universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„The oncobiome in preclinical *in vivo* models"
(Critical issues and challenges)

verfasst von / submitted by

Daphne del Carmen Kolland BSc.

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2019 / Vienna 2019

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 863
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Biologische Chemie
Betreut von / Supervisor:	Dr. Wolfgang Sommergruber

Abstract

The human microbiome is a fast growing field of interest in pharmacology. Recent studies have identified intratumoural bacteria and how it can alter anticancer treatment.
Our research aimed to establish preclinical oncobiome mouse models as a tool to enable future investigations. 16S rRNA sequencing, the method we chose for our analysis is the standard method for microbiome analysis. 11 different CDX mouse models were generated and their bacterial profile was examined.
Upon experimental procedure, we came across different challenges, the main one being the prevalence of bacteria that was not supposed to be present. Deeper investigation could identify them as contamination, as a consequence we decided to execute an *in silico* analysis, to evaluate the method in literature and proceeded with troubleshooting suggestions. Consequently we conclude that 16S sequencing is very sensitive and special measurements have to be taken to perform a clean library process such that no contamination is generated. Upon this conclusion we performed a simple *in situ* hybridization experiment using a general eubacteria probe to answer the main question whether there are bacteria present in tumour tissue in in–house grown CDX mouse models or not. Our results strongly suggest that the bacterial load of CDX samples might just be under a threshold such that no significant and reliable data can be yielded. Nevertheless, we would suggest repeating the experimental procedure with PDX samples, since they might harbour more bacteria thus, producing more reliable data.

Das menschliche Mikrobiom ist ein stark wachsendes Forschungsgebiet in der Pharmakologie. Studien berichten von intratumoralen Bakterien, die fähig sind Krebstherapien zu beeinflussen. In unserem Projekt wollten wir preklinische Onkobiom Maus Modelle etablieren um sie für spätere Onkobiom Analysen verwenden zu können. Als Analysemethode um das bakterielle Profil der Modelle zu analysieren, wählten wir 16S rRNA Sequenzierung für 11 ausgewählte CDX Maus Modelle.
Während des experimentellen Prozesses mussten wir jedoch feststellen, dass Bakterien detektiert wurden, die eigentlich nicht präsent sein dürften in den jeweiligen Proben. Aus diesem Grund analysierten wir zusätzlich *in silico* vorhandene Literatur zu dem Thema um die Methode zu evaluieren gefolgt von einer Fehlerdiagnose. Diese Analysen brachten uns zu dem Fazit, dass die 16S Sequenzierung eine äußerst sensitive Methode ist, welche genau etabliert und validiert werden muss um keine Kontaminationen zu verursachen und den Prozess möglichst rein durchführen zu können. Aus diesem Grund führten wir noch ein *in situ* Hybridisation Experiment durch um grundsätzlich die Präsenz von intratumoralen Bakterien in in–Haus generierten Mausmodellen zu verifizieren.
Unsere Resultate weisen stark darauf hin, dass die Bakterienanzahl von CDX Modellen sich unter einer signifikanten Nachweisgrenze befinden und die generierten Resultate deswegen nicht reproduzierbar sind. Deswegen würden wir empfehlen, den experimentellen Prozess mit PDX Modellen zu wiederholen, da solche Proben wahrscheinlich mehr Bakterienpräsenz aufweisen und aufgrund dessen, zuverlässigere Ergebnisse ergeben würden.

Acknowledgments

The first person I need to thank is Dr. Melanie Wurm, my supervisor. She was the one who entrusted me with this project and believed in me from the first minute. I really appreciate all the trouble you endured to help and support me throughout this project. You were committed to me and the project from the beginning and have supported me through the good and bad times always advising me and guiding me through. Thank you very much for everything you have done for me.

I would also like to thank Dr. Wolfgang Sommergruber, who is a constant inspiration to me. You are truly a one–of–kind scientist and a fountain of wisdom. I could learn so much in every single meeting I had with you (there were a lot!). Thank you for always taking your time and guiding me through this work. You always brought new and innovative ideas to the table and without you; this thesis would have not reached the level it did. I know it was not always easy but you really helped me overcome the obstacles and would re–challenge me.

A big thanks goes to Dr. Markus Bauer, the bioinformatics wizard. Your work is astonishing and I could learn so much from our talks. I really appreciate the time and effort that you have invested to turn this thing around and all the pep–talks you have given me when it got hard. Thank you very much for having lent me your skills, which were very much needed.

Next, comes an enormous amount of people that have just been present during this work; supporting me emotionally and mentally. Thank you for the great inputs and inspiration to extraordinary ideas. Biggest thanks to; Teresa Zanin, Sonja Porits, Astrid Jeschko, Sandra Petrovic, Mirko Raskovic, Rebecca Langlois, Stefan Fischer, Regina Ruzicka, Christoph Albrecht, Robin Jacob, Florian Mark, Roland Varecka, Loretta Hasana, Oliver Bergner, Herwig Machat, Sabine Kallenda, Alice Schoenanger, Katrin Gretzinger, Fabian Miller, Lisa– Marie Polster, Bettina Krenn, Klaus–Peter Künkele, Norbert Schweifer, Jürgen Moll, Francesca Trapani and many more. I have probably forgotten a million people. But this is just a little acknowledgement that will stay forever on this paper. Remembering, that you have helped a struggling Master student. Thank you!

