DISSERTATION

Technical confounders in microbiome studies – examples from fecal and sputum microbiome research

Technische Störfaktoren in Mikrobiomstudien - Beispiele aus der Fäkal- und Sputum Mikrobiomforschung

zur Erlangung des akademischen Grades
Doctor of Philosophy (PhD)

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von

Theda Ulrike Patricia Bartolomaeus

Erstbetreuung: PD Dr. Enno Klussmann

Datum der Promotion: 28.02.2025

Table of contents

List of abbreviations

DNA: Deoxyribonucleic Acid

HIV: Human Immunodeficiency Virus

COPD: Chronic Obstructive Pulmonary Disease

UK: United Kingdom

PDE3A: Phosphodiesterase 3A

FFQs: Food-Frequency Questionnaires

PCR: Polymerase Chain Reaction

RNA: Ribonucleic Acid

16S rRNA: 16S Ribosomal RNA

EtOH: Ethanol

PLWH: People Living With HIV

ART: Antiviral therapy

FDR: False Discovery Rate

BH-FDR: Benjamini-Hochberg False Discovery Rate

URT: Upper Respiratory Tract

PCoA: Principal Coordinate Analysis

OTU: Operational Taxonomic Unit

NCBI: National Center for Biotechnology Information

RCTs: Randomized Clinical Trials

DEL: Functional Deletion PDE3A gene

WT: Wild Type

Δ3aa: Consecutive Active PDE3A gene

List of figures

Abstract

Technological advances have greatly enhanced our ability to study the microbiome and its interactions with the host on a genomic level. However, inconsistencies in wet lab protocols and analysis pipelines can impact the comparability and reproducibility of research findings, making it difficult to draw reliable conclusions and replicate experiments. Moreover, confounding factors such as diet, lifestyle, and antibiotic use hinder establishing links between microbiome features and disease signals.

Using different case studies for addressing technical and analytical problems, this dissertation underscores the critical importance of standardization, confounder analysis, and translational animal models in advancing our understanding of the microbiome and its implications for human health. This thesis employed qualitative and quantitative methods to address these challenges. These methods provided insights into the factors contributing to variability. Moreover, statistical tools were applied to disentangle populations and investigate the effects of clinical, lifestyle, and experimental factors on gut and sputum microbiota. Furthermore, the translatability of knowledge gained from human studies to animal models was demonstrated through the presented PDE3A mutations model.

Key findings from this research demonstrate that preanalytical factors, such as DNA extraction and collection protocols, can significantly influence the diversity and distribution of bacterial taxa. Extraction protocols emerged as a significant technical confounder akin to individual donor identity. Human read contamination and adapting pipelines for low biomass samples like sputum are also considered. A study on the sputum microbiome exemplifies the need for confounder-aware statistics in a clinical cohort, in which the effects of HIV and COPD status, antiviral treatment, medication, and lifestyle were disentangled. The developed protocols and applied tools hold potential for diverse study settings from preclinical to clinical.

The established animal model, combined with the acquired knowledge of preanalytical factors and confounder analysis, poses a robust platform for future research on the role of the microbiota in hypertension.

Zusammenfassung

Die jüngsten technologischen Fortschritte haben unsere Möglichkeiten zur Untersuchung der Wirts-Mikrobiom-Interaktionen erheblich verbessert. Unstimmigkeiten in den Arbeitsabläufen sowohl im Labor als auch bei der Analyse stellen jedoch Herausforderungen für Forscher*innen dar. Diese treten besonders im Hinblick auf die Vergleichbarkeit und Reproduzierbarkeit von Forschungsergebnissen auf. Zusätzlich erschweren Störfaktoren wie z.B. die Medikation der Patient*innen die verlässliche Identifikation von Zusammenhängen zwischen mikrobiellen Taxa und Krankheitssignalen.

Daher ist das Ziel dieser Dissertation, die Bedeutung der Standardisierung in der Mikrobiomforschung anhand von zwei unterschiedlichen Probenmatrizen (Feces und Sputum) hervorzuheben. Darüber hinaus wird in dieser Arbeit, das komplexe Thema der Störfaktoren in der Mikrobiomforschung behandelt. Durch eine Kombination qualitativer und quantitativer Methoden werden Einblicke in die Faktoren, die zur Variabilität im Mikrobiom beitragen, ermöglicht. Darüber hinaus bieten diese Methoden Instrumente, die zur Entflechtung der Faktoren genutzt werden können. Zuletzt wurde die Übertragbarkeit, der aus Humanstudien gewonnenen Erkenntnisse auf Tiermodelle anhand des vorgestellten PDE3A-Mutationsmodells demonstriert.

Die zentralen Ergebnisse dieser Dissertation zeigen, dass präanalytische Faktoren wie Extraktions- und Sammelprotokolle die Vielfalt und Verteilung der Bakterientaxa erheblich beeinflussen. Die Extraktionsprotokolle erwiesen sich als wichtige technische Störfaktoren, ähnlich der Identität der Patient*innen. Statistische Methoden zur Kontrolle dieser Störfaktoren erwies sich als wirksam, um die komplexen Einflüsse auf das Mikrobiom gesunder Personen zu entschlüsseln. Ein weiteres wichtiges Ergebnis zeigt, dass bei der Analyse von Proben mit geringer Biomasse, wie z. B. Sputum Kontamination durch menschliche DNA berücksichtigt werden müssen. Der Einsatz von statistischen Methoden zur Störfaktor Analyse erwies sich ebenfalls als essentiell um die Auswirkungen des HIV- und COPD-Status, der antiviralen Behandlung und der Medikation auf das Sputum Mikrobiome zu entflechten. Die entwickelten Protokolle haben das Potenzial, als universelle Methoden für unterschiedliche Fragestellungen eingesetzt zu werden, die sowohl präklinische als auch klinische Studien umfassen. Die in der Dissertation etablierten Tiermodelle bieten in Verbindung mit dem erworbenen Wissen über präanalytische Faktoren

und der Analyse von Störfaktoren eine solide Plattform für die künftige Erforschung des Bluthochdrucks. Es bietet die Möglichkeit, neue therapeutische Ansätze zu identifizieren und die komplizierten Beziehungen zwischen Wirtsgenetik, Umweltfaktoren und Mikrobiom zu erforschen.

Diese Dissertation zeigt somit die entscheidende Bedeutung von Standardisierung, Störfaktorenanalyse und translationalen Tiermodellen, um unser Verständnis der Wirts-Mikrobiom-Interaktionen voranzutreiben.

1 Introduction

The human microbiome is the complex and dynamic community of microorganisms like bacteria and fungi residing on and in human tissues and biofluids along their corresponding anatomical side. Microbial cells are found in and on the human body in comparable numbers to human cells [1]. The human microbiota maintains a symbiotic relationship with its host by contributing to metabolic functions, protecting against pathogens, and training and developing the immune system [2,3]. Gut microbiota are not just passive bystanders but actively interact with their host, modulating host gene expression and immune function [4]. Many diseases are thus associated with microbial dysbiosis [3], including both communicable diseases such as tuberculosis [5] and syphilis [6], and non-communicable diseases like inflammatory bowel disease (Crohn's disease [7] and ulcerative colitis [8]), colorectal cancer [9], as well as other diseases such as obesity [10] and diabetes [11], atherosclerosis [12], or non-alcoholic steatohepatitis [13].

Due to the advances in sequencing technologies, microbiome characterization has drastically changed, leading to the rapid discovery of novel microbes associated with human health [14]. The inception of metagenomic studies can be traced back to environmental research, while clinical research, in its early stages, predominantly concentrated on exploring the gut microbiome through stool sampling [15]. Researchers have started studying other microbial communities and their associations with the human host; for example, the sputum microbiome, which is increasingly relevant in respiratory diseases such as COPD [16] and asthma [17]. Sputum is a secretion produced in the lungs and bronchi and is considered an acceptable surrogate for characterizing the mucosal surface microbiota of the URT [18]. In this relatively recent field of research, standardization is required to define microbiome dysbiosis and its association with diseases [19]. Reliability and comparability across studies form significant barriers that ultimately hamper the development of novel microbiome-centered therapeutic approaches [20,21].

Preanalytical factors refer to the collection, processing, and storage methods of biological samples before sequencing, preprocessing and statistical analysis, which have been shown to significantly impact the microbial community [20,22]. The sample processing method, DNA extraction, can influence the microbiome composition to such an extent that it mimics biologically relevant signals [23]. Moreover, researchers have shown that the DNA extraction method can profoundly affect detected diversity, leading to problematic interpretations [23]. The collection method, however, seems to impact microbial diversity to a

lesser extent [22]. Deviations in the processing of sequencing raw reads, finally, are easier to manage as these processes are less time and resource-intensive [24]. However, when comparing data sets, it is essential to remember that all data should be reprocessed using the same pipelines and annotation tools [25,26].

The standardization and analysis of the sputum microbiome are challenging for two reasons. First, due to low biomass and subsequently more host contamination, considerable background noise can be detected in sequencing data and requires adjustment of the analytical methods [27,28]. Second, the lung is exposed to different environmental factors, and the sputum contains various aerobic microorganisms, starkly contrasting with the gut. The exposure of the lung to air and its role as a small interface between air and blood fluid cause special conditions. The mucous membrane of the respiratory tract surfaces must have a low bacterial load to facilitate gas exchange [19]. Mechanisms such as mucociliary clearance, coughing, and host immune response effectively minimize bacterial load in the respiratory tract [29]. These facts and surface-specific nature concerning multiple environmental factors, such as oxygen partial pressure, protein content, and surface-active agents, result in a highly active and dynamic respiratory microbiome [19]. This dynamic situation in the respiratory tract, the periodic subjection to environmental microbes followed by their rapid elimination, explains the result of several cross-sectional studies in humans showing that traces of microbial DNA in all sections of the respiratory tract as well as the oral cavity are typical in the healthy respiratory microbiome [19,29–31]. Analogous to the core microbiome in the gut, several core genera characterizing the URT microbiome are defined and can be found in healthy populations [32]. The most abundant genera include *Staphylococcus*, *Streptococcus*, *Corynebacterium*, *Prevotella*, *Veillonella*, *Propionibacterium*, *Fusobacterium*, and *Moraxella* [32,33]. This healthy core microbiome is significantly altered by smoking [34], antibiotic use [35], and chronic lung diseases [19]. Defining and understanding the composition and function of the sputum microbiome as an easy assessable substitute for the URT microbiota is crucial for developing targeted treatments and interventions to improve respiratory health [36,37].

Sequencing technologies and biostatistical methods have enabled us to disentangle different factors shaping the gut microbiome compositions, such as diet, medication, and lifestyle, from those associated with pathogenic states [3,38]. However, these are, of course, limited by the quality of data produced. Thus, much effort has been made to standardize experimental protocols [23].

Confounder awareness is an emerging challenge in research, wherein confounders within this context pertain to covariates exhibiting correlations with the examined variables [39]. To address this, multivariable analysis or matching can be used to adjust for confounding variables between a hypothesized exposure and outcome [40]. For instance, matching disease subjects to controls based on confounding variables contributing to microbiota heterogeneity reduces observed differences [39,40]. However, this approach requires knowledge of these variables, risking the possibility of overlooking potential confounders [41]. In RCTs, randomization helps balance potential confounders between groups [41]. In observational studies, randomization doesn't work; thus, post-hoc statistical methods like regressions are needed to address confounders effectively and isolate the hypothesized exposure effects [40,41].

In a broader sense, more than this method is needed when we compare different cohorts from different studies since post-hoc matching is usually impossible to realize [42]. DNA extraction methods can significantly impact observed microbiome variability and are therefore considered technical confounders [23]. Therefore, the need to quantify and examine the role of such variables in microbiome studies is crucial in understanding variables that may lead to different outcomes in outwardly similar study designs [43,44].

Moreover, with emerging new fields of microbiomes to be investigated, we need to account for their uniqueness.

In summary, microbiome researchers face two significant challenges - comparability and reproducibility - leading to weak gut and sputum microbiome research robustness if not managed correctly. By carefully describing and addressing these challenges, the knowledge gained from these two fields of microbiome research can be expanded even further. This is why I focused my research on developing standardization of methods and using statistics to identify confounding factors and human variability in microbiome research. These standardizations and statistical methods are essential to improve the accuracy of microbiome studies so that comparability and reproducibility of the studies are established and the robustness of microbiome research increases. To achieve this, I focused on the confounder-aware statistical analysis and applied it in three different contexts of microbiome research. In the following sections of this dissertation, I first present the methods used to obtain my data, then explore the results, and finally discuss the findings in the context of current research.

This work addresses the causes of variability between microbiome studies to develop a consensus on conducting microbiome studies in various settings. To this end, I combine

qualitative and quantitative methods to provide insights into factors contributing to this variability. I present methods to disentangle populations and examine the effects of clinical, lifestyle, and preanalytical factors on the microbiome of two different compartments, the gut, and the sputum microbiome. Furthermore, I will explore how animal models could provide a powerful tool to generate fresh hypotheses and treatment strategies for hypertension and hypertension-associated end-organ damage. The presented work explores the potential of the developed protocols and tools for a universal method for different study settings for both preclinical and clinical studies. It is outlined in three papers, tracing the evolution from clinical to preclinical study designs.

1. Bartolomaeus TUP et al. 2020. Quantifying technical confounders in microbiome studies [43]. Hereafter called **Study 1**.
2. Kayongo A*, Bartolomaeus TUP*, Birkner T* et al. 2023. Sputum microbiome and chronic obstructive pulmonary disease in a rural Ugandan cohort of well-controlled HIV infection [37]. Hereafter called **Study 2**.
3. Ecru M, …. Bartolomaeus TUP et al. 2022. Mutant Phosphodiesterase 3A Protects From Hypertension-Induced Cardiac Damage [45]. Hereafter called **Study 3**

The first project is a clinical observational cohort study on the gut microbiome using different preanalytical methods. This project constitutes the baseline of my work since the main concepts have been developed during this project and applied to the second and third projects. The second project is a clinical study on the sputum microbiome, building on the knowledge gained from the gut microbiome project and using confounder-aware statistics first described in **Study 1** [43]. The third project is preclinical studies on PDE3A mutated rats, again building on the knowledge gained from the abovementioned studies, transferring the workflows and statistics from human to animal studies. In the following, I present the wet lab methods of all three projects and then show how I applied statistical models to the clinical and preclinical data.

