [PubMed]

84. Aoki, T.; Nagaoka, T.; Kobayashi, N.; Kurahashi, M.; Tsuji, C.; Takiguchi, H.; Tomomatsu, K.; Oguma, T.; Kobayashi, N.; Magatani, K.; et al. Editor's Highlight: Prospective Analyses of Volatile Organic Compounds in Obstructive Sleep Apnea Patients. *Toxicol. Sci. Off. J. Soc. Toxicol.* **2017**, *156*, 362–374. [CrossRef]

85. Garner, C.E.; Smith, S.; de Lacy Costello, B.; White, P.; Spencer, R.; Probert, C.S.; Ratcliffe, N.M. Volatile organic compounds from feces and their potential for diagnosis of gastrointestinal disease. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **2007**, *21*, 1675–1688. [CrossRef] [PubMed]

86. Raman, M.; Ahmed, I.; Gillevet, P.M.; Probert, C.S.; Ratcliffe, N.M.; Smith, S.; Greenwood, R.; Sikaroodi, M.; Lam, V.; Crotty, P.; et al. Fecal microbiome and volatile organic compound metabolome in obese humans with nonalcoholic fatty liver disease. *Clin. Gastroenterol. Hepatol. Off. Clin. Pract. J. Am. Gastroenterol. Assoc.* **2013**, *11*, 868–875. [CrossRef]

87. Rojo, F. Degradation of alkanes by bacteria. *Environ. Microbiol.* **2009**, *11*, 2477–2490. [CrossRef]

88. Chen, Q.; Janssen, D.B.; Witholt, B. Growth on octane alters the membrane lipid fatty acids of Pseudomonas oleovorans due to the induction of alkB and synthesis of octanol. *J. Bacteriol.* **1995**, *177*, 6894–6901. [CrossRef]

89. Merino, J.; Dashti, H.S.; Li, S.X.; Sarnowski, C.; Justice, A.E.; Graff, M.; Papoutsakis, C.; Smith, C.E.; Dedoussis, G.V.; Lemaitre, R.N.; et al. Genome-wide meta-analysis of macronutrient intake of 91,114 European ancestry participants from the cohorts for heart and aging research in genomic epidemiology consortium. *Mol. Psychiatry* **2019**, *24*, 1920–1932. [CrossRef]

90. Liu, H.; Wang, J.; He, T.; Becker, S.; Zhang, G.; Li, D.; Ma, X. Butyrate: A Double-Edged Sword for Health? *Adv. Nutr.* **2018**, *9*, 21–29. [CrossRef]

91. O'Donovan, C.M.; Madigan, S.M.; Garcia-Perez, I.; Rankin, A.; O' Sullivan, O.; Cotter, P.D. Distinct microbiome composition and metabolome exists across subgroups of elite Irish athletes. *J. Sci. Med. Sport* **2020**, *23*, 63–68. [CrossRef]

92. Pasolli, E.; Asnicar, F.; Manara, S.; Zolfo, M.; Karcher, N.; Armanini, F.; Beghini, F.; Manghi, P.; Tett, A.; Ghensi, P.; et al. Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. *Cell* **2019**, *176*, 649–662. [CrossRef]

93. Feng, Q.; Liu, Z.; Zhong, S.; Li, R.; Xia, H.; Jie, Z.; Wen, B.; Chen, X.; Yan, W.; Fan, Y.; et al. Integrated metabolomics and metagenomics analysis of plasma and urine identified microbial metabolites associated with coronary heart disease. *Sci. Rep.* **2016**, *6*, 22525. [CrossRef] [PubMed]

94. Wirbel, J.; Pyl, P.T.; Kartal, E.; Zych, K.; Kashani, A.; Milanese, A.; Fleck, J.S.; Voigt, A.Y.; Palleja, A.; Ponnudurai, R.; et al. Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nat. Med.* **2019**, *25*, 679–689. [CrossRef] [PubMed]

95. Ussher, J.R.; Lopaschuk, G.D.; Arduini, A. Gut microbiota metabolism of L-carnitine and cardiovascular risk. *Atherosclerosis* **2013**, *231*, 456–461. [CrossRef] [PubMed]

96. Koeth, R.A.; Wang, Z.; Levison, B.S.; Buffa, J.A.; Org, E.; Sheehy, B.T.; Britt, E.B.; Fu, X.; Wu, Y.; Li, L.; et al. Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nat. Med.* **2013**, *19*, 576–585. [CrossRef] [PubMed]

97. Tang, W.H.; Wang, Z.; Levison, B.S.; Koeth, R.A.; Britt, E.B.; Fu, X.; Wu, Y.; Hazen, S.L. Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *N. Engl. J. Med.* **2013**, *368*, 1575–1584. [CrossRef] [PubMed]

98. Wang, Z.; Klipfell, E.; Bennett, B.J.; Koeth, R.; Levison, B.S.; Dugar, B.; Feldstein, A.E.; Britt, E.B.; Fu, X.; Chung, Y.M.; et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **2011**, *472*, 57–63. [CrossRef]

99. Jameson, E.; Quareshy, M.; Chen, Y. Methodological considerations for the identification of choline and carnitine-degrading bacteria in the gut. *Methods* **2018**, *149*, 42–48. [CrossRef]

100. Day-Walsh, P.; Shehata, E.; Saha, S.; Savva, G.M.; Nemeckova, B.; Speranza, J.; Kellingray, L.; Narbad, A.; Kroon, P.A. The use of an in-vitro batch fermentation (human colon) model for investigating mechanisms of TMA production from choline, L-carnitine and related precursors by the human gut microbiota. *Eur. J. Nutr.* **2021**, *60*, 3987–3999. [CrossRef]

101. Kitada, Y.; Muramatsu, K.; Toju, H.; Kibe, R.; Benno, Y.; Kurihara, S.; Matsumoto, M. Bioactive polyamine production by a novel hybrid system comprising multiple indigenous gut bacterial strategies. *Sci. Adv.* **2018**, *4*, eaat0062. [CrossRef]

102. Nakamura, A.; Ooga, T.; Matsumoto, M. Intestinal luminal putrescine is produced by collective biosynthetic pathways of the commensal microbiome. *Gut Microbes* **2019**, *10*, 159–171. [CrossRef]

103. Yoshimoto, S.; Mitsuyama, E.; Yoshida, K.; Odamaki, T.; Xiao, J.Z. Enriched metabolites that potentially promote age-associated diseases in subjects with an elderly-type gut microbiota. *Gut Microbes* **2021**, *13*, 1865705. [CrossRef] [PubMed]

104. Grosheva, I.; Zheng, D.; Levy, M.; Polansky, O.; Lichtenstein, A.; Golani, O.; Dori-Bachash, M.; Moresi, C.; Shapiro, H.; Del Mare-Roumani, S.; et al. High-Throughput Screen Identifies Host and Microbiota Regulators of Intestinal Barrier Function. *Gastroenterology* **2020**, *159*, 1807–1823. [CrossRef]

105. Lee, J.S.; Kim, S.Y.; Chun, Y.S.; Chun, Y.J.; Shin, S.Y.; Choi, C.H.; Choi, H.K. Characteristics of fecal metabolic profiles in patients with irritable bowel syndrome with predominant diarrhea investigated using 1 H-NMR coupled with multivariate statistical analysis. *Neurogastroenterol. Motil. Off. J. Eur. Gastrointest. Motil. Soc.* **2020**, *32*, e13830. [CrossRef] [PubMed]

106. Johnson, C.H.; Dejea, C.M.; Edler, D.; Hoang, L.T.; Santidrian, A.F.; Felding, B.H.; Ivanisevic, J.; Cho, K.; Wick, E.C.; Hechenbleikner, E.M.; et al. Metabolism links bacterial biofilms and colon carcinogenesis. *Cell Metab.* **2015**, *21*, 891–897. [CrossRef] [PubMed]

107. Wishart, D.S.; Feunang, Y.D.; Marcu, A.; Guo, A.C.; Liang, K.; Vazquez-Fresno, R.; Sajed, T.; Johnson, D.; Li, C.; Karu, N.; et al. HMDB 4.0: The human metabolome database for 2018. *Nucleic Acids Res.* **2018**, *46*, D608–D617. [CrossRef] [PubMed]

108. Oliphant, K.; Allen-Vercoe, E. Macronutrient metabolism by the human gut microbiome: Major fermentation by-products and their impact on host health. *Microbiome* **2019**, *7*, 91. [CrossRef]

109. Le Gall, G.; Noor, S.O.; Ridgway, K.; Scovell, L.; Jamieson, C.; Johnson, I.T.; Colquhoun, I.J.; Kemsley, E.K.; Narbad, A. Metabolomics of fecal extracts detects altered metabolic activity of gut microbiota in ulcerative colitis and irritable bowel syndrome. *J. Proteome Res.* **2011**, *10*, 4208–4218. [CrossRef]

110. Wahlström, A.; Kovatcheva-Datchary, P.; Ståhlman, M.; Bäckhed, F.; Marschall, H.U. Crosstalk between Bile Acids and Gut Microbiota and Its Impact on Farnesoid X Receptor Signalling. *Dig. Dis.* **2017**, *35*, 246–250. [CrossRef]

111. Sayin, S.I.; Wahlström, A.; Felin, J.; Jäntti, S.; Marschall, H.U.; Bamberg, K.; Angelin, B.; Hyötyläinen, T.; Orešič, M.; Bäckhed, F. Gut microbiota regulates bile acid metabolism by reducing the levels of tauro-beta-muricholic acid, a naturally occurring FXR antagonist. *Cell Metab.* **2013**, *17*, 225–235. [CrossRef]

131

112. Hylemon, P.B.; Zhou, H.; Pandak, W.M.; Ren, S.; Gil, G.; Dent, P. Bile acids as regulatory molecules. *J. Lipid Res.* **2009**, *50*, 1509–1520. [CrossRef]

113. Staley, C.; Weingarden, A.R.; Khoruts, A.; Sadowsky, M.J. Interaction of gut microbiota with bile acid metabolism and its influence on disease states. *Appl. Microbiol. Biotechnol.* **2017**, *101*, 47–64. [CrossRef] [PubMed]

114. Heinken, A.; Ravcheev, D.A.; Baldini, F.; Heirendt, L.; Fleming, R.M.T.; Thiele, I. Systematic assessment of secondary bile acid metabolism in gut microbes reveals distinct metabolic capabilities in inflammatory bowel disease. *Microbiome* **2019**, *7*, 75. [CrossRef] [PubMed]

115. Fukiya, S.; Arata, M.; Kawashima, H.; Yoshida, D.; Kaneko, M.; Minamida, K.; Watanabe, J.; Ogura, Y.; Uchida, K.; Itoh, K.; et al. Conversion of cholic acid and chenodeoxycholic acid into their 7-oxo derivatives by Bacteroides intestinalis AM-1 isolated from human feces. *FEMS Microbiol. Lett.* **2009**, *293*, 263–270. [CrossRef] [PubMed]

116. Coman, V.; Vodnar, D.C. Hydroxycinnamic acids and human health: Recent advances. *J. Sci. Food Agric.* **2020**, *100*, 483–499. [CrossRef] [PubMed]

117. Santoru, M.L.; Piras, C.; Murgia, A.; Palmas, V.; Camboni, T.; Liggi, S.; Ibba, I.; Lai, M.A.; Orrù, S.; Blois, S.; et al. Cross sectional evaluation of the gut-microbiome metabolome axis in an Italian cohort of IBD patients. *Sci. Rep.* **2017**, *7*, 9523. [CrossRef]

118. Tso, S.C.; Qi, X.; Gui, W.J.; Chuang, J.L.; Morlock, L.K.; Wallace, A.L.; Ahmed, K.; Laxman, S.; Campeau, P.M.; Lee, B.H.; et al. Structure-based design and mechanisms of allosteric inhibitors for mitochondrial branched-chain α-ketoacid dehydrogenase kinase. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 9728–9733. [CrossRef] [PubMed]

119. Pedersen, H.K.; Gudmundsdottir, V.; Nielsen, H.B.; Hyotylainen, T.; Nielsen, T.; Jensen, B.A.; Forslund, K.; Hildebrand, F.; Prifti, E.; Falony, G.; et al. Human gut microbes impact host serum metabolome and insulin sensitivity. *Nature* **2016**, *535*, 376–381. [CrossRef]

120. Liu, S.; Zhao, W.; Liu, X.; Cheng, L. Metagenomic analysis of the gut microbiome in atherosclerosis patients identify cross-cohort microbial signatures and potential therapeutic target. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **2020**, *34*, 14166–14181. [CrossRef]

121. Beckonert, O.; Keun, H.C.; Ebbels, T.M.; Bundy, J.; Holmes, E.; Lindon, J.C.; Nicholson, J.K. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat. Protoc.* **2007**, *2*, 2692–2703. [CrossRef]

122. Murovec, B.; Makuc, D.; Kolbl Repinc, S.; Prevorsek, Z.; Zavec, D.; Sket, R.; Pecnik, K.; Plavec, J.; Stres, B. (1)H NMR metabolomics of microbial metabolites in the four MW agricultural biogas plant reactors: A case study of inhibition mirroring the acute rumen acidosis symptoms. *J. Environ. Manag.* **2018**, *222*, 428–435. [CrossRef]

123. Heaton, K.W.; Radvan, J.; Cripps, H.; Mountford, R.A.; Braddon, F.E.M.; Hughes, A.O. Defecation frequency and timing, and stool form in the general population: A prospective study. *Gut* **1992**, *33*, 818–824. [CrossRef] [PubMed]

124. Stres, B.; Sul, W.J.; Murovec, B.; Tiedje, J.M. Recently deglaciated high-altitude soils of the Himalaya: Diverse environments, heterogenous bacterial communities and long-range dust inputs from the upper troposphere. *PLoS ONE* **2013**, *8*, e76440. [CrossRef] [PubMed]

125. Kolbl, S.; Paloczi, A.; Panjan, J.; Stres, B. Addressing case specific biogas plant tasks: Industry oriented methane yields derived from 5L Automatic Methane Potential Test Systems in batch or semi-continuous tests using realistic inocula, substrate particle sizes and organic loading. *Bioresour. Technol.* **2014**, *153*, 180–188. [CrossRef] [PubMed]

126. Lever, M. Carbohydrate determination with 4-hydroxybenzoic acid hydrazide (PAHBAH): Effect of bismuth on the reaction. *Anal. Biochem.* **1977**, *81*, 21–27. [CrossRef]

127. Twardowski, M.S.; Boss, E.; Sullivan, J.M.; Donaghay, P.L. Modeling the spectral shape of absorption by chromophoric dissolved organic matter. *Mar. Chem.* **2004**, *89*, 69–88. [CrossRef]

128. Zhang, Y.; Liu, X.; Osburn, C.L.; Wang, M.; Qin, B.; Zhou, Y. Photobleaching response of different sources of chromophoric dissolved organic matter exposed to natural solar radiation using absorption and excitation-emission matrix spectra. *PLoS ONE* **2013**, *8*, e77515. [CrossRef]

129. Hammer, O.; Harper, D.A.T.; Ryan, P.D. PAST: Paleontological statistics software package for education and data analysis. *Palaeontol. Electron.* **2001**, *1*, 9.

130. Legendre, P.; Legendre, L.F.J. *Numerical Ecology*, 3rd ed.; Elsevier: Amsterdam, The Netherlands, 2012; Volume 24, p. 1006.

2.2 ADDITIONAL SCIENTIFIC WORK

2.2.1 Metagenomes assembled genomes from the PreTerm project

2.2.1.1 Introduction

In the context of taxonomic and functional analysis of the microbiome community, the third option is to assemble short-read sequences obtained with modern sequencing technologies into fully recovered genomes from the microbiome using available tools. This process is used to assemble new metagenome-assembled genomes. There are many genome assemblers specifically designed for metagenomic data, but none of them are perfect. A whole range of specialised tools have been developed to solve the problems of metagenomic assembly caused by the properties of the collected data. Depending on the length of the generated reads, assemblers are based on different approaches, from overlap-layout-consensus tools based on overlap strategies to those using de Bruijn graphs to work with data. It is important to note that it is not only the efficiency and quality of work that influence the popularity of assemblers, but also the ease of use of the tool, the existence of a simple, detailed and easy-to-understand manual, the continuous development of the tool, and the speed and quality of feedback form the tool's support team (Lapidus and Korobeynikov, 2021). For this reason, we have combined the multitude of different tools needed for metagenome assembly into the MAGO pipeline (Section 2.1.1). This approach can lead to the discovery of new species that cannot be cultured and that have become increasingly important in recent years (Fricker et al., 2019; Nayfach et al., 2019; Murovec et al., 2020; Lapidus and Korobeynikov, 2021). Sequence assembly can be divided into two necessary steps, all of which are already included in the MAGO pipeline:

1. metagenomic assembly (assembly of short read sequences (250 base pairs) into longer contigs).
2. binning (grouping of contigs with the same sequences into their taxon ID (e.g., closely related organisms)).

This process can also produce some artefacts in de novo assembled sequences, such as "bulges" or "tips", which are often artefacts due to sequencing errors (Zerbino and Birney, 2008). For this reason, MAGs need to be validated. For this purpose, the CheckM tool (Parks et al., 2015) is used to check the completeness and contamination of the assembled genomes. MAGs can be divided into high- and medium-quality groups according to the standards for minimum information about a metagenome-assembled genome (MIMAG). MAGs in the high-quality group contain < 5% contamination and are > 90% complete. Medium-quality MAGs contain < 10% contamination and are > 50% complete (Parks et al., 2015; Bowers et al., 2017).

The hypothesis from section 1.4.1 were partly discussed in this chapter (table 2).

2.2.1.2 Materials and methods

For this work, we used sequences from the PreTerm project and assembled the MAGs from the preterm and full-term groups individually. The main purpose was to obtain some characteristic species belonging to preterm group's adolescents involved in the PreTerm project.

Sequences from the Preterm Project (Deutsch et al., 2022b) were used to compile characteristic MAGs for the preterm and full-term groups of participants. Sequences from the preterm and control groups were assembled separately using the MAGO Singularity Image on the Leo4 HPC cluster (University of Innsbruck, Austria). Fastp (Chen et al., 2018) and FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) were used for quality control and pre-processing. Three different assemblers were used for assembly: metaSPAdes (Nurk et al., 2017), MEGAHIT (Li et al., 2015) and IDBA-UD (Peng et al., 2012). Contigs were binned and bins were improved using the tools BinSanity (Graham, Heidelberg and Tully, 2017), CONCOCT (Alneberg et al., 2014), MetaBAT (Kang et al., 2015) and MaxBin (Wu et al., 2016). In the end, DAStool (Sieber et al., 2018) was used to refine and dereplicate the resulting bins to obtain near-complete MAGs, which were then checked for completeness and contamination level using the ChekM (Parks et al., 2015) tool. High-quality MAGs from both groups were used for average amino acid identity calculation with ezTree (Wu, 2018), genome annotation with Prokka (Seeman, 2014), pan- and core-genome analysis with Roary (Page et al., 2015), and high-throughput average nucleotide identity calculation with FastANI (Jain et al., 2018), all of which were integrated into the MAGO tool (Murovec et al., 2020). JSpeciesWS Online Service (Richter et al., 2016) was used to determine taxonomic thresholds with tetra-correlation search (Teeling et al., 2004) by comparing our high-quality MAGs with the reference genome database (GenomesDB). A mosaic plot was generated using Past software (Hammer et al., 2001).

2.2.1.3 Results

The total number of sequencing reads was lower in the preterm group (491 million total reads compared to 531 million reads in the control group). After filtering with fastp, 494.6 million reads were obtained in the control group and 486 million reads in the preterm group. Other reads were removed because they were of poor quality or contained too many Ns (it was not possible to basecall for these bases). The remainder of the sequences were used for metagenome assembly; 320 MAGs were assembled in the preterm group, and 27 of these MAGs belonged to the MAGs in the high-quality group (average completeness was 93.93±2.9% and contamination was 2.9±1.43%). In the control group, 124 MAGs were assembled, 24 of which belonged to the high-quality group (average completeness was 95.4±2.8% and contamination was 2.5±1.4%). MAGs from the preterm groups were approximately 1 Mb larger and counted almost twice as many contigs. Preterm MAGs also had a higher percentage of GC base pairs (5% higher on average). All high-quality MAGs were submitted to the online service JSpeciesWS for a tetra-correlation search with the genome reference database GenomesDB, which contains more than 55,000 genomes. No significant differences were observed

between preterm and control group (Figure 10, Figure 11). Approximately the same number of high-, medium-, and low-quality MAGs were assembled in both groups.
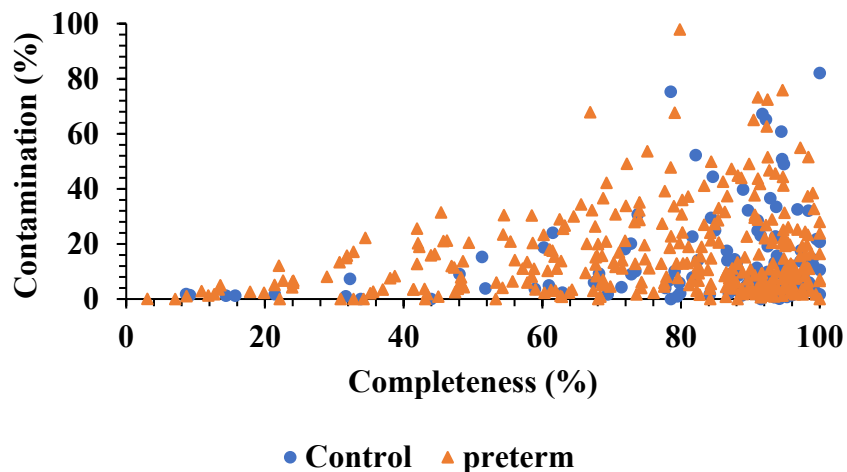
Figure 10:   Relationship between completeness and contamination of MAGs in control and preterm group.

Slika 10:   Odnos med popolnostjo in kontaminacijami na novo sestavljenih metagenomov v kontrolni in preterm skupini.
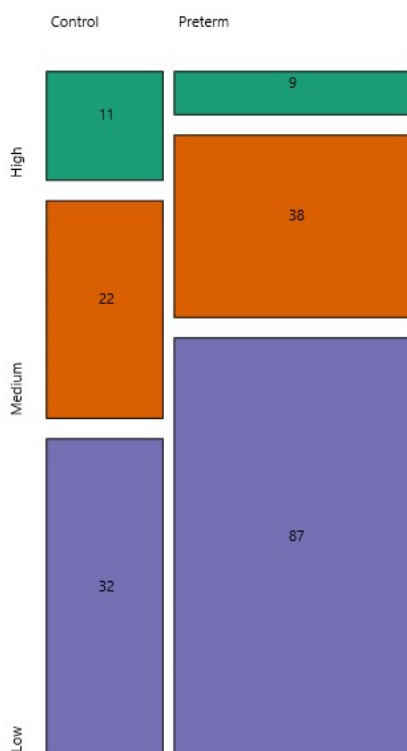
Figure 11:   Number of MAGs per both groups and their quality.
Number of high (completeness>95%, contamination<5%), medium (completeness>75%, contamination<10%) and low (completeness>50%, contamination<25%) quality MAGS in the preterm and control groups.

Slika 11:   Število na novo sestavljenih metagenomov med skupinami in njihova kvaliteta.
Število na novo sestavljenih metagenomov visoke (popolnost>95 %, kontaminacija<5 %), srednje (popolnost>75 %, kontaminacija<10 %) in nizke (popolnost>50 %, kontaminacija<25 %) v preterm in kontrolni skupini.

2.2.1.4 Discussion

Sequences from the preterm and full-term (control) groups were assembled separately in order to search for group-specific MAGs that could lead to discovery of taxonomic differences that were not observed in the previously published metaBakery analysis. Although a greater number of high-, medium-, and low quality MAGs were assembled in the preterm group according to the MIMAG standard (Bowers et al., 2017), we did not observe MAGs specific to the preterm group. The quality of the sequences was comparable and not significantly different in both groups. The higher number of MAGs is consistent with higher diversity indices in the PreTerm group, as previously observed (Deutsch et al., 2022). One of the most important parts of the de novo MAGs assembly is the ability to detect the "uncultured majority", which is also what we hoped to detect, especially in the preterm group. Based on these results, we can conclude that preterm and adult full-term born adults are not different in terms of microbial taxonomy, albeit due to the unequal variance within the groups. In contrast, we have shown that the functionality of the microbial worlds differs between adult preterm compared to adult full-term groups in terms of enzymatic reactions, metabolic pathways, and predicted metabolites (Deutsch et al., 2022b). This once again shows the higher relevance and

importance of microbial functionality relative to microbial taxonomic composition for the inference of relationships with human phenotypic characteristics through the production of various metabolites.

2.2.2 Data integration

2.2.2.1 Introduction

To properly understand the complexity of biological systems, well-being, and diseases, various 'omics high-throughput technologies (e.g., sequencing, various types of spectrometry, etc.) have been used and are becoming more affordable for scientists (Zitnik et al., 2019). It soon became clear very that we cannot capture the whole understanding of the system based on only one level of datasets. "Top-down" approach is the term that was evaluated in the context of systems biology research. In general, this means that we measure a set of parameters at the system level and then make inferences about the overall functionality of the system (Kohl et al., 2010; Price et al., 2017). Without 'omics methods, all domains relied strictly on single types of data that could not explain the entire system. 'Omics methods enabled the development of modern statistical approaches (data reduction methods) and data integration. With these methods, it became easier to draw conclusions based on thousands of parameters that could be measured with these methods (Zitnik et al., 2019). These approaches, along with machine learning, are converging into precision medicine, which is composed of four words (also referred to as P4 for short): predictive, preventive, personalized, and participatory precision medicine. The combination of all four terms leads us to maintain our health longer and prevent noncommunicable diseases (Hood and Friend, 2011; Hood and Flores, 2012; Price et al., 2017). With the combination of 'omics methods, developed models, and evaluation of these methods in practical medicine, future health policies will also change and the chances of detecting diseases as early as possible and before it is too late for effective treatment will also increase. However, there is also a need for caution in introducing this approach into daily use, especially in data protection and better and more secure computing infrastructure (Thapa and Camtepe, 2021).

The hypothesis from section 1.4.3 were assessed in this chapter (table 2).

2.2.2.2 Materials and methods

Total urinary NMR metabolomes collected from five different projects-Slovenian NMR database (PlanHab (Debevec et al., 2014; Sket et al., 2017a; Sket et al., 2017b; Sket et al., 2018; Šket et al., 2020), X-Adapt (Deutsch et al., 2022a), healthy women and men, SMA (Deutsch et al., 2020), PreTerm (Deutsch et al., 2022b)) were integrated with the aim to build up Slovenian NMR database (manuscript in preparation). We used the DIABLO (Singh et al., 2019) and PLSDA (Wang and Lê Cao, 2020) methods, which are integrated into the miXomics R package (Rohart et al., 2017).

2.2.2.3 Results

All urinary metabolites collected in five different projects were utilized: PlanHab (522 samples), PreTerm (183 samples), Spinal Muscular Atrophy (48 samples), X-Adapt (239 samples), Healthy Women and their daughters (94 samples), and Healthy Men and their sons (133 samples); 185 samples were included in the low physical activity group (bedrest part of the form the PlanHab study and spinal muscular atrophy participants), 919 samples were included into medium physical activity group (healthy women and men, start of the PlanHab and Hamb end from the PlanHab study, preterm and full-term born participants from the PreTerm project, untrained participants of the X-Adapt study), and 115 samples were included in the high physical activity group (trained X-Adapt study participants) (Figure 12). The largest area under the curve was observed when comparing the low activity group (AUC=0.91) with the others and the lowest when comparing the moderate activity group with the others (AUC=0.75) (Figure 13).
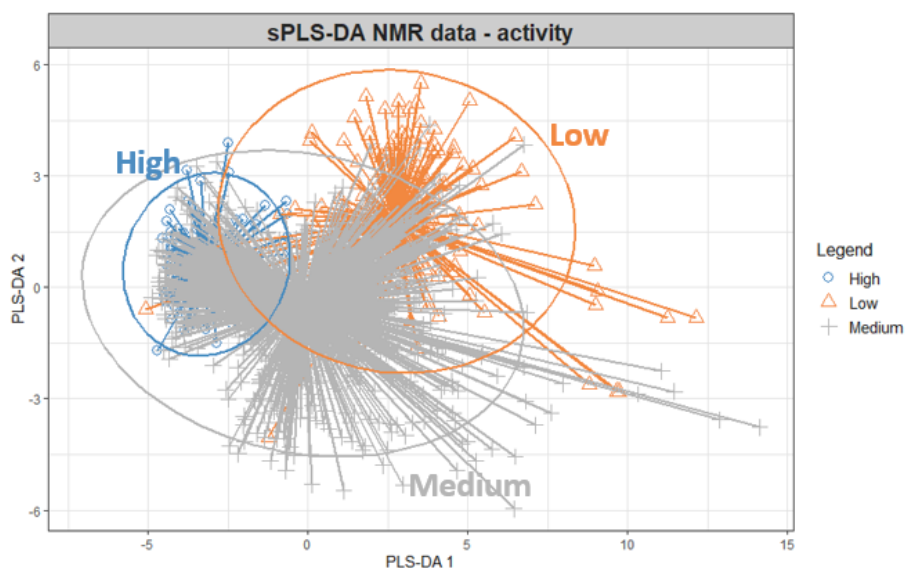
Figure 12:   PSLDA of all metabolomes stratified by activity.
The sample plot representing PLSDA centroids of all 1200 metabolomes obtained in five different dataset and corresponding to their level of physical activity.

Slika 12:   Rezultati analize PLSDA vse metabolomov glede na aktivnost.
Graf prikazuje centroide PLSDA vseh 1200 zbranih metabolomov v petih različnih študijah in razdeljenih glede na nivo njihove fizične aktivnosti preiskovancev.
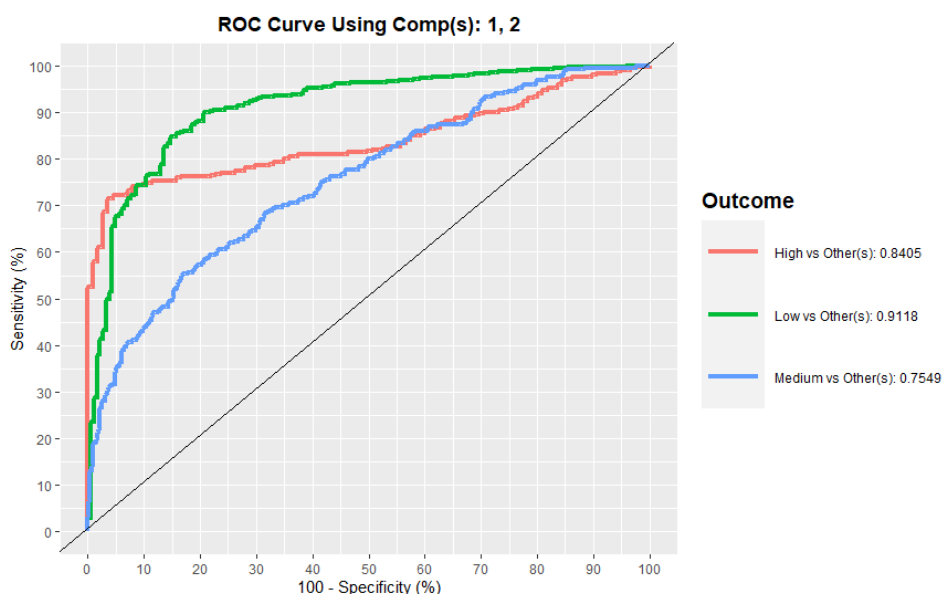
Figure 13:   The success of classification with PLSDA.
ROC curves and with the accompanying AUC values representing the success of classification of metabolomes between three different levels of activity.