Lastly, I want to thank my family which is the best anyone could wish for; my mom and dad and my sister who have supported me without a doubt from the beginning till the very bitter end. They would endure my moodiness and sometimes my desperation and would always have comforting words for me and keep my head up. I love you beyond words, thank you for being the best I could wish for.
Thank you to Jasmin Qiu and Lü Wei, my two best friends who would listen to my worries day in and out and would always help me find a solution for every single problem I had. You have been there from the beginning and let's hope we will stick together till the end. I could never have done this without you. My last thanks goes to my boyfriend Arunava Acharayya who would calm me down and live this adventure beside me. I would probably never have dared to apply at my dream company without you. Thank you for throwing (more like pushing) me out of my comfort zone.

Table of Contents

Abbreviation List

ATCC	American Type Culture Collection
BCG	Bacille Calmette–Guerin
bp	Base pair
BomTac	BomholdgardTaconic
CDD_L	Cytidine Deaminase
CDX	Cell line Derived Xenografts
CO_2	Carbon dioxide
CTLA–4	Cytotoxic T–Lymphocyte–Associated Antigen 4
DB	Database
DNA	Deoxyribuncleic Acid
DNA Ex.	DNA Extraction
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen
EB	Eubacteria
FACS	Fluorescence–activated cell sorting
FCS	Fetal Cow Serum
FDA	Food And Drug Administration
FELASA	Federation For Laboratory Animal Science Associations
FFPE	Formaline–Fixed, Paraffin–Embedded
FISH	Fluorescence *in situ* hybridization
FWD	Forward Primer
GBM	Glioblastoma Multiforme
gDNA	Genomic DNA
GI	Gastrointestinal
HMP	Human Microbiome Project
Hsa	Homo sapiens
IARC	International Agency For Research On Cancer
IBD	Inflammatory Bowel Disease
ISH	*In situ* hybridization
LCM	Laser capture microdissection
Lib. Prep.	Library Preparation
MALDI-TOF MS	Matrix-assisted laser desorption mass spectrometry
MEM	Minimum Essential Medium
MetaHIT	Metagenomics Of The Human Intestinal Tract
Min	Minute
Mmu	Mus musculus
NaPy	Sodium Pyruvate
NCBI	National Center for Biotechnology Information
NEAA	Non Essential Aminoacids
NGS	Next Generation Sequencing
OTU	Operational Taxonomic Unit
PDX	Patient Derived Xenograft
PBS	Phosphate–Buffered Saline
PCR	Polymerase Chain Reaction

PD–1	Programmed Cell Death–1
PDL–1	Programmed Cell Death Ligand–1
QC	Quality Control
Res	Result
REV	Reverse Primer
RH	Relative Humidity
ROS	Reactive Oxygen Species
rRNA	Ribosomal Ribonucleic Acid
RT	Room Temperature
s.c.	Subcutaneous
Sec	Seconds
Seq.	Sequencing
Th1	T helper cells 1
TMR–Dextran	Tetramethylrhodamine–Dextran
Tregs	Regulatory T–Cells
WHO	World Health Organization
WT	Wildtype

1. Introduction

I. The Human Microbiome

a. *Defining The Microbiome*

Microorganisms live in closely with *"other organisms such as plants, animals and humans by establishing commensal, ammensal, mutualistic, parasitic and/or pathogenic relationships with the host"*. The human microbiome is defined as the collection of genes contained within a community of microbes (composed of bacteria, bacteriophages, fungi, archea, protists and viruses). Nowadays the term is also used to refer to the organisms themselves (*Knight et al. 2017*). In the early stages of the microbiology field, researchers were struggling with connecting the taxa of microbial communities with the phylogenetic context. One of the first attempts to classify bacteria was a morphology–based microbial systematics (*Bergey & Etris, 1936)*. But, even when adding cell staining characteristics, physiological properties, capacity for biochemical reactions and other anecdotal features to classify microbiota, it failed to identify a membership or evolutionary relatedness between the phylogenetic assembles. The microbiologist and biophysicist Carl Woese was the first to attempt using RNA components of the ribosome to build up the phylogeny tree (*Carl R.* Woese, 1967). Today he is considered to have paved the way for today's microbiome studies. His research progressively shifted the definition of taxa into the molecular level. "*He reasoned that the earliest life forms must have developed protein synthesis with RNA as components, that would be locked inflexibly with their binding partner which therefore, evolve slowly*". The fact that ribosomal RNAs evolve 100–fold more slowly than protein–coding regions in bacteria genomes, thus, confirming Woese's hypothesis has been demonstrated through DNA and RNA competition hybridization experiments (*Pace & Campbell, 1971*) long before DNA and RNA sequencing technologies were even developed. From that moment on fragmentary 16S rRNA sequence information was compared and used to provide a phylogenetic framework (*Sogin & Woese, 1971; C. R. Woese & Fox, 1977*). Even nowadays it is the state–of–art technique to identify microbiota and the number one tool used to investigate the microbiome.