2 Methods

2.1 Experimental Lab Methods

In the Wet Lab methods part, I describe the methods used in all three studies: stool and sputum collection, DNA extraction, 16S sequencing, clinical metadata collection like HIV and COPD assessment, continuous blood pressure measurement, and gene expression analysis.

2.1.1 Human studies preanalytical factors (**Study 1** [43])

Study 1 [43] was designed as an observative cohort study to elucidate the effect of different preanalytical methods on the gut microbiome composition and describes the baseline of my research journey since the knowledge gained is used in all following projects. The study was conducted in healthy subjects. By focusing on the gut microbiome, we could ensure that already established methods for sequencing and processing samples would also keep the variation low.

The Charité – Universitätsmedizin Berlin institutional review board approved the observational cohort study. The study was registered on clinicaltrials.gov (NCT04082221) [43]. Written informed consent was obtained before entering the study. Subject recruitment was performed from Campus Berlin-Buch of Charité – Universitätsmedizin Berlin and Max Delbrück Center for Molecular Medicine [43]. Inclusion, exclusion criteria, and subject screening were performed as described in **Study 1** [43].

<u>Specimen collection</u>

Stool sample collection and storage were performed using six different methods as described in **Study 1** [43]: immediate freezing (faeces tube, SARSTEDT) with Dry Ice (naïve sample), OMNIgene GUT (DNA Genotek), 95% EtOH, (Carl Roth GmbH & Co.), RNALater (MERCK), Zymo DNA/RNA Shield (Zymo Research) and Zymo DNA/RNA Shield (Zymo Research) with Dry Ice.

<u>DNA extraction and 16S Sequencing:</u>

DNA extraction using the ZymoBIOMICS DNA Miniprep Kit (Zymo Research) and the QIAamp PowerFecal DNA Kit (Qiagen), followed by 16S sequencing using Illumina MiSeq targeting the V3-V4 region (300-bp read length) was performed as described in **Study 1** [43].

2.1.2 Sputum Microbiome (**Study 2** [37])

Study 2 [37] is characterized by more complex data, allowing me to challenge my approaches with data from low biomass microbiota, intending to challenge the analytical pipelines developed to analyze the gut microbiome and disentangle confounding effects between microbial features and clinical patient data. I applied these pipelines to study the sputum microbiome in a cross-sectional clinical cohort with underlying comorbidities (HIV and/or COPD) and medication regimes.

This cross-sectional study was approved by Mulago Hospital Research and Ethics Committee (MHREC) (no.1996) and Uganda National Council for Science and Technology (UNCST) (no. HS2375) in Kampala, Uganda [37].

Recruitment and examination of subjects as well as sample collection were performed as described in **Study 2** [37] summarized as follows:

HIV assessment

Subjects were retested for HIV using the ART clinical guidelines in Uganda [37]. Their medical history was recorded, including viral load, ART drug use, treatment with Cotrimoxazole, and occurrence of opportunistic infections or tuberculosis [37]. Furthermore, sociodemographic and other clinical characteristics were collected [37].

Sputum induction and sample collection

Before sputum induction, the subjects underwent a three-step cleansing routine using sterile water to minimize contamination with saliva or postnasal drip [37]. Each sputum sample was collected through deep coughing and expectoration [37]. If the collection procedure and sample quality were sufficient, the sputum samples were transferred into the DNA collection, preservation, and isolation kit (Norgen Biotek Corporation) [37].

DNA extraction and 16S sequencing

DNA extraction and 16S sequencing using Illumina MiSeq targeting the V3-V4 region (300-bp read length) were performed as described in **Study 2** [37]

2.1.3 Animal Experiments (**Study 3** [45])

Transgenic animal models are powerful tools for clinical research, enabling researchers to study individual genes in their complex natural environment with lower confounding factors than the human model [46,47]. In this project, we used a rat model with human

PDE3A mutations that exhibit hypertensive phenotypes to demonstrate the universal applicability of our analytical pipelines. Furthermore, this introduced animal model provides a tool to generate fresh hypotheses and treatment strategies for hypertension and hypertension-associated end-organ damage [48].

The State of Berlin approved the rat studies (license no. G 0435/17) [49]. The rat model with a PDE3A mutation similar to the human T445N mutation was engineered by pronuclear microinjection of Sprague-Dawley rat zygotes using CRISPR-Cas9 as described by Ercu et al. 2020 [49]. Isoproterenol challenge was performed using age-matched animals described in Study 3 [45]. Continuous telemetry blood pressure measurements, echocardiography, tomography, and RNA sequencing identified phenotypes of PDE3A performed as described in **Study 3** [45].

Continuous telemetric blood pressure monitoring

Telemetric blood pressure measurement was performed at least one week after recovery from pressure-sensing device implantation, as described by Ercu et al. 2020 [49]. The DATAQUEST software recorded blood pressure and heart rate (A.R.T. 2.1, Data Sciences International) [49]. Signals were recorded continuously at 5 min intervals for 10 s, day and night, in freely moving animals [49]. An initial echocardiographic examination was conducted after at least six days of basal blood pressure measurements [49].

Gene expression analysis

Total RNA was extracted from the kidney, heart and aorta tissue using the RNeasy Mini Kit (Qiagen). Frozen tissue (about 20mg) was put into a tube containing Qiazol (Qiagen) and six zirconium beads 2.8mm in diameter (Bertin Corp.) and homogenized on Precellys 24 tissue homogenizer (Bertin Corp.). The following steps differ from the manufacturer's standard protocol. To eliminate contaminations caused by remaining DNA molecules, enzymatic degradation of DNA was performed using the RNase-Free DNase Set (Qiagen). The cleaned RNA was eluted using 30µL of RNase-free water without repetition. RNA concentration was measured for the reverse transcription into cDNA—high Sensitivity RNA-Bioanalyzer (Agilent, 5067-5581). The transcription was done with a High-Capacity cDNA Transcription kit (Applied Biosystems) according to the manufacturer protocol [50]. The qPCR was performed using TaqMan Fast Universal PCR Master Mix (Applied Biosystems) in QuantStudio 3 Real-Time PCR System (Applied Biosystems) for the target

genes. A total of 45 amplification cycles were used as a cut-off. The relative gene expression was normalized to the 18S gene.

2.2 Biostatistics

In the Biostatistical methods part, I describe the methods used in all three studies.

2.2.1 Power calculation

Making power calculation a critical step in the study design process, two tools were used to determine the optimal sample size to ensure adequate power to detect statistical significance. The optimal sample size was determined a priori for **Study 1** [43]. The power calculation was performed using G*power as described in **Study 1** [43]. Previously published results of collection methods on alpha-div indices were used to estimate the effect size and calculate the required sample size plus a 15% drop-out rate [43,51].

Post-hoc power calculation for **Study 3** [37] was performed using the R package pwr (version 1.3-0) [52], which yielded a probability of 77% for determining the significance of each major respiratory bacteria [37].

2.2.2 Statistics and pipelines

Statistical analysis allows researchers to make sense of the data collected, draw valid conclusions, identify patterns and relationships, and make predictions or generalizations about populations [53]. R programming is a popular language for data analysis and statistical computing. It provides various packages and functions for data manipulation, visualization, and modelling. R (R 4.0.0–4.2.3) was used for the statistical analysis and visualizations. The rtk package (version 0.2.6.1) [54] for alpha diversity measurements, the vegan package (version 2.6-2) [55] for beta diversity and the permutational multivariate analysis of variance (PERMANOVA) [56].

Univariate analysis was performed using the confounder-aware statistical tool developed for this purpose (metadeconfoundR, version 0.2.9) [43,57]. The metadeconfoundR pipeline operates as follows: initial analyses involve testing naïve associations of each individual feature and covariate using rank-based tests. Subsequently, the relative redundancies between covariate associations to the same feature are modelled through nested linear likelihood ratio tests. Finally, univariate associations with clinical variables undergo Ben-

jamini-Hochberg adjustment for an FDR of 0.1 in cases where clear confounders are absent [37,43]. Major taxa and OTUs detected in at least 10% of samples after rarefaction were included [37]. As data were not normally distributed, non-parametric tests were used for all association tests [37,43]. For discrete variables, Wilcoxon test was applied [37,43]. A nonparametric Spearman correlation test was used for pairs of continuous variables [37,43]. For **Study 1** [43] Enterotype assignment was done using the online tool https://enterotypes.org/ [43]. Hierarchical clustering using the R package "Dirichlet multinomial" (version 1.36.0) [58] was applied to create patterns of the different study samples in **Study 2** [37] Telemetry's continuous blood pressure data were analyzed as described by Ercu et al. 2020 [49] by building mixed-effects models using the lme4 R package (version 1.1-29) [59]. The analysis included random effects for animal identity and time points within a day and fixed effects for genotype, animal sex, and day/night status at the time of measurement [49]. The model parameters (slope and intercept) were used to determine the scope of the effects [49]. The significance of each factor was determined by comparing the full model to a simpler model that omitted genotype as a predictor [49]. Likelihood ratio tests were performed using the lmtest R package (version 0.9-40) [60]. 1-way ANOVA, corrected for multiple testing using Tukey's test, was used to identify gene expression differences between the tested groups [45].

2.2.3 Bioinformatic preprocessing of 16S data

16S amplicon sequencing is a powerful, cost- and labor-efficient tool to characterize bacterial communities in different biosamples [61]. This approach has been used for taxonomic profiling in stool samples for decades. However, for low biomass samples (e.g., sputum), it has been shown that misamplification can lead to spurious results [62]. To avoid misamplification due to higher amounts of human contamination, slight modifications to standard bioinformatic preprocessing were made. This approach was established during this project and is now included in our day-to-day preprocessing of 16S data.

Human contamination filtering was performed as described in **Study 2** [37]**,** briefly raw sequences were processed to remove potential human contamination via masking the human genome (https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000001405.39/) with the microbial genome database Pro-Genomes2 (https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkz1002/5606617/) [37]. Through masking and filtering, raw reads were matched to the human genome and discarded at 95% [37]. Any human reads found

were validated through alignment with the NCBI nucleotide database, yielding only human hits [37]. Cleaned raw sputum, reads were processed using LotuS (1.62) [37,63]. Briefly, Poisson binomial model-based [64] read filtering followed by OTU clustering (UPARSE) [65] based on sequence similarity of 97% was performed while SILVA [66] was used stepwise as a database for taxonomic assignment using a taxonomic lambda similarity search [37].

The fecal raw reads were processed analogously. In addition to SILVA [66], GreenGenes [67] and HITdb47 [68] were incrementally used as databases for taxonomic assignment [43]. Classification (genus threshold at 95% identity) was performed using a custom Perl script, where unassigned taxonomic levels were matched to the last known taxonomic rank and sequentially numbered [37]. Unavoidable differences in library size (number of fragmented reads) between samples were corrected by rarefaction [37,43]. To obtain relative microbiota abundances, OTU counts were randomly excluded and rarefied to the smallest remaining sample size (stool: 1350 -, sputum: 8278 - raw reads) [37,43].

2.2.4 Sputum cohort comparisons

We linked our sputum microbiome data to a comparable cohort in the UK [69]. The UK [69] cohort comprised 64 PLWH (median blood CD4 count, 676 cells/mL) and 38 HIV-negative subjects [69]. Subject inclusion and exclusion criteria, sample handling, processing and 16S sequencing using Illumina MiSeq were performed as described by Rofael et. al 2020 [69]. The raw data were processed along with the Ugandan samples to minimize processing bias [37].

To summarize, while conducting the three studies in both preclinical and clinical designs, in the first step, I applied various wet lab methods ranging from preanalytical methods, sample collection, sample processing, and gene expression analysis to gain data. In the second step, I applied statistical models such as power calculation, bioinformatic analysis of 16S data, and mixed effect models to the clinical and preclinical data gained from the wet lab. These studies have led to numerous results that I highlight in the following section.

3 Results

Researchers generally face two critical challenges: comparability and reproducibility. These two challenges may lead to weak microbiome research robustness. This work aims to carefully describe and address the causes of weak comparability and reproducibility, to develop standardization of methods, and to use statistics to identify confounding factors and human variability in microbiome research.

I have described wet lab and biostatistical methods to identify confounding factors and human variability in microbiome research in three different study settings, from preclinical to clinical. In this section, I describe the findings of the presented studies. I first established the intestinal microbiota and preanalytical factors. I secondly highlighted low-biomass research and the effects of disease signatures and covariates on the sputum microbiota. Third, I present preclinical research in an animal model for PDE3A hyperactivation.

3.1 Intestinal microbiota and preanalytical factors

Study 1 [43] determined the effects of preanalytical parameters (extraction and collection protocols) and all available biological variables (diet, lifestyle, and socio-demographics) on the gut microbiome signature (graphical abstract for **Study 1** [43] in Figure 1). The study resulted in three key findings. (1) The collection method has an impact but is less severe. (2) The extraction method is a significant technical confounder similar to individual donor identity. (3) Firstly introduced, the confounder aware analysis, shows the massive effect of technical factors. In this study, I could provide strong evidence for the need to carefully consider preanalytical variables and standardize methods to ensure generalizability with other clinical studies when conducting and comparing microbiome studies.