Slika 13:   Uspeh klasifikacije z metodo PLSDA.
Krivulje ROC s pripadajočimi vrednostmi AUC, ki prikazujejo uspešnost klasifikacije metabolom glede na nivo fizične aktivnosti.

2.2.2.4 Discussion

We combined more than 1200 collected samples of urine ^{1}H-NMR metabolomes into the Slovenian urine NMR database. Information from all our previous projects (PlanHab, spinal muscular atrophy, X-Adapt, PreTerm, healthy women and men) were integrated. All measured spectra were analysed with the same procedure of spectral deconvolution to obtain metabolites in all projects. We have shown that we can distinguish between the different levels of physical activity based on the metabolites in urine. Future integration of additional data on various diseases with medical diagnoses could provide basis for the development of a pre-screening tool amenable for routine information gathering at clinical setting.

Such a large integrations of metabolomics data into a single database are also susceptible to several sources of systematic error that can lead to lack of reproducibility and poor data quality. To minimize this, all samples were processed in the same way using our in-house processing pipeline (Sket et al., 2017a; Sket et al., 2017b; Sket et al., 2018; Šket et al., 2020; Deutsch et al., 2021a; Deutsch et al., 2021b, Deutsch et al., 2022a; Deutsch et al., 2022b), alongside commercially available software for targeted spectral deconvolution analysis utilizing the same version of the Human Metabolome Database 4.0. Our pipeline is therefore generic and accessible to other interested researchers making

repeated exploration of the same data a reality. In addition, significant extensions with novel data can be made every year with reasonable effort. This should lead to database updates as the Human Metabolome Database has grown from a few thousand metabolites in the first edition (Wishart et al., 2007) to 217,000 metabolites in the latest edition, published in 2021 (Wishart et al., 2021). The data recorded in the past can be effectively reanalysed for novel insight and increased percent of explained spectral information.

Second, standardized analytical protocols in our laboratory allowed us to minimize the systematic errors that normally occur due to batch effects. However, there is still room for improvement. Batch effects need to be eliminated in the integration and construction of databases (Ding et al., 2022). There are already approaches to eliminate batch effects, usually developed in other 'omics domains. These approaches include Dirichlet-multinomial regression (Dai et al., 2019), percentile-normalization methods (Gibbons et al., 2018), quantile regression methods (Ling et al., 2021), the ComBat Bayesian approach (Johnson et al., 2007), Norm ISWSVR (Ding et al., 2022), and the sPLSDA (Wang and Lê Cao, 2020), which was implemented in miXomics. Batch effects can occur when comparing different studies for biological reasons (uniqueness of each biological system due to health status, diet, or lifestyle in general), technical reasons (different batches of the same buffers, different vendors, protocols, NMR devices), or computational reasons (use of different parameters and different software) (Wang and Lê Cao, 2020). Another question is which normalization method is the best for the data being analysed. In the field of metabolomics, the NOREVA software was developed to overcome this challenge. The only limitation is that it is not suitable for NMR metabolomics and was developed for MS metabolomics (Yang et al., 2020). In our case, Box-Cox normalization and sPLSDA approach were used to integrate all metabolomes. This method showed competitive performance in removing batch effects on one side, but still preserves variations due to lifestyle or other biological metadata categories (Wang and Lê Cao, 2020).

We have shown that urinary metabolic fingerprinting has the potential to reveal an individual's metabolic status and provide a snapshot of health and disease (Azad and Shulaev, 2019; Mussap et al., 2021). Metabolomics in general involves the systematic identification of metabolites in the human body. To increase its use in daily medical practise, all levels of metabolomics research should be standardised (sampling, wet lab analysis, and also analytical approaches at the level of algorithms) (Ashrafian et al., 2021). Building a national database will improve the understanding of the Slovenian metabolome and the identification of metabolites specific to particular disease or physical condition. This approach was demonstrated in the Netherlands based on 26,000 collected blood metabolomes in the Dutch Biobanking and BioMolecular Resources and Research Infrastructure (Bizzarri et al., 2022). They showed that ^{1}H-NMR metabolomics can capture a wide range of conventional clinical variables in epidemiological studies and that it is possible to generate predictors for discriminating between different diseases such as diabetes, metabolic syndrome, insulin resistance, inflammation (Crohn disease, ulcerative colitis) based on machine learning. Top-down interpretation of metabolomic datasets consisting of different studies is impossible using simple approaches due to the enormous amount of data (Lakrisenko and Weindl, 2021. In addition, and in line with the above,

metabolomics accounted for the majority of funding and potential research between all 'omics fields. However, it was also noted that the main problem is the lack of standardisation for integrating different metabolomics datasets and that this could be important in the future to increase the confidence of metabolite identification in large datasets but also to address the variability within and between different 'omics fields (Yu et al., 2022). For this reason, newly developed methods that were tailored to specifically address these problems in statistically sound way should be used.

Due to the complexity of the data linked to metadata of patients and/or participants, computational models are needed to understand these data in different ways, such as machine-learning methods (Bizzarri et al., 2022), metabolic networks (Töpfer et al., 2015), constraint-based and kinetic models (Volkova et al., 2020; Lakrisenko and Weindl, 2021). Our database already provides one implementation of the above considerations into sound and effective approach transforming the ^{1}H-NMR urine data into a form amenable for building machine-learning models in the very near future for their use in medical diagnostics. Unknown urine samples could easily be classified as members of either healthy or various disease groups. With this work, we aim to stimulate the interest of other researchers in the field of biomedicine to include NMR metabolomics in their research process in order to complement our newly established database with their concise descriptions of medical conditions in order to reach some 10,000 samples at national scale. This is of relevance due to the central European geographic location of the Republic of Slovenia and its local genetic characteristics coupled to lifestyle habits, dietary characteristics, and environmental conditions.

To summarize, the assembly and modelling of these data to create ML models is a viable approach that can be used in medical practise to distinguish between various disease phenotypes and healthy groups. Taking this approach is one step closer to precision data-driven medicine that would improve health care approach on a national scale. A Slovenian urine NMR database paper is currently in preparation.

3 DISCUSSION AND CONCLUSIONS

3.1 DISCUSSION

In this chapter, we summarise the developments presented within this doctoral thesis in a more comprehensive interrelated manner. First, we focus on "3.1 Developed tools for data integration" then "3.2 Physico-chemical characteristics of microbial world in the gut" and continue with the most important review of the data and findings produced within the four projects "3.3 Metabolomics in the PlanHab study", "3.4 Spinal muscular atrophy", "3.5 X-Adapt project – the influence of short term training on inactive individuals" and "3.6 metabolomes and microbial metagenomes can distinguish pre-term and full-term born adults". Finally, we focus on the most informative part of "3.7 data integration" with concluding remarks "3.8 What about the future?" and extensions of the presented work.

Table 2 lists my personal contributions to each paper published within four years of this PhD.

Table 2:   My contributions to published and unpublished work and postulated hypothesis in the frame of this PhD.

Preglednica 2:   Moj doprinos k objavljenim člankom in postavljene hipoteze v okviru doktorata.

Published or additional work	Leon Deutsch contributions	Postulated hypothesis in PhD proposal
Murovec B., **Deutsch L.**, Stres B. 2019. Computational framework for high-quality production and large-scale evolutionary analysis of metagenome assembled genomes. Molecular Biology and Evolution, 37, 2: 593-598	Conceptualization of analysis, B.S.; **methodology**, B.M., L.D., B.S.; **formal analysis**, L.D., B.S., B.M.; **data curation**, L.D., B.S., B.M.; **writing— original draft preparation**, B.S., L.D.; **visualization** L.D., B.S.; project administration, B.S.; funding acquisition, B.S., B.M.	
Murovec B., **Deutsch L.**, Stres B. 2021. General unified microbiome profiling pipeline (GUMPP) for large scale, streamlined and reproducible analysis of bacterial 16S rRNA data to predicted microbial metagenomes, enzymatic reactions and metabolic pathways. Metabolites, 11, 6: 336, doi: https://doi.org/10.3390/metabo11060336, 14 p.	Conceptualization of analysis, B.S.; **methodology**, B.M., L.D., B.S.; **formal analysis**, L.D., B.S., B.M.; **data curation**, L.D., B.S., B.M.; **writing— original draft preparation**, B.S., L.D.; **visualization** L.D., B.S.; project administration, B.S.; funding acquisition, B.S., B.M.	

Table 2 (continued)

Published or additional work	Leon Deutsch contributions	Postulated hypothesis in PhD proposal
Deutsch L., Osredkar D., Plavec J., Stres B. 2021. Spinal muscular atrophy after nusinersen therapy: improved physiology in pediatric patients with no significant change in urine, serum, and liquor ¹H-NMR metabolomes in comparison to an age-matched, healthy cohort. Metabolites, 11, 4: 206, doi.https://doi.org/10.3390/metabo11040206, 15 p.	Conceptualization for metabolomic analysis, B.S.; **samples collection**, L.D., B.S. and D.O.; **metabolome analysis**, L.D. and B.S.; methodology, D.O., B.S. and J.P.; **formal analysis**, L.D., B.S., D.O. and J.P.; **data curation**, L.D. and B.S.; **writing—original draft preparation**, L.D., B.S., D.O. and J.P.; **visualization** L.D. and B.S.; project administration, D.O. and B.S.; funding acquisition, D.O. and B.S. A	**H0:** There are no significant differences in metabolomes before and after treatment. **H1:** There are significant differences in urine (systemic) and liquor (local) metabolomes before and after treatment with gene therapy, enabling identification of characteristic metabolic pathways discerning the two groups.
Deutsch L., Stres B. 2021. The importance of objective stool classification in fecal 1H-NMR metabolomics: exponential increase in stool crosslinking is mirrored in systemic inflammation and associated to fecal acetate and methionine. Metabolites, 11, 3: 172, doi. https://doi.org/10.3390/metabo11030172, 16 p.	Conception and design of the study (B.S.), **data collection** (L.D., B.S.), **data preparation and analysis** (L.D., B.S.), **writing and critical revision of the manuscript** (L.D., B.S.). A	
Šket R., **Deutsch L.**, Prevoršek Z., Mekjavić I.B., Plavec J., Rittweger, J., Debevec T., Eiken O., Stres B. 2020. Deutsch L., Stres B. 2021. Systems view of deconditioning during spaceflight simulation in the PlanHab Project: The departure of urine 1H-NMR metabolomes from healthy state in young males subjected to bedrest inactivity and hypoxia. Frontiers in Physiology, 11: 532271, doi. https://doi.org/10.3389/fphys.2020.532271, 15 p.	BS provided the concept for metabolome analysis and drafted the manuscript. TD and JR collected the samples. BS, RŠ, and JP designed the metabolome analyses. RŠ, BS, ZP, LD, OE, and IM **conducted the research**. RŠ, BS, and LD **analyzed the data**. RŠ and BS provided necessary code to streamline 1H-NMR spectra analyses and provided statistical analyses.	
Deutsch L., Sotiridis A., Murovec B., Plavec J., Mekjavić I., Debevec T., Stres B. 2022. Exercise and interorgan communication: short-term exercise training blunts differences in consecutive daily urine 1H-NMR metabolomic signatures between physically active and inactive individuals. Metabolites, 12,6: 473, doi. https://doi.org/10.3390/metabo12060473, 18 p.	Conceptualization, T.D. and B.S.; **methodology**, L.D. and B.S.; conceptualization for metabolomic analysis, J.P. and B.S., **formal analysis**, L.D. and B.S.; **data curation**, B.M., L.D. and B.S; exercise database, T.D., A.S., I.M; **writing-original draft preparation**, L.D. and B.S.; **visualization**, L.D. and B.S.; supervision, B.S.; project administration, A.S., T.D., I.M., B.S.; funding acquisition, T.D., I.M., B.S.	

Table 2 (continued)

Published or additional work	Leon Deutsch contributions	Postulated hypothesis in PhD proposal
Deutsch L., Debevec T., Millet G.P.., Osredkar D., Opara S., Šket R., Murovec B., Mramor M., Plavec J. Stres B. 2022. Urine and fecal 1H-NMR metabolomes differ significantly between pre-term and full-term born physically fit healthy adult males. Metabolites, 12: 6, doi. https://doi.org/10.3390/metabo12060536, 23 p.	Conceptualization, T.D., D.O., G.P.M. and B.S.; **data collection**, M.M., S.O., R.Š., L.D. and B.S.; **methodology**, L.D., R.Š. and B.S.; conceptualization for metabolomic analysis, J.P. and B.S., **formal analysis**, L.D. and B.S.; data curation, B.M.; **writing—original draft preparation**, L.D. and B.S.; **visualization**, L.D. and B.S.; supervision, B.S.; project administration, T.D., G.P.M., D.O. and B.S.; funding acquisition, T.D., D.O. and B.S.	**H0:** No significant difference exists between preterm and term groups of participants at the levels of faecal or urine metabolomes or faecal metagenomes. **H1:** There are significant differences between preterm and term groups of participants in faecal and urine metabolomes that can be linked to their physical performance in experiments and physiological data at exercise and rest. **H2:** There are significant differences at the level of metagenomics makeup of both groups, giving rise to identification of specific metabolic pathways differing between the two groups and their gut environment characteristics. **H3:** Term and preterm gut samples contain specific MAGs associated with differences in gut environmental conditions between the two groups.
MAGs assembly	Data collection, formal analysis, visualisation, writing	
Data integration	Data collection, formal analysis, visualization, writing	**H0:** There is no significant difference between metabolomes of prematurely born, born on time, before SMA treatment and post SMA treatment groups. **H1:** There are significant differences in urine metabolomes that enable identification of biomarker pools and metabolic pathways delineating various groups under investigation.

144

3.1.1 Developed tools for data integration

Microbial species play important roles in diverse environments characterised by a wide range of organismal complexity (Murovec et al., 2020). Microbes living in the gut are in constant not only bidirectional interactions with the host but also multidirectional interaction with their microbial counterparts through the production of various molecules that can improve the health status of the host or, in contrast, lead to the development of a noncommunicable disease or its progression (Murovec et al., 2020). Disease progression can manifest as mild gastrointestinal symptoms or as serious diseases such as inflammatory bowel disease, colon cancer, or liver cancer.

It has to be kept in mind that specific proteins and peptides next to metabolites from metabolic reactions mediate the crosstalk between gut, brain, and other peripheral metabolic organs in order to maintain energy homeostasis. The multidirectional interactions between metabolic organs and the central nervous system have evolved in parallel with the multicellularity of organisms to maintain whole-body energy homeostasis and ensure the organism's adaptation to external environmental parameters. These interactions become severely affected in pathological conditions of noncommunicable diseases, such as obesity, insulin resistance, metabolic syndrome or type2 diabetes. Bioactive peptides and proteins next to hormones and cytokines, produced by both peripheral organs and the central nervous system, plus molecules from muscle wear and tear including metabolites from microbiome and energy production/consumption are key messengers in this inter-organ communication (Castillo-Armengol et al., 2019).

A number of diseases were linked to metabolic imbalances that are partially or completely related to the gut microbiome (from metabolic syndrome and obesity to autoimmune diseases, infections, and mental disorders (Murovec et al., 2021). The discovery of sequencing technologies enabled the study of microbes that cannot be cultured. It quickly became clear that most microbes (i.e., 99%) cannot be cultured in the laboratory environment, but we can sequence their genetic material and see which microbes are present in the sample. Based on amplicon sequencing (e.g., 16S rRNA) or whole metagenome sequencing, we can determine which microbes are present in the samples (microbiota) and if coupled to their genetic potential through inference (based on 16S rRNA coupled to nearest genome sequences) or analyse all the genes (based on whole metagenome) that are present in the sample. Based on their genetic potential, we can infer the microbial functionality of the sample (what these microbes most likely can do), enzymatic reactions that they support, next to the metabolic pathways that result from enzymatic reactions and metabolites that are most likely the result of all these numerous transformations (Berg et al., 2020).

A number of different methods were developed for the analysis of sequences in the context of microbiome research. Based on 16S rRNA, Mothur (Schloss et al., 2009) can be used to analyse amplicon sequence material at three different levels: (i) genus (Rühlemann et al., 2021), (ii) 97% 16S rRNA identity operational taxonomic units (Mysara et al., 2017), or (iii) amplicon sequence variants (Callahan et al., 2017; Schloss, 2021). In addition, another set of tools was developed for predicting

microbial functionality based on amplicon sequences: PICRUSt (Langille et al. 2013a), PICRUSt2 (Douglas et al., 2020), Tax4Fun (Aßhauser et al., 2015), Tax4Fun2 (Wemheuer et al., 2020), and Piphillin (Narayan et al., 2020). These tools link 16S rRNA sequence information to reference genome sequences and predict microbial potential based on metagenomic functional gene content (Sun et al., 2020). We have developed GUMPP for large-scale, streamlined, and reproducible analysis of bacterial amplicon data and prediction of their functional potential (Murovec et al., 2021), consisting of Mothur (Schloss et al., 2009), PICRUSt2 (Douglas et al., 2020), and piphillin (Narayan et al., 2020) in order to support large scale data analyses. Thus far, more than 600 samples from 32 studies amounting to 120 million reads were analysed in meta-analysis project (Klammsteiner, University of Innsbruck, in preparation).

The more objective analysis of functionality of microbes cannot be studied without sequencing the entire metagenome directly. Whole metagenome sequencing involves the untargeted sequencing of a random subset of all sequences to certain read depth, not like in targeted (amplicon) sequencing in which only a small portion of a specific gene is sequenced. BioBakery (McIver et al., 2018; Beghini et al., 2021) is the workflow for whole metagenome sequence analysis that combines different tools for quality analysis, taxonomic analysis (MetaPhlAn), functional genes, enzymatic reactions, and metabolic pathways of interest in the microbial community (HUMAn3). In addition, the extension of this method by utilizing training on actual metagenomes coupled to lipid-soluble and water-soluble metabolomes determined through mass spectrometry allows prediction of microbial metabolites on metagenome information alone and hence describing the metabolomes that might be produced in this community (MelonnPan). Another positive aspect of whole-genome sequencing is that information on genetic material can be obtained from different taxonomic groups (archaea, bacteria, protozoa, fungi, DNA viruses, (also human DNA)), which can improve the understanding of the complexity and interactions between different taxonomic layers. We are in the process of publishing the developed metaBakery workflow (manuscript in preparation), which is a re-implementation of the BioBakery workflow, with the addition of the sequence QC steps, extended with diversity calculators implemented within Mothur, guided by our in-house skeleton application, and implemented as Singularity container for large-scale, streamlined, and reproducible analyses at HPC setting.

The next step in whole metagenome sequencing is the possibility of de-novo metagenome assembly. This is a process in which reads are screened for quality, assembled, and binned together to yield assembled metagenomes. This process can lead to the discovery of entirely new species. However, care must be taken in this process regarding the completeness and contamination of the newly assembled genomes. According to the MIMAG standard (Bowers et al., 2019), we should all strive to assemble the most complete (> 95%) and least contaminated (< 5%) MAGs. These will enable the next stage of evolutionary analysis and hopefully provide new ideas on how microbes interact with human beings as their host. We have developed the Metagenome-Assembled Genomes Orchestra (MAGO (Murovec et al., 2020)) from highly successful tools for quality analysis (FastQC, fastp (Chen et al., 2018)), assembly (IDBA-UD (Peng et al, 2012), metaSPAdes (Nurk et al, 2017) and megaHIT (Li et al., 2015)) and binning (maxBin (Wu et al., 2016), MetaBAT (Kang et al., 2015),

CONCOT (Alneberg et al., 2014), BinSanity (Graham et al., 2017), and DAStool (Sieber et al., 2018)). In conjunction, the CheckM tool (Parks et al., 2015) is used to filter out which MAGs are of high quality according to the MIMAG standards (Bowers et al., 2019). The MAGO tool also allows the user to analyse the evolution of the MAGs obtained using ezTree (Wu, 2018), average amino acid identity (AAI) enables insight into species cut-off values, Prokka (Seemann, 2014) serves for genome annotation while Roary (Page et al., 2015) provides pan- and core-genome analysis, and FastANI enables nucleotide identity analysis of genomes. The resulting bins are then selected based on their completeness and contamination according to MIMAG standard and analysed subsequently using other tools (Castro et al., 2018; Rodriguez-R et al., 2018; Ruiz-Perez et al., 2021).

The three tools for large scale data analyses presented in our work (MAGO, GUMPP, metaBakery (manuscript in preparation)) were prepared as a skeleton framework consisting of more than 10,000 lines of code written in Python, which orchestrates the execution of each part and takes care of the execution of programs and the creation of their command lines (Murovec et al., 2020; Murovec et al., 2021). The parameters for the execution of the workflow are entirely in the hands of the user. All tools were developed as Singularity images (Kurtzer et al., 2017) prepared for straightforward deployment on HPC for large-scale, straightforward analysis of 10,000 samples as well as for educational purposes. Both, metaBakery and MAGO tools were used for metagenomic sequence analysis in the PreTerm project (Deutsch et al., 2022b). Both tools are under the CC-BY 4.0 open-source license and are open to any extensions, thus providing the opportunity to develop further and become standardized workflows for microbial analysis on a global scale. metaBakery (in preparation) will be used in the future project that is part of the Million Microbiomes from Human Project (MMHP, (Fang et al., 2018; Han et al., 2018; Patterson et al., 2019)) and will provide insight into the Slovenian gut microbiome. Currently, 5000 deep sequencing samples (10 mio reads/sample) encompassing 13 gastrointestinal diseases including depression next to healthy state (14 conditions) from 22 states were processed utilizing 1.2 million CPUh on a VEGA supercomputer (in preparation), providing thus another well represented dataset amenable for ML exploration.

3.1.2 Physicochemical characteristics of microbial world in the gut

The peristaltic waves that create the contractile patterns of the small intestine create an environment that is constantly changing. The constant mixing of faecal material results in changes in environmental conditions for the microbes living in the gut, such as pH, which can affect microbial growth (Ehrlein and Schemann, 2005; Johnson et al., 2012; Cremer et al., 2016; Glover et al., 2016; Cremer et al., 2017; Sket et al., 2017a). A number of studies have linked stool consistency, the microbial living environment, to the richness of the gut microbiota, its composition, enterotypes, elevated inflammatory levels, lipopolysaccharides, and bacterial growth rates (Tigchelaar et al., 2016; Vandeputte et al., 2016). Stool consistency was mostly assessed with BSS method (Heaton et al., 1992; Lewis and Heaton, 1997). Lower BSS scores were associated with longer colonic transit time, higher microbial richness, and protein catabolism (Roager et al., 2016). Alteration of the microbiota and the occurrence of local inflammation was previously found to be correlated with BSS. High intra-

and inter-rater variance was observed in the assessment of the BSS (Derrien et al., 2010; Chumpitazi et al., 2016). The assessment of the BSS is based on self-assessment, which may be biased. Therefore, only a well-trained expert can draw medically important conclusions based on the BSS alone (Matsuda et al., 2021).

Faecal materials are semisolid materials (i.e., pastes) in terms of material physics (Grillet et al., 2012), which places them between viscoelastic materials (semipermanent deformation in response to external forces) and plastic materials (permanent deformation). This way of thinking led us to the evaluation of the minimal pressure approach for the less biased and high-throughput evaluation of the consistency of faecal material (Deutsch and Stres, 2021). Minimal pressure, expressed as force per unit area, is the pressure required to cause permanent deformation of faecal material. We have shown that MP increases exponentially compared to decreasing values of BSS, regardless of the sex of the individuals (Deutsch and Stres, 2021). We demonstrated that there is a nonlinear (asymptomatic) and complex relationship between dry matter and MP. Longitudinal mapping of the surface MP over the entire length of a single stool sample revealed that various fine-grained internal, local differences existed. In addition, despite the BSS uniform scoring of lower BSS values, our analysis showed that a more resistant stool surface layer was followed by softer internal structures, resulting in lower MP values associated with approximately healthy stool consistency (Deutsch and Stres, 2021).

We found a boundary that may distinguish between healthy state (MP < 75) or constipation (MP > 75) (Blake et al., 2016; Sket et al., 2017b; Sket et al., 2018). MP < 30 corresponded to aqueous stool samples. MP approach introduced the continuous scale, which can be measured to overcome the problems of BSS assessment errors in BSS around 3 and 4, which are difficult to determine based on visual inspection, despite the training and visual support in classification (Deutsch and Stres, 2021).

MP was measured on the samples collected within the PlanHab (Šket et al., 2020) and the PreTerm study (Deutsch et al., 2022b). Notably, the past studies demonstrated that blockage of faecal surface pores and mucus retention were associated with selective pressure on the gut microbiome, its gene expression, and metabolic activity, leading to local inflammation (Vandeputte et al., 2016; Sket et al., 2017a; Sket et al., 2017b; Sket et al., 2018; Aron-Wisnewsky et al., 2019). Thus, we showed that the MP approach can accurately describe the clinical significance of stool consistency (Deutsch and Stres, 2021). In addition, the MP approach does not require the pre-treatment of samples and allows for ease of measurement without expensive equipment, as well as reproducibility of these measurements with different samples (fresh vs. frozen; male vs. female), with simple correction for the temperature of measurement. We also found that MP correlates with faecal methionine and acetate based on ^{1}H-NMR measurements. Based on these two metabolites, we can distinguish three different groups of faecal consistency (MP < 30, 30<MP<75, MP>75). Methionine was previously associated with oxidative stress and was elevated in inactive individuals, while acetate correlated negatively with insulin sensitivity, indicating that different stool consistencies may have an impact on the biological system of the host. The observed differences in methionine and acetate associated with MP, were thus apparently consequence of inactivity coupled with Western diet as based on the samples collected

within the PlanHab project. The MP approach enabled us to show that with the measurement of some physicochemical parameters and 'omics methods, a completely new level of understanding of complex biological systems can be commenced and explored (Deutsch and Stres, 2021).

3.1.3 Metabolomics in the PlanHab study

The PlanHab study was the first study by our group to examine the metabolomics of human urine (Šket et al., 2020). The run-in and the following three 21-day interventions (NBR, HBR, and HAmb) in a crossover manner) were performed. Morning urine samples were collected throughout the experimental setup (Sket et al., 2017a; Sket et al., 2017b; Sket et al., 2018). The unique crossover design allowed us to consider the responses of the same participants to all three experimental variants under controlled dietary, environmental, and experimental conditions. A total of 523 urine samples were collected and prepared for ^{1}H-NMR measurements. Participants in the bed rest group (NBR and HBR) had specific metabolic compositions compared with the HAmb group. We concluded that the decision of the host to minimize physical activity under hypoxic conditions can be detected within a few days at the level of the urine metabolome measured by NMR. Under normoxic bed rest conditions, these metabolic changes became detectable within the first ten days. The metabolites identified in this study were associated with a number of different diseases: (i) chronic obstructive pulmonary disease (Adamko et al., 2015; Ząbek et al., 2015) and (ii) cardiovascular disease associated with tissue hypoxia, which can also lead to type 2 diabetes, depression, and osteoporosis (Wang et al., 2011a; Senn et al., 2012; Adamko et al., 2015; Ząbek et al., 2015). The PlanHab study utilizing urine ^{1}H-NMR metabolomes led us to conclude that there is no simple metabolic biomarker that could distinguish between different states (healthy vs. sick, active vs. inactive; active vs. sedentary). Complex multivariate descriptions of metabolism were needed to capture commonalities in human physiology, interpersonal variability, and temporal variability. This concept was utilized in all other subsequent studies. For instance, a metabolite could be up- or down-regulated depending on the metabolic pathway. Overall, inactivity alone or in combination with hypoxia resulted in decreased systemic metabolic diversity, increased number of metabolic pathways affected, and more rapid metabolic deconditioning leading to the development of negative physiological symptoms such as insulin resistance, low-level systemic inflammation, constipation, depression, and metabolic syndrome (Sket et al., 2017a; Sket et al., 2017b; Sket et al., 2018). The results of the PlanHab study encouraged us to continue our research utilizing samples from other studies involving different levels of inactivity, such as X-Adapt (differences between trained and untrained individuals), spinal muscular atrophy, and the PreTerm project (Figure 14), which compares different times of exposure to hypoxia, physical activity, and time of exposure to different conditions (Šket et al., 2020).
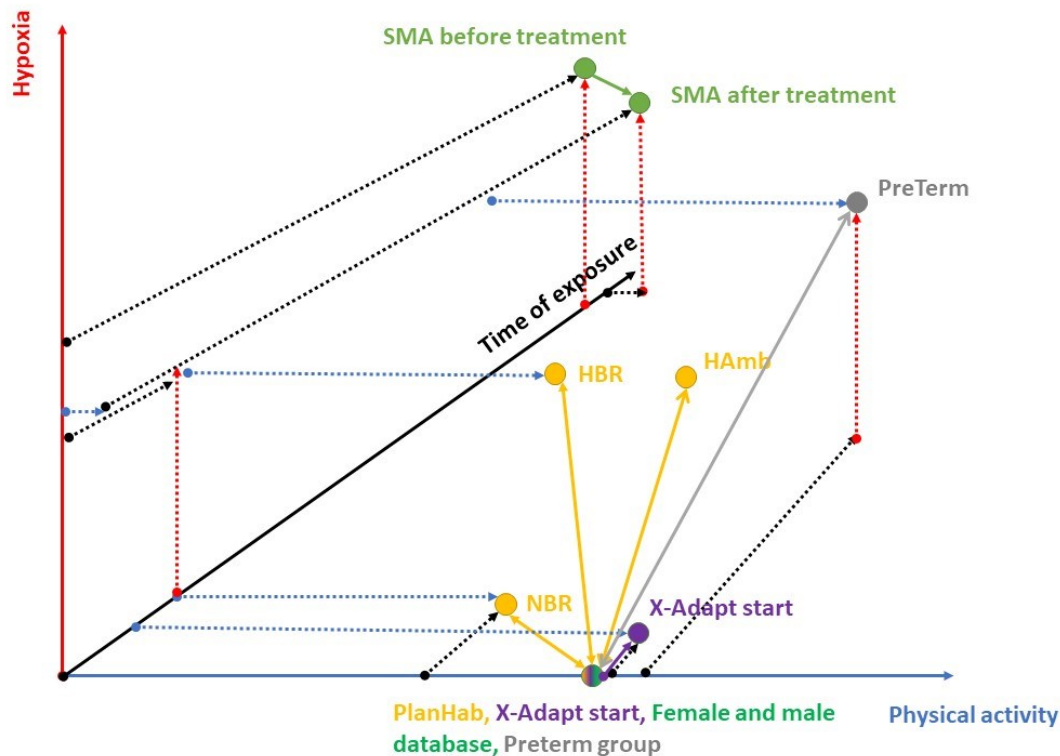
Figure 14:   Representation of studies involved in this work
Representation of studies involved in this work with the relation to physical activity and hypoxia exposure.

Slika 14:   Prikaz študij udeleženih v tem delu
Prikaz študij udeleženih v tem delu glede na stopnjo fizikalne aktivnosti na eni strani in izpostavitve hipoksiji na drugi.