b. *The Human Microbiome*

In 1977 the microbiologist Dwayne Savage stated that the human body is harboured by 100 billion microbiota cells while it only consists out of 10 trillion human cells. That report led to the widely discussed but also prominent statement that the bacteria–to–human–cells ratio is 1:10. This finding emphasized that such a quantity must have an effect on the human body and triggered a new research field, the human microbiome. This superstition has been falsely cited and quoted by many till date. Finally, Sender and his team revised the hypothesis in 2016 (*Sender, Fuchs & Milo, 2016*). They integrated the latest information and recalculated the numbers. Concluding that the human body consists of around 3.8×10^{13} bacteria cells and 3.0×10^{13} human cells in the 70 kg "reference man" (*Sender et al. 2016*). Since most studies focus on bacteria there is little data about how many microbiota cells in total make up the human microbiome. But, when considering only bacteria cells and comparing them to the amount of human cells, one can say that the ratio is more or less in the same order (*Sender*

et al. 2016). Nevertheless, at the end of their study, Sender does emphasize on the fact that even though the ratio might be less than previously assumed, it does not take away the biological importance of the human microbiome.

Already in the early 2000s many scientific groups started to study the microbiota living on and within the human body. In 2006 Gill and his team performed the first shotgun metagenomics analysis on the human distal gut using Sanger sequencing (*Gill et al. 2006*). Those early analysis already comprehend that the human microbiome was diverse among individuals (*Eckburg et al. 2005*) and determined many differences even between body sites (*Gill et al. 2006*). Until then, the research of the microbiome was still considered an exotic field. Especially, because calculating the estimate of richness and evenness of bacterial species in a community proved to be bioinformatically challenging.
The first giant human microbiome project to be launched was MetaHIT (Metagenomics of the Human Intestinal Tract) which was financed by the European Commission. This work included several institutions from all over the world that meant to establish a reference gene set of the microbes in the human GI (gastrointestinal) tract by performing metagenomics and genomic sequencing. Above everything, the goal of this project was to overcome the bioinformatical challenges and to establish the methodology necessary to characterize individual intestinal metagenomes thus, facilitating the research of bacterial genes associated with human diseases. Stool samples from 124 individuals (85 Danes and 39 Spaniards, between the age of 18 and 70) were taken (*Qin et al. 2010)*. "*1150 common bacterial species in the human gut were identified and 3.3 million genes were found. Only 75 of these microorganisms and 294 000 of the bacterial genes were shared by more than half of the subjects"*. However, it has to be taken in account that some results are influenced by the method chosen for sample collection and DNA extraction. Consequently to MetaHIT, an even bigger project was executed; the Human Microbiome Project (HMP). The findings of that project caused a breakthrough and an explosion in the human microbiome research field. The data verified the already earlier suggestions (*Eckburg et al. 2005; Gill et al. 2006*) that the microbiome was more complex and diverse than previously assumed. The statue of the HMP was similar to The Human Genome Project, emphasizing on the interest in that upcoming field. Moreover, was the human microbiome referred as our second genome in literature. The HMP was launched in 2007 and was a National Institute of Health– funded, multimillion–dollar research with many components. The project focused on identifying the human microbial flora, by collecting 15 body sites–from the skin (four sites), mouth and throat (nine sites), vagina (three sites), nostrils and faeces from 242 healthy humans both women and men between the age of 18 and 40. This data was published in two companion papers in *Nature* in 2012 (*Methé et al. 2012; The Human Microbiome Project, 2012)*. The findings revealed different patterns in the body sites. For example the communities of the stool and oral possess the most diverse microbial communities. On the other hand did the project exhibit that some body sites carried signature bacterial communities, for instance did "*all locations in the oral cavity harbour Streptococcus as major taxonomic group, but the second–most–dominant group differed between the buccal mucosa, supragingival plaque and subgingival plaque"* (*Methé et al. 2012; The Human Microbiome Project, 2012)*.