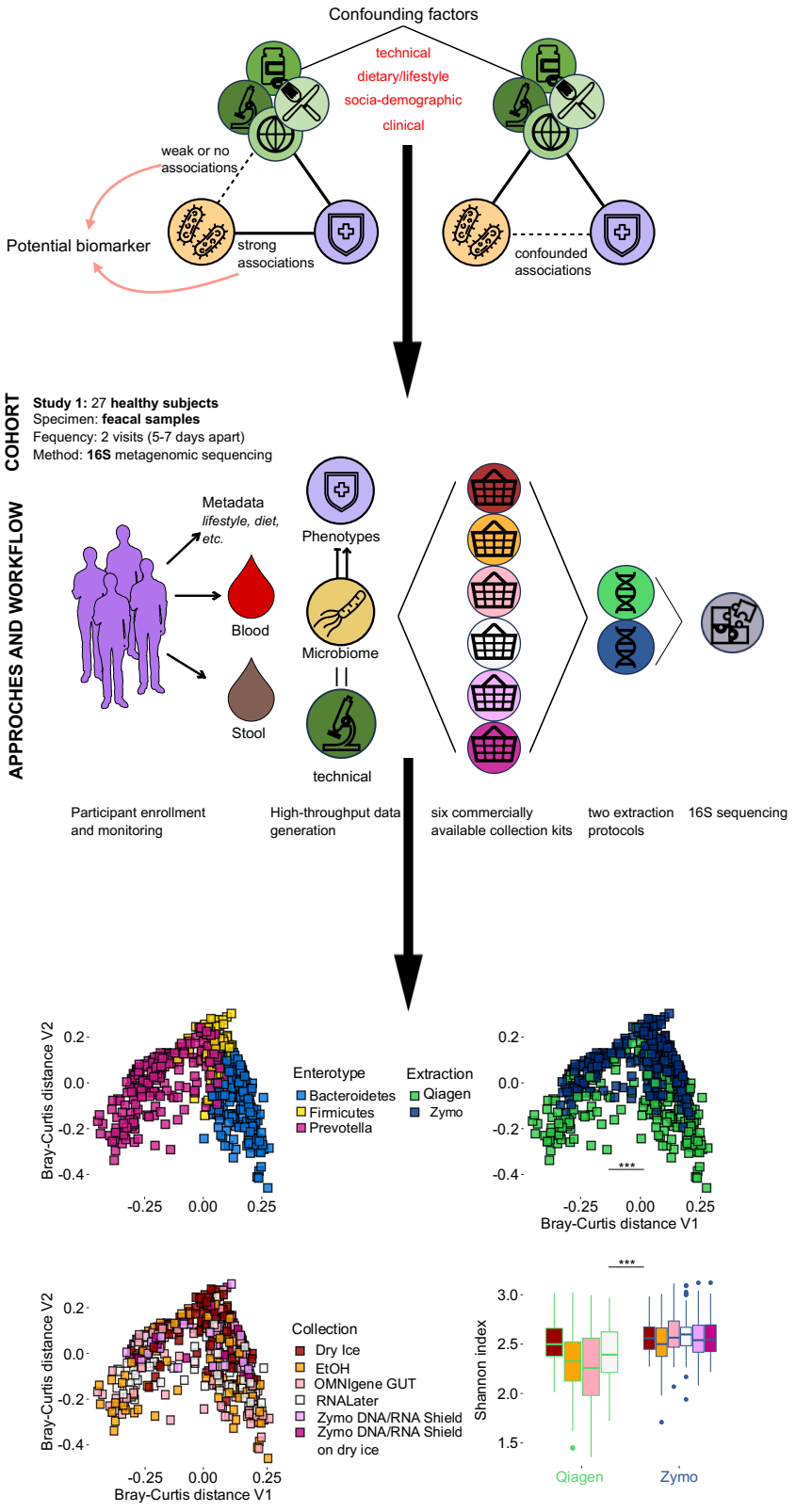

Figure 1 Graphical abstract Study 1 [43] (created by the author, TUP Bartolomaeus): Impact of preanalytical methods on fecal amplicon sequencing results. Healthy volunteers provided fecal samples using six collection methods, and DNA was isolated using two different commercial kits. Amplicon sequencing revealed a significant and profound impact of DNA isolation on the results.

Collection methods also exhibited an effect, although less pronounced. Moreover, DNA isolation effects mimicked the impact of clinical factors.

3.1.1 Alpha and beta diversity analysis

Typical microbiome analysis pipelines were applied to ensure generalizability with other clinical microbiome studies. We calculated alpha diversity using the Shannon index to measure species diversity by accounting for the number of different species and their relative abundance [70]. The Shannon indices for the extraction and collection methods were compared using a pairwise Wilcoxon test corrected for multiple testing using Benjamini-Hochberg adjustment. Comparing both extraction methods, Zymo (n = 321) and Qiagen (n = 215), the Shannon index was significantly lower in all samples extracted with the Qiagen kit (q-value < 0.001, BH-FDR) [43]. For the collection method, we detected a significantly reduced Shannon index in all samples collected using EtOH (n = 108) (q-value < 0.01, BH-FDR) and OMNIgene Gut (n = 106) (q-value < 0.1, BH-FDR) kits compared to the dry ice gold standard (n = 108) and all other collection methods like RNALater (n = 108), Zymo Shield (n = 53) and Zymo Shield on Dry Ice (n = 53) [43]. Next, we applied beta diversity measurements by summarizing the high-dimensional data into a reduced dimensional space to assess differences between spatially distributed communities [37,70]. A dissimilarity matrix (Bray-Curtis distance) between the different preanalytical methods was calculated and visualized using PCoA. Both DNA kits form clusters and are significantly separate in the PCoA (PERMANOVA, p < 0.0005) [43]. We determined enterotypes for each subject and looked at how the different preanalytical methods affected the distribution of these enterotypes. The DNA extraction kit has a substantial impact, with the Qiagen kit showing a decrease in Firmicutes-dominant samples (8%) and an increase in Bacteroides-dominant samples (52%) and the Zymo kit displaying a more even distribution between these two enterotypes (Firmicutes 24% and Bacteroides 31%) [43]. We could observe a similar skew towards reduced Firmicutes-dominant samples with all collection methods compared to the Dry Ice gold standard (32%), particularly with the OMNIgene Gut (14%), RNALater (12%), and EtOH (13%) collection methods [43].

Furthermore, we used a linear model to analyze variance and determine the relationship between intersample distance and sample pair congruence [43]. We considered several factors: donor ID, DNA extraction protocol, sample collection protocol, donor sex, and time point [43]. Our analysis showed that donor ID and preanalytical protocols significantly predicted intersample distance (likelihood ratio test, p < 0.001) [43]. The collection method

accounted for only 0.6% of the variation, while donor ID explained 7.4% and extraction protocol explained 5.7% [43]. Most of the variability was due to transient fluctuations of the gut microbiome [43]. Our findings indicate that the extraction protocol is a significant technical confounder, having a similar impact as individual donor identity [43].

3.1.2 Univariate analysis

We introduced confounder-aware analysis to clarify if the preanalytical effect we had found in the diversity indices was reflected in the relative abundances of specific bacterial taxa. Univariate analysis was performed on all bacterial taxa present in > 0.9% of the samples [43]. Associations between microbial characteristics and metadata such as technical, socio-demographic, dietary, and clinical variables are often difficult to disentangle. To make a statement about the effects of these variables on bacterial taxa, these covariates must be separated. Significant associations at both the phylum and genus levels with the preanalytical methods were often the most robust signal observed. Associations between taxa and the two extraction kits showed substantial and significant enrichment in Bacteroidetes (q-value <0.001, BH-FDR) stains for all samples extracted with the Qiagen kit [43]. The Zymo kit, in turn, offers significant enrichment in Firmicutes strains like *Blautia* alongside others (q-value < 0.001, BH-FDR) [43]. Corresponding signals are observed at other taxonomic levels. These tendencies from Bacteroidetes to Firmicutes are observed to a lesser extent for RNALater, OMNIgene Gut, and EtOH collection protocols [43]. Along with the DNA extraction procedure and the stool collection protocol, the presence or absence of some taxa is also influenced by clinical, dietary, and socio-demographic factors such as age, BMI and plasma tocopherol levels [43].

Taken together, a difference in technical protocol between identical metagenome datasets will most likely result in an apparent difference in taxonomic composition, which could either exaggerate or mask typical biological effects (such as those linked to cardiovascular, metabolic, or inflammatory diseases), leading to false results or failure when trying to validate previous results [43].

In conclusion, when conducting and comparing microbiome studies, it is necessary to carefully consider preanalytical variables and standardize methods to ensure generalizability with other clinical studies.

3.2 Sputum microbiota, low-biomass research, and disease signatures

In the second paper [37], we expanded the study into other microbial communities and their associations with the human host (graphical abstract Figure 2). In this section, the sputum microbiome from a cohort of HIV and/or COPD-positive subjects was used to understand how complex multimorbidities in a human-treated setting impact reliability, interpretability, and reproduction of microbiome findings. We provide insight into the relevance of addressing the unique features of low biomass samples and the applicability of confounder-aware statistical tools to disentangle the effect of HIV and COPD status, antiviral treatment, medication, and socio-demographic variables on the sputum microbiome composition. This study identified human contamination in low biomass samples, and confounder-aware statistics showed collinearity with antiviral treatment/duration thereof. When disentangled, HIV status had a more profound impact than COPD. However, the COPD status is associated with a significantly increased abundance of *Selenomonas* and a significant decrease in *Gammaproteobacteria* [37]. When comparing to a similar cohort from the UK [69], we could identify a systematic difference between the Gram state of the bacteria – like seen in **Study 1** [43] (also here, the two sputum microbiome cohorts used two different extraction protocols).

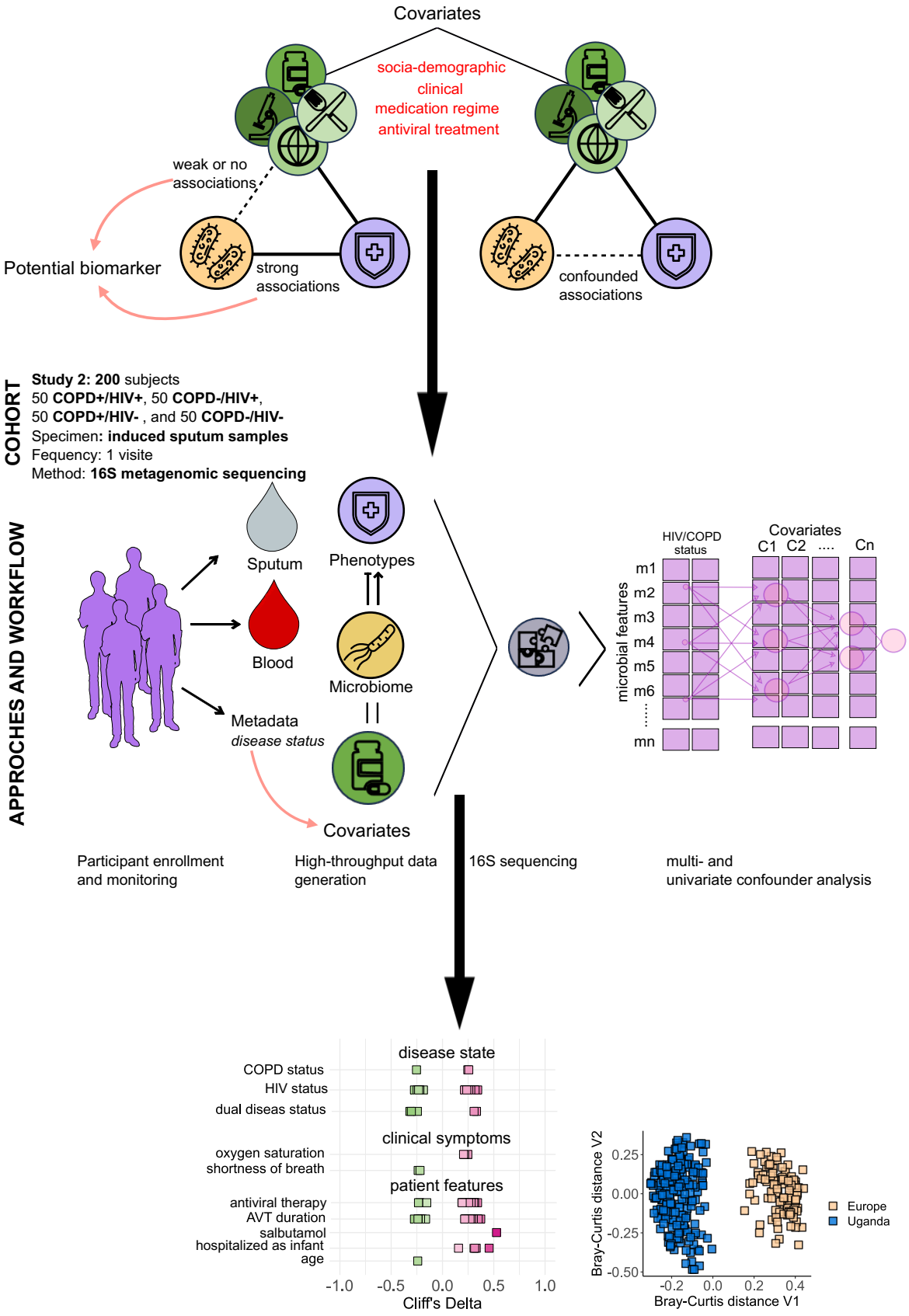

Figure 2: Graphical abstract Study 2 [37] (created by the author, TUP Bartolomaeus): Analysis of Sputum Microbiome in COPD and/or HIV positive Patients. Sputum microbiome from patients with COPD, HIV, or both analyzed by amplicon sequencing. Computational analysis revealed

features associated with disease, clinical symptoms, or patient features. Interestingly, we observed a strong effect of geographic location on the sputum microbiome (comparison of our Ugandan[37] versus published UK [69] cohort).

3.2.1 Human read contamination and multivariate analysis of the sputum microbiome

In **Study 2** [37], we found up to 40% of human read contamination in the sputum samples, with a mean contamination rate of 10%. Higher amounts of human read contamination were found in the COPD+ samples compared to the COPD- samples (q- value < 0.01 BH- FDR). After removing the human reads with a newly developed in-house Perl script, the impact of the disease status (HIV-/COPD-, HIV+/COPD+, HIV+/COPD- and HIV-/COPD+) on the sputum microbial composition was determined. Alpha diversity indices showed lower microbial richness and Chao1 index in the COPD+/HIV+ group compared to the COPD+/HIV- one (q-value < 0.01 BH-FDR) [37]. And an increase in microbial richness for the COPD-/HIV- group compared to the COPD+/HIV- participants (q-value < 0.01 BH-FDR) [37]. We observed the same increase in Chao1 index in the COPD+/HIV- group relative to the COPD-/HIV- and COPD-/HIV+ subjects, respectively.