3.1.4 Spinal muscular atrophy

Spinal muscular atrophy is a neuromuscular disease that manifests as progressive atrophy and weakening of skeletal muscle due to progressive loss of motor neurons and also affects a number of other organ systems (Melki, 2017; Yeo and Darras, 2020). With an incidence of 1 per 11,000 births, it is still considered the most common genetic cause of child deaths (Sugarman et al., 2012). In SMA patients, mutations in the centromeric SMN2 gene lead to the formation of unstable proteins and, at the same time, the expression of the telomeric SMN1 gene is also impaired due to deletion (Lefebvre et al., 1995; Lorson and Androphy, 2000; Lunn and Wang, 2008; Smeriglio et al., 2020). In recent years, new therapies have been developed for the treatment of SMA. These therapies alter the natural course of the disease by changing the expression of or replacing mutated genes involved in the development of SMA (Chiriboga et al., 2016). Nusinersen was the first drug approved by the Food and Drug Administration in the United States and by the European Medicines Agency for SMA. Nusinersen is an antisense oligonucleotide that modifies mRNA splicing, resulting in an active SMN 2 protein and thus better SMA outcomes (Chiriboga et al., 2016; Corey, 2017; Ramdas and Servais, 2020). It must be administered intrathecally because it cannot cross the blood-brain barrier (Faber et al., 2007; Rigo et al., 2012).

Urine, liquor, and serum samples from SMA patients were collected before treatment and after the 4th application of nusinersen. Medical examination at the 4th application showed improvement in mobility. The application of nusinersen resulted in better movement, easier writing and sitting or standing, and increase in strength, all as measured by the Children's Hospital of Philadelphia Infant Test of Neuromuscular Disorders (CHOP INTEND (Glanzman et al., 2010), the Hammersmith Functional Motor Scale (HMFS (Pera et al., 2017)), the Expanded Hammersmith Functional Motor Scale (HMFSE (Pera et al., 2017)), or the Motor Function Measurement (MFM (Bérard et al., 2005)) tests. Patients showed improvement in wheelchair control, ambulation, fatigue, hygiene, speech, and sleep after the 4th application of nusinersen (Osredkar et al., 2021).

In contrast to the physical examinations, we could not establish that based on the npMANOVA test on all metabolic matrices (urine, liquor, serum) regardless of gender or data transformation (Deutsch et al., 2021a). In this context, we could not reject the null hypothesis from section 1.3.2 and table 2, which states that there are no significant differences before and after treatment. Perhaps these differences could be confirmed after 10 applications of nusinersen, but this would take too much additional time to collect the samples and to complete within the timeframe of this doctoral thesis. These results show that the efficacy of nusinersen can be seen with the medical examinations and the assessment test. Perhaps the use of other metabolomics methods such as mass spectrometry, which is more sensitive to nanomolar concentrations compared with NMR, would lead to the detection of biomarkers that could be used as biomarkers for monitoring nusinersen treatment (Emwas et al., 2019).

In addition, a series of urine samples were collected from the matched healthy cohort to compare the metabolomes of SMA patients with the metabolomes of healthy individuals. This comparison led to the observation of a significant metabolic difference between females and males ($p=0.0001$), as well as the healthy cohort and the SMA patients. The npMANOVA showed the importance of gender ($F=54.9$; $p=0.0001$) and SMA status before and after treatment ($F=20.7$; $p=0.0001$) to be significant. Both methods, PLSDA and Random Forest, showed significant differences between female and male metabolomes, and we also detected different metabolic diversity when comparing SMA patients to a comparable healthy cohort. A significant reduction in the cumulative concentration of metabolites was observed in SMA patients ($p < 0.05$). The reduction in the number of metabolites was also observed in healthy females compared to healthy males. This was the first report describing the existence of differences between males and females. Because of these differences, it is important for future studies to include a larger number of females in studies such as this one to determine the important differences between female and male metabolic makeups and pathways. There are some preliminary parallels with studies of exercise showing that metabolite counts may increase after exercise (Nieman et al., 2013; Schranner et al., 2020) or studies of bed rest (e.g., PlanHab), which also showed a 30% reduction in metabolite counts after three weeks of bed rest (Sket et al., 2017a; Sket et al., 2017b; Sket et al., 2018). Symptoms such as insulin resistance, bone and muscle loss,

changes in lipid metabolism were all detected in bedrest studies, and all of these symptoms can be observed also on the list of conditions associated with SMA.

We used urine metabolomes from SMA patients and healthy individuals to create a classification model to distinguish between these two conditions. For this purpose, the JADBIO machine learning was used (Tsamardinos et al., 2022), and logistic ridge regression was selected with an AUC value of 0.958 as the best model to distinguish SMA patients and healthy controls. Creatinine was the key metabolite separating healthy from SMA-affected participants as was also reported a few months before our publication in another study that monitored the SMA progression of denervation with elevated levels of creatinine in more severe forms of SMA disease (Alves et al., 2020). Creatinine concentrations did not change significantly in SMA patients before and after the 4th application of nusinersen. The increased creatinine levels were also observed in urine samples from our bed rest studies (PlanHab (Šket et al., 2020)). The reintroduction of exercise completely reversed the adverse effects in these studies (Sket et al., 2017a; Sket et al., 2017b; Sket et al., 2018; Šket et al., 2020). Immobilized patients receiving vibration therapies benefited compared with controls and may represent a potential step in the physical activation of SMA patients after nusinersen therapy (Hoff et al., 2015) in the future due to involuntary contractions of muscles during balancing (Figure 15).
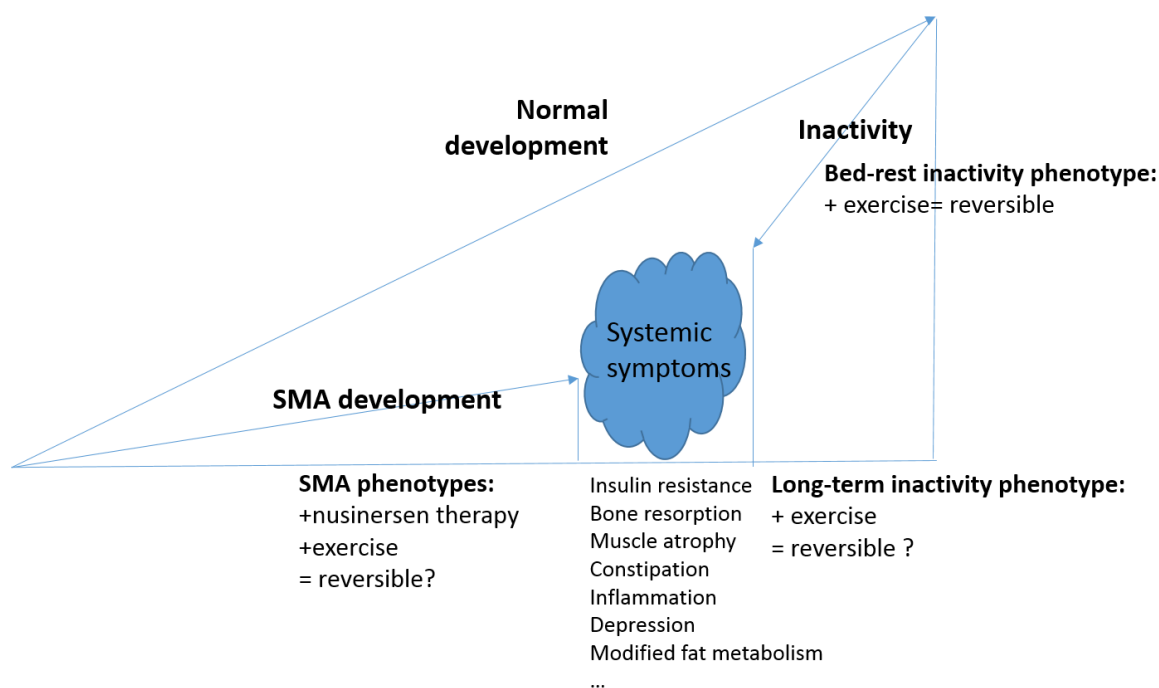
Figure 15:   Model representing results of inactivity.
Model representing the general results of long-term inactivity due to illness or bed-rest studies. All levels of inactivity can result in systemic symptoms leading to noncommunicable diseases. Resuming or reintroducing physical activity can reduce these symptoms and lead to better treatment and health outcomes.

Slika 15:   Model predstavlja rezultate neaktivnosti.
Model predstavlja rezultate dolgočasne neaktivnosti nastale zaradi bolezni ali študij ležanja. Ne glede na razlog, vse vrste neaktivnosti, vodijo v pojav sistemskih simptomov, ki se kažejo kot kronične bolezni. Povečana fizikalna aktivnost, lahko izboljša zdravje ali zdravljenje takih bolezni.

3.1.5 X-Adapt project – the influence of short-term training on inactive individuals

We investigated complete inactivity within the context of the SMA project. However, in the 21st century, it is becoming increasingly clear that physical inactivity, which is the consequence of a sedentary lifestyle and physically less challenging working conditions, is also a global problem that poses a risk for the development of chronic noncommunicable diseases and increased global mortality (Kelly et al., 2020b). It was showed that minimizing sedentary time can reduce the risk of chronic diseases such as coronary heart disease, type 2 diabetes, metabolic syndrome, etc. (Sallis et al., 2016). The goal of the X-Adapt project was to examine the differences between physically active (trained participants) and inactive individuals (Sotiridis et al., 2018; Sotiridis, 2019b; Sotiridis et al., 2019; Sotiridis et al., 2020). The project pre-screened the participants and enrolled 10 active and matching 10 inactive male participants in the 10-day training protocol, which consisted of daily training on a cycle ergometer at 50% of maximal pedalling power under normoxic and normobaric (~1000 hPA) conditions at 24°C ambient temperature. Before participating in the 10 days of training, all participants (active and inactive) underwent the three-day testing under thermoneutral normoxic and hypoxic conditions next to hot normoxic conditions. Study participants were classified as trained or

untrained based on their maximal oxygen output (untrained $VO_2max < 45$ mL-kg-1-min-1, trained $VO_2max > 55$ mL-kg-1-min-1) (Jay et al., 2011; Montero and Lundby, 2017). The trained participants practiced their activities several times per week (running, swimming, cycling) and the untrained participants were asked not to participate in organized sports but were allowed to be active because of commuting (cycling to work). Urine was collected from all participants before the start of the study, at pre-testing, after 10 days of training and after the study (Armstrong and Barker, 2011; Sotiridis et al., 2018; Sotiridis, 2019a; Sotiridis et al., 2019; Sotiridis et al., 2020; Deutsch et al., 2022a).

The measurements directed at human physiology showed that there were some nearly significant and statistically significant differences between trained and untrained subjects at pretesting, and that there were nearly significant (but still insignificant) differences even after only 10 days of training when comparing pre- and post-training, suggesting that some characteristics may be observed in subjects leading an active lifestyle. The differences between the condition before and after training were larger in the untrained groups and based on the measurements of VO_2max before training and its change during the 10 days of training the rate of adaptation to training is greater in untrained individuals. Based on physiological measurements, we observed that the untrained and trained groups became synchronized in terms of the measured training parameters (Sotiridis et al., 2018; Sotiridis, 2019a; Sotiridis et al., 2019; Sotiridis et al., 2020; Deutsch et al., 2022a).

Based on urine metabolomics, no significant difference could be detected between urine samples before and after 10 days of training. However, differences were observed between trained and untrained urine ^{1}H-NMR metabolomes. In addition, urine physicochemical properties (pH, total dissolved solids, salinity and conductivity) also differed significantly between these two groups. For example, pH was decreased in untrained individuals, a condition previously associated with metabolic syndrome and chronic heart failure (Maalouf et al., 2007; Otaki et al., 2013; Kraut and Madias, 2016; Shimodaira et al., 2017).

Metabolites (cholate, tartrate, cadaverine, lysine, N6-acetilysine, methanol, N-acetylglucosamine, butanone, and caprate) were identified as metabolites responsible for differentiation between trained and untrained group using multivariate statistics and machine learning. All metabolites were previously observed in studies related to muscle damage, hormone receptor levels, recovery after resistance training, lower cardiovascular risk (tartrate) (Abramowicz and Galloway, 2005; Spiering et al., 2008) or atrophic state in myotubes, and obesity (cholate) (Li et al., 2020; Abrigo et al., 2021; Alamoudi et al., 2021; Mercer et al., 2021; Pushpass et al., 2021; Zheng et al., 2021). Cholate is a primary bile acid that was enriched in the untrained group, which was previously associated with the development of cancer. Incidentally, increased concentrations of primary bile acids in the bloodstream were observed in less fit women, and a single training run may decrease the amount of these compounds (Danese et al., 2017; Maurer et al., 2020).

Lysine and cadaverine are polyamines previously associated with metabolic syndrome and colon or liver cancer (cadaverine was elevated in the untrained group). Lysine is involved in aminoacyl-tRNA biosynthesis, a metabolic pathway enriched in the trained group and previously correlated with higher physical activity, which may be due to changes in protein synthesis in active subjects (Robinson et al., 2017; Castro et al., 2019; Tabone et al., 2021; Tian et al., 2021). 2-hydroxy-3-methyl valerate was increased in untrained participants, which may affect energy metabolism via PPAR-α, as previously shown in older, functionally impaired adults (Coen et al., 2013; Lustgarten et al., 2014).

Using this approach, we showed that the entire system in active subjects was significantly different from that of inactive subjects (p=0.003). After 10 days of training, the significance of difference disappeared at the end of the campaign (p=0.226) (Figure 16). It became clear that minor metabolomic differences existed between the metabolomes of trained and untrained subjects, which remained physiologically completely different with respect to their physical capabilities. Therefore, lifelong training would be required to maintain a healthy metabolome phenotype. Our study showed that an exercise load 5 times higher than the 75–150 minutes per week recommended by WHO is effective (Sallis et al., 2016; Kelly et al., 2020b). In addition, this experiment has shown that 3-day morning urine samples provide a good biological matrix for discriminating active from inactive individuals, which cannot be observed in a 1-day sampling because of diurnal variability. Systemic homeostasis depends on a number of different parameters and involves communication between different organs through which metabolic pathways affected by a metabolite in one organ can affect other metabolic pathways in another organ. A sedentary lifestyle can disrupt this communication between organs, leading to the manifestation of various diseases. Higher levels of exercise can restore interorgan communication in physically inactive individuals towards that of healthy and active individuals (Di Liegro et al., 2019; Deutsch et al., 2022a).
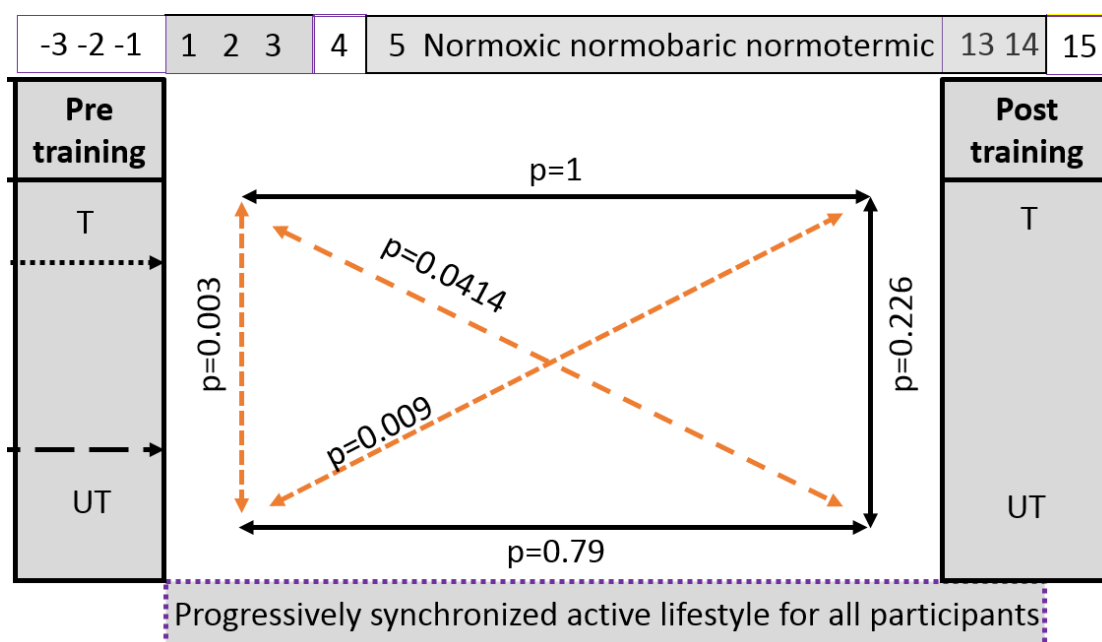
Figure 16: Change between trained (T) and untrained (UT) participants of X-Adapt study.
The entire system in active subjects was significantly different from that of inactive subjects (p=0.003) before the X-Adapt study. After 10 days of training, the significance of difference disappeared at the end of the campaign (p=0.226).

Slika 16: Sprememba med treniranimi (T) in netreniranimi (UT) udeleženci študije X-Adapt.
Na začetku kampanje je bil celoten sistem treniranih udeležencev študije X-Adapt drugačen od netreniranih udeležencev (p=0.003). Po 10-dnevnem treniranju je ta razlika na nivoju celotnega Sistema izginila (p=0.226).

3.1.6 Metabolomes and microbial metagenomes can distinguish preterm and full-term born adults

Preterm birth is defined as a birth before 37 weeks gestation; approximately 10% of births are preterm worldwide, and it is still one of the leading causes of death in children under 5 years of age. Preterm birth increases the risk of developing various chronic diseases such as cardiovascular, endocrine/metabolic, renal, neurological, and psychiatric disorders. One of the main causes of these disorders is increased oxidative stress in the first weeks of life (Moutquin, 2003; Magalhães et al., 2004; Pialoux et al., 2009; Blencowe et al., 2012; Lushchak, 2014; Liu et al., 2015; Manley et al., 2015; Debevec et al., 2017; Crump, 2020; Tingleff et al., 2021). There is a high probability that some clinical parameters such as body fat mass, arterial blood pressure, fasting glucose and cholesterol may be elevated (Kerkhof et al., 2012; Markopoulou et al., 2019; Crump, 2020). All of these characteristics were shown in various studies to be different between preterm and full-term born adults and that these differences are particularly related to the production of reactive oxygen species, and can be observed in association of different levels of exercise or physical activity (Magalhães et al., 2004; Powers et al., 2011; Filippone et al., 2012; Debevec et al., 2017; Martin et al., 2018).

The aim of the PreTerm project was to investigate whether differences of blunted ventilatory response (HVR) exist in physically fit young men (born preterm and full-term) under hypoxic and normoxic

environmental conditions at rest and during physical activity (Debevec et al., 2019; Debevec et al., 2022). In addition, a high-throughput analytical approach consisting of urine and faecal metabolomics and faecal metagenomics was used to describe the complexity of the human body and gut microbiome in its response to increased oxidative stress levels at rest and during exercise in normoxia and hypoxia (Martin et al., 2020).

A total of 37 men were enrolled in this study (15 born full- term and 22 born preterm). Incremental cycling in normoxia and hypoxia were shown to increase levels of oxidative proteins, catalase, superoxide dismutase, and nitrosative markers in both groups immediately after exercise (Martin et al., 2020). Participants in the preterm group showed lower exercise capacity in normoxia compared with the full-term group and had lower HVR, whereas no such difference was observed in hypoxia (Vrijlandt et al., 2006; Lovering et al., 2013; Svedenkrans et al., 2013; Bates et al., 2014; Clemm et al., 2014; Farrell et al., 2015; Debevec et al., 2019). These results indicate that preterm infants may have increased oxidative stress during acute exercise in normoxia, whereas such a response was not observed in hypoxia (Martin et al., 2020).

We measured 25 physicochemical variables in the stool samples (including the MP approach described above), and no significant differences were found between the preterm and full-term groups. These results indicate that there were no differences in gut environment parameters between preterm and full-term infants, regardless of environment (hypoxia vs. normoxia). Faecal and urine samples were collected three days before and three days after the hypoxic and normoxic tests. Multivariate statistics based on 1- and 2-way PERMANOVA showed that there were significant differences between preterm and full-term participants based on faecal and urine metabolome, but not between pre-test and post-test in normoxia and hypoxia (Deutsch et al., 2022b).

Acetone, tartrate, and trans-aconitate were metabolites that were decreased in the preterm group according to the MetaboAnalyst's results. These metabolites are associated with exercise, fasting, or diabetes mellitus (Paradis et al., 2015; Crump et al., 2019; Perrone et al., 2021). Based on the urinary metabolome, the most interesting enriched metabolic pathway (D-arginine and D-ornithine metabolism) was described previously in association with systemic or tissue hypoxia (Qiu et al., 2017; Haraldsdottir et al., 2019). The differences appear to be due to impaired autonomic function because heart rate recovers more slowly in preterm adults, which could lead to anoxia and increase their cardiovascular risk, as previously suggested (Sonntag et al., 2007; Ten, 2017).

Faecal metabolomes also differed between preterm and full-term participants. Lactate, serotonin, and tyrosine were the major metabolites that accounted for the difference between the preterm and full-term groups. The first two were increased in the preterm group, which, together with the enriched metabolic pathway (Warburg effect), shows that some metabolic changes can be observed in preterm infants. The Warburg effect was described previously in preterm infants and associated with mitochondrial dysfunction (McIver et al., 2018). These findings may represent the first evidence that systemic differences due to lifelong exposure to oxidative stress do indeed exist and raise the question

157

of whether these differences are associated with minute differences generated on the part of the preterm host or on the part of the microbiome responding to these environmental signals or with their mutual interaction in the form of a complex biochemical network in the steady state (Deutsch et al., 2022b).

The collected faecal samples were used for shotgun sequencing to investigate whether the observed differences in faecal metabolomes exist at the microbial level. No significant differences were observed at the taxonomic level, but the relative abundances of archaea and viruses were higher in the preterm group and would deserve further, more detailed inspection using larger sample collections. The calculation of Shannon and other diversity indices showed that microbial diversity was higher in preterm group. In the previous decade, it became clear that the more important question in the study of the microbiome is what the microbes in our gut are doing. For this reason, HUMAnN3 and metaPHLAnN3 were used to determine which gene families, enzymatic reactions, metabolic pathways, and predicted metabolites can be used to distinguish between preterm and full-term born adults. Machine learning was used to build classification models for this purpose utilizing JADBio (Deutsch et al., 2022b).

No significant differences were detected based on gene families, but we did detect some differences based on enzymatic reactions, metabolic pathways, and predicted metabolites. The previously described RXN-15378 enzymatic reaction of succinate dehydrogenase was increased in the preterm group. Succinate itself is a microbial metabolite and can accumulate in the intestinal tract during inflammation or microbial imbalances. It has tissue-specific but also pro-inflammatory properties and is also a source of propionate production by *Bacteroides* spp. and *Prevotella* sp. Succinate was shown to accumulate in cells under low-oxygen conditions and represents the metabolic signature of hypoxia. Excessive uptake of microbially produced succinate was shown to lead to higher levels of intracellular succinate, which slowed down prolyl- hydroxylase activity through product inhibition and lead to additional activation and stabilization of HIF-1α beyond the response to hypoxia itself, which significantly enhanced LPS-induced expression of proinflammatory cytokines in human cells (Rubic et al., 2008; Ariza et al., 2012; Tannahill et al., 2013; Akram, 2014; Littlewood-Evans et al., 2016; Connors et al., 2018).

PWY-7456 (β-(1,4)-mannan degradation), PWY-7323 (superpathway of GDP-mannose-derived O-antigen building blocks biosynthesis) and GLYCOLY-SIS-TCA-GLYOX-BYPASS (superpathway of glycolysis, pyruvate dehydrogenase, TCA, and glyoxylate bypass), P221-PWY (octane oxidation), PWY-5173 (unclassified) were pathways that were increased in the preterm group. Some of them may be beneficial and strive for mucosal integrity and host nutrition (β-(1,4)-mannan degradation) or significantly increase energy production, which would be important in the case of oxidative stress as in preterm individuals (super-pathway of glycolysis, pyruvate dehydrogenase, TCA, and glyoxylate bypass). Acetyl-CoA biosynthesis may also lead to increased production of butyrate via the production of acetyl-CoA. In contrast, some pathways have a more negative effect and were also increased in the preterm group. These pathways were shown to be involved in lipopolysaccharide

LPS production (GDP-mannose-derived O-antigen building blocks biosynthesis) associated with Gram-negative bacteria and causative agent of different degrees of inflammation (Samuel and Reeves, 2003; Wolfs et al., 2010; Shah et al., 2015; Kim et al., 2016; La Rosa et al., 2019; Lindstad et al., 2021). However, octane oxidation, previously described in the context of westernization of the human gut and associated with liver disease, was also observed (Deutsch et al., 2022b). All these differences can be associated with the physiologically significant deficits observed between both groups (Martin et al., 2018; Martin et al., 2020).

Seventeen predicted metabolites were also detected by the ^{1}H-NMR approach, none of which were considered important for differentiation in the machine learning. Significant differences were detected in the urine and faecal metabolomes in addition to predicted metabolites, suggesting that systemic differences between the two groups exist. Elevated metabolites were previously associated with cardiovascular disease (carnitine), increased intestinal permeability, elevated levels of inflammatory cytokines, metabolic syndrome, or cancer growth (putrescine and diacetylspermine). In contrast, some predicted metabolites were decreased in the preterm group. Deoxycholate is a secondary bile acid and a known promoter of colon cancer. The decreased levels of this molecule were generally observed due to the increased urinary excretion of cholate observed in urine metabolomics. Given the physiological differences between the two groups examined in this study, it seems plausible that there were also differences in the extent of utilization of these polyphenols in the preterm group. Hydrocinnamic acid was observed to a lesser extent in the preterm group. The lower content of reducing sugars (fructose, glucose, and galactose) in the preterm group corresponded with a greater capacity to form short-chain fatty acids (Fukiya et al., 2009; Wang et al., 2011b; Koeth et al., 2013; Tang et al., 2013; Ussher et al., 2013; Staley et al., 2017; Heinken et al., 2019; Wirbel et al., 2019).

In addition, de novo MAGs were assembled from the same sequences using our MAGO tool (see above). No significant differences were found at the level of MAGs, which corresponds to the same result at the level of taxonomic data obtained with Metaphlan. This is consistent with our observation that there are no significant taxonomic differences between the microbiota of the preterm and the control groups. By introducing a controlled diet, a controlled water intake, and a controlled circadian rhythm as previously described (Sket et al., 2017a; Sket et al., 2017b; Sket et al., 2018; Šket et al., 2020), it should be possible in future experiments with sufficient sample size to first establish the existence of significant differences in the microbiome (or the lack thereof) and then focus on the assembly of MAGs. The metabolic responses and predicted metabolites indicated that the microbiome of preterm group has greater metabolic flux compared with the full-term group, suggesting the existence of minor, yet unmeasured, but apparently significant environmental differences in the preterm gut relative to controls.

With the results described above (Figure 17), we can confirm two alternative hypotheses from section 1.4.1 and table 2. The first confirmed hypothesis states that there are significant differences between the preterm and full-term groups of participants in faecal and urine metabolomes that can be linked

to their physical performance in experiments and physiological data at exercise and rest. The second hypothesis states that there are significant differences at the level of metagenomics makeup of both groups, giving rise to identification of specific metabolic pathways differing between two groups and their gut environment characteristics. In the frame of taxonomic descriptions, we could not reject the null hypothesis of no difference between the groups as no significant differences were observed.

		Control	Preterm	Observed changes
Physiology	**Exercise/Rest**	●	▲	Pt significantly different from T and different responses to rest and some to exercise
NMR	**Urine**	●	▲	Pt significantly different from T at the level of urine metabolomes
	Feces	●	▲	Pt significantly different from T at the level of fecal metabolomes
Microbiome	**Taxonomy**	●	▲	No characteristic change in microbiome taxonomy
	Functional genes	●	▲	No characteristic change in microbiome functional genes
	Enzymatic reactions	●	▲	Distinct use of enzymatic reactions
	Metabolic pathways	●	▲	Distinct use of metabolic pathways
	Predicted metabolites	●	▲	Distinct use of predicted metabolites
	MAGs	●	▲	No characteristic change in de novo assembled metagenomes

Figure 17:   A summary of observed changes in PreTerm study.
A summary of observed changes at various information levels showing that significant differences exist between the preterm and full-term adult urine metabolomes, faecal metabolomes, and microbial metabolic reactions and pathways. Taken together, these results show that host and its microbiome behave measurably different in healthy physically fit young males in comparison to matched full-term controls.

Slika 17:   Povzetek opaženih razlik v študiji PreTerm.
Povzetek opaženih razlik na različnih nivojih informacij, ki kažejo na signifikantne razlike med predčasno in pravočasno rojenimi odraslimi na podlagi metabolomov urina in fekalnih vzorcev ter mikrobnih metabolnih reakcij in poti. Če povzamemo, ti rezultati nakazujejo, da se gostitelj in mikrobiom različno odzivata med predčasno in pravočasno rojenimi odraslimi.

3.1.7 Data integration

We summarized more than 1200 collected samples in the creation of the Slovenian urine ^{1}H-NMR database. Metabolomics data from all projects (PlanHab, spinal muscular atrophy, X-Adapt, PreTerm, healthy women and men) were integrated. All measured spectra were analysed with the same procedure to obtain the same metabolites in all projects. We showed that at this level of physiological

data characteristics distinction is possible between the different activity levels based on the metabolites in urine. All samples were processed in the same way and can be reprocessed again utilizing future database updates using our in-house processing pipeline (Sket et al., 2017a; Sket et al., 2017b; Sket et al., 2018; Šket et al., 2020; Deutsch et al. 2021a; Deutsch et al., 2021b; Deutsch et al., 2022a; Deutsch et al., 2022b) alongside commercially available software for targeted ^{1}H-NMR spectral deconvolution. For instance, the same spectra can be rerun with future database updates of the Human Metabolome Database (HMDB) as it grew from a few thousand metabolites in the first edition (Wishart et al., 2007) to 217,000 metabolites in the latest edition in 2021 (Wishart et al., 2021). The standardized analytical protocols established in our laboratory enabled minimizing systematic errors that usually occur due to batch effects or contributions by various NMR experts. The Box-Cox normalization and the sPLSDA approach utilized to integrate all metabolomes in our study showed competitive performance in removing batch effects, but still preserved variations due to lifestyle or other biological reasons (Wang and Lê Cao, 2020). This approach also allowed us to partly confirm the alternative hypothesis from section 1.4.3 and table 2 confirming significant differences in urinary metabolomes that allow the identification of biomarker pools and metabolic pathways that delineate different groups under study. The identification of biomarker pools should be confirmed on larger dataset.

We showed that urinary metabolic fingerprinting has the potential provide a snapshot of metabolic status relevant and related to health and activity status (Azad and Shulaev, 2019; Mussap et al., 2021). In general, metabolomics involves the systematic identification of metabolites in the human body (Ashrafian et al., 2021). The development of a national database should improve the understanding of the Slovenian metabolome in comparison to studies from other European countries and the identification of metabolites specific to various diseases or physical conditions. With an enlarged database, we avoid problems with small sample sizes as observed in individual studies described above. We would need cohorts at least two orders of magnitude larger to confirm the final results of these studies. ^{1}H-NMR metabolomics has the potential to capture a wide range of conventional clinical variables in epidemiological studies, including missing variables for patient metadata, and makes it possible to generate predictors of discrimination between different diseases based on machine learning. Top-down interpretation of metabolomic datasets, particularly urine that can be collected noninvasively, can provide sufficient data to draw conclusions about how samples should be classified into different groups. We hope to generate interest from other researchers to incorporate NMR metabolomics into their research to expand our established database to approximately 10,000 samples on a national scale. The modelling of such data collection represents unique avenue to create ML models that can be used in medical practice at least tentatively to distinguish between healthy and unhealthy metabolic states next to between different diseases. Thus, this approach represents a step closer to data-driven precision medicine that has the potential to inform health on a national scale. The publication of the Slovenian urine NMR database is in preparation.