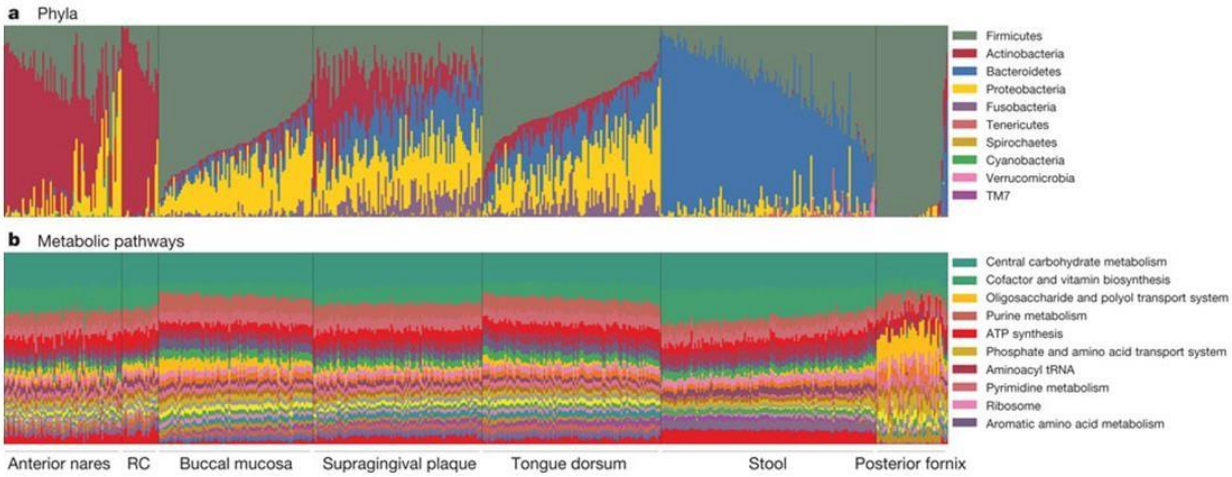

Figure 1. "*Carriage of microbial taxa varies while metabolic pathways remain stable within a healthy population: a,b,* *Vertical bars represent microbiome samples by body habitat in the seven locations with both shotgun and 16S data; bars indicate relative abundances colored by microbial phyla from binned OTUs* ***(a)*** *and metabolic modules* ***(b)****. The data indicates most abundant phyla/pathways by average within one or more body habitats; RC, retroauricular crease. A plurality of most communities' memberships consists of a single dominant phylum but this is universal neither to all body habitats nor to all individuals. Conversely, most metabolic pathways are evenly distributed and prevalent across both individuals and body habitats*" (The Human Microbiome Project, 2012).

Besides, due to the unique differences of bacterial taxa among individuals, which are sufficiently large and reproducible, the human microbiome may be able to become a powerful tool such as in forensic investigations in the future (*Fierer et al. 2010; Franzosa et al. 2015*), since the microbiome amongst cohabiting family members (*Song et al. 2013*) and even sexual partners (*Zozaya et al. 2016*) match up. Even though, the human microbiome displays a wide variety in taxonomic communities a constant could be observed across body sites and individuals when looking at functional metabolic pathways (**Figure 1.**) (The Human Microbiome Project, 2012). Also, when combining the results of MetaHIT with HMP it can be stated that the universal set of housekeeping genes in every body sites stays the same (The Human Microbiome Project, 2012) (*Qin et al. 2010*). This indicates that even though differences can be seen among individuals, there is still a pattern across the human microbiome. Furthermore, it appeared that at the level of taxonomy and function, humans closely resemble other omnivorous primates such as chimpanzees and in general mammals under similar conditions. **Figure 2** exhibit the data collected by the HMP as a map of the diversity in the human microbiome where the four most abundant phyla in the human microbiome are *Actinobacteria*, *Bacteroidetes*, *Firmictues* and *Proteobacteria*, their abundances differing between habitats.
Due to the differences of the outcome of the HMP and MetaHIT in samples choice (Nationality of subjects and age range), it can be said, that the human microbiome depends on age, geography, dietary, developmental, environmental and genetic factors.