3.2.2 Univariate analysis of factors influencing the sputum microbiome

Genus level comparison showed that *Streptococcus*, *Prevotella*, *Neisseria*, and *Veillonella* predominated in all cohort groups. Again, we applied univariate biomarker analysis using the metadeconfoundR package [43,57] for all taxa present in > 0.9% of the samples to assess their association with the collected patients' characteristics such as age, sex, BMI, socio-demographic features and clinical features like O2 saturation, systolic and diastolic blood pressure, as well as the use of medication while testing for confounding effects between these variables [37]. We could show that HIV status, its associated antiviral treatment, and the duration of such antiviral therapy display a similar pattern in influencing the sputum microbiome [37]. In general, HIV status affects bacterial composition at phylum and genus levels more than COPD status [37]. There has been a significant increase in the abundance of taxa belonging to the Firmicutes Phylum such as Clostridia (q-value = 0.02 BH-FDR), *Butyrivibrio* (q-value = 0.05 BH-FDR), Veilloneaceae (q-value = 0.05 BH-FDR), Erysipelotrichaceae (q-value = 0.006 BH-FDR) and *Solobacterium* (q-value = 0.006 BH-FDR) [37]. An increase in taxa belonging to the phylum Actinobacteria, namely Actinomyces (q-value = 0.05 BH-FDR) and *Atopobium* (q-value = 0.003 BH-FDR) [37]. Further, the taxa

Lentimicrobium (q-value = 0.07 BH-FDR), *Filifactor* (q-value = 0.007 BH-FDR), as well as *Pseudopropionibacterium* (q-value = 0.05 BH-FDR) were significantly decreased in association with HIV status [37].

In contrast, COPD status is associated with a significant reduction in Gammaproteobacteria (q-value = 0.09 BH-FDR) belonging to phylum Proteobacteria and a significantly increased abundance in *Selenomonas* (q-value = 0.09 BH-FDR), a member of the Firmicutes phylum [37].

A significant association was also found between COPD status and enriched Campylobacter, but this signal is determined by the "dual disease status" (COPD+/HIV+ group) [37]. More specifically, the COPD+/HIV+ comorbidity (dual disease status) showed a significant increase in Campylobacter (q-value = 0.04 BH-FDR) and a decrease in Staphylococcales (q-value = 0.01 BH-FDR) as well as Negativicutes (q-value = 0.03 BH-FDR) [37].

We identified other significant associations between the sputum microbiome and sociodemographic, clinical, and lifestyle characteristics [37]. Perhaps most interesting were the associations with the variable "hospitalization as an infant," the frequency of bacteria from the phylum Proteobacteria, namely Pseudomonas, and an increase in the frequency of *Corynebacterium* associated with the use of salbutamol [37]. Other host variables such as BMI, pulmonary tuberculosis, treatment with co-trimoxazole, prednisone, and smoking were not significantly associated with alterations in bacterial composition. Consequently, they were dropped from the model [37].

3.2.3 Cohort comparison with sputum data from the UK [69]

To validate the generalizability of our results, we compared our results with sputum microbiome data from a UK [69] HIV cohort [37]. Although the recruitment criteria and sequencing methods were similar, we observed differences in the preanalytical techniques [37]. To avoid further biases, computational preprocessing was performed on all samples together [37]. Interestingly, we detected significantly lower microbial richness (Shannon index) in the UK cohort [69] than in the Ugandan cohort [37](with q value < 0.0001 BH-FDR) [37].

Furthermore, beta diversity measurement revealed that both cohorts are building separate clusters (PERMANOVA $p < 0.001$) [37]. Applying univariate analysis on the combined dataset unveiled a systematic difference between the two cohorts regarding the Gram state of the bacteria [37]. The Ugandan cohort [37] shows significantly higher abundances in Gram-negative phyla like Proteobacteria, Fuscobacteria, and Bacteroides than the UK cohort [69].

3.3 PDE3A-mutant models for studying hypertension

As outlined in the method section, I provide insight into the universal applicability of the developed pipelines and statistical approaches in the two abovementioned projects. For this purpose, I will use these approaches to describe the effects of PDE3A mutations on hypertension in an animal model and show the translatability of the knowledge gained in human studies to unlock multi-omics studies in animal models [45,48,49].

3.3.1 PDE3A model generation and phenotyping

Consecutive active mutant PDE3A causes hypertension using novel mouse and rat models [49]. Using CRISPR-Cas9 engineering, we generated rat models with hyperactivity of the PDE3A Enzyme (Δ3aa) and its functional deletion (DEL) (graphical abstract Figure 3) [49]. The DEL rats fail to express the PDE3A enzyme. Telemetric blood pressure measurements were applied for long-term documentation. With the PDE3A project, we generated a new and promising hypertensive animal model, harboring new hypotheses and results in hypertensive-associated end-organ damage and in multi-omics effect of hypertension. Moving forward, this model is an established and controllable system that allows for systematic manipulation and study of microbiota. This system would allow implementation interventions such as administering probiotics, prebiotics, antibiotics, or fecal microbiota transplantation to test their effects on hypertension development and treatment in the future.

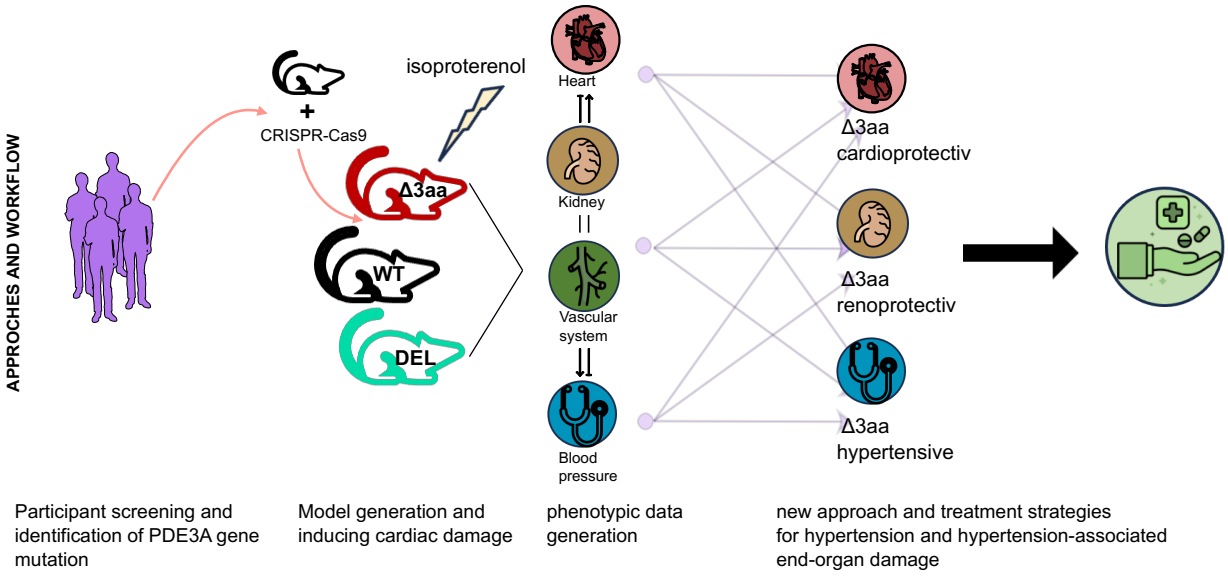

Figure *3*: Graphical abstract of Study 3 [37] (created by the author, TUP Bartolomaeus): Comparison of Rat Models with PDE3A Gene Mutations to WT littermates Rat models mimicking known human mutations in their PDE3A gene consecutive active (Δ3aa) and functional deletion (DEL),

were compared to WT rats. Despite exhibiting higher blood pressure, Δ3aa did not aggravate cardiovascular damage in these rats.

3.3.2 Effect of PDE3A mutation on Blood pressure and renal function

We found significant differences in mean systolic and diastolic blood pressure values between the tested groups [49]. The Δ3aa rats exhibited elevated systolic and diastolic blood pressure at 148/105 mm Hg, compared to 115/84 mm Hg for the DEL rats and 125/88 mm Hg in the WT control group [49]. Notably, the systolic blood and diastolic blood pressure differences between all groups were statistically significant [49]. To further evaluate the impact of the mutant PDE3A on renal function, we examined serum and urine parameters in our animal models [49]. Surprisingly, no proteinuria or disturbances in renal function were observed in the mutant (Δ3aa, DEL) animals [49]. Serum levels of albumin, creatinine, and urine/albumin creatinine ratios were not significantly different compared to WT animals [49].

Similarly, angiotensin-II levels, angiotensinogen, aldosterone, and plasma renin concentrations showed no significant differences between the control and mutant rats [49].

3.3.3 Cardioprotective effect of PDE3A mutation

To elucidate the potential cardioprotective effect of HTNB-causing Δ3aa mutations without antihypertensive treatment, we conducted experiments to determine if the mutant PDE3A could confer cardioprotection under the fused effects of preexisting hypertension and chronic β-adrenergic stimulation (isoproterenol) [45]. Isoproterenol typically induces cardiac damage, including hypertrophy in rats, by increasing heart rate and lowering blood pressure [45,71]. Interestingly, only Δ3aa rats showed a significant increase in cardiac output and fractional shortening, even though they had a similar heart rate to WT animals [45]. In contrast, the DEL animals exhibited a similar response to isoproterenol as the WT rats [45]. Furthermore, the expression of hypertrophy and fibrosis markers showed inconsistent results [45]. In summary, our findings indicate that the HTNB-causing Δ3aa mutant exerts a cardioprotective effect, as it does not exacerbate isoproterenol-induced cardiac damage as expected in the context of preexisting hypertension [45]. These two novel animal models provide valuable insights and offer a refined and potent approach for generating new hypotheses and treatment strategies for hypertension and hypertension-associated end-organ damage.

3.4 Result summary

In conclusion, my PhD thesis demonstrates the importance of considering potential pitfalls in microbiome analysis, including the impact of preanalytical experimental methods, sample type, and confounding factors to generalize results across independent studies. It thereby underscores the need for standardization of experimental workflows in microbiome research to ensure accurate and reliable results.

Moreover, we demonstrated that animal models are powerful tools in biomedical research. Aside from microbiome studies, we urgently need awareness of very different confounding factors when comparing human or animal studies. The combination of hypertensive animal models and microbiota manipulation provides a powerful platform for advancing our understanding of hypertension, identifying novel therapeutic approaches, and exploring the intricate relationship between host genetics, environmental factors, and the microbiome.

4 Discussion

In previous years, many microbiome studies have been conducted using various tools, packages, and algorithms to analyze microbiome data [14,72]. This causes issues with reproducibility and considerable variability between studies [72,73]. To solve these challenges this thesis aims to increase awareness of the importance of standardization in research and to propose standardizing suitable methods, particularly in studying the microbiome. To this end, I conducted three studies ranging from preclinical to clinical. This work combined wet lab and statistical models to demonstrate the importance of a systematic analysis of confounding factors that allows disentangling the effects of disease status, treatment, medications, and socio-demographics on the gut and sputum microbiome. Finally, this work demonstrates the utility of these tools in various study settings, from humans to rats, indicating the universal use of standardized workflows and tools. These findings will be discussed in the following section.

4.1 Short summary of results

This works findings underscore the three main goals of my dissertation, i.e., (1) to raise awareness of the importance of standardization when conducting research in the microbiome, (2) the need to thoroughly plan unique sample matrices and adapt pipelines, and (3) to shed light on the perplexing issue of co-occurring factors in microbiome research, since many known and unknown factors influence this ecosystem.

To illustrate the benefit of standardization, I showed that DNA extraction methods strongly and significantly impact the observed microbiome composition. However, the variance introduced by using the commercial kits is lower than those introduced by inadequate dry ice-based sampling, given that handling the former is much easier for patients.

The need to plan and adapt pipelines is shown in the case of contamination with host reads in low biomass sputum samples. Such contamination results in amplified host DNA sequences clustered and falsely assigned to bacterial taxa during the preprocessing, leading to false microbiome signatures of various health and nutritional factors. By systematic confounder analysis, we demonstrated that it is possible to disentangle the effects of HIV status, antiviral treatment, medication, and socio-demographics on the sputum microbiome.

Finally, I provide insight into a model with consecutive active mutant PDE3A that causes hypertension, showing that this model is a powerful tool for generating new hypotheses and treatment strategies for hypertension and hypertension-associated end-organ damage.

4.2 Interpretation of results

By addressing the above-named issues in microbiome research, i.e., standardization, uniqueness of sample matrices, and confounding factors, this dissertation offers three main findings:

The results of **Study 1** [43] underscore that preanalytical methods have a dominant effect on the detected microbial signals. To this point, sample collection using immediate freezing (Dry Ice) without preservatives was and continues to be widely used as a benchmark method for microbiome analysis because it preserves microbiome samples at a composition nearly identical to that of a fresh sample while avoiding any potential influence of preservative solutions [22,43,74]. However, alternative sample collection methods that do not require freezing can be informative for gut microbiome analysis [22]. Differences seen in the multivariate and univariate analyses between sample collection methods might be explained by the effect of buffer solutions used to preserve genetic integrity and expression profiles of samples [75,76]. Some taxa are more sensitive to changes and collection and storage conditions [22,74,77–79]. This skew in sensitivity is likely due to differences in the cellular composition of Gram-positive and Gram-negative bacteria, resulting in differences in their ability to preserve their DNA [78,80,81]. Freezing may improve DNA extraction from some Gram-positive genera [43,81]. This could explain why samples collected on Dry Ice show a slightly higher proportion of samples classified into the Firmicutes enterotype than other collection methods [43]. Overall, in line with the current research, the results suggest that on-market kits for microbial DNA stabilization are suitable for metagenomic analyses, especially for large epidemiological applications of microbiome profiling, and are easy to handle for study participants [43]. However, the chemistry of existing stabilizing solutions can impact the recovery of microbial genomic content, resulting in a source of bias [79]. Therefore, it is vital to identify the best-performing DNA stabilizers for a given ecosystem and to account for this when comparing different studies [79]. Extraction and collection protocols significantly impact the diversity and distribution of bacterial taxa. Of all factors quantified in the first study, variation in DNA extraction protocol has the most

considerable effect on the observed microbiome next to interindividual differences. This effect may involve several mechanisms, but the primary factors are likely mechanical lysis and bead-beating [21,82–85]. Mechanical lysis primarily impacts Gram-positive bacteria yield [21,82]. Their cell walls have significantly higher mechanical strength [21,82,86,87]. They are harder to break down, perhaps leading to an overestimate of Bacteroidetes over Firmicutes, an underestimate of alpha diversity, and a resulting skew in enterotype assignment in the presented study [43]. This skew mimics a microbiome similar to that more often found in obese or insulin-resistant subjects compared to healthy controls [11] and is thus a primary potential source of error in any cross-cohort comparative study [43]. Within this study, we first applied confounder-aware analysis to separate associations between microbial taxa and metadata, revealing that technical protocols substantially impact bacterial taxa composition.