3.1.8 What about the future?

The beauty of 'omics research is that it generates thousands of different variables in different data matrices. All of the datasets obtained in this work were analysed in different ways, but there is always room for improvement of the use of different methods on the wet lab side or on the computational side. For instance, efforts are being directed towards inclusion of ongoing projects focusing on urine metabolomics as part of Slovenian urine ^{1}H-NMR database. In line with these, (i) a total of 320 samples from the PreAlti project (extension of the PreTerm project) were collected and measured, (ii) extension of SMA is currently in the phase of ongoing sample collection, (iii) samples are also being collected from two clinical cohorts from the University Clinical Centre of Ljubljana including the Children's Hospital (tics, anorexia), while (iv) clinical cohorts associated with Million Microbiomes from Humans Project are aiming at collecting more than 1000 faecal and urine samples for metagenomics and metabolomics analyses. All these projects are on the way to generate thousands of gigabytes of molecular data accompanied by participants metadata in accordance with GDPR and ethical considerations as governed by the Ethics Commission of the Republic of Slovenia in ongoing efforts to improve the understanding of the Slovenian microbiome, metabolome, and physiology by creating better and more appropriate models and networks that will be characteristic of different diseases and/or physical conditions. Maintaining systemic homeostasis and responding to nutritional and environmental challenges requires the coordination of a variety of organs and tissues. To respond to diverse metabolic demands, the human body integrates a system of interorgan communication through which one tissue can influence metabolic pathways in a distant tissue. Dysregulation of these communication pathways through lack of exercise (sedentary lifestyle) and high-energy diets contributes to diseases such as obesity, diabetes, liver disease, and atherosclerosis. For timely interventions, we should think about using body fluids (such as urine) that allow for non-invasive sampling but are sensitive enough to differentiate between a range of biomarkers (Figure 18).

The ability to effectively conduct quality control of incoming datasets, the pre-processing of sequencing or metabolomics raw data files to organized data matrices, the pre-processing of missing values, standardization and normalization procedures, in addition to the batch corrections established in this study coupled with data integration approaches enable the syncing of metagenomics, metabolomics and metadata for the same participants in the future, integrating the information about different states in the complexity of human body. This enables a better understanding of inter-organ communication, which acts as a gatekeeper for metabolic health, as multidirectional interactions between metabolic organs and the central nervous system mediate crosstalk between the gut, brain, and other peripheral metabolic organs to maintain energy homeostasis. This enables the search for new therapeutic strategies and promotes a healthy lifestyle to counteract metabolic disorders and other diseases.
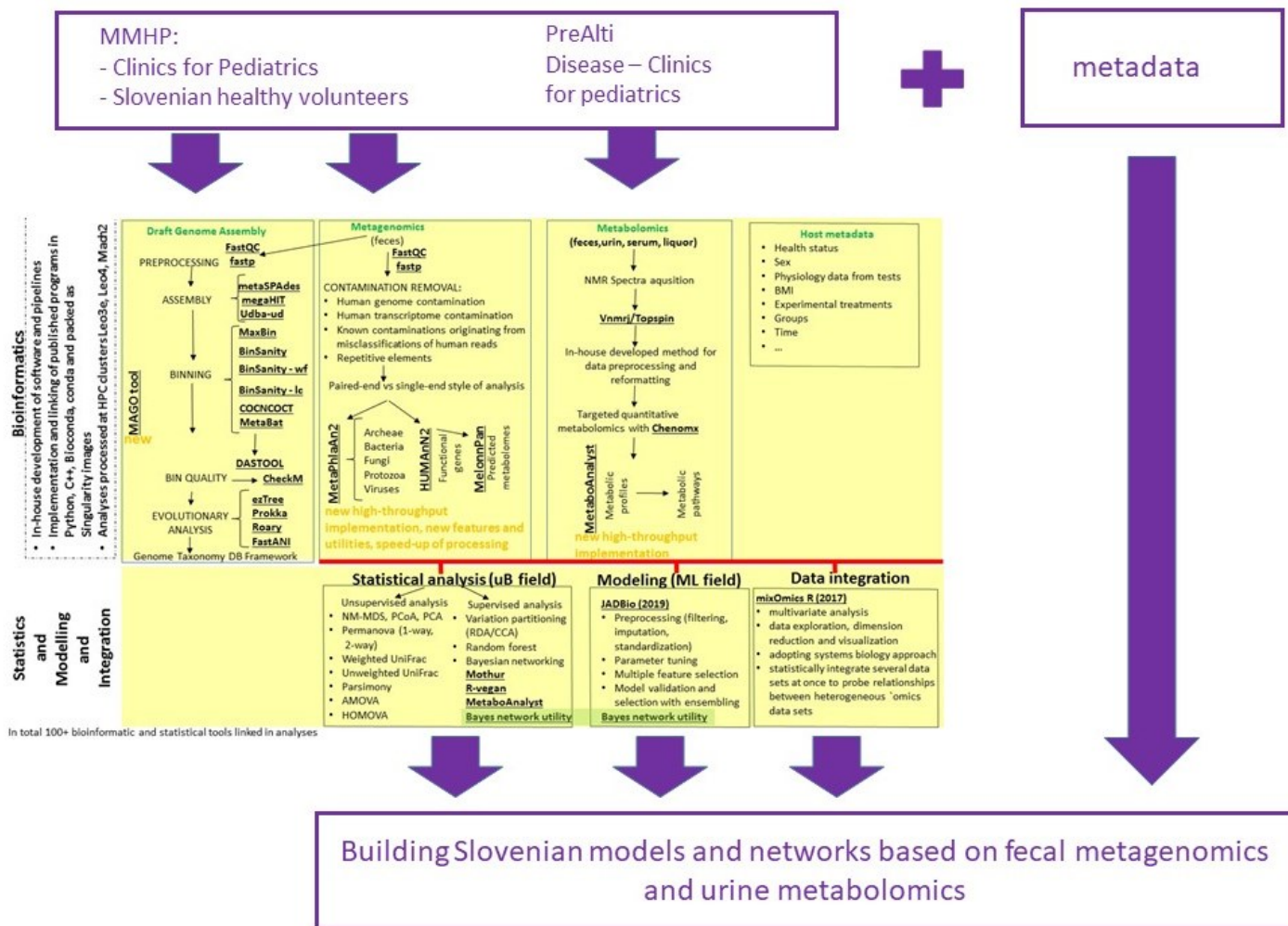
Figure 18:   The continuation of the projects, described in this work.

Slika 18:   Nadaljevanje projektov, opisanih v tem delu.

3.9 CONCLUSIONS

- The GUMPP and MAGO tools were developed, and the metaBakery tool is under development for high-throughput amplicon and shotgun sequencing analysis. All tools are available as Singularity image containers prepared for deployment on the HPC cluster and can be further developed by all users. All tools were developed under the open-source license CC-BY 4.0.

- By utilizing the newly developed MP approach on faecal samples, we showed that measuring some physicochemical parameters and using 'omics methods can lead to a completely new understanding of complex biological systems. The MP approach was a less-biased and fine-scale approach to measure faecal hardness compared to the previously used BSS approach, including the interior of samples.

- Participants of the bed rest group (NBR and HBR) from the PlanHab study had a specific metabolic composition compared with the HAmb group. We concluded that the host decision to minimize physical activity under hypoxic conditions can be detected within a few days at the level of the urine metabolome measured by NMR.

- When urine, serum, and liquor samples from SMA patients were compared before and after the 4th application of drug nusinersen, no differences were observed. However, urine creatinine was observed as a possible biomarker to distinguish healthy individuals from SMA patients.

- SMA study allowed us to observe some differences between healthy male and female urine metabolomes, which shows the importance of including women in biomedical and physiological studies.

- Urinary metabolomes of untrained individuals differed from metabolomes collected from trained participants in the X-Adapt study. After 10 days of training, these differences disappeared, demonstrating the importance of physical activity for humans.

- It was shown that consecutive 3-day urine collection can enable better understanding of morning metabolomes representing a systemic description of the state of human body.

- Urinary and faecal metabolomes of preterm and full-term born individuals of the PreTerm study were different. Microbial functionality observed on shotgun sequencing of stool samples was also different in the two groups. However, no significant taxonomic differences could be observed due to unequal variance at this information level.

- The integration of urine metabolomes from five different projects enabled creating a Slovenian NMR database that has the potential for the future to include more samples from different specimens and to create classification models to discriminate between different diseases or activity levels.

4 SUMMARY (POVZETEK)

4.1 SUMMARY

Modern human are increasingly threatened by the daily sedentary lifestyle and by serious diseases. The effects of short-term inactivity lead to maladaptations in body physiology, gut microbiota, and metabolic profiles, resulting in increased inflammation, depression, insulin resistance similar to metabolic syndrome, and type 2 diabetes symptoms. However, the effects of long-term physical inactivity, lack of oxygenation, and large muscle signalling are not well understood, although they have direct and widespread biomedical significance for preterm birth and/or genetic disorders, such as SMA, obesity, cardiovascular deconditioning, and chronic obstructive pulmonary disease. To address these issues, three projects analysed a variety of samples: i) physiological responses in adulthood as a consequence of preterm birth (PreTerm project; ARRS J3-7536; EU project https://recap-preterm.eu/); ii) spinal muscular atrophy (project within the University Clinical Centre of Ljubljana) as an extreme case of physical inactivity; and iii) cross-adaptation between heat and hypoxia: a novel strategy for performance and work-ability enhancement in various environments (X-Adapt; research project ARRS J5-9350). The SMA and PreTerm projects addressed lifelong exposure to systemic effects of reduced physical activity: i) intermittent episodes of systemic hypoxia at rest/sleep (PreTerm) and ii) continuous systemic hypoxia due to reduced host physical activity and relief of hypoxia after therapy. The X-Adapt project addressed the impact of regular 10-day training on the physiology of healthy trained and untrained individuals. In addition, little is known about the existence of differences in the human-gut microbiome relationship due to lifelong exposure to hypoxic episodes in preterm versus full-term born adolescents (The PreTerm project), which could impact the functionalities and metabolism of the microbiome in these hosts.

For a better understanding, especially of the microbiome, the appropriate tools for high-throughput big data analysis were developed on our side. The GUMPP workflow was developed for amplicon sequencing at three different levels (i) genus, (ii) OTU, or (iii) ASV. The GUMPP workflow consists of the most commonly cited tools for amplicon sequence analysis (Mothur) and microbial functionality prediction (PICRUSt2 and piphilin). The metaBakery workflow is prepared for shotgun sequence analysis and also consists of BioBakery tools (MetaPhlAn (taxonomic analysis), HUMANn3 (analysis of functional genes, enzymatic reactions, and metabolic pathways) and MelonnPan (prediction of microbial metabolites). The manuscript of the metaBakery tool is currently in the preparation phase. The third tool developed is a MAGO tool that uses the most advanced methods for microbiome analysis and consists of the main quality control tools (FastQC, fastp), assemblers (IDBA-UD, metaSPAdes, megahit) and binners (maxBin, MetaBAT, CONCOT, BinSanity and DAStool). CheckM tools were integrated throughout the pipeline to select assembled MAGs based on completeness and contamination according to the MIMAG standard. All tools were prepared as a skeleton framework consisting of 10,000 lines of code written in Python and packaged as a singularity image ready for use on HPC clusters. All tools were developed under the CC-BY 4.0 license and are released for development by other researchers.

In the context of the microbial world in the human gut, physicochemical parameters are important for both microbial homeostasis and human homeostasis. The BSS was previously used to assess gut health based on a visual assessment. One problem with this assessment was personal bias. In this work, we developed a new method for high-throughput assessment of faecal consistency, which we called minimal pressure (MP), which is expressed as the force per unit area required to cause permanent deformation of faeces. MP showed correlation with BSS, but provides the true assessment on a continuous scale. The correlation between MP and faecal methionine and acetate showed with different MP values. Both metabolites were previously associated with Western diet and inactivity, such as the sedentary lifestyle. With MP, a new approach for measuring physicochemical parameters was introduced, which, together with the 'omics method, provides another level of understanding of the microbial world in the human gut.

The PlanHab study was the first study by our group to investigate the problems of inactivity and hypoxia from the perspective of ^{1}H-NMR metabolomics. It was a crossover study with three different 21-day experiments (i) hypoxic bed rest, (ii) normoxic bed rest, and (iii) hypoxic ambulation. In both bed rest studies, detectable metabolic changes were observed based on morning urine. The identified metabolites were previously associated with various chronic diseases (chronic obstructive pulmonary disease, cardiovascular disease, etc.). Overall, inactivity alone or in combination with hypoxia resulted in decreased systemic metabolic diversity, increased the number of metabolic pathways affected, and accelerated metabolic deconditioning, leading to the development of negative physiological symptoms associated with these chronic diseases.

The results of the PlanHab project allowed us to join the spinal muscular atrophy project. In this project, we were able to analyse the metabolomes of atrophic patients in three different samples (serum, liquor, and urine) before treatment and after the 4th application of nusinersen, the first treatment approved by the EMA and FDA for the treatment of SMA. We found no significant differences between metabolomes. In parallel, we also collected urine samples from healthy Slovenian patients who matched the SMA patients in age and sex. Using machine-learning methods, we were able to determine urine creatinine to be a potential biomarker for the diagnosis of SMA.

The SMA project studied complete disease-related inactivity. The X-Adapt project allowed us to understand the impact of a 10-day exercise regimen on the metabolome of trained and untrained participants in the study. It was showed before that minimal activity can reduce the likelihood of metabolic syndrome due to a sedentary lifestyle. Participants were tested before and after the 10 days of training. Urine samples were collected at four different time points. Urine samples were collected over three days to reduce day-to-day variation. Briefly, some metabolites were found to be important in discriminating between trained and untrained subjects, but the significant differences disappeared after 10 days of training when trained and untrained subjects became more metabolically synchronised. In general, we showed that there is little difference between the two groups and that a lifelong active lifestyle is necessary to maintain a healthy metabolome.

In the PreTerm project, adult preterm and full-term adults participated to observe differences in metabolome (faecal and urine) and microbial metagenome when exposed to hypoxic or normoxic conditions (cycling on an ergometer). No significant differences were observed based on 25 measured physicochemical parameters in faeces (including the MP approach). In addition, some metabolic differences were observed in faecal and urine samples, some of which were previously associated with the development of noncommunicable diseases, particularly in preterm born adults. In addition, shotgun sequencing of the faecal samples was performed. We demonstrated that the taxonomic composition of preterm and full-term groups was the same, based on analysis of sequences and de novo MAGs, but microbial functions were different, once again demonstrating the importance of studying microbial functionality. Metabolic responses and predicted metabolites indicated that the microbiome of the preterm group had greater metabolic flux than that of the full-term group, suggesting minor, previously unmeasured, but apparently significant environmental differences in the preterm gut compared with controls.

The final step was completed with data integration. More than 1200 metabolomes from all projects (PlanHab, X-Adapt, SMA, PreTerm and healthy comparison group) were integrated with the miXomics package. We have shown that there is a possibility that we can use urine NMR metabolomes to differentiate between different groups (diseased vs healthy, active vs inactive) in the future. Top-down interpretation of metabolomic datasets, especially urine that can be collected noninvasively, may provide sufficient data to draw conclusions about how samples should be classified into different groups. We hope to stimulate the interest of other researchers to incorporate NMR metabolomics into their research in order to expand our established database to approximately 10,000 samples on a national scale. The manuscript of the Slovenian NMR database is currently under preparation.

In addition, the expansion of our NMR database continues: 320 samples from the Prealti project (continuation of the PreTerm project) were already collected and measured, the SMA project was extended and sample collection continues, two additional clinical cohorts are being collected (tics, anorexia), and more than 1000 faecal and urine samples will be collected as part of the Million Microbiomes from Humans project. All of these projects are on track to generate thousands of gigabytes of molecular data accompanied by participant metadata. This is being done in compliance with the General Data Protection Regulation (GDPR) and ethical considerations as defined by the Ethics Committee of the Republic of Slovenia to improve the understanding of the Slovenian microbiome, metabolome, and human physiological states. To respond to diverse metabolic demands, the human body integrates a system of interorgan communication through which one tissue can influence metabolic pathways in a distant tissue. Dysregulation of these communication pathways through lack of exercise (sedentary lifestyle) and high-energy diets contributes to human diseases such as obesity, diabetes, liver disease, and atherosclerosis. For timely interventions, body fluids (such as urine) represent logical choice and allow for non-invasive sampling but are sensitive enough to differentiate between a range of biomarkers.

167

4.2 POVZETEK

V človeškem prebavnem traktu živi 10^{13} mikrobni celic, ki proizvajajo, spreminjajo in porabljajo na tisoče kemijskih spojin, ki vplivajo na mikrobno sestavo in zdravje ljudi. Tehnologije sekvenciranja (metagenomika) in druge 'omske metode (metabolomika, proteomika, lipomika) so nam poleg sodobnih biostatističnih in strojnih metod učenja omogočile globlje razumevanje in nova spoznanja o kompleksnosti in vzročnosti med mikrobioto in njenim gostiteljem pri raziskavah bolnih in zdravih kohort preiskovancev skozi čas. Pomembna povezava med tema dvema skupinama je stanje metabolnega okolja, ki odraža medsebojni vpliv fiziologije gostitelja in mikrobioma (Schmidt, 2021).

V prejšnjih študijah smo v okviru projekta PlanHab raziskovali posledice zmanjšane fizične aktivnosti in zmanjšane vadbe pri gostitelju (človeku) (Debevec in sod., 2014; Sket in sod., 2017a; Sket in sod., 2017b; Sket, 2018; Sket in sod., 2018). Posledice kratkotrajne neaktivnosti so povzročile nepravilnosti v telesni fiziologiji, črevesni mikrobioti in metabolomskih profilih, kar je povzročilo povečano sistemsko vnetje, depresijo, inzulinsko rezistenco, pojave, ki so podobni začetkom pri metabolnem sindromu in diabetesu tipa 2. Po drugi strani pa učinki dolgotrajne telesne neaktivnosti, pomanjkanja kisika in signalov velikih mišic v primeru posledic prezgodnjega poroda in/ali genetskih motenj, kot so spinalna mišična atrofija (SMA), debelost, srčno popuščanje in kronična obstruktivna pljučna bolezen, kljub neposrednemu in velikemu biomedicinskem pomenu niso dobro razumljeni.

Da bi raziskali ta problem, smo zbrali raznovrstno paleto vzorcev v okviru treh kontroliranih in natančno vodenih projektov: i) fiziološki odzivi v odraslosti kot posledica prezgodnjih porodov (projekt PreTerm; ARRS J3-7536; projekt EU https: //recap-preterm.eu/); ii) spinalna mišična atrofija (SMA KCLJ) in iii) navzkrižna adaptacija na vročino in hipoksijo – nova strategija za pripravljenost in povečanje netreniranosti v različnih okoljih (X-Adapt; projekt ARRS projekt J5-9350). Vsi projekti obravnavajo vseživljenjsko izpostavljenost sistemskim učinkom zmanjšane telesne aktivnosti: i) prekinjajoče epizode sistemske hipoksije v mirovanju / spanju (PreTerm), ii) kontinuirano sistemsko hipoksijo zaradi zmanjšane telesne aktivnosti gostitelja zaradi genetskega defekta in lajšanje hipoksije po genetski terapiji, ali iii) primerjavo treniranih in netreniranih zdravih, mladih moških. Opravili smo biokemijsko karakterizacijo telesnih tekočin, zbranih v okviru vseh projektov in jih uporabili za raziskovanje biokemijske sestave (metaboliti) in njihovih interakcij (metabolne poti).

Mikrobne vrste igrajo pomembno vlogo v raznolikih okoljih, za katera je značilen širok spekter kompleksnosti organizmov (Murovec in sod., 2019). Mikrobi, ki živijo v črevesju, so v stalni interakciji z gostiteljem in večsmerni interakciji s svojimi mikrobnimi sorodniki s proizvodnjo različnih molekul, ki lahko izboljšajo zdravstveno stanje gostitelja ali po drugi strani vodijo v razvoj nenalezljive (kronične) bolezni ali njeno napredovanje (Murovec in sod., 2020). Napredovanje bolezni se lahko kaže kot blagi gastrointestinalni simptomi na eni strani ali resne bolezni, kot so vnetna črevesna bolezen, rak debelega črevesa ali rak jeter na drugi strani. Številne bolezni so bile povezane s presnovnimi neravnovesji, ki so delno ali v celoti povezana s črevesnim mikrobiomom (od metabolnega sindroma in debelosti do avtoimunskih bolezni, okužb in duševnih motenj (Murovec

in sod., 2021)). Po razvoju in izboljšavah tehnologij sekvenciranja je hitro postalo jasno, da večine mikrobov (npr. 99 %) ni mogoče gojiti v laboratorijskem okolju, da ločevanje sevov na podlagi biološko relevantnih lastnosti ni enostavno izvedljivo, vseeno pa lahko na podlagi njihovega genetskega materiala vidimo, kateri mikrobi so prisotni v vzorcu. Na podlagi njihovega genetskega materiala lahko tudi sklepamo o mikrobni funkcionalnosti vzorca (kaj lahko ti mikrobi naredijo (funkcionalni geni, encimske reakcije, ali metabolne poti (Murovec in sod., 2020))).

Za zanesljivo in ponovljivo analizo obsežnih mikrobnih podatkov smo razvili tri orodja. Prvo je orodje General Unified Microbiome Profiling Pipeline (GUMPP), ki je namenjeno obsežni, poenostavljeni in ponovljivi analizi bakterijskih amplikonskih podatkov (na nivoju rodu, operacijskih taksonomskih enot in razlik v sekvenčni variantah) in napovedovanje njihovega funkcionalnega potenciala (Murovec in sod., 2021), ki ga sestavljajo Mothur (Schloss in sod., 2009), PICRUSt2 (Douglas in sod., 2020) in piphillin (Narayan in sod., 2020).

Sekvenciranje celotnega zaporedja genomov vključuje netarčno sekvenciranje naključne podmnožice vseh zaporedij do določene globine sekvenciranja, ne kot pri tarčnem (amplikonskem) sekvenciranju, kjer je posekvenciran le majhen del specifičnega gena. BioBakery je orodje za analizo zaporedja celotnega metagenoma, ki združuje različna orodja za analizo kakovosti, taksonomsko analizo (MetaPhlAn), funkcionalne gene, encimske reakcije in presnovne poti, ki so prisotne v mikrobni združbi (HUMAn3 (Beghini et al., 2021)). Poleg tega omogoča napovedovanje mikrobnih metabolitov samo na podlagi metagenomskih informacij in s tem vpogled v potencialno sestavo mikrobnih metabolomov, ki bi lahko bili prisotni v tej združbi (MelonnPan (Mallick in sod., 2019)). Pozitiven vidik sekvenciranja celotnega genoma je tudi ta, da lahko pridobimo informacije o genskem materialu iz različnih taksonomskih skupin (arheje, bakterije, protozoji, glive, virusi, tudi človeška DNA), kar lahko izboljša razumevanje kompleksnosti in interakcij med različnimi taksonomskimi nivoji. To nas pripelje do drugega orodja, ki je bilo razvito iz naše strani (metaBakery - v pripravi), ki je reimplementacija orodja BioBakery, z dodatkom, ki omogočajo kvalitativne analize in razširjeno z algoritmi za izračun mikrobne pestrosti.

Naslednji korak pri analizi celotnega metagenoma je možnost de-novo sestavljanja metagenoma (MAG). To je postopek, pri katerem se sekvenčni odčitki pregledajo glede kakovosti, sestavijo in združijo skupaj, da dobimo sestavljene metagenome. To je proces, ki lahko vodi do odkritja popolnoma novih mikrobnih vrst, saj 99 % mikrobnih vrst ne moremo gojiti v laboratorijskih pogojih. Za namene obsežnih, poenostavljenih in ponovljivih analiz smo razvili orodje MAGO (Murovec in sod., 2020). To sestoji iz zelo uspešnih orodij za analizo kakovosti (FastQC, fastp (Chen in sod., 2018)), orodij za sestavljanje (IDBA-UD (Peng in sod., 2012), metaSPAdes (Nurk in sod., 2017) in megaHIT (Li in sod., 2015) in združevanje (maxBin (Wu in sod., 2016), MetaBAT (Kang in sod., 2015), CONCOT (Alneberg in sod., 2016), BinSanity (Graham in sod., 2017) in DAStool (Sieber in sod., 2018)). V nadaljevanju se uporablja orodje CheckM (Parks in sod., 2015) za filtriranje, kateri MAG so visokokakovostni v skladu s standardi MIMAG (Bowers in sod., 2019) (glede na popolnost

in kontaminacijo). Orodje MAGO uporabniku omogoča tudi evolucijsko analizo z orodji ezTree (Wu, 2018), Prokka (Seeman, 2014), Roary (Page in sod., 2015) in FastANI (Jain in sod., 2018).

Vsa razvita orodja so bila pripravljena v programskem jeziku Python in sestavljena iz več kot 10.000 vrstic kode. Parametri za izvedbo poteka analiz so v celoti v rokah uporabnika. Vsa orodja so bila razvita kot slike Singularity (Kurtzer in sod., 2017), pripravljene za preprosto uporabo na visoko zmogljivih računalniških grozdih (HPC) za obsežne in preproste analize 10.000 vzorcev na eni strani in za izobraževalne namene na drugi strani (Murovec in sod., 2020; Murovec in sod., 2021). Obe orodji sta pod odprtokodno licenco CC-BY 4.0 in sta odprti za vse razširitve, s čimer nudita priložnost za nadaljnji razvoj in postaneta standardizirani za mikrobno analizo v svetovnem merilu.

Peristaltični valovi, ki ustvarjajo kontraktilne vzorce tankega črevesa, in s tem ustvarjajo nenehno se spreminjajoče se okolje. Konstantno mešanje fekalnega materiala povzroča prostorske in kemijske spremembe okoljskih pogojev skozi čas za mikrobe, ki živijo v črevesju, kar lahko vpliva na njihovo aktivnost, ekspresijo genov, rast in številčnost posameznih skupin mikrobov (Ehrlein and Schemann, 2005; Johnson in sod., 2012; Cremer in sod., 2016; Glover in sod., 2016; Cremer in sod., 2017; Sket in sod., 2017a). Številne študije v preteklosti so povezale konsistenco blata z bogastvom črevesne mikrobiote, njeno sestavo, enterotipi, povišanimi nivoji vnetja, lipopolisaharidi in hitrostjo rasti bakterij (Tigchelaar in sod., 2016; Vandeputte in sod., 2016). Konsistenca blata je bila v preteklosti ocenjena z bristolsko lestvico (ang. Bristol Stool Scale (BSS)) (Heaton in sod., 1992; Lewis and Heaton, 1997). Ena od pomanjkljivosti metode BSS je, da prihaja do visokega odstopanja med ocenjevalci zaradi pristranskosti in vizualne ocene (Derrien in sod., 2010; Chumpitazi in sod., 2016). Zato lahko le dobro usposobljen strokovnjak pripravi medicinsko pomembne zaključke na podlagi ocene BSS (Matsuda in sod., 2021).

Fekalni materiali so po fiziki materialov poltrdni materiali (tj. paste) (Grillet in sod., 2012), ki jih umeščamo med viskoelastične materiale (poltrajna deformacija kot odziv na zunanje sile) na eni strani in plastične materiale (trajna deformacija) na drugi strani. Ta način razmišljanja nas je pripeljal do vrednotenja s pomočjo minimalnega tlaka (MP) kot metode za manj pristransko in visoko zmogljivo ocenjevanje konsistence fekalnega materiala (Deutsch in Stres, 2021). Minimalni tlak, izražen kot sila na enoto površine, je tlak, ki je potreben, da povzroči trajno deformacijo fekalnega materiala. Pokazali smo, da MP narašča eksponentno v primerjavi z linearno padajočimi vrednostmi BSS, ne glede na spol (Deutsch in Stres, 2021). Pokazali smo tudi, da obstaja nelinearna (asimptomatska) in kompleksna povezava med suho snovjo in MP. Vzdolžno kartiranje površinskega MP po celotni dolžini posameznega vzorca blata je pokazalo, da obstajajo različne drobnozrnate notranje, lokalne razlike. Poleg tega je kljub enotnemu točkovanju BSS pri nižjih vrednostih BSS naša analiza pokazala, da so bolj odpornim površinskim plastem blata sledile mehkejše notranje strukture, kar ima za posledico nižje vrednosti MP, povezane s približno zdravo konsistenco blata (Deutsch in Stres, 2021). Te lastnosti z uporabo BSS ne moremo ovrednotiti. Določili smo mejo, ki lahko razlikuje med zdravim stanjem (MP < 75) ali zaprtjem (MP > 75) (Blake in sod., 2016; Sket in sod., 2017b; Sket in sod., 2018). MP < 30 je ustrezalo vzorcem tekočega blata (driska). MP smo izmerili na vzorcih,

zbranih v okviru študij PlanHab (Sket in sod., 2017a; Sket in sod., 2017b; Sket in sod., 2018) in PreTerm (Deutsch in sod., 2022b). Predvsem prva je pokazala, da sta bili blokada por na fekalni površini in zadrževanje sluzi povezana s selektivnim pritiskom na mikrobiom črevesja, njegovo gensko ekspresijo in presnovno aktivnost, kar lahko vodi do lokalnega vnetja (Vandeputte in sod., 2016; Sket in sod., 2017a; Sket in sod., 2017b; Sket in sod., 2018; Aron-Wisnewsky in sod., 2019). S tem smo pokazali, da lahko pristop MP natančno opiše klinični pomen konsistence blata (Deutsch in Stres, 2021). Poleg tega pristop MP ne zahteva predhodne obdelave vzorcev in omogoča enostavno merjenje brez drage opreme, pa tudi ponovljivost teh meritev med različnimi vzorci (sveži proti zamrznjeni; moški proti ženski). Ugotovili smo tudi, da MP korelira s fekalnim metioninom in acetatom na podlagi meritev ^{1}H-NMR. Na podlagi teh dveh metabolitov lahko ločimo tri različne skupine fekalne konsistence (MP < 30, 30<MP<75, MP>75). Metionin je bil prej povezan z oksidativnim stresom in je bil povišan pri neaktivnih posameznikih, medtem ko je acetat negativno koreliral z občutljivostjo na inzulin (Martínez in sod., 2017; Müller in sod., 2019), kar kaže, da lahko različna konsistenca blata vpliva na biološki sistem gostitelja. Opažene razlike v metioninu in acetatu, povezanem z MP, so bile tako očitno posledica neaktivnosti v okviru projekta PlanHab v kombinaciji z zahodno prehrano. Pristop MP nam je omogočil, da smo z merjenjem nekaterih fizikalno-kemijskih parametrov na eni strani in z omskimi metodami na drugi strani lahko začeli in raziskali povsem novo raven razumevanja kompleksnih bioloških sistemov (Deutsch in Stres, 2021).