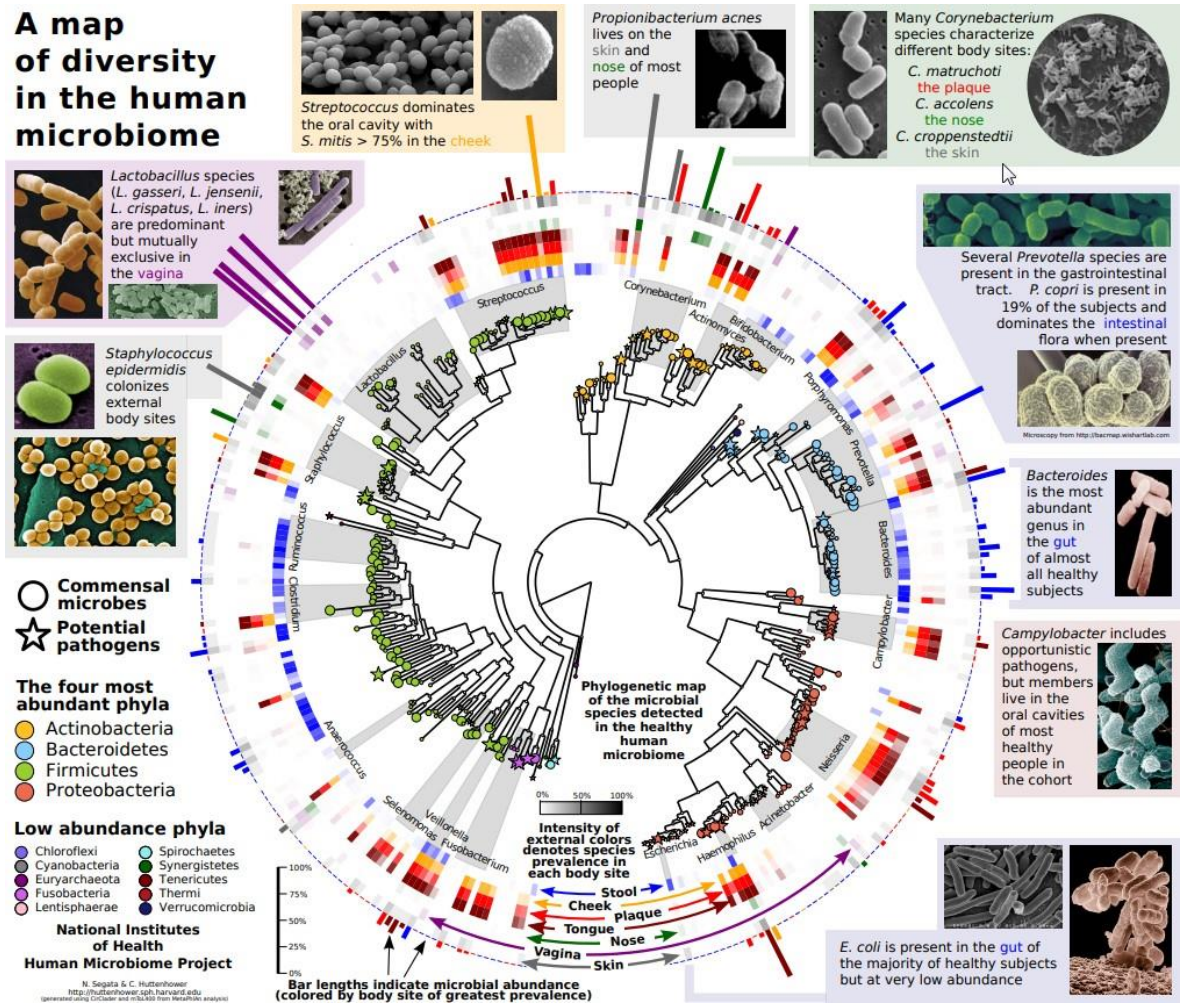

Figure 2. "*A map of diversity in the human microbiome: In the center is a phylogenetic tree of organisms abundant in the human microbiome. Commensal microbes are indicated by circles and potential pathogens by stars. The middle ring corresponds to body sites at which the various taxa are abundant and is color–coded by sites. The bar heights on the outside of the circle are proportional to taxa abundance at the body site of greatest prevalence. The intensity of external colors corresponds to species prevalence in each body site*". (*Morgan, Segata, & Huttenhower, 2013*)

The HMP project has paved the baseline for further microbiome investigation. After HMP other microbiome related studies have been executed like the American Project and Earth Microbiome Project to name a few. These findings yield to a rich picture of the commensal host–microbe and microbe–microbe interactions. The next frontier would be to harness this information by embarking on studies that explore how the microbiome affects human health.

c. *Microbiome And Health*

The HMP and MetaHIT managed to establish the profiling tools to make the search easier for the association of microbial genes within human health and disease easier. In the prior chapter, the importance of the human microbiome has been already emphasized upon and it has been indicated that it has an influence on human health due to the microbiota–host and microbiota–microbiota interactions. In this chapter we will go into detail of how the microbiota within and on the human body has been associated with human diseases.

Even though, abundances of individual members of bacteria fluctuate, the membership of the microbiome, especially in the gut, is strictly controlled. Breach of this delicate equilibrium can result in opportunistic infections. This so–called microbiome dysbiosis can cause invasion of exogenous bacteria or outgrowth of certain endogenous bacteria. Other classical infectious diseases are often developed due to exposure of susceptible hosts to pathogens like *Vibrio cholera* in cholera. However, these kinds of infections are not of importance for this study and therefore, will not be further discussed. In this study the focus relies on how the human microbiome itself can influence the health of the host.
Since the human microbiome co–evolves with its host, it is constantly influenced by a variety of factors such as the host's digestion and nutrition (*Hehemann et al. 2010; Turnbaugh & Gordon, 2009)* detoxification, body defence (*Relman, 2012)*, maturation of the immune system (*Turnbaugh & Gordon, 2009*) and disease mediation (*Baldridge et al. 2015; Hashimoto et al. 2012)*. Commensal bacteria have been demonstrated to crosstalk with its host intestinal epithelial and lymphatic cells therefore, bacteria harbouring the human body are able to regulate the host immune system's homeostasis (*Filyk & Osborne, 2016*). Trivial examples of the microbiome influencing human health in general would be dental caries and bacterial vaginosis. But also other diseases like acne vulgaris (*Fitz-Gibbon et al. 2013)*, atopic dermatitis (*Kobayashi et al. 2015*) and chronic skin wounds *(Bessa, Fazii, Di Giulio, & Cellini, 2015)* have been linked to the skin microbiome while astherosclerosis is associated with the translocation of oral microbes into atherosclerotic plaques (*Jonsson & Bäckhed, 2016)*. More complex diseases like Parkinson's disease, asthma and rheumatoid arthritis also seem to be influenced by the microbiome (*Abdel-Aziz et al.2019*; *Sampson et al. 2016;*
X. Zhang et al. 2015). Nevertheless, the gut microbiome remains the main focus of human microbiome research.

In **Table 1** a partial list of diseases which were linked to the human gut microbiome is stated. Due to the relevance of the microbe–gut–brain axis, it is of no surprise that even psychopathological diseases like anorexia (*Kleiman et al. 2015*) and depression *(Zheng et al. 2016*) can be linked to the microbiota in the human gut. The HMP also discovered novel OTUs (Operational Taxonomic Unit) in low abundances (< 2%) which belonged mostly to the little known *Barnesiella* genus and *Dorea, Oscillibactera* and *Desulfovibrio* genera. After further investigation of these low abundance genera they could be associated with colon cancer, dietary shifts and opportunistic infections (*Wylie et al. 2012*) thus, demonstrating that even low abundance bacteria might also play a role in influencing human's health once they overgrow.