Nevertheless, our univariate analysis discovered that diet significantly influences several bacterial genera. Our findings strongly overlap with recent research linking dietary fat intake to insulin resistance via the gut microbiome [88]. The findings highlight the need to carefully consider preanalytical methods when designing microbiome studies, especially those examining populations with altered microbiomes.

Study 2 [37] demonstrates the impact of complex comorbidities and the effect of a complex sample matrix, e.g. low biomass samples like sputum, on microbiome results' reliability, interpretability, and reproducibility. **Study 2** [37] illustrates the next major challenge in microbiome research: can we transfer knowledge from gut microbiome analysis to different areas of microbiome research? I will outline how the pipelines used in analyzing the gut microbiome of healthy subjects can be applied to samples from the sputum microbiome of a population suffering from HIV and/ or COPD. I will also elucidate the composition and dynamics of the sputum microbiome in HIV-associated COPD.

The results highlight the importance of considering the sample matrix and using confounder-aware statistics to disentangle the effects of clinical, socio-demographics, and disease status on the sputum microbiome. We found up to 40% of human read contamination in the sputum samples, with the highest contamination values in COPD+ samples. It is known that low microbial biomass samples are more prone to technical biases and can easily be masked by background DNA contamination [19,89–91]. However, researchers have focused on the contamination from reagents and kits and suggest using blank values as controls followed by post-hoc computational correction [92,93]. Within this second

study, I discovered another source of contamination, DNA from host cells. In 2022, this was first addressed briefly when introducing the new preprocessing pipeline LotuS [94] update. This approach cannot be used as filtering is performed after clustering raw sequences to OTUs. We discovered that raw reads of human contamination already bias OTU clustering, mimicking microbiome signatures for patient characteristics.
Moreover, we could show that systematic confounder analysis demonstrated the possibility to disentangle the effects of HIV and COPD status, antiviral treatment, medication, and socio-demographics, providing researchers with a valuable understanding of the specific dynamics in HIV-associated COPD. HIV status had a more profound impact on bacterial composition than COPD status [37]. However, our analysis showed the inability to disentangle associations between HIV seropositivity, ART treatment, and its duration [37]. Therefore, it currently requires more work to separate the individual impacts of these factors. Nevertheless, certain genera such as *Veillonella*, *Actinomyces*, *Atopobium*, and *Filifactor* were significantly enriched in HIV-positive individuals in line with previous reports [37,95–97]. These genera have been associated with proinflammatory cytokine production, which may contribute to airway dysbiosis in HIV subjects, potentially interacting with other COPD risk factors [37,96,98]. The COPD status is associated with a significantly increased abundance of *Selenomonas* and a significant decrease in Gammaproteobacteria [37]. *Selenomonas*, known for its anaerobic metabolism, could thrive in the anaerobic conditions of the airways in COPD patients [99]. It has been suggested that *Selenomonas* interacts positively with anaerobes like *Campylobacter* and Veillonellaceae, forming an anaerobic consortium [37,99]. We also identified significant associations between the sputum microbiome and sociodemographic, clinical, and lifestyle characteristics, breaking ground for potential new research hypotheses and exciting aspects to consider when conducting microbiome studies. A comparison of results with a UK-based HIV cohort [69] revealed systematic differences between the two cohorts regarding the Gram state of the bacteria, reinforcing the results of **Study 1** [43] that preanalytical factors (in this case, the extraction methods) significantly impact the study results.

The findings highlight the need to carefully consider the preprocessing of samples in accordance with their unique properties. Moreover, our systematic confounder analysis demonstrated the possibility of unravelling the effects of various factors on the sputum microbiome, such as HIV status, antiviral treatment, medication, and lifestyle. It provided a valuable tool for understanding the dynamics in HIV-associated COPD.

In **Study 3** [45], we demonstrated the universal applicability of our developed pipelines and tools by using them to investigate the effects of PDE3A mutations on hypertension in an animal model. Our findings contribute to the growing body of literature on translational research, bridging the knowledge gained from human studies to unlock multi-omics investigations in animal models. We successfully established a link between consecutive active mutant PDE3A and hypertension using novel rat models. CRISPR-Cas9 engineering allowed us to generate rat models with hyperactivity of the PDE3A enzyme and its functional deletion. The mutant PDE3A enzyme has been shown to cause hypertension in rats, resulting in significant differences in blood pressure values between groups. However, the mutant did not appear to affect renal function. These findings have led us to propose a new hypothesis regarding the renoprotective effect of PDE3A mutations in a rat kidney disease model. We are investigating this hypothesis in a chronic kidney disease model and have published our findings in a rat model having acute kidney disease [48]. We could furthermore show the cardioprotective effect of the mutant PDE3A enzyme.

In summary, researchers can obtain more accurate and informative data on the microbiome and its role in human health and disease by carefully controlling for preanalytical factors. The systematic confounder analysis demonstrated the possibility of disentangling the effects of various factors on the sputum microbiome, such as HIV status, antiviral treatment, medication, and lifestyle. This approach provides a valuable tool for understanding the effects of treatment, disease, and environment on the microbiome, allowing researchers to distinguish these factors more clearly. These findings highlight the need for meticulous consideration of preanalytical parameters, rigorous study design, and analysis to obtain reliable microbiome data.

4.3 Embedding the results into the current state of research

This work highlights the current state of microbiome analysis. Microbiome research has gained widespread attention, stretching from human contexts to other scientific fields to understand the microbiome's influence on various ecological niches. The vast amount of literature estimates how complex microbiome analysis is, considering the multitude of preanalytical and postanalytical approaches in the field. In summary, the findings presented in this work are of essential value to this expanding research field, supporting the

call for standardization and confounder consideration to obtain accurate microbiome associations [23,40]. With the increase of multi-omics approaches in all research fields, this work can be generalized to other omics approaches, allowing researchers to conduct more robust and standardized research.

We now have a more critical understanding of potential biases, giving us the knowledge to support research across multiple omics fields to conduct more robust analyses.

4.4 Strengths and weaknesses of the studies

To critically address the validity of the presented study results, I evaluated the accuracy and appropriateness of the study methodology. The provided points highlight some limitations that should be considered: **Study 1** [43] only tested two extraction methods commonly used in microbiome studies. While these methods are widely utilized, it is essential to acknowledge that other extraction methods may be available in the market, affecting generalizability. Moreover, wet lab methods, e.g., extraction and collection protocols, constantly evolve, introducing newer and more refined techniques. However, **Study 1** [43] acknowledges that standardization and commitment to a specific protocol may not be feasible. To address this evolution in wet lab methods and sequencing pipelines, confounder analysis can be conducted to account for these variations and their potential impact on the results. The results presented in **Study 1** [43] demonstrate the awareness of limitations and a proactive approach to addressing them.

Furthermore, we could show that the confounder analysis applies to the second presented project, **Study 2** [37]. We addressed the inability to disentangle HIV seropositivity from ART and its duration in our data, which we were transparent about. However, it is essential to consider generalizability limitations due to particular study characteristics, such as the specific cohort used and the potential impact of preanalytical factors (extraction methods) on the results. Moreover, we could detect systematic differences between the study cohort and a UK-based HIV cohort [69], particularly regarding the Gram state of bacteria. This reinforces the need to consider preanalytical factors and potential population variations when interpreting and generalizing the findings.

Finally, in **Study 3** [45], we demonstrate the accuracy and appropriate use of new methods to generate and phenotype a rat model with hyperactivity of the PDE3A enzyme and its functional deletion, allowing us to establish a link between consecutive active mutant PDE3A and hypertension. We assessed the blood pressure values between groups and

observed significant differences, indicating the impact of the mutant PDE3A enzyme on hypertension. This outcome strengthens the validity of our study and supports the association between PDE3A mutations and hypertension. Additionally, we investigated the cardiac and renal function in the rat models, providing valuable information and leading us to propose a new hypothesis regarding the renoprotective effect of PDE3A mutations in a rat kidney disease model.

In conclusion, while acknowledging the limitations and specific study contexts, the presented studies provide valuable insights and contribute to the scientific discourse in their respective fields.

4.5 Implications for practice and/or future research

The present work suggests that previous studies reported incorrect results, leading to erroneous conclusions and potentially misdirected research efforts. Future research in this area will need to take steps to reduce or eliminate host DNA contamination, such as using different sequencing technologies or increasing the sensitivity of bacterial detection methods. Failure to address this issue could lead to continued inaccuracies and delays in developing effective diagnostic and therapeutic interventions. Sharing this knowledge with other research fields targeting the microbiome, such as environmental ecology, ensures accurate and reliable results.

5 Conclusions

Comparability and reproducibility are major challenges facing microbiome research, which this work aimed to examine. This thesis aimed to develop standardization strategies for methods and utilize statistics to identify confounding factors and human variability in microbiome research.

The most important finding from this work is the importance of implementing standardized workflows. In situations where standardization proves infeasible or has not been realized, as is the case in observational studies, it is mandatory to thoroughly investigate pre-analytical factors as they have the potential to introduce confounding effects. This work shows the drastic impact of DNA extraction methods on observed microbiome variability and the influence of misamplification with host DNA on microbiome signatures.

Moreover, this thesis shows that to identify host-microbiome interaction correctly, researchers must be conscious of potential confounders and take steps to address them during experimental design or through post hoc analysis. This approach ensures more accurate and reliable results and is relevant to other omics fields, enabling researchers to improve the robustness of their research and obtain more accurate results.

Reference List

1. Sender R, Fuchs S, Milo R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biol*. 2016;14(8):e1002533. doi:10.1371/journal.pbio.1002533
2. Thursby E, Juge N. Introduction to the human gut microbiota. *Biochem J*. 2017;474(11):1823-1836. doi:10.1042/BCJ20160510
3. Ahn J, Hayes RB. Environmental Influences on the Human Microbiome and Implications for Noncommunicable Disease. *Annu Rev Public Health*. 2021;42:277-292. doi:10.1146/annurev-publhealth-012420-105020
4. Mody D, Verma V, Rani V. Modulating host gene expression via gut microbiome-microRNA interplay to treat human diseases. *Crit Rev Microbiol*. 2021;47(5):596-611. doi:10.1080/1040841X.2021.1907739
5. Smith I. Mycobacterium tuberculosis pathogenesis and molecular determinants of virulence. *Clin Microbiol Rev*. 2003;16(3):463-496. doi:10.1128/CMR.16.3.463-496.2003
6. Lafond RE, Lukehart SA. Biological basis for syphilis. *Clin Microbiol Rev*. 2006;19(1):29-49. doi:10.1128/CMR.19.1.29-49.2006
7. Manichanh C, Rigottier-Gois L, Bonnaud E, Gloux K, Pelletier E, Frangeul L, Nalin R, Jarrin C, Chardon P, Marteau P, Roca J, Dore J. Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut*. 2006;55(2):205-211. doi:10.1136/gut.2005.073817
8. Schirmer M, Denson L, Vlamakis H, Franzosa EA, Thomas S, Gotman NM, Rufo P, Baker SS, Sauer C, Markowitz J, Pfefferkorn M, Oliva-Hemker M, Rosh J, Otley A, Boyle B, Mack D, Baldassano R, Keljo D, LeLeiko N, Heyman M, Griffiths A, Patel AS, Noe J, Kugathasan S, Walters T, Huttenhower C, Hyams J, Xavier RJ. Compositional and Temporal Changes in the Gut Microbiome of Pediatric Ulcerative Colitis Patients Are Linked to Disease Course. *Cell Host Microbe*. 2018;24(4):600-610.e4. doi:10.1016/j.chom.2018.09.009
9. Zeller G, Tap J, Voigt AY, Sunagawa S, Kultima JR, Costea PI, Amiot A, Böhm J, Brunetti F, Habermann N, Hercog R, Koch M, Luciani A, Mende DR, Schneider MA,

Schrotz-King P, Tournigand C, Tran Van Nhieu J, Yamada T, Zimmermann J, Benes V, Kloor M, Ulrich CM, von Knebel Doeberitz M, Sobhani I, Bork P. Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol Syst Biol*. 2014;10:766. doi:10.15252/msb.20145645

10. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, Egholm M, Henrissat B, Heath AC, Knight R, Gordon JI. A core gut microbiome in obese and lean twins. *Nature*. 2009;457(7228):480-484. doi:10.1038/nature07540

11. Forslund K, Hildebrand F, Nielsen T, Falony G, Le Chatelier E, Sunagawa S, Prifti E, Vieira-Silva S, Gudmundsdottir V, Krogh Pedersen H, Arumugam M, Kristiansen K, Yvonne Voigt A, Vestergaard H, Hercog R, Igor Costea P, Roat Kultima J, Li J, Jørgensen T, Levenez F, Dore J, Bjørn Nielsen H, Brunak S, Raes J, Hansen T, Wang J, Dusko Ehrlich S, Bork P, Pedersen O. Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. *Nature*. 2015;528(7581):262-266. doi:10.1038/nature15766

12. Jie Z, Xia H, Zhong SL, Feng Q, Li S, Liang S, Zhong H, Liu Z, Gao Y, Zhao H, Zhang D, Su Z, Fang Z, Lan Z, Li J, Xiao L, Li J, Li R, Li X, Li F, Ren H, Huang Y, Peng Y, Li G, Wen B, Dong B, Chen JY, Geng QS, Zhang ZW, Yang H, Wang J, Wang J, Zhang X, Madsen L, Brix S, Ning G, Xu X, Liu X, Hou Y, Jia H, He K, Kristiansen K. The gut microbiome in atherosclerotic cardiovascular disease. *Nat Commun*. 2017;8(1):1-11. doi:10.1038/s41467-017-00900-1

13. Aron-Wisnewsky J, Vigliotti C, Witjes J, Le P, Holleboom AG, Verheij J, Nieuwdorp M, Clément K. Gut microbiota and human NAFLD: disentangling microbial signatures from metabolic disorders. *Nat Rev Gastroenterol Hepatol*. 2020;17(5):279-297. doi:10.1038/s41575-020-0269-9

14. Wensel CR, Pluznick JL, Salzberg SL, Sears CL. Next-generation sequencing: insights to advance clinical investigations of the microbiome. *J Clin Invest*. 2022;132(7). doi:10.1172/JCI154944

15. Wooley JC, Ye Y. Metagenomics: Facts and Artifacts, and Computational Challenges*. *J Comput Sci Technol*. 2009;25(1):71-81. doi:10.1007/s11390-010-9306-4