Študija PlanHab je bila prva študija naše skupine, ki je vključevala metabolomiko človeškega urina. Vzorce jutranjega urina smo zbirali skozi celoten eksperiment, ki je bil zamišljen kot navzkrižno oblikovan eksperiment (angl. Cross-over design). Vsi udeleženci študije so šli skozi vse tri oblike poskusa (21-dnevno ležanje v hipoksiji ali normoksiji ali pa gibanje v hipoksiji (Sket in sod., 2017a; Sket in sod., 2017b; Sket in sod., 2018, Šket in sod., 2020)). Edinstvena zasnova nam je omogočila, da smo upoštevali odzive istih udeležencev v vseh treh eksperimentalnih različicah pod nadzorovanimi prehranskimi, okoljskimi in eksperimentalnimi pogoji. Zbrali smo 523 vzorcev urina in jih pripravili za meritve ^{1}H-NMR. Udeleženci, ki so ležali (NBR in HBR), so imeli specifične metabolne značilnosti v primerjavi s skupino HAmb. Pokazalo se je, da je odločitev gostitelja, da zmanjša telesno aktivnost v hipoksičnih pogojih, mogoče zaznati v nekaj dneh na ravni urinskega ^{1}H-NMR metaboloma. V normoksičnih pogojih ležanja v postelji smo te metabolne spremembe zaznali šele v prvih desetih dneh. Metaboliti, opaženi v tej študiji, so bili povezani s številnimi različnimi boleznimi: (i) kronično obstruktivno pljučno boleznijo (Adamko, 2015; Zabek, 2015) in (ii) srčno-žilno boleznijo, povezano s tkivno hipoksijo, ki lahko vodi tudi do sladkorne bolezni tipa 2, depresijo in osteoporozo (Jones, 2014; Wang in sod., 2011; Senn in sod., 2012). Študija PlanHab z uporabo metabolomov ^{1}H-NMR v urinu nas je pripeljala do zaključka, da ni enostavnega metabolnega biomarkerja, ki bi lahko razlikoval med različnimi stanji (zdravo proti bolnemu, aktivno proti neaktivnemu; aktivno proti sedečemu). Za zajetje skupnih značilnosti človeške fiziologije, medosebne in časovne variabilnosti so bili potrebni kompleksni multivariatni opisi metaboloma. Ta koncept je bil uporabljen v vseh drugih nadaljnjih študijah. Na splošno je neaktivnost sama ali v kombinaciji s hipoksijo povzročila zmanjšano sistemsko metabolno raznolikost in povečano število prizadetih metabolnih poti, kar je povzročilo razvoj negativnih fizioloških simptomov, kot so

inzulinska rezistenca, nizka stopnja sistemskega vnetja, zaprtje, depresija in presnovni sindrom (Sket in sod., 2017a; Sket in sod., 2017b; Sket in sod., 2018, Šket in sod., 2020). Rezultati študije PlanHab so nas spodbudili, da nadaljujemo z raziskavami v drugih študijah, ki vključujejo različne stopnje neaktivnosti, kot so X-Adapt (razlike med treniranimi in netreniranimi posamezniki), spinalna mišična atrofija in projekt PreTerm, ki primerja različne čase izpostavljenosti hipoksiji, telesno aktivnost in čas izpostavljenosti različnim pogojem. Poleg tega smo dodatno zbrali več kot 200 vzorcev zdravih moških in žensk ter njihovih sinov in hčera (Schmidt, 2021).

Spinalna mišična atrofija je živčno-mišična bolezen, ki se kaže kot progresivna atrofija in oslabitev skeletnih mišic zaradi progresivne izgube motoričnih nevronov in prizadene številne druge organske sisteme (Melki, 2017; Yeo in Darras, 2020). Z incidenco 1 na 11.000 rojstev še vedno velja za najpogostejši genetski vzrok smrti otrok (Sugarman in sod., 2012). Pri bolnikih s SMA mutacije v centromernem genu SMN2 vodijo do tvorbe nestabilnih proteinov, hkrati pa je zaradi delecije motena tudi ekspresija telomernega gena SMN1 (Lefebvre in sod., 1995; Lorson in Androphy, 2000; Lunn in Wang, 2008; Smeriglio in sod., 2020). V zadnjih letih so se pojavile nove terapije za zdravljenje SMA. Te terapije spremenijo naravni potek bolezni s spremembo izražanja ali zamenjavo mutiranih genov, ki sodelujejo pri razvoju SMA (Chiriboga in sod., 2016). Nusinersen je bilo prvo zdravilo za zdravljenje SMA, ki sta ga odobrila Uprava za hrano in zdravila v Združenih državah Amerike in Evropska agencija za zdravila. Nusinersen je protismiselni oligonukleotid, ki vpliva na spajanje mRNA, kar ima za posledico aktiven protein SMN 2 in s tem boljše rezultate SMA (Chiriboga in sod., 2016; Corey, 2017; Ramdas in Servais, 2020). Nusinersen zahteva intratekalno aplikacijo, ker ne more prečkati krvno-možganske pregrade (Faber in sod., 2007; Rigo in sod., 2012).

Vzorci urina, likvorja in seruma bolnikov s SMA so bili zbrani pred zdravljenjem in po 4. aplikaciji zdravila nusinersen. Zdravniški pregled ob četrti aplikaciji zdravila je pokazal izboljšanje gibljivosti. Bolniki so pokazali izboljšanje nadzora nad invalidskim vozičkom, premikanja, utrujenosti, higiene, govora in spanja po 4. aplikaciji nusinersena (Deutsch in sod., 2021a, Osredkar in sod., 2021).

V nasprotju s fizičnimi pregledi, razlik nismo uspeli potrditi, na podlagi metabolomov urina, likvorja in seruma pred in po aplikaciji zdravila. V tem kontekstu ne moremo ovreči ničelne hipoteze iz poglavja 1.4.2, ki pravi, da ni bistvenih razlik pred in po zdravljenju. Morda bi te razlike lahko potrdili po 10 aplikacijah nusinersena, vendar bi to trajalo preveč dodatnega časa za zbiranje vzorcev in dokončanje v časovnem okviru tega doktorata. Ti rezultati kažejo, da je učinkovitost nusinersena mogoče ugotoviti z zdravniškimi pregledi in testi gibljivosti. Morda bi uporaba drugih metabolomskih metod, kot je masna spektrometrija, ki je bolj občutljiva (nM) v primerjavi z NMR (mM), privedla do odkrivanja biomarkerjev, ki bi jih lahko uporabili kot biomarkerje za spremljanje zdravljenja z nusinersenom. Lahko pa, da so signali iz izboljšanega metabolizma na račun večje fizične aktivnosti še premalo vidni in se pokažejo šele pri kasnejših aplikacijah (Deutsch in sod., 2020a).

Poleg vzorcev iz projekta SMA smo zbrali vzorce urina iz ujemajoče se zdrave kohorte, da bi primerjali metabolome bolnikov s SMA z metabolomi zdravih posameznikov. Ta primerjava je privedla do opazovanja pomembnih metabolnih razlik med ženskami in moškimi na eni strani (p=0,0001) in zdravo kohorto in bolniki s SMA na drugi. Vpliv spola in prisotnost bolezni je v obeh primerih bila statistično signifikantna. Obe metodi, PLSDA in Random Forest, sta pokazali pomembne razlike med ženskimi in moškimi metabolomi. Pri bolnikih s SMA smo opazili znatno zmanjšanje kumulativne koncentracije metabolitov (p < 0,05). Zmanjšanje števila metabolitov smo opazili tudi pri zdravih ženskah v primerjavi z zdravimi moškimi. Zaradi razlik med ženskami in moškimi je pomembno, da prihodnje študije vključijo večje število žensk v študije, kot je ta, da bi ugotovili pomembne razlike med ženskimi in moškimi metaboliti in njihovimi biokemijskimi potmi. Opazili smo nekaj vzporednic s predhodnimi študijami vadbe, ki kažejo, da se lahko število metabolitov poveča po vadbi (Nieman in sod., 2013; Schranner in sod., 2020) ali študijah ležanja v postelji (npr. PlanHab), ki so prav tako pokazale 30-odstotno zmanjšanje števila presnovkov po 3 tednih ležanja v postelji (Sket in sod., 2017a; Sket in sod., 2017b; Sket in sod., 2018). Simptomi, kot so inzulinska rezistenca, izguba kosti in mišic, spremembe v presnovi lipidov, so bili odkriti v študijah ležanja in vse te simptome je mogoče opaziti tudi na seznamu stanj, povezanih s SMA (Osredkar in sod., 2021).

Za namene sestavljanja klasifikacijskih modelov za razlikovanje med bolnimi in zdravimi, smo uporabili metabolome urina pri bolnikih s SMA in zdravih posameznikih. S pomočjo avtomatskega strojnega učenja smo kreirali model, ki uspešno loči med tema skupinama (AUC 0,958). Kreatinin je bil ključni metabolit, ki je ločil zdrave od pacientov s SMA, kot so poročali tudi nekaj mesecev pred našo objavo v drugi študiji, ki je spremljala napredovanje denervacije SMA s povišanimi ravnmi serumskega kreatinina pri hujših oblikah bolezni SMA (Alves in sod., 2020). Koncentracije kreatinina se pri bolnikih s SMA niso bistveno spremenile pred in po 4. aplikaciji nusinersena. Spremenjeno raven kreatinina so opazili tudi v vzorcih urina iz naših preteklih študij (PlanHab (Šket in sod., 2020)). Ponovna uvedba vadbe je v teh študijah popolnoma obrnila neželene učinke. Imobilizirani bolniki, ki so v preteklosti prejemali vibracijsko terapijo pri drugih boleznih, so imeli koristi v primerjavi s kontrolami in lahko predstavljajo potencialni korak pri fizični aktivaciji bolnikov s SMA po terapiji z nusinersenom (Deutsch in sod., 2021a).

V okviru projekta SMA smo raziskali stanje popolne neaktivnosti. Vendar pa je v 21. stoletju vse bolj jasno, da je telesna neaktivnost, ki je posledica sedečega načina življenja, tudi globalni problem, ki predstavlja tveganje za razvoj kroničnih nenalezljivih bolezni in povečano globalno smrtnost (Kelly in sod., 2020b). Pokazalo se je že, da lahko minimiziranje časa sedenja zmanjša tveganje za kronične bolezni, kot so koronarna bolezen srca, sladkorna bolezen tipa 2, metabolni sindrom itd. (Sallis in sod., 2016). Cilj projekta X-Adapt je bil preučiti razlike med fizično aktivnimi (treniranimi udeleženci) in neaktivnimi (netreniranimi) posamezniki (Sotiridis in sod., 2018; Sotiridis, 2019b; Sotiridis in sod., 2019; Sotiridis in sod., 2020). Projekt je vključeval 10 treniranih in 10 netreniranih moških v 10-dnevnem protokolu vadbe, ki je obsegal vsakodnevno vadbo na kolesarskem ergometru pri 50 % največje moči pedaliranja v normoksičnih in normobaričnih (~1000 hPA) pogojih pri 24°C.

Pred udeležbo in po 10 dneh vadbe so vsi udeleženci (aktivni in neaktivni) opravili tridnevno testiranje v termonevtralnih normoksičnih in hipoksičnih pogojih ter vročih normoksičnih pogojih. Udeleženci študije so bili razvrščeni kot trenirani ali netrenirani glede na njihovo maksimalno aerobno kapaciteto (netrenirani VO₂max < 45 mL-kg-1-min-1, trenirani VO₂max > 55 mL-kg-1-min-1) (Jay in sod., 2011;Montero in Lundby, 2017).

Meritve, usmerjene v človeško fiziologijo, so pokazale, da je obstajalo nekaj pomembnih razlik med treniranimi in netreniranimi preiskovanci. Razlike med stanjem pred in po treningu so bile večje v netreniranih skupinah. Na podlagi meritev VO₂max pred treningom in njegove spremembe v 10 dneh treninga je stopnja prilagajanja na trening največja pri netreniranih posameznikih (Sotiridis in sod., 2018; Sotiridis, 2019b; Sotiridis in sod., 2019; Sotiridis in sod., 2020).

Glede na metabolome urina ni bilo mogoče zaznati pomembnih razlik pred in po 10 dneh treninga. Vendar pa so bile opažene razlike pri primerjavi med urinskimi metabolomi med treniranimi in netreniranimi udeleženci. Poleg tega so se med tema dvema skupinama bistveno razlikovale tudi fizikalno-kemijske lastnosti urina (pH, skupne raztopljene trdne snovi, slanost in prevodnost). Na primer, pH se je znižal pri netreniranih posameznikih, kar je bilo prej povezano s presnovnim sindromom in kroničnim srčnim popuščanjem (Maalouf in sod., 2007; Otaki in sod., 2013; Kraut in Madias, 2016; Shimodaira in sod., 2017).

Metaboliti (holat, tartrat, kadaverin, lizin, N6-acetilizin, metanol, N-acetilglukozamin, butanon in kaprat) so bili identificirani s pomočjo multivariatne statistike in strojnega učenja kot metaboliti, ki so odgovorni za razlikovanje med trenirano in netrenirano skupino. Vse metabolite so predhodno opazili v študijah, povezanih s poškodbami mišic, ravnmi hormonskih receptorjev, okrevanjem po treningu z odpornostjo, nižjim kardiovaskularnim tveganjem (tartrat) (Abramowicz in Galloway, 2005; Spiering in sod., 2008) ali atrofičnim stanjem, debelostjo, razvojem raka, metabolnim sindromom (holat) (Li in sod., 2020; Abrigo in sod., 2021; Alamoudi in sod., 2021; Mercer in sod., 2021; Pushpass in sod., 2021; Zheng in sod., 2021).

S tem pristopom smo pokazali, da se celoten sistem pri aktivnih osebah bistveno razlikuje od tistega pri neaktivnih (p=0,003). Po 10 dneh treniranja so se celokupne razlike med treniranimi in netreniranimi zmanjšale (p=0,226). Naša študija je pokazala, da je vadba od 75-150 minut na teden, ki jih priporoča Svetovna zdravstvena organizacij, premalo učinkovita in da bi bila potrebna 5-krat večja vadba. Poleg tega je ta poskus pokazal, da 3-dnevni jutranji vzorci urina zagotavljajo dobro biološko matriko za razlikovanje aktivnih od neaktivnih posameznikov, ki jih ni mogoče opaziti pri dnevnem vzorčenju zaradi dnevnih variabilnosti posameznika. Sistemska homeostaza je odvisna od številnih različnih parametrov in vključuje komunikacijo med različnimi organi, prek katere lahko metabolne poti, na katere vplivajo metaboliti v enem organu, vplivajo na druge metabolne poti v drugem organu. Sedeči način življenja z odsotnostjo signalov velikih mišic in oksigenacije sistema ter porabe hranil lahko moti to komunikacijo med organi, kar vodi v manifestacijo različnih bolezni.

Višje ravni vadbe lahko obnovijo medorgansko komunikacijo do zdravih in fizično aktivnih posameznikov (Deutsch in sod., 2022a).

Prezgodnji porod je opredeljen kot rojstvo pred 37. tednom gestacije. Po vsem svetu je približno 10 % prezgodnjih porodov in je še vedno eden vodilnih vzrokov smrti pri otrocih, mlajših od 5 let. Prezgodnji porod povečuje tveganje za razvoj različnih kroničnih bolezni, kot so srčno-žilne, endokrine/metabolične, ledvične, nevrološke in psihiatrične motnje. Eden glavnih vzrokov za te motnje je povečan oksidativni stres v prvih tednih življenja (Moutquin, 2003; Magalhães in sod., 2004; Pialoux in sod., 2009; Blencowe in sod., 2012; Lushchak, 2014; Liu in sod., 2015; Manley in sod., 2015; Debevec in sod., 2017; Crump, 2020; Tingleff in sod., 2021). Obstaja velika verjetnost, da so nekateri klinični parametri, kot so telesna masa, arterijski krvni tlak, glukoza na tešče in holesterol, lahko povišani pri prezgodaj rojenih odraslih (Kerkhof in sod., 2012; Markopoulou in sod., 2019; Crump, 2020). Različne študije so pokazale, da se vse te značilnosti razlikujejo med prezgodaj in pravočasno rojenimi odraslimi in da so te razlike, zlasti povezane s proizvodnjo reaktivnih kisikovih vrst, in jih je mogoče opaziti v povezavi z različnimi stopnjami vadbe ali telesne dejavnosti (Magalhães in sod., 2004; Powers in sod., 2011; Filippone in sod., 2012; Debevec in sod., 2017; Martin in sod., 2018).

Namen projekta PreTerm je bil raziskati, ali obstajajo razlike med prezgodaj in pravočasno rojenimi mladimi moškimi v ventilacijskem odzivu (HVR) pri telesni aktivnosti ali mirovanju v hipoksičnih in normoksičnih okoljskih pogojih (Debevec in sod., 2019; Debevec in sod., 2022). Poleg tega je bil za opis kompleksnosti človeškega telesa in črevesnega mikrobioma pri njegovem odzivu na povečane ravni oksidativnega stresa v mirovanju in med vadbo pri normoksiji in hipoksiji uporabljen analitični pristop, ki je sestavljen iz metabolomike urina in fecesov ter fekalne metagenomike (Deutsch in sod., 2022b).

Pokazalo se je, da kolesarjenje pri normoksiji in hipoksiji zviša ravni oksidativnega stresa v obeh skupinah takoj po vadbi (Martin in sod., 2020). Udeleženci v skupini prezgodaj rojenih so pokazali nižjo vadbeno zmogljivost pri normoksiji v primerjavi s kontrolno skupino, in so imeli nižji HVR, medtem ko takšne razlike niso opazili pri hipoksiji (Vrijlandt in sod., 2006; Lovering in sod., 2013; Svedenkrans in sod., 2013; Bates in sod., 2014; Clemm in sod., 2014; Farrell in sod., 2015; Debevec in sod., 2019). Ti rezultati kažejo, da imajo lahko prezgodaj rojeni povečan oksidativni stres med akutno vadbo v normoksiji, medtem ko takšnega odziva pri hipoksiji niso opazili (Martin in sod., 2020).

V vzorcih blata smo izmerili 25 fizikalno-kemijskih spremenljivk (vključno z zgoraj opisanim pristopom MP), pri čemer med obema skupinama nismo ugotovili bistvenih razlik. Vzorci blata in urina so bili zbrani tri dni pred hipoksičnim in normoksičnim testom in tri dni po njem (Deutsch in sod., 2022b).

Aceton, tartrat in trans-akonitat so bili urinski metaboliti, ki so se glede na rezultate MetaboAnalyst zmanjšali v skupini prezgodaj rojenih in korelirajo z vadbo, postom ali diabetes melitusom (Paradis in sod., 2015; Crump in sod., 2019; Perrone in sod., 2021). Zdi se, da so razlike posledica oslabljene avtonomne funkcije, ker se srčni utrip pri prezgodaj rojenih odraslih obnavlja počasneje, kar bi lahko povzročilo anoksijo in povečalo srčno-žilno tveganje, kot je bilo že objavljeno (Qiu in sod., 2017; Haraldsdottir in sod., 2019).

Laktat, serotonin in tirozin so bili glavni fekalni metaboliti, ki so predstavljali razliko med prezgodaj in pravočasno rojeno skupino. Prva dva metabolita sta bila povečana v skupini prezgodaj rojenih, kar skupaj z obogateno metabolno potjo (Warburgov učinek) kaže, da lahko pri njih opazimo nekatere metabolne spremembe, ki ji lahko povezujemo z mitohondrijsko disfunkcijo (Sonntag in sod., 2007; Ten, 2017). Te ugotovitve lahko predstavljajo prvi dokaz, da sistemske razlike zaradi vseživljenjske izpostavljenosti oksidativnemu stresu res obstajajo in postavljajo vprašanje, ali so te razlike povezane z majhnimi razlikami, ki nastanejo na strani prezgodaj rojenega gostitelja ali na delu mikrobioma, ki se odziva zaradi teh okoljskih signalov drugače kot pri pravočasno rojenih (Deutsch in sod., 2022b).

Zbrani fekalni vzorci so bili uporabljeni za sekvenciranje, da bi raziskali, ali opažene razlike v fekalnih metabolomih korelirajo z razlikami na mikrobni ravni. Na taksonomski ravni nismo opazili bistvenih razlik, čeprav je bila relativna številčnost arhej in virusov višja v skupini prezgodaj rojenih. V zadnjem desetletju je postalo jasno, da je pri preučevanju mikrobioma pomembnejše vprašanje, kaj mikrobi v našem črevesju počnejo. Zato smo naredil analizo funkcionalnosti preučevanega mikrobioma z našim orodjem metaBakery. Strojno učenje je bilo uporabljeno za izdelavo klasifikacijskih modelov in identifikacijo potencialnih biomarkerjev (Deutsch in sod., 2022b).

Na podlagi genskih družin ni bilo odkritih bistvenih razlik, vendar smo zaznali nekaj razlik na podlagi encimskih reakcij, metabolnih poti in predvidenih metabolitov. Predhodno opisana encimska reakcija sukcinat dehidrogenaze (RXN-15378) je bila povečana v skupini prezgodaj rojenih. Sukcinat je sam po sebi mikrobni metabolit in se lahko kopiči v črevesnem traktu med vnetjem ali mikrobnim neravnovesjem. Ima tkivno specifične, a tudi protivnetne lastnosti in je tudi vir produkcije propionata s strani *Bacteroides* spp. in *Prevotella* sp. Pokazalo se je, da se sukcinat kopiči v celicah v pogojih z nizko vsebnostjo kisika in predstavlja metabolni podpis hipoksije. Pokazalo se je, da prekomerni privzem mikrobno proizvedenega sukcinata vodi do višjih ravni znotrajceličnega sukcinata, ki na koncu poveča odziv na samo hipoksijo in hkrati poveča LPS- inducirano ekspresijo proinflamatornih citokinov v človeških celicah (Rubic in sod., 2008; Ariza in sod., 2012; Tannahill in sod., 2013; Akram, 2014; Littlewood-Evans in sod., 2016; Connors in sod., 2018; Deutsch in sod., 2022b).

PWY-7456 (razgradnja β-(1,4)-manana), PWY-7323 (superpot biosinteze gradnikov O-antigena iz GDP-manoze) in GLYCOLY-SIS-TCA-GLYOX-BYPASS (superpot glikolize, piruvat dehidrogenaza, TCA in glioksilatni obvod), P221-PWY (oksidacija oktana), PWY-5173 (nerazvrščen) so bile poti, ki so bile povečane v skupini prezgodaj rojenih. Nekateri od njih so lahko koristne in si prizadevajo za celovitost sluznice in prehranjevanje gostitelja (razgradnja β-(1,4)-

manana) ali pa znatno povečajo proizvodnjo energije, kar bi bilo pomembno v primeru oksidativnega stresa kot pri prezgodaj rojenih posameznikih (superpot glikolize, piruvat dehidrogenaza, TCA in glioksilatni obvod). Biosinteza acetil-CoA lahko povzroči tudi povečano proizvodnjo butirata s proizvodnjo acetil-CoA. Po drugi strani pa imajo nekatere poti bolj negativen učinek in so bile povečane tudi v skupini prezgodaj rojenih. Izkazalo se je, da so te poti vključene v proizvodnjo lipopolisaharidov LPS (biosinteza gradnikov O-antigena iz GDP-manoze), povezane s po gramu negativnimi bakterijami in povzročitelji različnih stopenj vnetja (Samuel and Reeves, 2003; Wolfs in sod., 2010; Shah in sod., 2015; Kim in sod., 2016; La Rosa in sod., 2019; Lindstad in sod., 2021). Po drugi strani pa je bila opažena tudi oktanska oksidacija, ki je bila prej opisana v kontekstu zahodnjaškega načina prehranjevanja in povezana z boleznijo jeter. Vse te razlike je mogoče povezati s fiziološko pomembnimi primanjkljaji, opaženimi med obema skupinama (Martin in sod., 2018; Martin in sod., 2020; Schmidt, 2021).

S pristopom napovedovanja mikrobnih metabolitov je bilo odkritih tudi sedemnajst metabolitov, ki ločujejo med obema skupinama, vendar nobeden od njih ni bil zaznan v primeru fekalne metabolomike (Deutsch in sod., 2022b) s pomočjo strojnega učenja. Poleg metabolnih poti, ki izhajajo iz metagenomsko predvidenih metabolitov, so bile odkrite pomembne razlike v metabolitih v urinu in blatu, kar kaže, da obstajajo sistemske razlike med obema skupinama. Povišani metaboliti so bili prej povezani s srčno-žilnimi boleznimi (karnitin), povečano prepustnostjo črevesja, zvišanimi ravnmi vnetnih citokinov, metabolnim sindromom ali razvojem raka (putrescin in diacetilspermin). Po drugi strani so se nekateri predvideni presnovki zmanjšali v skupini prezgodaj rojenih. Deoksiholat je sekundarna žolčna kislina in znan promotor raka debelega črevesa. Zmanjšane ravni te molekule so na splošno opazili zaradi povečanega izločanja holata z urinom, opaženega pri metabolomiki urina. Nižja vsebnost redukcijskih sladkorjev (fruktoze, glukoze in galaktoze) v skupini prezgodaj rojenih je ustrezala večji sposobnosti tvorbe kratkoverižnih maščobnih kislin (Fukiya in sod., 2009; Wang in sod., 2011b; Koeth in sod., 2013; Tang in sod., 2013; Ussher in sod., 2013; Staley in sod., 2017; Heinken in sod., 2019; Wirbel in sod., 2019).

Na ravni na novo sestavljenih metagenomov nismo ugotovili razlik, kar sovpada rezultatom na ravni taksonomskih podatkov, pridobljenih s programom Metaphlan. To je skladno z našim opažanjem, da med mikrobioto prezgodaj in pravočasno rojenih ni pomembnih taksonomskih razlik (Deutsch in sod., 2022b).

Z zgoraj opisanimi rezultati lahko potrdimo dve alternativni hipotezi iz poglavja 1.4.1. Prva potrjena hipoteza navaja, da obstajajo pomembne razlike med prezgodaj rojenih in pravočasno rojenimi skupinami udeležencev v metabolitih fecesa in urina, ki jih je mogoče povezati z njihovo fizično zmogljivostjo v poskusih in fiziološkimi podatki med vadbo in mirovanjem. Druga hipoteza navaja, da obstajajo pomembne razlike na ravni metagenomske sestave obeh skupin, zaradi česar je mogoče identificirati specifične metabolne poti, ki se med skupinama razlikujejo, in značilnosti njihovega črevesnega okolja. Razlike med na novo sestavljenimi metagenomi med obema skupinama nismo

opazili, zato v tem primeru ne moremo ovreči ničelne hipoteze, ki pravi, da ni razlike med prezgodaj in pravočasno rojeno skupino.

Več kot 1200 zbranih vzorcev smo združili pri izdelavi slovenske baze podatkov ^{1}H-NMR urina. Vsi zbrani urinski vzorci iz 5 projektov (PlanHab, spinalna mišična atrofija, X-Adapt, PreTerm, zdrave ženske in moški) so bili integrirani. Vsi izmerjeni spektri so bili analizirani z enakim postopkom, da bi dobili enake metabolite v vseh skupinah. Pokazali smo, da je na tej ravni fizioloških podatkov mogoče razlikovati med različnimi stopnjami aktivnosti na podlagi metabolitov v urinu. Vsi vzorci so bili obdelani na enak način in jih je mogoče v prihodnosti ponovno obdelati z uporabo nadaljnjih posodobitev baze podatkov Human Metabolome Database (Wishart in sod. 2007; Wishart in sod, 2022) z uporabo naših lastnih orodij za obdelavo metabolomskih podatkov (Šket in sod., 2020; Murovec in sod., 2018; Deutsch in sod., 2021a; Deutsch in sod., 2021b; Deutsch in sod., 2022a; Deutsch in sod., 2022b) skupaj s komercialno dostopno programsko opremo za tarčno ^{1}H-NMR analizo. Na primer, iste spektre je mogoče ponovno analizirati s prihodnjimi posodobitvami baze podatkov o človeški metabolomski bazi (HMDB), saj je ta narasla z nekaj tisoč metabolitov v prvi izdaji (Wishart in sod., 2007) na 217.000 metabolitov v zadnji izdaji v 2021 (Wishart in sod., 2021). Standardizirani analitični protokoli, vzpostavljeni v našem laboratoriju, so nam omogočili, da smo zmanjšali sistematične napake. Box-Cox normalizacija in pristop sPLSDA, uporabljena za integracijo vseh metabolomov v naši študiji, sta pokazala uspešnost pri odstranjevanju učinkov različnih serij vzorcev na eni strani, hkrati pa pokaže še vedno ohranjene razlike zaradi življenjskega sloga ali drugih bioloških razlogov (Wang in La Cao, 2020). Ta pristop nam je omogočil tudi potrditev alternativne hipoteze iz razdelka 1.4.3, da obstajajo pomembne razlike v urinskih metabolomih, ki omogočajo identifikacijo naborov biomarkerjev in presnovnih poti, ki razmejujejo različne skupine, ki jih preučujemo (Schmidt, 2021).

Pokazali smo, da lahko metabolni prstni odtis v urinu omogoči posnetek metabolnega statusa celotnega sistema telesa, ki ga lahko povezujemo z zdravjem ali boleznijo (Azad in Shulaev, 2019; Mussap in sod., 2021). Metabolomika na splošno vključuje sistematično identifikacijo metabolitov v človeškem telesu (Ashrafian in sod., 2021). Razvoj nacionalne baze podatkov naj bi izboljšal razumevanje slovenskega metaboloma vzporedno s študijami iz drugih evropskih držav in identifikacijo metabolitov, specifičnih za različne bolezni ali fizična stanja. Metabolomika ^{1}H-NMR ima potencial za zajemanje širokega spektra običajnih kliničnih spremenljivk v epidemioloških študijah, vključno z manjkajočimi spremenljivkami za metapodatke o pacientih in omogoča ustvarjanje biomarkerjev za razlikovanje med različnimi boleznimi na podlagi strojnega učenja. Celostna razlaga metabolomskih podatkovnih nizov, zlasti urina, ki ga je mogoče zbrati neinvazivno, lahko zagotovi dovolj podatkov za sklepanje o tem, kako je treba vzorce razvrstiti v različne skupine. Upamo, da bomo spodbudili zanimanje drugih raziskovalcev za vključitev NMR metabolomike v svoje raziskave, da bi razširili našo uveljavljeno bazo podatkov na približno 10.000 vzorcev na nacionalni ravni. Modeliranje takšnega zbiranja podatkov predstavlja edinstveno pot za ustvarjanje modelov strojnega učenja, ki jih je mogoče vsaj okvirno uporabiti v medicinski praksi za razlikovanje med zdravimi in nezdravimi metabolomskimi stanji poleg različnih bolezni. Tako ta pristop

predstavlja korak bližje personalizirani medicini, ki temelji na podatkih in ima potencial za informiranje o zdravju na nacionalni ravni. Raziskovalni članek o slovenski NMR bazi je v pripravi.