A difficult challenge of any research that associates the microbiota in the human body to a disease is to address the question whether the alterations of the microbiome are a cause or an effect of the disease. Also, sometimes the microbiome might not directly cause a condition but subsequently, a shift in the community increases the risk of developing a type

of disease. A prominent example for such a case would be the bacterial species *Helicobacter pylori*. This species was the first bacterium ever to be classified as carcinogen in 2012 by the IARC and is till date, the only bacterium officially classified as carcinogen I (is defined as an carcinogenic agent to humans). However, only 1.7% of the population that is infected with this strain develops gastric or stomach carcinoma (*Fuccio, Zagari, Eusebi, & et al. 2009*). Hence, it is not completely trivial whether the species is a cause or a consequence of those types of cancer. The ability to transfer human phenotypes to germ–free mice has been proven to be an important method when determining causality. A well–known example was the research conducted on twins, where fecal microbiota from adult female twin pairs discordant for obesity was transplanted into germ–free fed low–fat mouse chow (*Ridaura et al. 2013)* which resulted in an obesity associated metabolic phenotype in mice. This effect was reversible. The result determined the causality of the microbiota for obesity. However, these kinds of experiments are not that easily executed with other diseases since many phenotypes are more complex. Nevertheless, it is seen in multiple cases that the microbiota on and within the human body have an effect on diseases, their development and treatment. These findings implicate that a dysbiosis of the human microbiome subsequently, impact our well–being, whether it is a cause or effect. Therefore, researchers are trying to profile the microbiome in healthy bodies compared to non–healthy ones for therapeutic applications. Microbiota could be used as biomarkers or indication for specific diseases just like the bacterial strain *Helicobacter pylori* (*Fuccio et al. 2009*) or might even be used as therapeutic agent, like transplantation of faeces to reverse obesity *(Ridaura et al. 2013*). "*One challenge in therapeutic applications of the microbiome is that microbiomes are unique to each individual and change over time. These properties complicate both diagnostic uses of the microbiome as well as its viability as a therapeutic agent*". On the other hand diseases such as *Clostridium difficile* infection exist, where the shift in the microbial community is so remarkable (*Weingarden et al. 2015*) that it actually can be used as a therapeutic target. Unfortunately, "*the picture is less clear for many other associations of the microbiome with disease. Just like in research on the human genome, it is necessary to understand the scope of diversity associated with different human populations"*. The research on finding the fundamental mechanistic behind certain phenotypes is very active since it is expected "*that some effects on the microbiome might be so large that they can be identified and exploited in a systematic way for any individual"*.

Table 1. A partial list of diseases that have been linked to the gut microbiome

Disease	Description	Reference
Acute anorexia	Changes in taxa abundances of patients with anorexia nervosa were found	(*Kleiman et al. 2015*)
Addiction	In mouse models of addiction, antibiotic treatment increased addictive behavior when receiving low–dose opioids	(*Kiraly et al. 2016*)
Asthma	Differences in microbiota composition between asthmatic and non-asthmatic patients	(*Abdel-Aziz et al.2019*)
Autism	Differences in gut microbial communities between children with autism and neurotypical controls	(*Ding, Taur, & Walkup, 2017*)
Cardiovascular disease	Diet and the gut microbiome were linked to cardiovascular disease risk	(*Tang et al. 2014*)
Depression	Dysbiosis of the gut microbiome showed to have a role in depressive–like behaviors	(*Zheng et al. 2016*)
Inflammatory bowel disease	Adherent enterobacteria may have promoted gut inflammation disease	(*Kostic, Xavier, & Gevers, 2014*)
Irratiable bowel syndrome	Patients with irratiable bowel syndrome showed mucosal and luminal gut microbial changes	(*Simrén et al. 2013*)
Childhood Malnutrition	Altered gut microbiome was strongly linked with childhood malnutrition	(*Kane et al. 2014*)
Multiple sclerosis	Gut microbiota changes may be related with pathology of multiple sclerosis	(*Jangi et al. 2016*)
Obesity (metabolic disease)	Gut microbiomes of obese individuals showed an increased capacity to harvest energy from a diet.	(*Turnbaugh et al. 2006*)
Osteoporosis	The gut microbiome had both direct and indirect effects on deregulated bone remodeling.	(*Hernandez et al. 2016*)

II. Mouse Models For Human Microbiome Research

Even though yeasts, worms and flies are excellent models for studying developmental processes, mice have still proven to be the better scientific tool for probing complex physiological systems that mammals share. Rodents in general, but especially mice, have become the most studied disease models for pharmaceutical research (*Vandamme, 2014*). Principally, because they possess a genome that can be easily manipulated, analysed and harbours genetic and physiological similarities to humans which make them the preferred model organism for *in vivo* studies.