16. Dicker AJ, Huang JTJ, Lonergan M, Keir HR, Fong CJ, Tan B, Cassidy AJ, Finch

S, Mullerova H, Miller BE, Tal-Singer R, Chalmers JD. The sputum microbiome, airway inflammation, and mortality in chronic obstructive pulmonary disease. *J Allergy Clin Immunol*. 2021;147(1):158-167. doi:10.1016/j.jaci.2020.02.040

17. Zhou Y, Jackson D, Bacharier LB, Mauger D, Boushey H, Castro M, Durack J, Huang Y, Lemanske RFJ, Storch GA, Weinstock GM, Wylie K, Covar R, Fitzpatrick AM, Phipatanakul W, Robison RG, Beigelman A. The upper-airway microbiota and loss of asthma control among asthmatic children. *Nat Commun*. 2019;10(1):5714. doi:10.1038/s41467-019-13698-x

18. Ditz B, Christenson S, Rossen J, Brightling C, Kerstjens HAM, van den Berge M, Faiz A. Sputum microbiome profiling in COPD: beyond singular pathogen detection. *Thorax*. 2020;75(4):338-344. doi:10.1136/thoraxjnl-2019-214168

19. Natalini JG, Singh S, Segal LN. The dynamic lung microbiome in health and disease. *Nat Rev Microbiol*. 2023;21(4):222-235. doi:10.1038/s41579-022-00821-x

20. Kennedy NA, Walker AW, Berry SH, Duncan SH, Farquarson FM, Louis P, Thomson JM, Satsangi J, Flint HJ, Parkhill J, Lees CW, Hold GL. The impact of different DNA extraction kits and laboratories upon the assessment of human gut microbiota composition by 16S rRNA gene sequencing. *PLoS One*. Published online 2014. doi:10.1371/journal.pone.0088982

21. Costea PI, Zeller G, Sunagawa S, Pelletier E, Alberti A, Levenez F, Tramontano M, Driessen M, Hercog R, Jung FE, Kultima JR, Hayward MR, Coelho LP, Allen-Vercoe E, Bertrand L, Blaut M, Brown JRM, Carton T, Cools-Portier S, Daigneault M, Derrien M, Druesne A, De Vos WM, Finlay BB, Flint HJ, Guarner F, Hattori M, Heilig H, Luna RA, Van Hylckama Vlieg J, Junick J, Klymiuk I, Langella P, Le Chatelier E, Mai V, Manichanh C, Martin JC, Mery C, Morita H, O'toole PW, Orvain C, Patil KR, Penders J, Persson S, Pons N, Popova M, Salonen A, Saulnier D, Scott KP, Singh B, Slezak K, Veiga P, Versalovic J, Zhao L, Zoetendal EG, Ehrlich SD, Dore J, Bork P. Towards standards for human fecal sample processing in metagenomic studies. *Nat Biotechnol*. 2017;35(11):1069-1076. doi:10.1038/nbt.3960

22. Wang Z, Zolnik CP, Qiu Y, Usyk M, Wang T, Strickler H, Isasi C, Kaplan RC, Kurland IJ, Qi Q, Burk RD. Comparison of Fecal Collection Methods for Microbiome

and Metabolomics Studies. *Front Cell Infect Microbiol*. 2018;8(August):301. doi:10.3389/FCIMB.2018.00301

23. Wesolowska-Andersen A, Bahl MI, Carvalho V, Kristiansen K, Sicheritz-Pontén T, Gupta R, Licht TR. Choice of bacterial DNA extraction method from fecal material influences community structure as evaluated by metagenomic analysis. *Microbiome*. 2014;2(1):1-11. doi:10.1186/2049-2618-2-19

24. He B, Zhu R, Yang H, Lu Q, Wang W, Song L, Sun X, Zhang G, Li S, Yang J, Tian G, Bing P, Lang J. Assessing the Impact of Data Preprocessing on Analyzing Next Generation Sequencing Data. *Front Bioeng Biotechnol*. 2020;8:817. doi:10.3389/fbioe.2020.00817

25. Allaband C, McDonald D, Vázquez-Baeza Y, Minich JJ, Tripathi A, Brenner DA, Loomba R, Smarr L, Sandborn WJ, Schnabl B, Dorrestein P, Zarrinpar A, Knight R. Microbiome 101: Studying, Analyzing, and Interpreting Gut Microbiome Data for Clinicians. *Clin Gastroenterol Hepatol Off Clin Pract J Am Gastroenterol Assoc*. 2019;17(2):218-230. doi:10.1016/j.cgh.2018.09.017

26. Ducarmon QR, Hornung BVH, Geelen AR, Kuijper EJ, Zwittink RD. Toward Standards in Clinical Microbiota Studies: Comparison of Three DNA Extraction Methods and Two Bioinformatic Pipelines. *mSystems*. 2020;5(1). doi:10.1128/mSystems.00547-19

27. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ, Walker AW. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol*. 2014;12:87. doi:10.1186/s12915-014-0087-z

28. Erb-Downward JR, Falkowski NR, D'Souza JC, McCloskey LM, McDonald RA, Brown CA, Shedden K, Dickson RP, Freeman CM, Stringer KA, Foxman B, Huffnagle GB, Curtis JL, Adar SD. Critical Relevance of Stochastic Effects on Low-Bacterial-Biomass 16S rRNA Gene Analysis. *MBio*. 2020;11(3). doi:10.1128/mBio.00258-20

29. Charlson ES, Bittinger K, Haas AR, Fitzgerald AS, Frank I, Yadav A, Bushman FD, Collman RG. Topographical continuity of bacterial populations in the healthy human respiratory tract. *Am J Respir Crit Care Med*. 2011;184(8):957-963. doi:10.1164/rccm.201104-0655OC

30. Dickson RP, Erb-Downward JR, Freeman CM, McCloskey L, Falkowski NR, Huffnagle GB, Curtis JL. Bacterial Topography of the Healthy Human Lower Respiratory Tract. *MBio*. 2017;8(1). doi:10.1128/mBio.02287-16

31. Segal LN, Clemente JC, Tsay JCJ, Koralov SB, Keller BC, Wu BG, Li Y, Shen N, Ghedin E, Morris A, Diaz P, Huang L, Wikoff WR, Ubeda C, Artacho A, Rom WN, Sterman DH, Collman RG, Blaser MJ, Weiden MD. Enrichment of the lung microbiome with oral taxa is associated with lung inflammation of a Th17 phenotype. *Nat Microbiol*. 2016;1:16031. doi:10.1038/nmicrobiol.2016.31

32. Clark SE. Commensal bacteria in the upper respiratory tract regulate susceptibility to infection. *Curr Opin Immunol*. 2020;66:42-49. doi:10.1016/j.coi.2020.03.010

33. Bassis CM, Erb-Downward JR, Dickson RP, Freeman CM, Schmidt TM, Young VB, Beck JM, Curtis JL, Huffnagle GB. Analysis of the upper respiratory tract microbiotas as the source of the lung and gastric microbiotas in healthy individuals. *MBio*. 2015;6(2):e00037. doi:10.1128/mBio.00037-15

34. Lim MY, Yoon HS, Rho M, Sung J, Song YM, Lee K, Ko G. Analysis of the association between host genetics, smoking, and sputum microbiota in healthy humans. *Sci Rep*. 2016;6(1):23745. doi:10.1038/srep23745

35. Nelson MT, Wolter DJ, Eng A, Weiss EJ, Vo AT, Brittnacher MJ, Hayden HS, Ravishankar S, Bautista G, Ratjen A, Blackledge M, McNamara S, Nay L, Majors C, Miller SI, Borenstein E, Simon RH, LiPuma JJ, Hoffman LR. Maintenance tobramycin primarily affects untargeted bacteria in the CF sputum microbiome. *Thorax*. 2020;75(9):780-790. doi:10.1136/thoraxjnl-2019-214187

36. Man WH, de Steenhuijsen Piters WAA, Bogaert D. The microbiota of the respiratory tract: gatekeeper to respiratory health. *Nat Rev Microbiol*. 2017;15(5):259-270. doi:10.1038/nrmicro.2017.14

37. Kayongo A, Bartolomaeus TUP, Birkner T, Markó L, Löber U, Kigozi E, Atugonza C, Munana R, Mawanda D, Sekibira R, Uwimaana E, Alupo P, Kalyesubula R, Knauf F, Siddharthan T, Bagaya BS, Kateete DP, Joloba ML, Sewankambo NK, Jjingo D, Kirenga B, Checkley W, Forslund SK. Sputum Microbiome and Chronic Obstructive Pulmonary Disease in a Rural Ugandan Cohort of Well-Controlled HIV Infection. *Microbiol Spectr*. 2023;11(2):e0213921. doi:10.1128/spectrum.02139-21

38. Rothschild D, Weissbrod O, Barkan E, Kurilshikov A, Korem T, Zeevi D, Costea PI, Godneva A, Kalka IN, Bar N, Shilo S, Lador D, Vila AV, Zmora N, Pevsner-Fischer M, Israeli D, Kosower N, Malka G, Wolf BC, Avnit-Sagi T, Lotan-Pompan M, Weinberger A, Halpern Z, Carmi S, Fu J, Wijmenga C, Zhernakova A, Elinav E, Segal E. Environment dominates over host genetics in shaping human gut microbiota. *Nature*. 2018;555(7695):210-215. doi:10.1038/nature25973

39. Greifer N, Stuart EA. Matching Methods for Confounder Adjustment: An Addition to the Epidemiologist's Toolbox. *Epidemiol Rev*. 2022;43(1):118-129. doi:10.1093/epirev/mxab003

40. Pourhoseingholi MA, Baghestani AR, Vahedi M. How to control confounding effects by statistical analysis. *Gastroenterol Hepatol from bed to bench*. 2012;5(2):79-83.

41. Kahlert J, Gribsholt SB, Gammelager H, Dekkers OM, Luta G. Control of confounding in the analysis phase - an overview for clinicians. *Clin Epidemiol*. 2017;9:195-204. doi:10.2147/CLEP.S129886

42. Ruxton GD, Beauchamp G. Time for some a priori thinking about post hoc testing. *Behav Ecol*. 2008;19(3):690-693. doi:10.1093/beheco/arn020

43. Bartolomaeus TUP, Birkner T, Bartolomaeus H, Löber U, Avery EG, Mähler A, Weber D, Kochlik B, Balogh A, Wilck N, Boschmann M, Müller DN, Markó L, Forslund SK. Quantifying technical confounders in microbiome studies. *Cardiovasc Res*. 2020;117(3):863-875. doi:10.1093/cvr/cvaa128

44. Shankar J. Insights into study design and statistical analyses in translational microbiome studies. *Ann Transl Med*. 2017;5(12):249. doi:10.21037/atm.2017.01.13

45. Ercu M, Mücke MB, Pallien T, Markó L, Sholokh A, Schächterle C, Aydin A, Kidd A, Walter S, Esmati Y, McMurray BJ, Lato DF, Yumi Sunaga-Franze D, Dierks PH, Flores BIM, Walker-Gray R, Gong M, Merticariu C, Zühlke K, Russwurm M, Liu T, Batolomaeus TUP, Pautz S, Schelenz S, Taube M, Napieczynska H, Heuser A, Eichhorst J, Lehmann M, Miller DC, Diecke S, Qadri F, Popova E, Langanki R, Movsesian MA, Herberg FW, Forslund SK, Müller DN, Borodina T, Maass PG, Bähring S, Hübner N, Bader M, Klussmann E. Mutant Phosphodiesterase 3A Protects From Hypertension-Induced Cardiac Damage. *Circulation*. 2022;146(23):1758-1778. doi:10.1161/CIRCULATIONAHA.122.060210

46. Costantini F. Transgenic Animals. In: Brenner S, Miller JHBTE of G, eds. Academic Press; 2001:1990-1998. doi:https://doi.org/10.1006/rwgn.2001.1305

47. Houdebine LM. Use of transgenic animals to improve human health and animal production. *Reprod Domest Anim*. 2005;40(4):269-281. doi:10.1111/j.1439-0531.2005.00596.x

48. Sholokh A, Walter S, Markó L, McMurray BJ, Sunaga-Franze DY, Xu M, Zühlke K, Russwurm M, Bartolomaeus TUP, Langanki R, Qadri F, Heuser A, Patzak A, Forslund SK, Bähring S, Borodina T, Persson PB, Maass PG, Bader M, Klussmann E. Mutant phosphodiesterase 3A protects the kidney from hypertension-induced damage. *Kidney Int*. 2023;104(2):388-393. doi:10.1016/j.kint.2023.04.026

49. Ercu M, Markó L, Schächterle C, Tsvetkov D, Cui Y, Maghsodi S, Bartolomaeus TUP, Maass PG, Zühlke K, Gregersen N, Hübner N, Hodge R, Mühl A, Pohl B, Illas RM, Geelhaar A, Walter S, Napieczynska H, Schelenz S, Taube M, Heuser A, Anistan YM, Qadri F, Todiras M, Plehm R, Popova E, Langanki R, Eichhorst J, Lehmann M, Wiesner B, Russwurm M, Forslund SK, Kamer I, Müller DN, Gollasch M, Aydin A, Bähring S, Bader M, Luft FC, Klussmann E. Phosphodiesterase 3A and Arterial Hypertension. *Circulation*. 2020;142(2):133-149. doi:10.1161/CIRCULATIONAHA.119.043061

50. Zhao G, Cardenas H, Matei D. Ovarian Cancer-Why Lipids Matter. *Cancers (Basel)*. 2019;11(12). doi:10.3390/cancers11121870

51. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods*. 2007;39(2):175-191. doi:10.3758/BF03193146

52. S.Champely, C.Ekstrom, P.Dalgaard, J.Gill, S.Weibelzahl, A.Anandkumar, C.Ford, R.Volcic HDR. Package “pwr.” Published online 2020. https://github.com/heliosdrm/pwr

53. Guetterman TC. Basics of statistics for primary care research. *Fam Med community Heal*. 2019;7(2). doi:10.1136/fmch-2018-000067

54. Saary P, Forslund K, Bork P, Hildebrand F. RTK: efficient rarefaction analysis of large datasets. *Bioinformatics*. 2017;33(16):2594-2595. doi:10.1093/bioinformatics/btx206