Zgodba o Slovenski NMR bazi teče naprej. V skladu s tem je bilo (i) zbranih in izmerjenih skupno 320 vzorcev iz projekta PreAlti (razširitev projekta PreTerm), (ii) razširitev SMA je trenutno v fazi zbiranja vzorcev, (iii) vzorci se zbirajo tudi iz dveh kliničnih kohort iz Univerzitetnega kliničnega centra Ljubljana v sodelovanju s Pediatrično kliniko (tiki, anoreksija), medtem ko v okviru (iv) klinične kohorte, povezane s projektom Million Microbiomes from Humans Project, nameravamo zbrati več kot 1000 vzorcev blata in urina za metagenomiko in metabolomske analize. S temi projekti smo na poti, da ustvarimo na tisoče gigabajtov molekularnih podatkov, ki bodo v bodoče uporabni tudi v vsakodnevni diagnostiki in so primerljivi največjim evropskim študijam. Ohranjanje sistemske homeostaze ter odzivanje na prehranske in okoljske izzive zahteva usklajevanje različnih organov in tkiv. Da bi odgovorili na različne presnovne zahteve, človeško telo integrira sistem medorganske komunikacije, prek katerega lahko eno tkivo vpliva na presnovne poti v oddaljenem tkivu. Porušitev teh komunikacijskih poti zaradi pomanjkanja vadbe (sedeči življenjski slog) ali vnosa visokokalorične prehrane prispeva k človeškim boleznim, kot so debelost, sladkorna bolezen, bolezni jeter in ateroskleroza. Za pravočasne posege bi morali razmišljati o uporabi telesnih tekočin (kot je urin), ki omogočajo neinvazivno vzorčenje, hkrati pa so dovolj občutljive, da razlikujejo med vrstami biomarkerjev (Schmidt, 2021).

To odpira prostor za boljše razumevanje medorganske komunikacije kot vratarja za metabolno zdravje, saj obstajajo večsmerne interakcije med organi in osrednjim živčnim sistemom, z namenom ohranjanja energijske homeostaze in omogočanja novih terapevtskih strategij in spodbujanja zdravega življenja za preprečevanje presnovnih motenj in drugih bolezni.

5 REFERENCES

Abramowicz W.N., Galloway S.D. 2005. Effects of acute versus chronic L-carnitine L-tartrate supplementation on metabolic responses to steady state exercise in males and females. International Journal of Sport Nutrition and Exercise Metabolism, 15, 4: 386-400

Abrigo J., Gonzalez F., Aguirre F., Tacchi F., Gonzalez A., Meza M.P., Simon F., Cabrera D., Arrese M., Karpen S., Cabello-Verrugio, C. 2021. Cholic acid and deoxycholic acid induce skeletal muscle atrophy through a mechanism dependent on TGR5 receptor. Journal of Cellular Physiology, 236, 1: 260-272

Adamko D.J., Nair P., Mayers I., Tsuyuki R.T., Regush S., Rowe, B.H. 2015. Metabolomic profiling of asthma and chronic obstructive pulmonary disease: A pilot study differentiating diseases. The Journal of Allergy and Clinical Immunology, 136, 3: 571-580

Akram M. 2014. Citric acid cycle and role of its intermediates in metabolism. Cell Biochemistry and Biophysics, 68, 3: 475-478

Alamoudi J.A., Li W., Gautam N., Olivera M., Meza J., Mukherjee S., Alnouti, Y. 2021. Bile acid indices as biomarkers for liver diseases I: diagnostic markers. World Journal of Hepatology, 13, 4: 433-455

Alneberg J., Bjarnason B.S., De Bruijn I., Schirmer M., Quick J., Ijaz U.Z., Lahti L., Loman N.J., Andersson A.F., Quince C. 2014. Binning metagenomic contigs by coverage and composition. Nature Methods, 11, 4: 1144-1146

Alves C.R.R., Zhang R., Johnstone A.J., Garner R., Nwe P.H., Siranosian J.J., Swoboda, K.J. 2020. Serum creatinine is a biomarker of progressive denervation in spinal muscular atrophy. Neurology, 94, 9: e921-e931

Amano H., Maruyama K., Naka M., Tanaka T. 2003. Target validation in hypoxia-induced vascular remodeling using transcriptome/metabolome analysis. The Pharmacogenomics Journal, 3, 3: 183-188

Anderson M.J., Walsh D.C.I. 2013. PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: What null hypothesis are you testing? Ecological Monographs, 83, 4: 557-574

Apweiler R., Beissbarth T., Berthold M.R., Blüthgen N., Burmeister Y., Dammann O., Deutsch A., Feuerhake F., Franke A., Hasenauer J., Hoffmann S., Höfer T., Jansen P.L., Kaderali L., Klingmüller U., Koch I., Kohlbacher O., Kuepfer L., Lammert F., Maier D., Pfeifer N., Radde N., Rehm M., Roeder I., Saez-Rodriguez J., Sax U., Schmeck B., Schuppert A., Seilheimer B., Theis F.J., Vera J., Wolkenhauer O. 2018. Whither systems medicine? Experimental & Molecular Medicine, 50, 3: e453, doi: 10.1038/emm.2017.290, 6 p.

Argmann C.A., Houten S.M., Zhu J., Schadt, E.E. 2016. A next generation multiscale view of inborn errors of metabolism. Cell Metabolism, 23, 1: 13-26

Ariza A.C., Deen P.M., Robben J.H. 2012. The succinate receptor as a novel therapeutic target for oxidative and metabolic stress-related conditions. Frontiers in Endocrinology, 3: 22, doi: 10.3389/fendo.2012.00022, 8 p.

Armstrong N., Barker A.R. 2011. Endurance training and elite young athletes. Medicine and Sport science, 56: 59-83

Aron-Wisnewsky J., Prifti E., Belda E., Ichou F., Kayser B.D., Dao M.C., Verger E.O., Hedjazi L., Bouillot J.L., Chevallier J.M., Pons N., Le Chatelier E., Levenez F., Ehrlich, S.D., Dore J., Zucker J.D., Clement K. 2019. Major microbiota dysbiosis in severe obesity: fate after bariatric surgery. Gut, 68, 1: 70-82

Aßhauer K.P., Wemheuer B., Daniel R., Meinicke P. 2015. Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data. Bioinformatics (Oxford, England), 31, 17: 2882-2884

Ashrafian H., Sounderajah V., Glen R., Ebbels T., Blaise B.J., Kalra D., Kultima K., Spjuth O., Tenori L., Salek R.M., Kale N., Haug K., Schober D., Rocca-Serra P., O'donovan C., Steinbeck C., Cano I., De Atauri P., Cascante, M. 2021. Metabolomics: the stethoscope for the twenty-first century. Medical Principles and Practice: International Journal of the Kuwait University, Health Science Centre, 30, 4: 301-310

Aureli T., Miccheli A., Di Cocco M.E., Ghirardi O., Giuliani A., Ramacci M.T., Conti F. 1994. Effect of acetyl-L-carnitine on recovery of brain phosphorus metabolites and lactic acid level during reperfusion after cerebral ischemia in the rat--study by ^{13}P- and ^{1}H-NMR spectroscopy. Brain Research, 643, 1-2: 92-99

Azad R.K., Shulaev V. 2019. Metabolomics technology and bioinformatics for precision medicine. Briefings in Bioinformatics, 20, 6: 1957-1971

Barnes S., Benton H.P., Casazza K., Cooper S.J., Cui X., Du X., Engler J., Kabarowski J.H., Li S., Pathmasiri W., Prasain J.K., Renfrow M.B., Tiwari, H.K. 2016. Training in metabolomics research. II. Processing and statistical analysis of metabolomics data, metabolite identification, pathway analysis, applications of metabolomics and its future. Journal of Mass Spectrometry: JMS, 51, 8: 535-548

Barr A.J. 2018. The biochemical basis of disease. Essays in Biochemistry, 62, 5: 619-642

Bates M.L., Farrell, E.T., Eldridge, M.W. 2014. Abnormal ventilatory responses in adults born prematurely. The New England Journal of Medicine, 370, 6: 584-585

Beghini F., Mciver L.J., Blanco-Míguez A., Dubois L., Asnicar F., Maharjan S., Mailyan A., Manghi P., Scholz M., Thomas A.M., Valles-Colomer M., Weingart G., Zhang Y., Zolfo M., Huttenhower C., Franzosa, E.A., Segata, N. 2021. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. ELife, 10: e65088, doi: 10.7554/eLife.65088, 42 p.

Beirnaert C., Meysman P., Vu T.N., Hermans N., Apers S., Pieters L., Covaci A., Laukens K. 2018. speaq 2.0: A complete workflow for high-throughput 1D NMR spectra processing and quantification. PLoS Computational Biology, 14, 3: e1006018, doi: 10.1371/journal.pcbi.1006018, 25 p.

Bender A., Scheiber J., Glick M., Davies J.W., Azzaoui K., Hamon J., Urban L., Whitebread S., Jenkins J.L. 2007. Analysis of pharmacology data and the prediction of adverse drug reactions and off-target effects from chemical structure. ChemMedChem, 2, 6: 861-873

Bérard C., Payan C., Hodgkinson I., Fermanian J. 2005. A motor function measure for neuromuscular diseases. Construction and validation study. Neuromuscular Disorders: NMD, 15, 463-470

Berg G., Rybakova D., Fischer D., Cernava T., Vergès M.C., Charles T., Chen X., Cocolin L., Eversole K., Corral G.H., Kazou M., Kinkel L., Lange L., Lima N., Loy A., Macklin J.A., Maguin E., Mauchline T., Mcclure R., Mitter B., Ryan M., Sarand I., Smidt H., Schelkle B., Roume H., Kiran G.S., Selvin J., Souza R.S.C., Van Overbeek L., Singh B.K., Wagner M., Walsh A., Sessitsch A., Schloter, M. 2020. Microbiome definition re-visited: old concepts and new challenges. Microbiome, 8, 1: 103, doi: 10.1186/s40168-020-00875-0, 22 p.

Bingol K. 2018. Recent advances in targeted and untargeted metabolomics by NMR and MS/NMR methods. High-Throughput, 7, 2: 9, doi: 10.3390/ht7020009, 11 p.

Biomarkers Definitions Working Group. 2001. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clinical Pharmacology and Therapeutics, 69, 3: 89-95

Bizzarri D., Reinders M.J.T., Beekman M., Slagboom P.E., Bbmri N.L., Van Den Akker E.B. 2022. 1H-NMR metabolomics-based surrogates to impute common clinical risk factors and endpoints. EBioMedicine, 75: 103764, doi: 10.1016/j.ebiom.2021.103764, 15 p.

Bjerrum J.T., Wang Y., Hao F., Coskun M., Ludwig C., Gunther U., Nielsen O.H. 2015. Metabonomics of human fecal extracts characterize ulcerative colitis, Crohn's disease and healthy individuals. Metabolomics, 11: 122-133

Blake M.R., Raker J.M., Whelan K. 2016. Validity and reliability of the Bristol stool form scale in healthy adults and patients with diarrhoea-predominant irritable bowel syndrome. Alimentary Pharmacology & Therapeutics, 44, 7: 693-703

Blencowe H., Cousens S., Oestergaard M.Z., Chou D., Moller A.B., Narwal R., Adler A., Vera Garcia C., Rohde S., Say L., Lawn, J.E. 2012. National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications. Lancet (London, England), 379, 9832: 2162-2172

Booth F.W., Roberts C.K., Laye M.J. 2012. Lack of exercise is a major cause of chronic diseases. Comprehensive Physiology, 2, 2: 1143-1211

Borkowski A.A., Wilson C.P., Borkowski S.A., Thomas L.B., Deland L.A., Grewe S.J., Mastorides S.M. 2019. Google Auto ML versus Apple Create ML for histopathologic cancer diagnosis; Which algorithms are better? Comprehensive Physiology, 2, 2: 1143-1211

Bousquet J., Anto J.M., Sterk P.J., Adcock I.M., Chung K.F., Roca J., Agusti A., Brightling C., Cambon-Thomsen A., Cesario A., Abdelhak S., Antonarakis S.E., Avignon A., Ballabio A., Baraldi E., Baranov A., Bieber T., Bockaert J., Brahmachari S., Brambilla C., Bringer J., Dauzat M., Ernberg I., Fabbri L., Froguel P., Galas D., Gojobori T., Hunter P., Jorgensen C., Kauffmann F., Kourilsky P., Kowalski M.L., Lancet D., Le Pen C., Mallet J., Mayosi B., Mercier J., Metspalu A., Nadeau J.H., Ninot G., Noble D., Ozturk M., Palkonen S., Prefaut C., Rabe K., Renard E., Roberts R.G., Samolinski B., Schunemann H.J., Simon H.U., Soares M.B., Superti-Furga G., Tegner J., Verjovski-Almeida S., Wellstead P., Wolkenhauer O., Wouters E., Balling R., Brookes A.J., Charron D., Pison C., Chen Z., Hood L., Auffray, C. 2011. Systems medicine and integrated care to combat chronic noncommunicable diseases. Genome Medicine, 3: 43, doi: 10.1186/gm259, 12 p.

Bowers R.M., Kyrpides N.C., Stepanauskas R., Harmon-Smith M., Doud D., Reddy T.B.K., Schulz F., Jarett J., Rivers A.R., Eloe-Fadrosh E.A., Tringe S.G., Ivanova N.N., Copeland A., Clum A.,

Becraft E.D., Malmstrom R.R., Birren B., Podar M., Bork P., Weinstock G.M., Garrity G.M., Dodsworth J.A., Yooseph S., Sutton G., Glöckner F.O., Gilbert J.A., Nelson W.C., Hallam S.J., Jungbluth S.P., Ettema T.J.G., Tighe S., Konstantinidis K.T., Liu W.T., Baker B.J., Rattei T., Eisen J.A., Hedlund B., Mcmahon K.D., Fierer N., Knight R., Finn R., Cochrane G., Karsch-Mizrachi I., Tyson G.W., Rinke C., Consortium G.S., Lapidus A., Meyer F., Yilmaz P., Parks D.H., Eren A.M., Schriml L., Banfield J.F., Hugenholtz P., Woyke, T. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nature Biotechnology, 35, 8: 725-731

Bro R., Kamstrup-Nielsen M.H., Engelsen S.B., Savorani F., Rasmussen M.A., Hansen L., Olsen A., Tjonneland A., Dragsted, L.O. 2015. Forecasting individual breast cancer risk using plasma metabolomics and biocontours. Metabolomics, 11, 5: 1376-1380

Brown C.T., Sharon I., Thomas B.C., Castelle C.J., Morowitz M.J., Banfield, J.F. 2013. Genome resolved analysis of a premature infant gut microbial community reveals a Varibaculum cambriense genome and a shift towards fermentation-based metabolism during the third week of life. Microbiome, 1: 30, doi: 10.1186/2049-2618-1-30, 19 p.

Callahan B.J., Mcmurdie P.J., Holmes, S.P. 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. The ISME Journal, 11, 12: 2639-2643

Cañueto D., Gómez J., Salek R.M., Correig X., Cañellas, N. 2018. rDolphin: a GUI R package for proficient automatic profiling of 1D 1 H-NMR spectra of study datasets. Metabolomics: Official Journal of the Metabolomic Society, 14, 3: 24, doi: 10.1007/s11306-018-1319-y, 5 p.

Castillo-Armengol J., Fajas L., Lopez-Mejia I.C. 2019. Inter-organ communication: a gatekeeper for metabolic health. EMBO Reports, 20, 9: e47903, doi: 10.15252/embr.201947903, 16 p.

Castro A., Duft R.G., Ferreira M.L.V., Andrade A.L.L., Gáspari A.F., Silva L.M., Oliveira-Nunes S.G., Cavaglieri C.R., Ghosh S., Bouchard C., Chacon-Mikahil M.P.T. 2019. Association of skeletal muscle and serum metabolites with maximum power output gains in response to continuous endurance or high-intensity interval training programs: The TIMES study - a randomized controlled trial. PloS One 14, 2: e0212115, doi: 10.1371/journal.pone.0212115, 32 p.

Castro J.C., Rodriguez-R L.M., Harvey W.T., Weigand M.R., Hatt J.K., Carter M.Q., Konstantinidis, K.T. 2018. imGLAD: accurate detection and quantification of target organisms in metagenomes. PeerJ, 6: e5882, doi: 10.7717/peerj.5882, 23 p.

Chen K., Zhang Q., Wang J., Liu F., Mi M., Xu H., Chen F., Zeng, K. 2009. Taurine protects transformed rat retinal ganglion cells from hypoxia-induced apoptosis by preventing mitochondrial dysfunction. Brain Research, 1279: 131-138

Chen P.C., Pan C., Gharibani P.M., Prentice H., Wu, J.Y. 2013. Taurine exerts robust protection against hypoxia and oxygen/glucose deprivation in human neuroblastoma cell culture. Advances in Experimental Medicine and Biology, 775: 167-175

Chen S., Zhou Y., Chen Y., Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics (Oxford, England), 34, 17: i884-i890

Chiriboga C.A., Swoboda K.J., Darras B.T., Iannaccone S.T., Montes J., De Vivo D.C., Norris D.A., Bennett C.F., Bishop K.M. 2016. Results from a phase 1 study of nusinersen (ISIS-SMN(Rx)) in children with spinal muscular atrophy. Neurology, 86, 10: 890-897

Chong J., Liu P., Zhou G., Xia J. 2020. Using MicrobiomeAnalyst for comprehensive statistical, functional, and meta-analysis of microbiome data. Nature Protocols, 15, 3: 799-821

Chong J., Soufan O., Li C., Caraus I., Li S., Bourque G., Wishart D.S., Xia J. 2018. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. Nucleic Acids Research, 46, W1: W486-W494

Chong J., Wishart D.S., Xia J. 2019. Using MetaboAnalyst 4.0 for comprehensive and integrative metabolomics data analysis. Current Protocols in Bioinformatics, 68, 1: e86, doi: 10.1002/cpbi.86, 128 p.

Chumpitazi B.P., Self M.M., Czyzewski D.I., Cejka S., Swank P.R., Shulman R.J. 2016. Bristol stool form scale reliability and agreement decreases when determining Rome III stool form designations. Neurogastroenterology and Motility: The Official Journal of the European Gastrointestinal Motility Society, 28, 3: 443-448

Clemm H.H., Vollsæter M., Røksund O.D., Eide G.E., Markestad T., Halvorsen, T. 2014. Exercise capacity after extremely preterm birth. Development from adolescence to adulthood. Annals of the American Thoracic Society, 11, 4: 537-545

Coen P.M., Jubrias S.A., Distefano G., Amati F., Mackey D.C., Glynn N.W., Manini T.M., Wohlgemuth S.E., Leeuwenburgh C., Cummings S.R., Newman A.B., Ferrucci L., Toledo F.G., Shankland E., Conley K.E., Goodpaster, B.H. 2013. Skeletal muscle mitochondrial energetics are associated with maximal aerobic capacity and walking speed in older adults. The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences, 68, 4: 447-455

Connors J., Dawe N., Van Limbergen J. 2018. The role of succinate in the regulation of intestinal inflammation. Nutrients, 11, 1: 25, doi: 10.3390/nu11010025, 12 p.

Corey D.R. 2017. Nusinersen, an antisense oligonucleotide drug for spinal muscular atrophy. Nature Neuroscience, 20, 4: 497- 499

Costea P.I., Zeller G., Sunagawa S., Pelletier E., Alberti A., Levenez F., Tramontano M., Driessen M., Hercog R., Jung F.E., Kultima J.R., Hayward M.R., Coelho L.P., Allen-Vercoe E., Bertrand L., Blaut M., Brown J.R.M., Carton T., Cools-Portier S., Daigneault M., Derrien M., Druesne A., De Vos W.M., Finlay B.B., Flint H.J., Guarner F., Hattori M., Heilig H., Luna R.A., Van Hylckama Vlieg J., Junick J., Klymiuk I., Langella P., Le Chatelier E., Mai V., Manichanh C., Martin J.C., Mery C., Morita H., O'toole P.W., Orvain C., Patil K.R., Penders J., Persson S., Pons N., Popova M., Salonen A., Saulnier D., Scott K.P., Singh B., Slezak K., Veiga P., Versalovic J., Zhao L., Zoetendal E.G., Ehrlich S.D., Dore J., Bork, P. 2017. Towards standards for human fecal sample processing in metagenomic studies. Nature Biotechnology, 35, 11: 1069-1076

Craig, J. 2008. Complex diseases: research and applications. Nature Education, 1: 184

Crass M.F., Lombardini J.B. 1977. Loss of cardiac muscle taurine after acute left ventricular ischemia. Life Sciences, 21, 7: 951-958

Cremer J., Segota I., Yang C.Y., Arnoldini M., Sauls J.T., Zhang Z., Gutierrez E., Groisman A., Hwa T. 2016. Effect of flow and peristaltic mixing on bacterial growth in a gut-like channel. Proceedings of the National Academy of Sciences of the United States of America, 113, 41: 11414-11419

Cremer J., Arnoldini M., Hwa T. 2017. Effect of water flow and chemical environment on microbiota growth and composition in the human colon. Proceedings of the National Academy of Sciences of the United States of America, 114, 25: 10069-10240

Cristianini N., Shawe-Taylor, J. 2000. An introduction to support vector machines and other kernel-based learning methods. Cambridge, Cambridge University Press: 189 p.

Crump C., Sundquist J., Winkleby M.A., Sundquist K. 2019. Gestational age at birth and mortality from infancy into mid-adulthood: a national cohort study. The Lancet. Child & Adolescent Health, 3, 6: 408-417

Crump C. 2020. An overview of adult health outcomes after preterm birth. Early Human Development, 150: 105187, doi: 10.1016/j.earlhumdev.2020.105187, 8 p.

D'amore R., Ijaz U.Z., Schirmer M., Kenny J.G., Gregory R., Darby A.C., Shakya M., Podar M., Quince C., Hall, N. 2016. A comprehensive benchmarking study of protocols and sequencing platforms for 16S rRNA community profiling. BMC Genomics, 17: 55, doi: 10.1186/s12864-015-2194-9, 20 p.

Da Silva R.R., Dorrestein P.C., Quinn R.A. 2015. Illuminating the dark matter in metabolomics. Proceedings of the National Academy of Sciences of the United States of America, 112, 41: 12549-12550

Dai Z., Wong S.H., Yu J., Wei Y. 2019. Batch effects correction for microbiome data with Dirichlet-multinomial regression. Bioinformatics (Oxford, England), 35, 5: 807-814

Danese E., Salvagno G.L., Tarperi C., Negrini D., Montagnana M., Festa L., Sanchis-Gomar F., Schena F., Lippi G. 2017. Middle-distance running acutely influences the concentration and composition of serum bile acids: Potential implications for cancer risk? Oncotarget, 8, 32: 52775-52782

Debevec T., Bali T.C., Simpson E.J., Macdonald I.A., Eiken O., Mekjavic I.B. 2014. Separate and combined effects of 21-day bed rest and hypoxic confinement on body composition. European Journal of Applied Physiology, 114, 11: 2411-2425

Debevec T., Millet G.P., Pialoux, V. 2017. Hypoxia-induced oxidative stress modulation with physical activity. Frontiers in Physiology, 8: 84, doi: 10.3389/fphys.2017.00084, 9 p.

Debevec T., Pialoux V., Millet G.P., Martin A., Mramor M., Osredkar D. 2019. Exercise overrides blunted hypoxic ventilatory response in prematurely born men. Frontiers in Physiology, 10: 437, doi: 10.3389/fphys.2019.00437, 10 p.

Debevec T., Poussel M., Osredkar D., Willis S.J., Sartori C., Millet, G.P. 2022. Post-exercise accumulation of interstitial lung water is greater in hypobaric than normobaric hypoxia in adults born prematurely. Respiratory Physiology & Neurobiology, 297: 34890833, doi: 10.1016/j.resp.2021.103828, 4 p.

Deo R.C. 2015. Machine learning in medicine. Circulation, 132, 20: 1920-1930

Derrien M., Van Passel M.W., Van De Bovenkamp J.H., Schipper R.G., De Vos W.M., Dekker J. 2010. Mucin-bacterial interactions in the human oral cavity and digestive tract. Gut Microbes, 1, 4: 254-268

Deutsch L., Osredkar D., Plavec J., Stres, B. 2021. Spinal muscular atrophy after nusinersen therapy: improved physiology in pediatric patients with no significant change in urine, serum, and liquor

^{1}H-NMR metabolomes in comparison to an age-matched, healthy cohort. Metabolites, 11, 4: 206, doi: 10.3390/metabo11040206, 15 p.

Deutsch L., Stres, B. 2021. The importance of objective stool classification in fecal ^{1}H-NMR metabolomics: exponential increase in stool crosslinking is mirrored in systemic inflammation and associated to fecal acetate and methionine. Metabolites, 11, 3: 172, doi: 10.3390/metabo11030172, 16 p.

Deutsch L., Soritirdis A., Murovec B., Plavec J., Mekjavic I., Debevec T., Stres B. 2022a. Exercise and interorgan communication: short-term exercise training blunts differences in consecutive daily urine ^{1}H-NMR metabolomic signatures between physically active and inactive individuals. Metabolites, 12, 6: 473, doi: https://doi:org/10.3390/metabo12060473, 16 p.

Deutsch L., Debevec T., Millet G. P., Osredkar D., Opara S., Šket R., Murovec B., Mramor M, Plavec J., Stres B. 2022b. Urine and fecal 1H-NMR metabolomes differ significantly between pre-term and full-term born physically fit healthy adults. Metabolites, 12, 6: 536, doi: https://doi.org/10.3390/metabo12060536, 23 p.

Dhariwal A., Chong J., Habib S., King I.L., Agellon L.B., Xia J. 2017. MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. Nucleic Acids Research, 45, W1: W180-W188

Di Liegro C.M., Schiera G., Proia P., Di Liegro I. 2019. Physical activity and brain health. Genes, 10, 9: 720, doi: 10.3390/genes10090720, 40 p.

Ding X., Yang F., Chen Y., Xu J., He J., Zhang R., Abliz Z. 2022. Norm ISWSVR: A data integration and normalization approach for large-scale metabolomics. Analytical Chemistry, 78, 13: 4281-4290

Dona A.C., Kyriakides M., Scott F., Shephard E.A., Varshavi D., Veselkov K., Everett J.R. 2016. A guide to the identification of metabolites in NMR-based metabonomics/metabolomics experiments. Computational and Structural Biotechnology Journal, 14: 135-153

Douglas G.M., Maffei V.J., Zaneveld J.R., Yurgel S.N., Brown J.R., Taylor C.M., Huttenhower C., Langille M.G.I. 2020. PICRUSt2 for prediction of metagenome functions. Nature Biotechnology, 38, 6: 685-688

Dumas M.E., Kinross J., Nicholson J.K. 2014. Metabolic phenotyping and systems biology approaches to understanding metabolic syndrome and fatty liver disease. Gastroenterology, 146, 1: 46-62

Dunstan D.W., Dogra S., Carter S.E., Owen, N. 2021. Sit less and move more for cardiovascular health: emerging insights and opportunities. Nature Reviews. Cardiology, 18, 9: 637-648

Ebbels T.M.D., Lindon J.C., Coen M. 2013. Processing and modelling of nuclear magnetic resonance (NMR). In: Metabolic profiling. Metz T. O. (ed). London, Humana Press: 365-388

Ehrlein H.J., Schemann M. 2005. Gastrointestinal motility. Technische Universität München, 26 p.

Ekins S., Puhl A.C., Zorn K.M., Lane T.R., Russo D.P., Klein J.J., Hickey A.J., Clark A.M. 2019. Exploiting machine learning for end-to-end drug discovery and development. Nature Materials, 18, 5: 435-441

Elliott P., Posma J.M., Chan Q., Garcia-Perez I., Wijeyesekera A., Bictash M., Ebbels T.M., Ueshima H., Zhao L., Van Horn L., Daviglus M., Stamler J., Holmes E., Nicholson J.K. 2015. Urinary

metabolic signatures of human adiposity. Science Translational Medicine, 7, 285: 285ra62, doi: 10.1126/scitranslmed.aaa5680, 16 p.

Emwas A.H. 2015. The strengths and weaknesses of NMR spectroscopy and mass spectrometry with particular focus on metabolomics research. Methods in Molecular Biology (Clifton, N.J.), 1277: 161-193

Emwas A.H., Roy R., Mckay R.T., Ryan D., Brennan L., Tenori L., Luchinat C., Gao X., Zeri A.C., Gowda G.A., Raftery D., Steinbeck C., Salek R.M., Wishart D.S. 2016. Recommendations and standardization of biomarker quantification using NMR-based metabolomics with particular focus on urinary analysis. Journal of Proteome Research, 15, 2: 360-373

Emwas A.H., Saccenti E., Gao X., Mckay R.T., Dos Santos V.A.P.M., Roy R., Wishart D.S. 2018. Recommended strategies for spectral processing and post-processing of 1D 1 H-NMR data of biofluids with a particular focus on urine. Metabolomics: Official Journal of the Metabolomic Society, 14, 3: 31, doi: 10.1007/s11306-018-1321-4, 23 p.

Emwas A.H., Roy R., Mckay R.T., Tenori L., Saccenti E., Gowda G.A.N., Raftery D., Alahmari F., Jaremko L., Jaremko M., Wishart, D.S. 2019. NMR spectroscopy for metabolomics research. Metabolites, 9, 7: 123, doi: 10.3390/metabo9070123, 39 p.

Faber J.E., Szymeczek C.L., Cotecchia S., Thomas S.A., Tanoue A., Tsujimoto G., Zhang H. 2007. Alpha1-adrenoceptor-dependent vascular hypertrophy and remodeling in murine hypoxic pulmonary hypertension. American Journal of Physiology. Heart and Circulatory Physiology, 292, 5: H2316-H2323

Fang C., Zhong H., Lin Y., Chen B., Han M., Ren H., Lu H., Luber J.M., Xia M., Li W., Stein S., Xu X., Zhang W., Drmanac R., Wang J., Yang H., Hammarström L., Kostic A.D., Kristiansen K., Li J. 2018. Assessment of the cPAS-based BGISEQ-500 platform for metagenomic sequencing. GigaScience, 7, 3: gix133, doi: 10.1093/gigascience/gix133, 8 p.

Fanos, V. 2016. Metabolomics and microbiomics personalized - medicine from the fetus to the adult. London, Academic Press: 144 p.

Farrell E.T., Bates M.L., Pegelow D.F., Palta M., Eickhoff J.C., O'brien M.J., Eldridge M.W. 2015. Pulmonary gas exchange and exercise capacity in adults born preterm. Annals of the American Thoracic Society, 12, 8: 1130-1137

Feurer M., Eggensperger K., Falkner S., Lindauer M., Hutter F. 2021. Auto-Sklearn 2.0: hands-free AutoML via meta-Learning. ArXiv, doi: https://doi:org/10.48550/arXiv.2007.04074, 56 p.

Filippone M., Bonetto G., Corradi M., Frigo A.C., Baraldi E. 2012. Evidence of unexpected oxidative stress in airways of adolescents born very pre-term. The European Respiratory Journal, 40, 5: 1253-1259

Franconi F., Stendardi I., Failli P., Matucci R., Baccaro C., Montorsi L., Bandinelli R., Giotti A. 1985. The protective effects of taurine on hypoxia (performed in the absence of glucose) and on reoxygenation (in the presence of glucose) in guinea-pig heart. Biochemical Pharmacology, 34, 15: 2611-2615

Frank D.N., St Amand A.L., Feldman R.A., Boedeker E.C., Harpaz N., Pace N.R. 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: 13780-13785

Fricker A.M., Podlesny D., Fricke W.F. 2019. What is new and relevant for sequencing-based microbiome research? A mini-review. Journal of Advanced Research, 19: 105-112

Fukiya S., Arata M., Kawashima H., Yoshida D., Kaneko M., Minamida K., Watanabe J., Ogura Y., Uchida K., Itoh K., Wada M., Ito S., Yokota A. 2009. Conversion of cholic acid and chenodeoxycholic acid into their 7-oxo derivatives by Bacteroides intestinalis AM-1 isolated from human feces. FEMS Microbiology Letters, 293, 2: 263-270

Gallo Cantafio M.E., Grillone K., Caracciolo D., Scionti F., Arbitrio M., Barbieri V., Pensabene L., Guzzi P.H., Di Martino M.T. 2018. From single level analysis to multi-omics integrative approaches: a powerful strategy towards the precision oncology. High-throughput, 7, 4: 33, doi: 10.3390/ht7040033, 20 p.