Studies of local microbiota at different locations of humans but especially the human intestinal tract have already been conducted. These studies normally require rather invasive sampling methods which cannot be scaled for practical and ethnical reasons (*Gill et al. 2006*) (*Zoetendal et al. 2012*) (The Human Microbiome Project, 2012). Even though, the ethnical consideration also applies to mouse models, they provide an easy way to collect many samples from different sites and enable multiple comparisons at larger scale. Also, since the animal breeding and keeping is intensively controlled, the data collected from *in vivo* mouse models are far more consistent, reproducible and easier supervised. Factors like geographical, dietary, developmental differences etc. are standardized and self–managed. Furthermore, mouse models allow perturbations in gut microbiota thus, helping in assessing causalities of the complex host–microbiota interactions and even developing mechanistic hypotheses.
However, when conducting microbiome studies which include mice, the biggest challenge is the translation of those studies to humans. Since, even though there are major spatio–temporal similarities, differences between the human and mouse microbiome exist which need to be considered. Both organisms have similarities when it comes to physiology, and both harbour common bacteria genera, but differences in abundances. This is why the genetic background of the chosen mouse model is very crucial and needs to be regarded.
Another reason why mouse models are crucial for microbiome studies is that the majority of bacterial species are not routinely culturable since they rely on host–microbiota and microbiota–microbiota interactions to grow (*Raymond et al. 2015*). This is why when investigating the microbiome most studies rely on faeces isolation and transplantation (a technique called faeces microbiota transplantation) in order to transfer the microbiota. However, when applying that technique one mostly considers the gut microbiome. But studies on mouse models enable to investigate the microbiome as a whole since host– microbiota and microbiota–interactions can directly be investigated thus, yielding more direct results.
Even though the mouse has been used for microbiome studies there is no “Mouse Microbiome” published yet since, every mouse strain harbours a distinct microbiome (*Elderman et al. 2018*). In fact, since the main interest is the gut microbiome that of human and mouse has been compared and studied exclusively (*Floor Hugenholtz & De Vos, 2017; Nguyen et al. 2015; Xiao et al. 2015)*. Hugenholtz and his team compared physiology and anatomy of mouse and human. There are big differences concerning the mucus layer, intestine weight, speed of digestion, pH differences in the stomach and the anatomy of the tract which mostly can be tracked back to the fact that mice are herbivores while humans are omnivores. These differences are crucial since they lead to the notable discrepancy of bacteria genera between both organisms. However, human and mice resemble at phyla level, for example in both organisms *Proteobacteria, Firmicutes* and *Bacteroidetes* make up more than 70% of the gut microbiota (*Xiao et al. 2015*). This seems to apply also to many other mammals, herbivores and carnivores (*Ley, Turnbaugh, Klein, & Gordon, 2006*). This phenomenon is not new, since already after the Human Genome Project was executed and compared to the mouse and chimp genome, similarities could be found (*"Initial sequence of the chimpanzee genome and comparison with the human genome," 2005*) . On the other hand there are phyla which do not exist in both organisms. A bacterial phylum which makes up a large amount in mice intestine but barely appears in the human intestine is *Deferribacteres*, (*Bik et al. 2006; Eppig et al. 2015*) . Mice also harbour a specific bacterial strain called *Candidatus arthromitus* which belongs to the *Firmicutes* phylum and is part of the *Lachnospiraceae* family (*SNEL et al. 1995*). The strain which humans lack is a

segmented filamentous bacterium which has a pronounced effect on the maturation of the innate immune system.
Bacterial species differences between human and mice can result in the failure of some drug evaluations, as it already happened in the past. "*Anaerobic species of Bifidobacterium longum, lactic acid bacteria, Clostridium novyi and Salmonella typhimurium have been used as an anticancer protein–delivery agent to repress tumour growth for their selective growth in the hypoxic environment of large solid tumours"* (*Fujimori, 2008; Low et al. 1999; Zeuthen, Christensen, & Frøkiær, 2006; Zhao et al. 2007*). However, the clinical trials of *Salmonella* failed, since it failed to fully colonize human tumours even though it worked in murine models. Some researchers have indicated that functional differences might be the reason for the failure of *Salmonella* (*Nemunaitis et al. 2003; Toso et al. 2002)*. Therefore, it is crucial to investigate the bacterial profile of the used *in vivo* models to make them comparable with humans.

Overall, it can be said that for investigating basic initial host–microbe interactions, germ–free mice provide an adequate model (*Faith et al. 2010*).

III. The Oncobiome

a. Cancer

Cancer is the second most common death cause worldwide. According to the WHO 8.8 million people died worldwide from cancer in 2015 (http://www.who.int/news-room/fact-sheets/detail/cancer). Moreover, is the number of people harbouring cancer expected to increase due to the growth and aging of the world population. Cancer is a complex disease which involves abnormal cell growth and bears the potential to invade and spread to other parts of the body. Cancer arises when cells acquire enough driver mutations to enter into a tumourigenic state and gain an uncontrolled replicative potential. The disease is influenced by interplay between host genetic, overweight and environmental factors among other external agents like physical, chemical or biological carcinogens which can cause cancer.