55. Oksanen AJ, Blanchet FG, Friendly M, Kindt R, Legendre P, Mcglinn D, Minchin PR, Hara RBO, Simpson GL, Solymos P, Stevens MHH SE. Package “vegan.” Published online 2016. https://cran.r-project.org/

56. Zapala MA, Schork NJ. Multivariate regression analysis of distance matrices for testing associations between gene expression patterns and related variables. *Proc Natl Acad Sci U S A*. 2006;103(51):19430-19435. doi:10.1073/pnas.0609333103

57. Forslund SK, Chakaroun R, Zimmermann-Kogadeeva M, Markó L, Aron-Wisnewsky J, Nielsen T, Moitinho-Silva L, Schmidt TSB, Falony G, Vieira-Silva S, Adriouch S, Alves RJ, Assmann K, Bastard JP, Birkner T, Caesar R, Chilloux J, Coelho LP, Fezeu L, Galleron N, Helft G, Isnard R, Ji B, Kuhn M, Le Chatelier E, Myridakis A, Olsson L, Pons N, Prifti E, Quinquis B, Roume H, Salem JE, Sokolovska N, Tremaroli V, Valles-Colomer M, Lewinter C, Søndertoft NB, Pedersen HK, Hansen TH, Gøtze JP, Køber L, Vestergaard H, Hansen T, Zucker JD, Hercberg S, Oppert JM, Letunic I, Nielsen J, Bäckhed F, Ehrlich SD, Dumas ME, Raes J, Pedersen O, Clément K, Stumvoll M, Bork P. Combinatorial, additive and dose-dependent drug-microbiome associations. *Nature*. 2021;600(7889):500-505. doi:10.1038/s41586-021-04177-9

58. Holmes I, Harris K, Quince C. Dirichlet multinomial mixtures: generative models for microbial metagenomics. *PLoS One*. 2012;7(2):e30126. doi:10.1371/journal.pone.0030126

59. D.Bates, M.Maechler, B.Bolker, S.Walker, R.H. Bojesen Christensen , H.Singmann, B.Dai, F.Scheipl, G.Grothendieck, P.Green, J.Fox, A.Bauer PNK. Package “lme4.” Published online 2023. https://github.com/lme4/lme4

60. T.Hothorn, A.Zeileis, R.W.Farebrother, C.Cummins, G.Millo DM. Package “lmtest.” Published online 2022. https://github.com/cran/lmtest

61. Beaudry MS, Wang J, Kieran TJ, Thomas J, Bayona-Vásquez NJ, Gao B, Devault A, Brunelle B, Lu K, Wang JS, Rhodes OEJ, Glenn TC. Improved Microbial Community Characterization of 16S rRNA via Metagenome Hybridization Capture Enrichment. *Front Microbiol*. 2021;12:644662. doi:10.3389/fmicb.2021.644662

62. Hasrat R, Kool J, de Steenhuijsen Piters WAA, Chu MLJN, Kuiling S, Groot JA, van Logchem EM, Fuentes S, Franz E, Bogaert D, Bosch T. Benchmarking laboratory processes to characterise low-biomass respiratory microbiota. *Sci Rep*.

2021;11(1):17148. doi:10.1038/s41598-021-96556-5

63. Hildebrand F, Tadeo R, Voigt AY, Bork P, Raes J. LotuS: an efficient and user-friendly OTU processing pipeline. *Microbiome*. 2014;2(1):30. doi:10.1186/2049-2618-2-30

64. Puente-Sánchez F, Aguirre J, Parro V. A novel conceptual approach to read-filtering in high-throughput amplicon sequencing studies. *Nucleic Acids Res*. 2015;44(4). doi:10.1093/nar/gkv1113

65. Edgar RC. UPARSE: Highly accurate OTU sequences from microbial amplicon reads. *Nat Methods*. 2013;10(10):996-998. doi:10.1038/nmeth.2604

66. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*. 2012;41(D1):D590-D596. doi:10.1093/nar/gks1219

67. McDonald D, Price MN, Goodrich J, Nawrocki EP, Desantis TZ, Probst A, Andersen GL, Knight R, Hugenholtz P. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J*. 2012;6(3):610-618. doi:10.1038/ismej.2011.139

68. Ritari J, Salojärvi J, Lahti L, de Vos WM. Improved taxonomic assignment of human intestinal 16S rRNA sequences by a dedicated reference database. *BMC Genomics*. 2015;16(1). doi:10.1186/s12864-015-2265-y

69. Rofael SAD, Brown J, Pickett E, Johnson M, Hurst JR, Spratt D, Lipman M, McHugh TD. Enrichment of the airway microbiome in people living with HIV with potential pathogenic bacteria despite antiretroviral therapy. *EClinicalMedicine*. 2020;24:100427. doi:10.1016/j.eclinm.2020.100427

70. Kers JG, Saccenti E. The Power of Microbiome Studies: Some Considerations on Which Alpha and Beta Metrics to Use and How to Report Results. *Front Microbiol*. 2021;12:796025. doi:10.3389/fmicb.2021.796025

71. Krenek P, Kmecova J, Kucerova D, Bajuszova Z, Musil P, Gazova A, Ochodnicky P, Klimas J, Kyselovic J. Isoproterenol-induced heart failure in the rat is associated with nitric oxide-dependent functional alterations of cardiac function. *Eur J Heart Fail*. 2009;11(2):140-146. doi:10.1093/eurjhf/hfn026

72. Galloway-Peña J, Hanson B. Tools for Analysis of the Microbiome. *Dig Dis Sci*. 2020;65(3):674-685. doi:10.1007/s10620-020-06091-y

73. Schloss PD. Identifying and Overcoming Threats to Reproducibility, Replicability, Robustness, and Generalizability in Microbiome Research. *MBio*. 2018;9(3). doi:10.1128/mBio.00525-18

74. Choo JM, Leong LEX, Rogers GB. Sample storage conditions significantly influence faecal microbiome profiles. *Sci Rep*. 2015;5:1-10. doi:10.1038/srep16350

75. Hallmaier-Wacker LK, Lueert S, Roos C, Knauf S. The impact of storage buffer, DNA extraction method, and polymerase on microbial analysis. *Sci Rep*. 2018;8(1):1-9. doi:10.1038/s41598-018-24573-y

76. Panek M, Čipčić Paljetak H, Barešić A, Perić M, Matijašić M, Lojkić I, Vranešić Bender D, Krznarić Ž, Verbanac D. Methodology challenges in studying human gut microbiota - effects of collection, storage, DNA extraction and next generation sequencing technologies. *Sci Rep*. 2018;8(1):5143. doi:10.1038/s41598-018-23296-4

77. Thomas V, Clark J, Doré J. Fecal microbiota analysis : An overview of sample collection methods and sequencing strategies. *Futur Microbiol Rev*. 2015;10:1485-1504. doi:10.2217/fmb.15.87

78. Bahl MI, Bergström A, Licht TR. Freezing fecal samples prior to DNA extraction affects the Firmicutes to Bacteroidetes ratio determined by downstream quantitative PCR analysis. *FEMS Microbiol Lett*. 2012;329(2):193-197. doi:10.1111/j.1574-6968.2012.02523.x

79. Plauzolles A, Toumi E, Bonnet M, Pénaranda G, Bidaut G, Chiche L, Allardet-Servent J, Retornaz F, Goutorbe B, Halfon P. Human Stool Preservation Impacts Taxonomic Profiles in 16S Metagenomics Studies. *Front Cell Infect Microbiol*. 2022;12:722886. doi:10.3389/fcimb.2022.722886

80. Nearing JT, Comeau AM, Langille MGI. Identifying biases and their potential solutions in human microbiome studies. *Microbiome*. 2021;9(1):113. doi:10.1186/s40168-021-01059-0

81. Fouhy F, Deane J, Rea MC, O'Sullivan Ó, Ross RP, O'Callaghan G, Plant BJ, Stanton C. The effects of freezing on faecal microbiota as determined using miseq

sequencing and culture-based investigations. *PLoS One*. 2015;10(3):1-12. doi:10.1371/journal.pone.0119355

82. Li X, Bosch-Tijhof CJ, Wei X, de Soet JJ, Crielaard W, Loveren C van, Deng DM. Efficiency of chemical versus mechanical disruption methods of DNA extraction for the identification of oral Gram-positive and Gram-negative bacteria. *J Int Med Res*. 2020;48(5):300060520925594. doi:10.1177/0300060520925594

83. Yuan S, Cohen DB, Ravel J, Abdo Z, Forney LJ. Evaluation of methods for the extraction and purification of DNA from the human microbiome. *PLoS One*. 2012;7(3). doi:10.1371/journal.pone.0033865

84. Knudsen BE, Bergmark L, Munk P, Lukjancenko O, Priemé A, Aarestrup FM, Pamp SJ. Impact of Sample Type and DNA Isolation Procedure on Genomic Inference of Microbiome Composition. *mSystems*. Published online 2016. doi:10.1128/msystems.00095-16

85. Tourlousse DM, Narita K, Miura T, Sakamoto M, Ohashi A, Shiina K, Matsuda M, Miura D, Shimamura M, Ohyama Y, Yamazoe A, Uchino Y, Kameyama K, Arioka S, Kataoka J, Hisada T, Fujii K, Takahashi S, Kuroiwa M, Rokushima M, Nishiyama M, Tanaka Y, Fuchikami T, Aoki H, Kira S, Koyanagi R, Naito T, Nishiwaki M, Kumagai H, Konda M, Kasahara K, Ohkuma M, Kawasaki H, Sekiguchi Y, Terauchi J. Validation and standardization of DNA extraction and library construction methods for metagenomics-based human fecal microbiome measurements. *Microbiome*. 2021;9(1):95. doi:10.1186/s40168-021-01048-3

86. Smith B. Optimising Bacterial DNA Extraction from Faecal Samples: Comparison of Three Methods. *Open Microbiol J*. 2011;5(1):14-17. doi:10.2174/1874285801105010014

87. Maukonen J, Simões C, Saarela M. The currently used commercial DNA-extraction methods give different results of clostridial and actinobacterial populations derived from human fecal samples. *FEMS Microbiol Ecol*. 2012;79(3):697-708. doi:10.1111/j.1574-6941.2011.01257.x

88. Caesar R, Tremaroli V, Kovatcheva-Datchary P, Cani PD, Bäckhed F. Crosstalk between gut microbiota and dietary lipids aggravates WAT inflammation through TLR signaling. *Cell Metab*. 2015;22(4):658-668. doi:10.1016/j.cmet.2015.07.026

89. Karstens L, Asquith M, Davin S, Fair D, Gregory WT, Wolfe AJ, Braun J, McWeeney S. Controlling for Contaminants in Low-Biomass 16S rRNA Gene Sequencing Experiments. *mSystems*. 2019;4(4):10.1128/msystems.00290-19. doi:10.1128/msystems.00290-19

90. Eisenhofer R, Minich JJ, Marotz C, Cooper A, Knight R, Weyrich LS. Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations. *Trends Microbiol*. 2019;27(2):105-117. doi:https://doi.org/10.1016/j.tim.2018.11.003

91. Selway CA, Eisenhofer R, Weyrich LS. Microbiome applications for pathology: challenges of low microbial biomass samples during diagnostic testing. *J Pathol Clin Res*. 2020;6(2):97-106. doi:10.1002/cjp2.151

92. Kim D, Hofstaedter CE, Zhao C, Mattei L, Tanes C, Clarke E, Lauder A, Sherrill-Mix S, Chehoud C, Kelsen J, Conrad M, Collman RG, Baldassano R, Bushman FD, Bittinger K. Optimizing methods and dodging pitfalls in microbiome research. *Microbiome*. 2017;5(1):52. doi:10.1186/s40168-017-0267-5

93. Drengenes C, Wiker HG, Kalananthan T, Nordeide E, Eagan TML, Nielsen R. Laboratory contamination in airway microbiome studies. *BMC Microbiol*. 2019;19(1):187. doi:10.1186/s12866-019-1560-1

94. Özkurt E, Fritscher J, Soranzo N, Ng DYK, Davey RP, Bahram M, Hildebrand F. LotuS2: an ultrafast and highly accurate tool for amplicon sequencing analysis. *Microbiome*. 2022;10(1):176. doi:10.1186/s40168-022-01365-1

95. Bhadriraju S, Fadrosh DW, Shenoy MK, Lin DL, Lynch K V, McCauley K, Ferrand RA, Majonga ED, McHugh G, Huang L, Lynch S V, Metcalfe JZ. Distinct lung microbiota associate with HIV-associated chronic lung disease in children. *Sci Rep*. 2020;10(1):16186. doi:10.1038/s41598-020-73085-1

96. Lawani MB, Morris A. The respiratory microbiome of HIV-infected individuals. *Expert Rev Anti Infect Ther*. 2016;14(8):719-729. doi:10.1080/14787210.2016.1206469

97. Ueckermann V, Lebre P, Geldenhuys J, Hoosien E, Cowan D, van Rensburg LJ, Ehlers M. The lung microbiome in HIV-positive patients with active pulmonary tuberculosis. *Sci Rep*. 2022;12(1):8975. doi:10.1038/s41598-022-12970-3

98. Zevin AS, McKinnon L, Burgener A, Klatt NR. Microbial translocation and microbiome dysbiosis in HIV-associated immune activation. *Curr Opin HIV AIDS*. 2016;11(2):182-190. doi:10.1097/COH.0000000000000234

99. Ticlla MR, Hella J, Hiza H, Sasamalo M, Mhimbira F, Rutaihwa LK, Droz S, Schaller S, Reither K, Hilty M, Comas I, Beisel C, Schmid CD, Fenner L, Gagneux S. The Sputum Microbiome in Pulmonary Tuberculosis and Its Association With Disease Manifestations: A Cross-Sectional Study. *Front Microbiol*. 2021;12:633396. doi:10.3389/fmicb.2021.633396

Statutory Declaration

“I, Theda Ulrike Patricia Bartolomaeus, by personally signing this document in lieu of an oath, hereby affirm that I prepared the submitted dissertation on the topic “Technical confounders in microbiome studies – examples from fecal and sputum microbiome research/ Technische Störfaktoren in Mikrobiomstudien - Beispiele aus der Fäkal- und Sputum Mikrobiomforschung”, independently and without the support of third parties, and that I used no other sources and aids than those stated.
All parts which are based on the publications or presentations of other authors, either in letter or in spirit, are specified as such in accordance with the citing guidelines. The sections on methodology (in particular regarding practical work, laboratory regulations, statistical processing) and results (in particular regarding figures, charts and tables) are exclusively my responsibility.