Garud N.R., Good B.H., Hallatschek O., Pollard K.S. 2019. Evolutionary dynamics of bacteria in the gut microbiome within and across hosts. PLoS Biology, 17,1: e3000102, doi: 10.1371/journal.pbio.3000102, 29 p.

Gibbons S.M., Duvallet C., Alm E.J. 2018. Correcting for batch effects in case-control microbiome studies. PLoS Computational Biology, 14, 4: e1006102, doi: 10.1371/journal.pcbi.1006102, 17 p.

Giongo A., Gano K.A., Crabb D.B., Mukherjee N., Novelo L.L., Casella G., Drew J.C., Ilonen J., Knip M., Hyoty H., Veijola R., Simell T., Simell O., Neu J., Wasserfall C.H., Schatz D., Atkinson M.A., Triplett E.W. 2011. Toward defining the autoimmune microbiome for type 1 diabetes. The ISME Journal, 5, 1: 82-91

Giraudeau P., Silvestre V., Akoka S. 2015. Optimizing water suppression for quantitative NMR-based metabolomics: a tutorial review. Metabolomics, 11, 5: 1041-1055

Glanzman A.M., Mazzone E., Main M., Pelliccioni M., Wood J., Swoboda K.J., Scott C., Pane M., Messina S., Bertini E., Mercuri E., Finkel R.S. 2010. The children's hospital of philadelphia infant test of neuromuscular disorders (CHOP INTEND): test development and reliability. Neuromuscular Disorders: NMD, 20, 3: 155-161

Glover L.E., Lee J.S., Colgan, S.P. 2016. Oxygen metabolism and barrier regulation in the intestinal mucosa. The Journal of Clinical Investigation, 126, 10: 3680-3688

Graham E.D., Heidelberg J.F., Tully, B.J. 2017. BinSanity: unsupervised clustering of environmental microbial assemblies using coverage and affinity propagation. PeerJ, 5: e3035, doi: 10.7717/peerj.3035, 19 p.

Grillet A.M., Wyatt N.B., Gloe L.M. 2012. Polymer gel rheology and adhesion. In: Rheology. De Vicente J. (ed). London, IntechOpen: 59-80

Hammer O., Harper D.A.T., Ryan P.D. 2001. PAST: Paleontological statistics software package for education and data analysis. Palaeontologia Electronica, 1, 9: 4, doi: http://palaeo-electronica.org/2001_1/past/issue1_01.htm, 9 p.

Han M., Hao L., Lin Y., Li F., Wang J., Yang H., Xiao L., Kristiansen K., Jia H., Li J. 2018. A novel affordable reagent for room temperature storage and transport of fecal samples for metagenomic analyses. Microbiome, 6, 1: 43, doi: 10.1186/s40168-018-0429-0, 7 p.

Haraldsdottir K., Watson A.M., Beshish A.G., Pegelow D.F., Palta M., Tetri L.H., Brix M.D., Centanni R.M., Goss K.N., Eldridge M.W. 2019. Heart rate recovery after maximal exercise is impaired in healthy young adults born preterm. European Journal of Applied Physiology, 119, 4: 857-866

Hasin Y., Seldin M., Lusis A. 2017. Multi-omics approaches to disease. Genome Biology, 18,1: 83, doi: 10.1186/s13059-017-1215-1, 15 p.

Heaton K.W., Radvan J., Cripps H., Mountford R.A., Braddon F.E.M., Hughes A.O. 1992. Defecation frequency and timing, and stool form in the general population: a prospective study. Gut, 33, 6: 818-824

Heinken A., Ravcheev D.A., Baldini F., Heirendt L., Fleming R.M.T., Thiele I. 2019. Systematic assessment of secondary bile acid metabolism in gut microbes reveals distinct metabolic capabilities in inflammatory bowel disease. Microbiome, 7, 1: 75, doi: 10.1186/s40168-019-0689-3, 18 p.

Hoff P., Belavý D.L., Huscher D., Lang A., Hahne M., Kuhlmey A.K., Maschmeyer P., Armbrecht G., Fitzner R., Perschel F.H., Gaber T., Burmester G.R., Straub R.H., Felsenberg D., Buttgereit F. 2015. Effects of 60-day bed rest with and without exercise on cellular and humoral immunological parameters. Cellular & Molecular Immunology, 12, 4: 483-492

Holmes E., Loo R.L., Stamler J., Bictash M., Yap I.K., Chan Q., Ebbels T., De Iorio M., Brown I.J., Veselkov K.A., Daviglus M.L., Kesteloot H., Ueshima H., Zhao L., Nicholson J.K., Elliott P. 2008a. Human metabolic phenotype diversity and its association with diet and blood pressure. Nature, 453, 7193: 396-400

Holmes E., Wilson I.D., Nicholson J.K. 2008b. Metabolic phenotyping in health and disease. Cell, 134, 5: 714-717

Hood L., Friend S.H. 2011. Predictive, personalized, preventive, participatory (P4) cancer medicine. Nature reviews. Clinical Oncology, 8, 3: 184-187

Hood L., Flores M. 2012. A personal view on systems medicine and the emergence of proactive P4 medicine: predictive, preventive, personalized and participatory. New Biotechnology, 29, 6: 613-624

Hunter P., Nielsen P. 2005. A strategy for integrative computational physiology. Physiology (Bethesda, Md.), 20: 316-325

Hutter F., Kotthoff L., Vanschoren J. 2019. Automated machine learning. Cham, Switzerland, Springer: 219 p.

Jain C., Rodriguez-R L.M., Phillippy A.M., Konstantinidis K.T., Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nature Communications, 9 1: 5114, doi: 10.1038/s41467-018-07641-9, 8p.

Jay O., Bain A.R., Deren T.M., Sacheli M., Cramer M.N. 2011. Large differences in peak oxygen uptake do not independently alter changes in core temperature and sweating during exercise. American journal of physiology. Regulatory, Integrative and Comparative physiology, 301, 3: R832-R841

Johnson L.R., Ghishan F.K., Kaunitz J.D., Merchant J.L., Said H.M., Wood J.D. 2012. Physiology of the gastrointestinal tract. London, Academic Press: 2308 p.

Johnson W.E., Li C., Rabinovic A. 2007. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics (Oxford, England), 8, 1: 118-127

Karaglani M., Gourlia K., Tsamardinos I., Chatzaki E. 2020. Accurate blood-based diagnostic biosignatures for Alzheimer's disease via automated machine learning. Journal of Clinical Medicine, 9, 9: 3016, doi: 10.3390/jcm9093016, 14 p.

Kelly R.S., Kelly M.P., Kelly P. 2020a. Metabolomics, physical activity, exercise and health: A review of the current evidence. Biochimica et Biophysica Acta (BBA) Molecular Basis of Disease, 1866, 12: 165936, doi: https://doi:org/10.1016/j.bbadis.2020.165936, 17 p.

Kerkhof G.F., Breukhoven P.E., Leunissen R.W., Willemsen R.H., Hokken-Koelega A.C. 2012. Does preterm birth influence cardiovascular risk in early adulthood? The Journal of Pediatrics, 161, 3: 390-396

Keun H.C., Athersuch T.J. 2022. Nuclear magnetic resonance (NMR) - based metabolomics. In: Metabolic Profiling. Metz T. O. (ed). London, Humana Press: 321-334

Kim K.A., Jeong J.J., Yoo S.Y., Kim D.H. 2016. Gut microbiota lipopolysaccharide accelerates inflamm-aging in mice. BMC Microbiology, 16: 9, doi: 10.1186/s12866-016-0625-7, 9 p.

Klein M.S. 2021. Affine transformation of negative values for NMR metabolomics using the mrbin R package. Journal of Proteome Research, 20, 2: 1397-1404

Knight R., Vrbanac A., Taylor B.C., Aksenov A., Callewaert C., Debelius J., Gonzalez A., Kosciolek T., Mccall L.I., Mcdonald D., Melnik A.V., Morton J.T., Navas J., Quinn R.A., Sanders J.G., Swafford A.D., Thompson L.R., Tripathi A., Xu Z.Z., Zaneveld J.R., Zhu Q., Caporaso J.G., Dorrestein P.C. 2018. Best practices for analysing microbiomes. Nature Reviews. Microbiology, 16, 7: 410-422

Koeth R.A., Wang Z., Levison B.S., Buffa J.A., Org E., Sheehy B.T., Britt E.B., Fu X., Wu Y., Li L., Smith J.D., Didonato J.A., Chen J., Li H., Wu G.D., Lewis J.D., Warrier M., Brown J.M., Krauss R.M., Tang W.H., Bushman F.D., Lusis A.J., Hazen S.L. 2013. Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. Nature Medicine, 19, 5: 576-585

Kohl H.W., Craig C.L., Lambert E.V., Inoue S., Alkandari J.R., Leetongin G., Kahlmeier S., Group L.P.A.S.W. 2012. The pandemic of physical inactivity: global action for public health. Lancet (London, England), 380, 9838: 294-305

Kohl P., Crampin E.J., Quinn T.A., Noble D. 2010. Systems biology: an approach. Clinical Pharmacology and Therapeutics, 88, 1: 25-33

Kostic A.D., Gevers D., Pedamallu C.S., Michaud M., Duke F., Earl A.M., Ojesina A.I., Jung J., Bass A.J., Tabernero J., Baselga J., Liu C., Shivdasani R.A., Ogino S., Birren B.W., Huttenhower C., Garrett W.S., Meyerson M. 2012. Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. Genome Research, 22, 2: 292-298

Kotthoff L., Thornton C., Hoos H.H., Hutter F., Leyton-Brown K. 2017. Auto-WEKA 2.0: Automatic model selection and hyperparameter optimization in WEKA. Journal of Machine Learning Research, 18, 25: 1-5

Kraut J.A., Madias N.E. 2016. Metabolic acidosis of CKD: an update. American Journal of Kidney Diseases: The Official Journal of the National Kidney Foundation, 67, 2: 307-317.

Kurtzer G.M., Sochat V., Bauer M.W. 2017. Singularity: scientific containers for mobility of compute. PloS One, 12, 5: e0177459, doi: 10.1371/journal.pone.0177459, 20 p.

La Rosa S.L., Leth M.L., Michalak L., Hansen M.E., Pudlo N.A., Glowacki R., Pereira G., Workman C.T., Arntzen M.Ø., Pope P.B., Martens E.C., Hachem M.A., Westereng B. 2019. The human gut Firmicute Roseburia intestinalis is a primary degrader of dietary β-mannans. Nature Communications, 10, 1: 905, doi: 10.1038/s41467-019-08812-y, 14 p.

Lakrisenko P., Weindl D. 2021. Dynamic models for metabolomics data integration. Current Opinion in Systems Biology, 28: 100358, doi: https://doi:org/10.1016/j.coisb.2021.100358, 7 p.

Langille M.G., Zaneveld J., Caporaso J.G., Mcdonald D., Knights D., Reyes J.A., Clemente J.C., Burkepile D.E., Vega Thurber R.L., Knight R., Beiko R.G., Huttenhower C. 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nature Biotechnology, 31, 9: 814-821

Lapidus A.L., Korobeynikov A.I. 2021. Metagenomic data assembly - the way of decoding unknown microorganisms. Frontiers in Microbiology, 12: 613791, doi: https://doi:org/10.3389/fmicb.2021.613791, 16 p.

Lee E.C., Fragala M.S., Kavouras S.A., Queen R.M., Pryor J.L., Casa D.J. 2017. Biomarkers in sports and exercise: tracking health, performance, and recovery in athletes. Journal of Strength and Conditioning Research, 31, 10: 2920-2937

Lefebvre S., Bürglen L., Reboullet S., Clermont O., Burlet P., Viollet L., Benichou B., Cruaud C., Millasseau P., Zeviani M. 1995. Identification and characterization of a spinal muscular atrophy-determining gene. Cell, 80, 1: 155-165

Legendre P., Legendre L.F.J. 2012. Numerical Ecology. Amsterdam, Elsevier: 1006 p.

Lent-Schochet D., Mclaughlin M., Ramakrishnan N., Jialal I. 2019. Exploratory metabolomics of metabolic syndrome: A status report. World journal of diabetes, 10, 1: 23-36

Lewis S.J., Heaton K.W. 1997. Stool form scale as a useful guide to intestinal transit time. Scandinavian Journal of Gastroenterology, 32, 9: 920-924

Li D., Liu C.M., Luo R., Sadakane K., Lam T.W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics (Oxford, England), 31, 10: 1674-1676

Li S., Ung T.T., Nguyen T.T., Sah D.K., Park S.Y., Jung Y.D. 2020. Cholic acid stimulates MMP-9 in human colon cancer cells via activation of MAPK, AP-1, and NF-κB Activity. International Journal of Molecular Sciences, 21, 10: 3420, doi: 10.3390/ijms21103420, 16 p.

Lin W., Djukovic A., Mathur D., Xavier J.B. 2021. Listening in on the conversation between the human gut microbiome and its host. Current Opinion in Microbiology, 63: 150-157

Lindon J.C., Holmes E., Nicholson J.K. 2003. So what's the deal with metabonomics? Analytical Chemistry, 75, 17: 384a-391a

Lindstad L.J., Lo G., Leivers S., Lu Z., Michalak L., Pereira G.V., Røhr Å.K., Martens E.C., Mckee L.S., Louis P., Duncan S.H., Westereng B., Pope P.B., La Rosa S.L. 2021. Human gut Faecalibacterium prausnitzii deploys a highly efficient conserved system to cross-feed on β-mannan-derived oligosaccharides. MBio, 12, 3: e0362820, doi: 10.1128/mBio.03628-20, 18 p.

Ling W., Zhao N., Lulla A., Plantinga A.M., Fu W., Zhang A., Liu H., Li Z., Chen J., Randolph T., Koay W.L.A., White J.R., Launer L.J., Fodor A.A., Meyer K.A., Wu M.C. 2021. Batch effects removal for microbiome data via conditional quantile regression (ConQuR). BioRxiv: 10.1101/2021.09.23.461592, 29 p.

Littlewood-Evans A., Sarret S., Apfel V., Loesle P., Dawson J., Zhang J., Muller A., Tigani B., Kneuer R., Patel S., Valeaux S., Gommermann N., Rubic-Schneider T., Junt T., Carballido J.M. 2016. GPR91 senses extracellular succinate released from inflammatory macrophages and exacerbates rheumatoid arthritis. The Journal of Experimental Medicine, 213, 9: 1655-1662

Liu L., Oza S., Hogan D., Perin J., Rudan I., Lawn J.E., Cousens S., Mathers C., Black R.E. 2015. Global, regional, and national causes of child mortality in 2000-13, with projections to inform post-2015 priorities: an updated systematic analysis. Lancet (London, England), 385, 9966: 430-440

Lorson C.L., Androphy E.J. 2000. An exonic enhancer is required for inclusion of an essential exon in the SMA-determining gene SMN. Human Molecular Genetics, 9, 2: 259-265

Lovering A.T., Laurie S.S., Elliott J.E., Beasley K.M., Yang X., Gust C.E., Mangum T.S., Goodman R.D., Hawn J.A., Gladstone I.M. 2013. Normal pulmonary gas exchange efficiency and absence of exercise-induced arterial hypoxemia in adults with bronchopulmonary dysplasia. Journal of Applied Physiology (Bethesda, Md.: 1985), 115, 7: 1050-1056

Lunn M.R., Wang C.H. 2008. Spinal muscular atrophy. Lancet (London, England), 371, 9630: 2120-2133

Lushchak V.I. 2014. Free radicals, reactive oxygen species, oxidative stress and its classification. Chemico-Biological Interactions, 224: 164-175

Lustgarten M.S., Price L.L., Chalé A., Fielding R.A. 2014. Metabolites related to gut bacterial metabolism, peroxisome proliferator-activated receptor-alpha activation, and insulin sensitivity are associated with physical function in functionally-limited older adults. Aging Cell, 13, 5: 918-925

Ma S., Tong M., Yuan S., Liu H. 2019. Responses of the microbial community structure in Fe (II)-bearing sediments to oxygenation: the role of reactive oxygen species. ACS Earth and Space Chemistry, 3, 5: 738-747

Maalouf N.M., Cameron M.A., Moe O.W., Adams-Huet B., Sakhaee K. 2007. Low urine pH: a novel feature of the metabolic syndrome. Clinical Journal of the American Society of Nephrology: CJASN, 2, 5: 883-888

Madrid-Gambin F., Oller-Moreno S., Fernandez L., Bartova S., Giner M.P., Joyce C., Ferraro F., Montoliu I., Moco S., Marco S. 2020. AlpsNMR: An R package for signal processing of fully untargeted NMR-based metabolomics. Bioinformatics (Oxford, England), 36, 9: 2943-2945

Magalhães J., Ascensão A., Viscor G., Soares J., Oliveira J., Marques F., Duarte J. 2004. Oxidative stress in humans during and after 4 hours of hypoxia at a simulated altitude of 5500 m. Aviation, Space, and Environmental Medicine, 75, 1: 16-22

Maguire M.L. 2014. An introduction to metabolomics and systems biology. In: Metabolomics and systems biology in human health and medicine. Oliver J. A. H. (ed.). Oxfordshire, United Kingdom, Cabi: 1-19

Malcangio M., Bartolini A., Ghelardini C., Bennardini F., Malmberg-Aiello P., Franconi F., Giotti A. 1989. Effect of ICV taurine on the impairment of learning, convulsions and death caused by hypoxia. Psychopharmacology, 98, 3: 316-320

Mallick H., Franzosa E.A., McIver L.J., Banerjee S., Sirota-Madi A., Kostic A.D., Clish C.B., Vlamakis H., Xavier R.J., Huttenhower C. 2019. Predictive metabolomic profiling of microbial communities using amplicon or metagenomic sequences. Nature Communications, 10: 3136, doi: 10.1038/s41467-019-10927-1, 11 p.

Manley B.J., Doyle L.W., Davies M.W., Davis P.G. 2015. Fifty years in neonatology. Journal of Paediatrics and Child Health, 51, 1: 118-121

Marin L., Miguelez E.M., Villar C.J., Lombo F. 2015. Bioavailability of dietary polyphenols and gut microbiota metabolism: antimicrobial properties. BioMed Research International: 905215, doi: 10.1155/2015/905215, 18 p.

Markopoulou P., Papanikolaou E., Analytis A., Zoumakis E., Siahanidou T. 2019. Preterm birth as a risk factor for metabolic syndrome and cardiovascular disease in adult life: a systematic review and meta-analysis. The Journal of Pediatrics, 210: 69-80

Martin A., Faes C., Debevec T., Rytz C., Millet G., Pialoux V. 2018. Preterm birth and oxidative stress: Effects of acute physical exercise and hypoxia physiological responses. Redox Biology, 17: 315-322

Martin A., Millet G., Osredkar D., Mramor M., Faes C., Gouraud E., Debevec T., Pialoux V. 2020. Effect of pre-term birth on oxidative stress responses to normoxic and hypoxic exercise. Redox Biology, 32: 101497, doi: 10.1016/j.redox.2020.101497, 7 p.

Martínez Y., Li X., Liu G., Bin P., Yan W., Más D., Valdivié M., Hu C.A., Ren W., Yin Y. 2017. The role of methionine on metabolism, oxidative stress, and diseases. Amino Acids, 49, 12: 2091-2098

Matsuda K., Akiyama T., Tsujibe S., Oki K., Gawad A., Fujimoto J. 2021. Direct measurement of stool consistency by texture analyzer and calculation of reference value in Belgian general population. Scientific Reports, 11, 1: 2400, doi: 10.1038/s41598-021-81783-7, 7p.

Maurer A., Ward J.L., Dean K., Billinger S.A., Lin H., Mercer K.E., Adams S.H., Thyfault J.P. 2020. Divergence in aerobic capacity impacts bile acid metabolism in young women. Journal of Applied Physiology (Bethesda, Md.: 1985), 129, 4: 768-778

McIver L.J., Abu-Ali G., Franzosa E.A., Schwager R., Morgan X.C., Waldron L., Segata N., Huttenhower C. 2018. bioBakery: a meta'omic analysis environment. Bioinformatics 34, 7: 1235-1237

Melki J. 2017. Advances in Spinal Muscular Atrophy Research. In: Spinal muscular atrophy - Disease mechanisms and therapy. Sumner C.J., Paushkin S., Ko C (eds). London, Academic Press: xxiii - xxiv

Mercer K.E., Maurer A., Pack L.M., Ono-Moore K., Spray B.J., Campbell C., Chandler C.J., Burnett D., Souza E., Casazza G., Keim N., Newman J., Hunter G., Fernadez J., Garvey W.T., Harper M.E., Hoppel C., Adams S.H., Thyfault J. 2021. Exercise training and diet-induced weight loss increase markers of hepatic bile acid (BA) synthesis and reduce serum total BA concentrations in

obese women. American Journal of Physiology. Endocrinology and Metabolism, 320, 5: E864-E873

Michalk D.V., Wingenfeld P., Licht C. 1997. Protection against cell damage due to hypoxia and reoxygenation: the role of taurine and the involved mechanisms. Amino Acids, 13, 3-4: 337-346

Mizeranschi A., Groen D., Borgdorff J., Hoekstra A.G., Chopard B., Dubitzky W. 2016. Anatomy and physiology of multiscale modeling and simulation in systems medicine. In: Systems medicine. Methods in molecular biology. Schmitz U. (ed). New York, USA, Humana Press: 375-405

Montero D., Lundby C. 2017. Refuting the myth of non-response to exercise training: 'non-responders' do respond to higher dose of training. The Journal of Physiology, 595, 11: 3377-3387

Moreno-Indias I., Lahti L., Nedyalkova M., Elbere I., Roshchupkin G., Adilovic M., Aydemir O., Bakir-Gungor B., Santa Pau E.C., D'elia D., Desai M.S., Falquet L., Gundogdu A., Hron K., Klammsteiner T., Lopes M.B., Marcos-Zambrano L.J., Marques C., Mason M., May P., Pašić L., Pio G., Pongor S., Promponas V.J., Przymus P., Saez-Rodriguez J., Sampri A., Shigdel R., Stres B., Suharoschi R., Truu J., Truică C.O., Vilne B., Vlachakis D., Yilmaz E., Zeller G., Zomer A.L., Gómez-Cabrero D., Claesson M.J. 2021. Statistical and machine learning techniques in human microbiome studies: contemporary challenges and solutions. Frontiers in Microbiology, 12: 635781, doi: 10.3389/fmicb.2021.635781, 9 p.

Moutquin J.M. 2003. Classification and heterogeneity of preterm birth. BJOG: An International Journal of Obstetrics and Gynaecology, 110, 20: 30-33

Murovec B., Deutsch L., Stres B. 2020. Computational framework for high-quality production and large-scale evolutionary analysis of metagenome assembled genomes. Molecular Biology and Evolution, 37, 2: 593-598

Murovec B., Makuc D., Kolbl Repinc S., Prevorsek Z., Zavec D., Sket R., Pecnik K., Plavec J., Stres B. 2018. (1)H NMR metabolomics of microbial metabolites in the four MW agricultural biogas plant reactors: A case study of inhibition mirroring the acute rumen acidosis symptoms. Journal of Environmental Management, 222: 428-435

Mussap M., Noto A., Piras C., Atzori L., Fanos V. 2021. Slotting metabolomics into routine precision medicine. Expert Review of Precision Medicine and Drug Development, 6, 3: 173-187

Mustafa A., Rahimi Azghadi M. 2021. Automated machine learning for healthcare and clinical notes analysis. Computers 10, 2: 24, doi: 10.3390/computers10020024, 31 p.

Mysara M., Vandamme P., Props R., Kerckhof F.M., Leys N., Boon N., Raes J., Monsieurs P. 2017. Reconciliation between operational taxonomic units and species boundaries. FEMS Microbiology Ecology, 93, 4: fix029, doi: 10.1093/femsec/fix029, 12 p.

Müller M., Hernández M.A.G., Goossens G.H., Reijnders D., Holst J.J., Jocken J.W.E., Van Eijk H., Canfora E.E., Blaak E.E. 2019. Circulating but not faecal short-chain fatty acids are related to insulin sensitivity, lipolysis and GLP-1 concentrations in humans. Scientific Reports, 9:12515, https://doi:org/10.1038/s41598-019-48775-0, 9 p.

Narayan N.R., Weinmaier T., Laserna-Mendieta E.J., Claesson M.J., Shanahan F., Dabbagh K., Iwai S., Desantis T.Z. 2020. Piphillin predicts metagenomic composition and dynamics from DADA2-

corrected 16S rDNA sequences. BMC Genomics, 21, 1: 56, doi: 10.1186/s12864-019-6427-1, 12 p.

Nayfach S., Pollard, K.S. 2016. Toward accurate and quantitative comparative metagenomics. Cell, 166, 5: 1103-1116

Nayfach S., Shi Z.J., Seshadri R., Pollard K.S., Kyrpides N.C. 2019. New insights from uncultivated genomes of the global human gut microbiome. Nature, 568, 7753: 505-510

Nieman D.C., Shanely R.A., Gillitt N.D., Pappan K.L., Lila M.A. 2013. Serum metabolic signatures induced by a three-day intensified exercise period persist after 14 h of recovery in runners. Journal of Proteome Research, 12, 10: 4577-4584

Noble D. 2002. Modeling the heart--from genes to cells to the whole organ. Science (New York, N.Y.), 295, 5560: 1678-1682

Nurk S., Meleshko D., Korobeynikov A., Pevzner P.A. 2017. metaSPAdes: a new versatile metagenomic assembler. Genome Research, 27, 5: 824-834

Osredkar D., Jílková M., Butenko T., Loboda T., Golli T., Fuchsová P., Rohlenová M., Haberlova J. 2021. Children and young adults with spinal muscular atrophy treated with nusinersen. European Journal of Paediatric Neurology, 30: 1 - 8

Otaki Y., Watanabe T., Takahashi H., Hasegawa H., Honda S., Funayama A., Netsu S., Ishino M., Arimoto T., Shishido T., Miyashita T., Miyamoto T., Konta T., Kubota I. 2013. Acidic urine is associated with poor prognosis in patients with chronic heart failure. Heart and Vessels, 28, 6: 735-741

Page A.J., Cummins C.A., Hunt M., Wong V.K., Reuter S., Holden M.T., Fookes M., Falush D., Keane J.A., Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics (Oxford, England), 31,22: 3691-3693

Pang Z., Chong J., Zhou G., De Lima Morais D.A., Chang L., Barrette M., Gauthier C., Jacques P.É., Li S., Xia, J. 2021. MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. Nucleic Acids Research, 49, W1: W388-W396

Paradis A.N., Gay M.S., Wilson C.G., Zhang L. 2015. Newborn hypoxia/anoxia inhibits cardiomyocyte proliferation and decreases cardiomyocyte endowment in the developing heart: role of endothelin-1. PloS One, 10, 2: e0116600, doi: 10.1371/journal.pone.0116600, 21p.

Parks D.H., Imelfort M., Skennerton C.T., Hugenholtz P., Tyson G.W. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Research, 25, 7: 1043-1055

Patterson J., Carpenter E.J., Zhu Z., An D., Liang X., Geng C., Drmanac R., Wong G.K. 2019. Impact of sequencing depth and technology on de novo RNA-Seq assembly. BMC Genomics, 20, 1: 604, doi: 10.1186/s12864-019-5965-x, 14 p.

Peisl B.Y.L., Schymanski E.L., Wilmes P. 2018. Dark matter in host-microbiome metabolomics: Tackling the unknowns-a review. Analytica Chimica Acta, 1037: 13-27

Peng Y., Leung H.C., Yiu S.M., Chin F.Y. 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics (Oxford, England), 28, 11: 1420-1428

Pera M.C., Coratti G., Forcina N., Mazzone E.S., Scoto M., Montes J., Pasternak A., Mayhew A., Messina S., Sframeli M., Main M., Lofra R.M., Duong T., Ramsey D., Dunaway S., Salazar R., Fanelli L., Civitello M., De Sanctis R., Antonaci L., Lapenta L., Lucibello S., Pane M., Day J., Darras B.T., De Vivo D.C., Muntoni F., Finkel R., Mercuri E. 2017. Content validity and clinical meaningfulness of the HFMSE in spinal muscular atrophy. BMC Neurology, 17, 1:39, doi: 10.1186/s12883-017-0790-9, 10 p.

Perrone S., Negro S., Laschi E., Calderisi M., Giordano M., De Bernardo G., Parigi G., Toni A.L., Esposito S., Buonocore G. 2021. Metabolomic profile of young adults born preterm. Metabolites, 11, 10: 697, doi: 10.3390/metabo11100697, 11 p.

Pialoux V., Mounier R., Rock E., Mazur A., Schmitt L., Richalet J.P., Robach P., Coudert J., Fellmann N. 2009. Effects of acute hypoxic exposure on prooxidant/antioxidant balance in elite endurance athletes. International Journal of Sports Medicine, 30, 2: 87-93

Powers S.K., Nelson W.B., Hudson M.B. 2011. Exercise-induced oxidative stress in humans: cause and consequences. Free Radical Biology & Medicine, 51, 5: 942-950

Price N.D., Magis A.T., Earls J.C., Glusman G., Levy R., Lausted C., Mcdonald D.T., Kusebauch U., Moss C.L., Zhou Y., Qin S., Moritz R.L., Brogaard K., Omenn G.S., Lovejoy J.C., Hood L. 2017. A wellness study of 108 individuals using personal, dense, dynamic data clouds. Nature Biotechnology, 35, 8: 747-756

Pushpass R.G., Alzoufairi S., Jackson K.G., Lovegrove J.A. 2021. Circulating bile acids as a link between the gut microbiota and cardiovascular health: impact of prebiotics, probiotics and polyphenol-rich foods. Nutrition Research Reviews: 1-20

Qin J., Li R., Raes J., Arumugam M., Burgdorf K.S., Manichanh C., Nielsen T., Pons N., Levenez F., Yamada T., Mende D.R., Li J., Xu J., Li S., Li D., Cao J., Wang B., Liang H., Zheng H., Xie Y., Tap J., Lepage P., Bertalan M., Batto J.M., Hansen T., Le Paslier D., Linneberg A., Nielsen H.B., Pelletier E., Renault P., Sicheritz-Ponten T., Turner K., Zhu H., Yu C., Li S., Jian M., Zhou Y., Li Y., Zhang X., Li S., Qin N., Yang H., Wang J., Brunak S., Doré J., Guarner F., Kristiansen K., Pedersen O., Parkhill J., Weissenbach J., Consortium M., Bork P., Ehrlich S.D., Wang, J. 2010. A human gut microbial gene catalogue established by metagenomic sequencing. Nature, 464: 59-65

Qin J., Li Y., Cai Z., Li S., Zhu J., Zhang F., Liang S., Zhang W., Guan Y., Shen D., Peng Y., Zhang D., Jie Z., Wu W., Qin Y., Xue W., Li J., Han L., Lu D., Wu P., Dai Y., Sun X., Li Z., Tang A., Zhong S., Li X., Chen W., Xu R., Wang M., Feng Q., Gong M., Yu J., Zhang Y., Zhang M., Hansen T., Sanchez G., Raes J., Falony G., Okuda S., Almeida M., Lechatelier E., Renault P., Pons N., Batto J.M., Zhang Z., Chen H., Yang R., Zheng W., Yang H., Wang J., Ehrlich S.D., Nielsen R., Pedersen O., Kristiansen K. 2012. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature, 490: 55-60

Qin N., Yang F., Li A., Prifti E., Chen Y., Shao L., Guo J., Le Chatelier E., Yao J., Wu L., Zhou J., Ni S., Liu L., Pons N., Batto J.M., Kennedy S.P., Leonard P., Yuan C., Ding W., Hu X., Zheng B., Qian G., Xu W., Ehrlich S.D., Zheng S., Li, L. 2014. Alterations of the human gut microbiome in liver cirrhosis. Nature, 513: 59-64

Qiu S., Cai X., Sun Z., Li L., Zuegel M., Steinacker J.M., Schumann U. 2017. Heart rate recovery and risk of cardiovascular events and all-cause mortality: A meta-analysis of prospective cohort studies. Journal of the American Heart Association, 6, 5: e005505, doi: 10.1161/JAHA.117.005505, 16 p.