Furthermore, I declare that I have correctly marked all of the data, the analyses, and the conclusions generated from data obtained in collaboration with other persons, and that I have correctly marked my own contribution and the contributions of other persons (cf. declaration of contribution). I have correctly marked all texts or parts of texts that were generated in collaboration with other persons.

My contributions to any publications to this dissertation correspond to those stated in the below joint declaration made together with the supervisor. All publications created within the scope of the dissertation comply with the guidelines of the ICMJE (International Committee of Medical Journal Editors; http://www.icmje.org) on authorship. In addition, I declare that I shall comply with the regulations of Charité – Universitätsmedizin Berlin on ensuring good scientific practice.

I declare that I have not yet submitted this dissertation in identical or similar form to another Faculty.

The significance of this statutory declaration and the consequences of a false statutory declaration under criminal law (Sections 156, 161 of the German Criminal Code) are known to me.”

Date Signature

Declaration of your own contribution to the publications

Theda Ulrike Patricia Bartolomaeus contributed the following to the below-listed publications:

Publication 1: T.U.P. Bartolomaeus, T. Birkner, H. Bartolomaeus, U. Löber, E.G. Avery, A. Mähler, D. Weber, B. Kochlik, A. Balogh, N. Wilck, M. Boschmann, D.N. Müller, L. Markó, S.K. Forslund, Quantifying technical confounders in microbiome studies, Cardiovascular Research, 2020
Contribution in detail:

- planning the study design (Fig. 1) with H. Bartolomaeus and L. Markó
- writing the ethics and performing power analysis
- creating food-frequency questionnaires (FFQ) on animal products with A. Mähler
- detailed patients' instructions for the study procedure
- executing DNA extraction (Fig. 1, Fig. S2 and S3)
- preparing/measuring samples micronutrients with D. Weber and B. Kochlik (Fig. S1A and S1B)
- evaluating FFQs
- preprocessing of 16S raw data U. Löber
- statistical Analysis (Fig. 2, 3, 4, 5, and 6) with T. Birkner and S.K. Forslund
- perform qPCR for specific Phyla (Fig. 6)
- Writing the manuscript together with S.K. Forslund, L. Markó, and H. Bartolomaeus
- Lead role during the revision process

Publication 2: A. Kayongo, T.U.P. Bartolomaeus, T. Birkner, L. Markó, U. Löber, E. Kigozi, C. Atugonza, R. Munana, D. Mawanda, R. Sekibria, E. Uwimaana, P. Alupo, R. Kalyesubula, F. Knauf, T. Siddharthan, B.S. Bagaya, D.P. Kateete, M.L. Joloba, N.K. Sewankambo, D. Jjingo, B. Kirenga, W. Checkley, S.K. Forslund, Sputum Microbiome and Chronic Obstructive Pulmonary Disease in a Rural Ugandan Cohort of Well-Controlled HIV Infection, Microbiology Spectrum, 2023
Contribution in detail:

- Preprocessing and human contamination removal of 16S raw data with U. Löber
- Functional prediction using PiCRUST2

- statistical Analysis (Fig. 2, 3, 4, 5, S1, S2, S3, S4, S5, and Table S1) and with T. Birkner and S.K. Forslund
- Writing of the manuscript together with A. Kayongo
- Lead role during the revision process

Publication 3: M. Ecru, M.B. Mücke, T. Pallien, L. Markó, A. Sholokh, C. Schächterle, A. Aydin, A. Kidd, S. Walter, Y. Esmati, B. J. McMurray, D.F. Lato, D. F. Sunaga-Franze, P.H. Dierks, B.I. Montesinos, R. Walker-Gray, M. Gong, C. Merticariu, K. Zühlke, M. Russwurm, T. Liu, T.U.P. Bartolomaeus, S. Pautz, S. Schelenz, M. Taube, H. Napieczyn-ska, A. Heuser, J. Eichhorst, M. Lehmann, D.C. Miller, S. Diecke, F. Qadri, E. Popova, R. Langanki, M.A. Movsesian, F.W. Herberg, S.K. Forslund, D. N. Müller, T. Borodina, P.G. Maass, S. Bähring, N. Hübner, M. Bader, E. Klussmann, Mutant Phosphodiesterase 3A Protects From Hypertension-Induced Cardiac Damage, 2022
Contribution in detail:

- Preparation, weighing and fixation of rat organs (heart, kidney, aorta and tibia) with K. Zühlke and L. Markó
- Blood pressure analysis with S.K. Forslund (Fig. 2, Fig. 3)
- executing RNA extraction
- DNA Transcription and gene expression analysis (qPCR) (S2)

_____

Signature, date and stamp of first supervising university professor / lecturer

_____

Signature of doctoral candidate

Printing copy(s) of the publication(s)

Study 1: Bartolomaeus TUP et al. 2020. Quantifying technical confounders in microbiome studies [43]

Bartolomaeus TUP, Birkner T, Bartolomaeus H, Löber U, Avery EG, Mähler A, Weber D, Kochlik B, Balogh A, Wilck N, Boschmann M, Müller DN, Markó L, Forslund SK. Quantifying technical confounders in microbiome studies. *Cardiovasc Res*. 2020;117(3):863-875.https://doi.org/10.1093/cvr/cvaa128

Study 2: Kayongo A*, Bartolomaeus TUP*, Birkner T* et al. 2023. Sputum microbiome and chronic obstructive pulmonary disease in a rural Ugandan cohort of well-controlled HIV infection [37]

Kayongo A, Bartolomaeus TUP, Birkner T, Markó L, Löber U, Kigozi E, Atugonza C, Munana R, Mawanda D, Sekibira R, Uwimaana E, Alupo P, Kalyesubula R, Knauf F, Siddharthan T, Bagaya BS, Kateete DP, Joloba ML, Sewankambo NK, Jjingo D, Kirenga B, Checkley W, Forslund SK. Sputum Microbiome and Chronic Obstructive Pulmonary Disease in a Rural Ugandan Cohort of Well-Controlled HIV Infection. *Microbiol Spectr*. 2023;11(2):e0213921. https://doi.org/10.1128/spectrum.02139-21

Study 3: Ecru M, …. Bartolomaeus TUP et al. 2022. Mutant Phosphodiesterase 3A Protects From Hypertension-Induced Cardiac Damage [45]

Ercu M, Mücke MB, Pallien T, Markó L, Sholokh A, Schächterle C, Aydin A, Kidd A, Walter S, Esmati Y, McMurray BJ, Lato DF, Yumi Sunaga-Franze D, Dierks PH, Flores BIM, Walker-Gray R, Gong M, Merticariu C, Zühlke K, Russwurm M, Liu T, Batolomaeus TUP, Pautz S, Schelenz S, Taube M, Napieczynska H, Heuser A, Eichhorst J, Lehmann M, Miller DC, Diecke S, Qadri F, Popova E, Langanki R, Movsesian MA, Herberg FW, Forslund SK, Müller DN, Borodina T, Maass PG, Bähring S, Hübner N, Bader M, Klussmann E. Mutant Phosphodiesterase 3A Protects From Hypertension-Induced Cardiac Damage. *Circulation*. 2022;146(23):1758-1778. https://doi.org/10.1161/CIRCULATIONAHA.122.060210

Curriculum Vitae

My curriculum vitae will not be published in the electronic version of my work for data protection reasons.

Publication list

PUBLICATIONS

In preparation

Ghorbani Z*, **Bartolomaeus TUP*** et al. Synbiotic Supplementation Could Be Effective in Improving Atherogenic Status in Nonalcoholic Fatty Liver Disease Patients: Post-Hoc Analysis of Two Randomized Controlled Trials *(in prep)*

Borum L.S*, Bartolomaeus TUP* et al. Probiotic ice cream influences gut and vaginal microbiota in high-risk pregnancies *(in prep)*

In revision

Cuthbertson, LF, Is-Horowicz, JS, Olanipekum, MT, Forslund, SK, Löber U, Birkner T, **Bartolomaeus TUP** et al. Genomic attributes of the airway microbiota. *Nature Communication* (in revision)

Kayongo A*, **Bartolomaeus TUP***, Birkner T*, Siddharthan, T, Markó L et al. Sputum microbiome and Chronic Obstructive Pulmonary Disease in a rural Ugandan cohort of well-controlled HIV infection. *Microbiology Spectrum* 15;11(2): e0213921. DOI: 10.1128/spectrum.02139-21

Samuelsson A*, **Bartolomaeus TUP**, Anandakumar H, Thowsen I, T, Nikpey E et al. VEGF-B hypertrophy predisposes to transition from diastolic to systolic heart failure in hypertensive rats. *Cardiovascular Research* 119(7):1553-1567. DOI: 10.1093/cvr/cvad040

2023

Noormohammad M, Ghorbani Z, Löber U, Mahdavi-Roshan M, **Bartolomaeus TUP** et al. (2023). The effect of probiotic and synbiotic supplementation on appetite-regulating hormones and desire to eat: A systematic review and meta-analysis of clinical trials. *Pharmacological Research* 187: 106614 DOI: 10.1016/j.phrs.2022.106614

2022

Ercu M, Mücke MB, Pallien T, Markó L, Sholokh A, Schächterle C, Aydin A, Kidd A, Walter S, Esmati Y, McMurray BJ, Lato DF, Sunaga-Franze DY, Dierks PH, Montesinos Flores BI, Walker-Gray R, Gong M, Merticariu C, Zühlke K, Russwurm M, Liu T, **Bartolomaeus TUP**, et al. (2022). Mutant Phosphodiesterase 3A Protects From Hypertension-Induced Cardiac Damage. *Circulation* 146:1758–1778. DOI: 10.1161/CIRCULATIONAHA.122.060210

Holle J., Bartolomaeus H., Löber U., Behrens F., **Bartolomaeus TUP** et al. (2022). Inflammation in children with chronic kidney disease linked to gut dysbiosis and metabolite imbalance. *JASN* Epub 2022 Aug 19. DOI: 10.1681/ASN.2022030378

Ghorbani Z, Kazemi A, **Bartolomaeus TUP** et al. (2022). The effect of probiotic and synbiotic supplementation on lipid parameters among patients with cardiometabolic risk factors: a systematic review and meta-analysis of clinical trials, *Cardiovascular Research*, cvac128, DOI: 10.1093/cvr/cvac128

Zheng Z., Tsvetkov D**, Bartolomaeus TUP** et al. (2022). Role of TRPC6 in Kidney Damage after Acute Ischemic Kidney Injury. *Scientific Reports* 12, 3038, DOI: 10.1038/s41598-022-06703-9

Bartolomaeus H, Avery EG, Löber U, **Bartolomaeus TUP** et al. (2022). Quantifying the impact of gut microbiota on inflammation and hypertensive organ damage. *Cardiovascular Research*, DOI: 10.1093/cvr/cvac121

Maifeld A, Wild J, Karlsen TV, Rakova N, Wistorf E, Linz P, Jung R, Birukov A, Gimenez-Rivera V-A, Wilck N, **Bartolomaeus TUP**, et al. (2022). Skin Sodium Accumulates in Psoriasis and Reflects Disease Severity, *Journal of Investigative Dermatology* 2022 Jan;142(1):166-178.e8**,** DOI: 10.1016/j.jid.2021.06.013

2021

Ercu M, Markó L, Schächterle C, Tsvetkov D, Cui Y, Maghsodi S, **Bartolomaeus TUP**, Maass PG, Zühlke K, et al. (2021). Phosphodiesterase 3A and Arterial Hypertension. *Circulation* 142: 133-149. DOI: 10.1161/CIRCULATIONAHA.119.043061

2020

Bartolomaeus TUP, Birkner T, Bartolomaeus H, Löber U, Avery EG, et al. (2020). Quantifying technical confounders in microbiome studies, *Cardiovascular Research* 117(3): 863–875, DOI: 10.1093/cvr/cvaa128

Bartolomaeus TUP, Forslund SK (2020). Hitchhiker's guide to microbiome studies, *Cardiovascular Research* 116 (3) e44–e47, DOI: 10.1093/cvr/cvz316

Bartolomaeus H, Avery EG, **Bartolomaeus TUP**, Kozhakhmetov S, Zhumadilov Z, et al. (2020). Blood pressure changes correlate with short-chain fatty acid production potential shifts under a synbiotic intervention. *Cardiovascular Research* 116 (7): 1252–1253, DOI: 10.1093/cvr/cvaa083

Acknowledgments

I would like to thank PD Dr. Enno Klussmann for entrusting me with the opportunity to work on the PDE3A project. I am also thankful to Prof. Dr. Sofia Forslund-Startceva for being an exceptional role model, providing me with the chance to grow and nurture my scientific curiosity, and empowering me to express myself.

Special thanks to my entire research group for their active engagement in fruitful discussions and the invaluable exchange of ideas on various projects. Your collective input and collaboration have played a pivotal role in the success of this work. In particular, I would like to thank Dr. Lajos Marko, who made working with the kidney such an enjoyable experience and provided unwavering support, to Till Birkner, for supporting me in improving my R skills and to Dr. Ulrike Löber for her patience in teaching me terminal commands and preprocessing samples. My appreciation extends to Morgen Essex and Christoph Bornemann for providing final comments and corrections.

Furthermore, I would like to thank my brother, Hendrik, for his continuous support in formulating research ideas and being my sparring's partner. I want to express my gratitude to my wife, for always supporting me in my work. Thanks to my father for always encouraging my fantasy and curiosity, and to my mother for her loving yet strict upbringing.

I extend my heartfelt appreciation to all the co-authors of the presented publications. Your expertise and contributions have significantly enriched the quality and depth of our research, and I am grateful for the opportunity to work with such a talented and dedicated team. A special mention goes to Dr. Alex Kayongo, for extending his research and providing access to the dataset which was crucial to the progress of this project. I am deeply thankful for this trust and for continuing to place confidence in me with new datasets.

Thank you all for being part of this journey and for making it a rewarding and fulfilling experience. Your collective efforts have made a significant impact on the advancement of our research, and I look forward to future collaborations and discoveries with this amazing team.

I want to conclude this work with a quote from Umberto Eco's "Das Foucaultsche Pendel": *"Whether I write it or not, it makes no difference. You would always seek another meaning, even in my silence. That's how you are. Blind to revelation. Malchuth is Malchuth, and that's it... He is so beautiful."*