Quince C., Walker A.W., Simpson J.T., Loman N.J., Segata N. 2017. Shotgun metagenomics, from sampling to analysis. Nature Biotechnology, 35, 12: 1211-1211

Ramdas S., Servais L. 2020. New treatments in spinal muscular atrophy: an overview of currently available data. Expert Opinion on Pharmacotherapy, 21,3: 307-315

Richter M., Rosselló-Móra R., Oliver Glöckner F., Peplies J. 2016. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. Bioinformatics (Oxford, England), 32, 6: 929-931

Rigo F., Hua Y., Krainer A.R., Bennett C.F. 2012. Antisense-based therapy for the treatment of spinal muscular atrophy. The Journal of Cell Biology, 199, 1: 21-25

Rittweger J., Debevec T., Frings-Meuthen P., Lau P., Mittag U., Ganse B., Ferstl P.G., Simpson E.J., Macdonald I.A., Eiken O., Mekjavic I.B. 2016. On the combined effects of normobaric hypoxia and bed rest upon bone and mineral metabolism: Results from the PlanHab study. Bone, 91: 130-138

Roager H.M., Hansen L.B.S., Bahl M.I., Frandsen H.L., Carvalho V., Gobel R.J., Dalgaard M.D., Plichta D.R., Sparholt M.H., Vestergaard H., Hansen T., Sicheritz-Ponten T., Nielsen H.B., Pedersen O., Lauritzen L., Kristensen M., Gupta R., Licht T.R. 2016. Colonic transit time is related to bacterial metabolism and mucosal turnover in the gut. Nature Microbiology, 1, 9:16093, doi: 10.1038/nmicrobiol.2016.93, 9 p.

Robinson M.M., Dasari S., Konopka A.R., Johnson M.L., Manjunatha S., Esponda R.R., Carter R.E., Lanza I.R., Nair, K.S. 2017. Enhanced protein translation underlies improved metabolic and physical adaptations to different exercise training modes in young and old humans. Cell Metabolism, 25, 3: 581-592

Rodriguez-R L.M., Gunturu S., Harvey W.T., Rosselló-Mora R., Tiedje J.M., Cole J.R., Konstantinidis K.T. 2018. The microbial genomes atlas (MiGA) webserver: taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level. Nucleic Acids Research, 46, W1:W282-W288

Rohart F., Gautier B., Singh A., Lê Cao K.A. 2017. mixOmics: An R package for 'omics feature selection and multiple data integration. PLoS Computational Biology. 13, 11: e1005752, doi: 10.1371/journal.pcbi.1005752, 19 p.

Rubic T., Lametschwandtner G., Jost S., Hinteregger S., Kund J., Carballido-Perrig N., Schwärzler C., Junt T., Voshol H., Meingassner J.G., Mao X., Werner G., Rot A., Carballido J.M. 2008. Triggering the succinate receptor GPR91 on dendritic cells enhances immunity. Nature Immunology, 9, 11: 1261-1269

Ruiz-Perez C.A., Conrad R.E., Konstantinidis K.T. 2021. MicrobeAnnotator: a user-friendly, comprehensive functional annotation pipeline for microbial genomes. BMC Bioinformatics, 22, 1: 11, doi: 10.1186/s12859-020-03940-5, 16 p.

Rühlemann M.C., Hermes B.M., Bang C., Doms S., Moitinho-Silva L., Thingholm L.B., Frost F., Degenhardt F., Wittig M., Kässens J., Weiss F.U., Peters A., Neuhaus K., Völker U., Völzke H., Homuth G., Weiss S., Grallert H., Laudes M., Lieb W., Haller D., Lerch M.M., Baines J.F., Franke A. 2021. Genome-wide association study in 8,956 German individuals identifies influence of ABO histo-blood groups on gut microbiome. Nature Genetics, 53, 2: 147-155

Sallis J.F., Bull F., Guthold R., Heath G.W., Inoue S., Kelly P., Oyeyemi A.L., Perez L.G., Richards J., Hallal P.C., Committee L.P.A.S.E. 2016. Progress in physical activity over the olympic quadrennium. Lancet (London, England), 388, 10051: 1325-1336

Samuel G., Reeves P. 2003. Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. Carbohydrate Research, 338, 23: 2503-2519

Scafidi S., Fiskum G., Lindauer S.L., Bamford P., Shi D., Hopkins I., Mckenna M.C. 2010. Metabolism of acetyl-L-carnitine for energy and neurotransmitter synthesis in the immature rat brain. Journal of Neurochemistry, 114, 3: 820-831

Scheer M., Bischoff A.M., Kruzliak P., Opatrilova R., Bovell D., Büsselberg D. 2016. Creatine and creatine pyruvate reduce hypoxia-induced effects on phrenic nerve activity in the juvenile mouse respiratory system. Experimental and Molecular Pathology, 101, 1: 157-162

Scheperjans F., Aho V., Pereira P.A., Koskinen K., Paulin L., Pekkonen E., Haapaniemi E., Kaakkola S., Eerola-Rautio J., Pohja M., Kinnunen E., Murros K., Auvinen P. 2015. Gut microbiota are related to Parkinson's disease and clinical phenotype. Movement Disorders: Official Journal of the Movement Disorder Society, 30, 3: 350-358

Scher J.U., Sczesnak A., Longman R.S., Segata N., Ubeda C., Bielski C., Rostron T., Cerundolo V., Pamer E.G., Abramson S.B., Huttenhower C., Littman D.R. 2013. Expansion of intestinal Prevotella copri correlates with enhanced susceptibility to arthritis. Elife, 2: e01202, doi: 10.7554/eLife.01202, 20 p.

Schloss P.D. 2021. Amplicon sequence variants artificially split bacterial genomes into separate clusters. MSphere, 6, 4: e0019121, doi: 10.1101/2021.02.26.433139, 6p.

Schloss P.D., Westcott S.L., Ryabin T., Hall J.R., Hartmann M., Hollister E.B., Lesniewski R.A., Oakley B.B., Parks D.H., Robinson C.J., Sahl J.W., Stres B., Thallinger G.G., Van Horn D.J., Weber C.F. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and Environmental Microbiology, 75, 23: 7537-7541

Schmidt H.H.H.W. 2021. The end of medicine as we know it - and why your health has a future. Springer Cham: 291 p.

Schmidt T.S.B., Raes J., Bork P. 2018. The human gut microbiome: from association to modulation. Cell, 172, 6: 1198-1215

Schranner D., Kastenmuller G., Schonfelder M., Romisch-Margl W., Wackerhage H. 2020. Metabolite concentration changes in humans after a bout of exercise: a systematic review of exercise metabolomics studies. Sports Medicine - Open, 6, 1: 11, doi: 10.1186/s40798-020-0238-4, 17 p.

Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics (Oxford, England) 30, 14: 2068-2069

Segata N., Waldron L., Ballarini A., Narasimhan V., Jousson O., Huttenhower C. 2012. Metagenomic microbial community profiling using unique clade-specific marker genes. Nature Methods, 9, 8: 811-814

Senn T., Hazen S.L., Tang W.H. 2012. Translating metabolomics to cardiovascular biomarkers. Progress in Cardiovascular Diseases, 55, 1: 70-76

Shah J., Jefferies A.L., Yoon E.W., Lee S.K., Shah P.S., Network C.N. 2015. Risk factors and outcomes of late-onset bacterial sepsis in preterm neonates born at. American Journal of Perinatology, 32, 7: 675-682

Shen M., Xiao Y., Golbraikh A., Gombar V.K., T Ropsha, A. 2003. Development and validation of k-nearest-neighbor QSPR models of metabolic stability of drug candidates. Journal of Medicinal Chemistry, 46, 14: 3013-3020

Shimodaira M., Okaniwa S., Nakayama T. 2017. Fasting single-spot urine pH is associated with metabolic syndrome in the Japanese population. Medical Principles and Practice: International Journal of the Kuwait University, Health Science Centre, 26, 433-437

Sibomana I., Foose D.P., Raymer M.L., Reo N.V., Karl J.P., Berryman C.E., Young A.J., Pasiakos S.M., Mauzy C.A. 2021. Urinary metabolites as predictors of acute mountain sickness severity. Frontiers in Physiology, 12: 709804, doi: 10.3389/fphys.2021.709804, 11 p.

Sidey-Gibbons J.A.M., Sidey-Gibbons C.J. 2019. Machine learning in medicine: a practical introduction. BMC Medical Research Methodology, 19, 1: 64, doi: 10.1186/s12874-019-0681-4, 18 p.

Sieber C.M.K., Probst A.J., Sharrar A., Thomas B.C., Hess M., Tringe S.G., Banfield J.F. 2018. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nature Microbiology, 3, 7: 836-843

Singh A., Shannon C.P., Gautier B., Rohart F., Vacher M., Tebbutt S.J., Lê Cao K.A. 2019. DIABLO: an integrative approach for identifying key molecular drivers from multi-omics assays. Bioinformatics (Oxford, England), 35, 17: 3055-3062

Sinha R., Abu-Ali G., Vogtmann E., Fodor A.A., Ren B., Amir A., Schwager E., Crabtree J., Ma S., Consortium M.Q.C.P., Abnet C.C., Knight R., White O., Huttenhower C. 2017. Assessment of variation in microbial community amplicon sequencing by the microbiome quality control (MBQC) project consortium. Nature Biotechnology, 35, 11: 1077-1086

Sket R., Debevec T., Kublik S., Schloter M., Schoeller A., Murovec B., Mikus K.V., Makuc D., Pecnik K., Plavec J., Mekjavic I.B., Eiken O., Prevorsek Z., Stres B. 2018. Intestinal metagenomes and metabolomes in healthy young males: inactivity and hypoxia generated negative physiological symptoms precede microbial dysbiosis. Frontiers in Physiology, 9: 198, doi:10.3389/fphys.2018.00198, 16 p.

Sket R., Treichel N., Debevec T., Eiken O., Mekjavic I., Schloter M., Vital M., Chandler J., Tiedje J.M., Murovec B., Prevorsek Z., Stres B. 2017a. Hypoxia and inactivity related physiological changes (constipation, inflammation) are not reflected at the level of gut metabolites and butyrate producing microbial community: The PlanHab study. Frontiers in Physiology, 8: 250, doi: 10.3389/fphys.2017.00250, 16 p.

Sket R., Treichel N., Kublik S., Debevec T., Eiken O., Mekjavic I., Schloter M., Vital M., Chandler J., Tiedje J.M., Murovec B., Prevorsek Z., Likar M., Stres, B. 2017b. Hypoxia and inactivity related physiological changes precede or take place in absence of significant rearrangements in bacterial community structure: The PlanHab randomized trial pilot study. Plos One, 12, 12: e0188556, 26 p.

Smeriglio P., Langard P., Querin G., Biferi M.G. 2020. The identification of novel biomarkers is required to improve adult SMA patient stratification, diagnosis and treatment. Journal of Personalized Medicine, 10, 3: 75, doi:10.3390/jpm10030075, 23 p.

Sonntag B., Stolze B., Heinecke A., Luegering A., Heidemann J., Lebiedz P., Rijcken E., Kiesel L., Domschke W., Kucharzik T., Maaser C. 2007. Preterm birth but not mode of delivery is associated with an increased risk of developing inflammatory bowel disease later in life. Inflammatory bowel diseases, 13, 11: 1385-1390

Sotiridis A. 2019. Independent and combined effects of heat and hypoxic acclimation on exercise performance in humans: with particular reference to cross-adaption. Doctoral Dissertation. Ljubljana, Institute Jozef Stefan: 190 p.

Sotiridis A., Debevec T., Ciuha U., Eiken O., Mekjavic I.B. 2019. Heat acclimation does not affect maximal aerobic power in thermoneutral normoxic or hypoxic conditions. Experimental Physiology, 104, 3: EP087268, doi: 10.1113/EP087268, 14 p.

Sotiridis A., Debevec T., Ciuha U., Mcdonnell A.C., Mlinar T., Royal J.T., Mekjavic I.B. 2020. Aerobic but not thermoregulatory gains following a 10-day moderate-intensity training protocol are fitness level dependent: A cross-adaptation perspective. Physiological Reports, 8, 3: e14355, doi: 10.14814/phy2.14355, 17 p.

Sotiridis A., Debevec T., Mcdonnell A.C., Ciuha U., Eiken O., Mekjavic I.B. 2018. Exercise cardiorespiratory and thermoregulatory responses in normoxic, hypoxic and hot environment following 10-day continuous hypoxic exposure. Journal of applied physiology (Bethesda, Md.: 1985), 125: 1284–1295

Spiering B.A., Kraemer W.J., Hatfield D.L., Vingren J.L., Fragala M.S., Ho J.Y., Thomas G.A., Häkkinen K., Volek J.S. 2008. Effects of L-carnitine L-tartrate supplementation on muscle oxygenation responses to resistance exercise. Journal of Strength and Conditioning Research, 22, 4: 1130-1135

Staley C., Weingarden A.R., Khoruts A., Sadowsky M.J. 2017. Interaction of gut microbiota with bile acid metabolism and its influence on disease states. Applied Microbiology and Biotechnology, 101, 1: 47-64

Stres B., Kronegger L. 2019. Shift in the paradigm towards next-generation microbiology. Fems Microbiology Letters, 366, 15: fnz159, doi:10.1093/femsle/fnz159, 9 p.

Sugarman E.A., Nagan N., Zhu H., Akmaev V.R., Zhou Z., Rohlfs E.M., Flynn K., Hendrickson B.C., Scholl T., Sirko-Osadsa D.A., Allitto B.A. 2012. Pan-ethnic carrier screening and prenatal diagnosis for spinal muscular atrophy: clinical laboratory analysis of >72,400 specimens. European journal of human genetics: EJHG, 20, 1: 27-32

Sun S., Jones R.B., Fodor A.A. 2020. Inference-based accuracy of metagenome prediction tools varies across sample types and functional categories. Microbiome, 8: 46, doi: https://doi:org/10.1186/s40168-020-00815-y, 9 p.

Susnow R.G., Dixon S.L. 2003. Use of robust classification techniques for the prediction of human cytochrome P450 2D6 inhibition. Journal of Chemical Information and Computer Sciences 43, 4: 1308-1315

Svedenkrans J., Henckel E., Kowalski J., Norman M., Bohlin K. 2013. Long-term impact of preterm birth on exercise capacity in healthy young men: a national population-based cohort study. PloS One, 8, 12: e80869, doi: 10.1371/journal.pone.0080869, 10 p.

Šket R., Deutsch L., Prevoršek Z., Mekjavić I.B., Plavec J., Rittweger J., Debevec T., Eiken O., Stres B. 2020. Systems view of deconditioning during spaceflight simulation in the PlanHab project: the departure of urine ^{1}H-NMR metabolomes from healthy state in young males subjected to bedrest inactivity and hypoxia. Frontiers in physiology, 11: 532271, doi: 10.3389/fphys.2020.532271, 15 p.

Tabone M., Bressa C., García-Merino J.A., Moreno-Pérez D., Van E.C., Castelli F.A., Fenaille F., Larrosa M. 2021. The effect of acute moderate-intensity exercise on the serum and fecal metabolomes and the gut microbiota of cross-country endurance athletes. Scientific Reports, 11, 1: 3558, doi: 10.1038/s41598-021-82947-1, 12 p.

Tang W.H., Wang Z., Levison B.S., Koeth R.A., Britt E.B., Fu X., Wu Y., Hazen S.L. 2013. Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. The New England Journal of Medicine, 368, 17: 1575-1584

Tannahill G.M., Curtis A.M., Adamik J., Palsson-Mcdermott E.M., Mcgettrick A.F., Goel G., Frezza C., Bernard N.J., Kelly B., Foley N.H., Zheng L., Gardet A., Tong Z., Jany S.S., Corr S.C., Haneklaus M., Caffrey B.E., Pierce K., Walmsley S., Beasley F.C., Cummins E., Nizet V., Whyte M., Taylor C.T., Lin H., Masters S.L., Gottlieb E., Kelly V.P., Clish C., Auron P.E., Xavier R.J., O'neill L.A. 2013. Succinate is an inflammatory signal that induces IL-1β through HIF-1α. Nature, 496: 238-242

Tebani A., Afonso C., Marret S., Bekri S. 2016. Omics-based strategies in precision medicine: toward a paradigm shift in inborn errors of metabolism investigations. International Journal of Molecular Sciences, 17, 9: 1555, doi: 10.3390/ijms17091555, 27 p.

Teeling H., Meyerdierks A., Bauer M., Amann R., Glöckner F.O. 2004. Application of tetranucleotide frequencies for the assignment of genomic fragments. Environmental Microbiology, 6, 9: 938-947

Ten V.S. 2017. Mitochondrial dysfunction in alveolar and white matter developmental failure in premature infants. Pediatric Research, 81, 2, 286-292

Teschendorff A.E. 2019. Avoiding common pitfalls in machine learning omic data science. Nature Materials, 18, 5: 422-427

Thapa C., Camtepe S. 2021. Precision health data: Requirements, challenges and existing techniques for data security and privacy. Computers in Biology and Medicine, 129: 104130, doi: 10.1016/j.compbiomed.2020.104130, 19 p.

Thornton C., Hutter F., Hoos H.H., Eyton-Brown K. 2013. Auto-WEKA: Combined selection and hyperparameter optimization of classification algorithms: 1208.3719v2, doi: https://doi:org/10.48550/arXiv.1208.3719, 9 p.

Tian Q., Corkum A.E., Moaddel R., Ferrucci L. 2021. Metabolomic profiles of being physically active and less sedentary: a critical review. Metabolomics: Official Journal of the Metabolomic Society, 17, 7: 68, doi: 10.1007/s11306-021-01818-y, 16 p.

Tigchelaar E.F., Bonder M.J., Jankipersadsing A., Fu J., Wijmenga C., Zhernakova A. 2016. Gut microbiota composition associated with stool consistency. Gut, 65, 3: 540-542

Tingleff T., Vikanes Å., Räisänen S., Sandvik L., Murzakanova G., Laine K. 2021. Risk of preterm birth in relation to history of preterm birth: a population-based registry study of 213 335 women in Norway. BJOG: An International Journal of Obstetrics and Gynaecology, 129: 900-907, doi: 10.1111/1471-0528.17013, 8 p.

Trygg J., Wold, S. 2002. Orthogonal projections to latent structures (O-PLS). Journal of Chemometrics, 16, 3: 119-128

Tsamardinos I., Charonyktakis P., Lakiotaki K., Borboudakis G., Zenklusen J.C., Juhl H., Chatzaki E., Lagani, V. 2020. Just add data: Automated predictive modeling and biosignature discovery. bioRxiv: doi: 10.1101/2020.05.04.075747, 46 p.

Tsamardinos I., Charonyktakis P., Papoutsoglou G., Borboudakis G., Lakiotaki K., Zenklusen J. C., Juhl H., Chatzaki E., Lagani V. 2022. Just add data: automated and predicitve modeling for knowledge discovery and feature selection. Npj Precision Oncology, 6: 38, doi: https://doi.org/10.1038/s41698-022-00274-8, 17 p.

Turner C.E., Byblow W.D., Gant N. 2015. Creatine supplementation enhances corticomotor excitability and cognitive performance during oxygen deprivation. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 35, 4: 1773-1780

Töpfer N., Kleessen S., Nikoloski Z. 2015. Integration of metabolomics data into metabolic networks. Frontiers in Plant Science, 6: 49, doi: https://doi:org/10.3389/fpls.2015.00049, 13 p.

Ussher J.R., Lopaschuk G.D., Arduini A. 2013. Gut microbiota metabolism of L-carnitine and cardiovascular risk. Atherosclerosis, 231, 2: 456-461

Vandeputte D., Falony G., Vieira-Silva S., Tito R.Y., Joossens M., Raes J. 2016. Stool consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial growth rates. Gut, 65, 1: 57-62

Vignoli A., Ghini V., Meoni G., Licari C., Takis P.G., Tenori L., Turano P., Luchinat C. 2019. High-throughput metabolomics by 1D NMR. Angewandte Chemie (International Ed. in English), 58, 4: 968-994

Volkova S., Matos M.R.A., Mattanovich M., Marín De Mas I. 2020. Metabolic modelling as a framework for metabolomics data integration and analysis. Metabolites, 10, 8:303, doi: 10.3390/metabo10080303, 27 p.

Vrijlandt E.J., Gerritsen J., Boezen H.M., Grevink R.G., Duiverman E.J. 2006. Lung function and exercise capacity in young adults born prematurely. American Journal of Respiratory and Critical Care Medicine, 173, 8: 890-896

Wang Q., Wang K., Wu W., Giannoulatou E., Ho J.W.K., Li L. 2019. Host and microbiome multi-omics integration: applications and methodologies. Biophysical Reviews, 11, 1: 55-65

Wang T.J., Larson M.G., Vasan R.S., Cheng S., Rhee E.P., Mccabe E., Lewis G.D., Fox C.S., Jacques P.F., Fernandez C., O'donnell C.J., Carr S.A., Mootha V.K., Florez J.C., Souza A., Melander O., Clish C.B., Gerszten R.E. 2011a. Metabolite profiles and the risk of developing diabetes. Nature Medicine 17, 4: 448-453

Wang Y., Lê Cao K.A. 2020. Managing batch effects in microbiome data. Briefings in Bioinformatics, 21, 6: 1954-1970

Wang Z., Klipfell E., Bennett B.J., Koeth R., Levison B.S., Dugar B., Feldstein A.E., Britt E.B., Fu X., Chung Y.M., Wu Y., Schauer P., Smith J.D., Allayee H., Tang W.H., Didonato J.A., Lusis A.J., Hazen S.L. 2011b. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. Nature, 472, 7341: 57-63

Waring J., Lindvall C., Umeton R. 2020. Automated machine learning: Review of the state-of-the-art and opportunities for healthcare. Artificial Intelligence in Medicine, 104: 101822, doi: 10.1016/j.artmed.2020.101822, 12 p.

Wemheuer F., Taylor J.A., Daniel R., Johnston E., Meinicke P., Thomas T., Wemheuer B. 2020. Tax4Fun2: prediction of habitat-specific functional profiles and functional redundancy based on 16S rRNA gene sequences. Environmental Microbiome, 15, 1: 11, doi: 10.1186/s40793-020-00358-7, 12 p.

Whipps J.M., Lewis K., Cooke R.C. 1988. Mycoparasitism and plant disease control. In: Fungi in Biological Control Systems. Burge N.M. (ed). Manchester, Manchester University Press: 161-187

Wilken B., Ramirez J.M., Richter D.W., Hanefeld F. 2022. The response to hypoxia is affected by creatine in the central respiratory network of mammals 251. Pediatric Research, 40, 3: 557-557

Wilkinson J.E., Franzosa E.A., Everett C., Li C., Trainees H.R.A., Investigators H., Hu F.B., Wirth D.F., Song M., Chan A.T., Rimm E., Garrett W.S., Huttenhower C. 2021. A framework for microbiome science in public health. Nature Medicine, 27, 5: 766-774

Wirbel J., Pyl P.T., Kartal E., Zych K., Kashani A., Milanese A., Fleck J.S., Voigt A.Y., Palleja A., Ponnudurai R., Sunagawa S., Coelho L.P., Schrotz-King P., Vogtmann E., Habermann N., Niméus E., Thomas A.M., Manghi P., Gandini S., Serrano D., Mizutani S., Shiroma H., Shiba S., Shibata T., Yachida S., Yamada T., Waldron L., Naccarati A., Segata N., Sinha R., Ulrich C.M., Brenner H., Arumugam M., Bork P., Zeller G. 2019. Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. Nature Medicine, 25, 4: 679-689

Wishart D.S. 2008. Quantitative metabolomics using NMR. Trac-Trends in Analytical Chemistry, 27, 3: 228-237

Wishart D.S. 2019. NMR metabolomics: A look ahead. Journal of Magnetic Resonance (San Diego, Calif.: 1997), 306: 155-161

Wishart, D.S., Feunang, Y.D., Marcu, A., Guo, A.C., Liang, K., Vazquez-Fresno, R., Sajed, T., Johnson, D., Li, C., Karu, N., Sayeeda, Z., Lo, E., Assempour, N., Berjanskii, M., Singhal, S., Arndt, D., Liang, Y., Badran, H., Grant, J., Serra-Cayuela, A., Liu, Y., Mandal, R., Neveu, V.,

Pon, A., Knox, C., Wilson, M., Manach, C., and Scalbert, A. 2018. HMDB 4.0: the human metabolome database for 2018. Nucleic Acids Research 46: D608-D617

Wishart D.S., Guo A., Oler E., Wang F., Anjum A., Peters H., Dizon R., Sayeeda Z., Tian S., Lee B.L., Berjanskii M., Mah R., Yamamoto M., Jovel J., Torres-Calzada C., Hiebert-Giesbrecht M., Lui V.W., Varshavi D., Varshavi D., Allen D., Arndt D., Khetarpal N., Sivakumaran A., Harford K., Sanford S., Yee K., Cao X., Budinski Z., Liigand J., Zhang L., Zheng J., Mandal R., Karu N., Dambrova M., Schiöth H.B., Greiner R., Gautam V. 2022. HMDB 5.0: the human metabolome database for 2022. Nucleic Acids Research, 50: D622-D631

Wishart D.S., Jewison T., Guo A.C., Wilson M., Knox C., Liu Y., Djoumbou Y., Mandal R., Aziat F., Dong E., Bouatra S., Sinelnikov I., Arndt D., Xia J., Liu P., Yallou F., Bjorndahl T., Perez-Pineiro R., Eisner R., Allen F., Neveu V., Greiner R., Scalbert A. 2013. HMDB 3.0--the human metabolome database in 2013. Nucleic Acids Research, 41: D801-D807

Wishart D.S., Tzur D., Knox C., Eisner R., Guo A.C., Young N., Cheng D., Jewell K., Arndt D., Sawhney S., Fung C., Nikolai L., Lewis M., Coutouly M.A., Forsythe I., Tang P., Shrivastava S., Jeroncic K., Stothard P., Amegbey G., Block D., Hau D.D., Wagner J., Miniaci J., Clements M., Gebremedhin M., Guo N., Zhang Y., Duggan G.E., Macinnis G.D., Weljie A.M., Dowlatabadi R., Bamforth F., Clive D., Greiner R., Li L., Marrie T., Sykes B.D., Vogel H.J., Querengesser, L. 2007. HMDB: the human metabolome database. Nucleic Acids Research, 35: D521-D526

Wold S., Sjostrom M., Eriksson L. 2001. PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems, 58, 2: 109-130

Wolfs T.G., Derikx J.P., Hodin C.M., Vanderlocht J., Driessen A., De Bruïne A.P., Bevins C.L., Lasitschka F., Gassler N., Van Gemert W.G., Buurman W.A. 2010. Localization of the lipopolysaccharide recognition complex in the human healthy and inflamed premature and adult gut. Inflammatory Bowel Diseases, 16, 1: 68-75

Wu Y.W. 2018. ezTree: an automated pipeline for identifying phylogenetic marker genes and inferring evolutionary relationships among uncultivated prokaryotic draft genomes. BMC Genomics, 19: 921, doi: 10.1186/s12864-017-4327-9, 10 p.

Wu Y.W., Simmons B.A., Singer, S.W. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics (Oxford, England), 32, 4: 605-607

Yang C., Chowdhury D., Zhang Z., Cheung W.K., Lu A., Bian Z., Zhang L. 2021. A review of computational tools for generating metagenome-assembled genomes from metagenomic sequencing data. Computational and Structural Biotechnology Journal, 19: 6301-6314

Yang Q., Wang Y., Zhang Y., Li F., Xia W., Zhou Y., Qiu Y., Li H., Zhu F. 2020. NOREVA: enhanced normalization and evaluation of time-course and multi-class metabolomic data. Nucleic Acids Research, 48: W436-W448

Yeo C.J.J., Darras B.T. 2020. Overturning the paradigm of spinal muscular atrophy as sust a motor neuron disease. Pediatric Neurology, 109: 12-19

Yu C.T., Chao B.N., Barajas R., Haznadar M., Maruvada P., Nicastro H.L., Ross S.A., Verma M., Rogers S., Zanetti K.A. 2022. An evaluation of the National Institutes of Health grants portfolio: identifying opportunities and challenges for multi-omics research that leverage metabolomics

data. Metabolomics: Official Journal of the Metabolomic Society, 18, 5: 29, doi: 10.1007/s11306-022-01878-8, 12 p.

Zerbino D.R., Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Research, 18, 5: 821-829

Zhang C., Yin A., Li H., Wang R., Wu G., Shen J., Zhang M., Wang L., Hou Y., Ouyang H., Zhang Y., Zheng Y., Wang J., Lv X., Wang Y., Zhang F., Zeng B., Li W., Yan F., Zhao Y., Pang X., Zhang X., Fu H., Chen F., Zhao N., Hamaker B.R., Bridgewater L.C., Weinkove D., Clement K., Dore J., Holmes E., Xiao H., Zhao G., Yang S., Bork P., Nicholson J.K., Wei H., Tang H., Zhao L. 2015. Dietary Modulation of gut microbiota contributes to alleviation of both genetic and simple obesity in children. EBioMedicine, 2, 8: 968-984

Zheng G., Price W.S. 2010. Solvent signal suppression in NMR. Progress in Nuclear Magnetic Resonance Spectroscopy, 56, 3: 267-288

Zheng X., Chen T., Zhao A., Ning Z., Kuang J., Wang S., You Y., Bao Y., Ma X., Yu H., Zhou J., Jiang M., Li M., Wang J., Ma X., Zhou S., Li Y., Ge K., Rajani C., Xie G., Hu C., Guo Y., Lu A., Jia W., Jia W. 2021. Hyocholic acid species as novel biomarkers for metabolic disorders. Nature Communications, 12, 1: 1487, doi: 10.1038/s41467-021-21744-w, 11 p.

Zhou G., Ewald J., Xia, J. 2021. OmicsAnalyst: a comprehensive web-based platform for visual analytics of multi-omics data. Nucleic Acids Research, 49: W476-W482

Zitnik M., Nguyen F., Wang B., Leskovec J., Goldenberg A., Hoffman M.M. 2019. Machine learning for integrating data in biology and medicine: principles, practice, and opportunities. An International Journal on Information Fusion, 50: 71-91

Ząbek A., Stanimirova I., Deja S., Barg W., Kowal A., Korzeniewska A., Orczyk-Pawiłowicz M., Baranowski D., Gdaniec Z., Jankowska R., Młynarz P. 2015. Fusion of the ^{1}H NMR data of serum, urine and exhaled breath condensate in order to discriminate chronic obstructive pulmonary disease and obstructive sleep apnea syndrome. Metabolomics: Official Journal of the Metabolomic Society, 11, 6: 1563-1574

ACKNOWLEDGEMENTS