A correct cancer diagnosis is crucial for effective treatment since every cancer type requires a specific treatment. The high complexity of cancer as well as the problem of acquired resistance makes therapeutic and diagnostic very demanding which is why there is a constant search for new approaches for treatment and diagnosis to improve already established ones. However, numerous studies have revealed that microorganisms can also influence cancer biology.

b. The Microbiome And Cancer

"In recent years there has been increasing recognition that humans harbour an extensive microbial community within and on the body, and that alterations of this biome can lead to profound alterations in health" (*Collins, Hogan, & Winter, 2011*). So far, different microorganisms were identified which have been associated with carcinogenesis. The most prominent examples are *Helicobacter pylori* and oral *Fusobacterium* which are mostly associated with colorectal and stomach cancer (*Collins, Hogan, & Winter, 2011*). The bacteria belong either to the oral microbiome in case

of *Fusobacterium* or gut microbiome in case of *Helicobacter pylori*. Other examples for bacteria that have been identified to play a role in carcinogenesis are stated in **Table 2**.

There is another aspect which connects microbiota to cancer. It seems that microbiota are not only able to influence the cancer development but also the therapeutic outcomes due to the ability of bacteria to metabolize drugs and modulate host inflammation as well as immune responses (*Pope et al. 2017*). Tumours respond differently to chemo–and immunotherapy and the therapeutic failure is often unknown. A deeper insight of this phenomenon is necessary to understand it better as well as to improve existing therapeutic regimes, while informing the development of new treatments. Different aspects of anticancer therapies ask for improvement. For example, do many anticancer drugs harbour a high toxicity and the incidence of resistance is increasingly becoming an issue. Multiple microbiota have been linked with various cancer types in a number of ways (*Goodman & Gardner, 2018*). The microorganisms within and on the host seem to interfere with host–target anticancer therapies in a direct or indirect matter, yielding to three clinical outcomes; either they facilitate drug efficacy, abrogate and compromise anticancer effect or mediate toxicity (*Alexander et al. 2017)*.

Table 2. Overview of bacteria identified in different tumour types

Cancer	Bacterium	Reference
Lung cancer	*Streptococcus mitis*	(*Apostolou et al. 2011*)
	Staphylococcus epidermis	
	Bacillus sp.	
	Mycoplasma sp.	(*Rogers, 2011*)
Breast cancer	*Staphylocccus epidermidis*	(*Cantwell & Kelso, 1981*)
	Mycoplasma sp.	(*Rogers, 2011*)
Oral cancer	*Prevotella sp.*	(*Hooper et al. 2007*)
	Fusobacterium naviforme	
	Mycoplasma sp.	
Gall–bladder carcinoma	*Salmonella typhi*	(*Nath, Singh, & Shukla, 1997*)
	H.pylori, *H. hepaticus*	(*de Martel et al. 2009*)
Ovarian cancer	*Chlamydia trachomatis*	(*Shanmughapriya et al. 2012*)
	Mycoplasma sp.	(*Rogers, 2011*)
Colorectal cancer	*Fusobacteriuma nucleatum*	(*Castellarin et al. 2012*)
	Faecalibacterium sp.	(*Marchesi et al. 2011*)
	Escherichia coli.	(*Hubbard et al. 1998*)
	H. pylori	
	Mycoplasma sp.	(*Rogers, 2011*)

Therefore, interactions between microbiota and anticancer drugs has been increasingly studied. An example of how microorganisms can influence the anticancer therapy outcome was studied by Lehouritis and his team. They exhibited that *Eschirichia coli* is able to increase the cytotoxicity of the alkylating agent CB 1954 (*Lehouritis et al. 2015)*. The increased effect is based on nitroreductase activity of *Eschirichia coli*.
Platinum compounds which are often used for chemotherapy and have been investigated in association with microorganisms (*Iida et al. 2013*), are another example of how microbiota are able to influence effects of anticancer drugs. This type of compounds form platinum–DNA adducts which block DNA replication and stimulate ROS (Reactive Oxygen Species)

production, which can cause oxidative damage or lead to stimulation of the immune response. Iida and his team treated mice bearing EL4 lymphoma with a cocktail of antibiotics additionally to oxaliplatin treatment yielding to reduction of cancer regression. However, when repeating the same experiment with cisplatin instead of oxaliplatin, the therapy failed. The failure was identified as a decreased microbiota–dependent ROS production, elucidating the key role of bacteria in chemotherapy (*Iida et al. 2013)*. This is especially interesting since oxaliplatin, in contrast to cisplatin, is able to induce immunogenic cell death in *in vitro* colorectal spheroids (*Wernitznig et al. 2019*).

Yet an example of the concept of how microbiota influences not the therapy itself but its side effects is the exemplification of the pro–drug irinotecan. This cytostatic drug is mainly used to treat advanced colorectal cancer and becomes activated once the SN–38 metabolite is inactivated into SN–38G in the liver by glucuronisidation and then, eliminated through biliary excretion. When reaching the intestine the bacterial metabolite β–glucuronidases is able to reactivate SN–38G to SN–38 thus, restoring the drug activity (*Wallace et al. 2010*) which subsequently, is responsible for the appearing intestinal toxicity. When animals are treated with irinotecan and a specific inhibitor of the bacterial β–glucuronidases, the normally appearing side effects like colonic damages and diarrhea are prevented (Wallace et al. 2010). More examples of how chemotherapeutic agents shape the microbiome are stated in **Table 3**.

Bacteria are not only able to perturb chemotherapy but also immunotherapy, due to their ability to modulate inflammation and immunity. Till date identifying the patients who are most likely to benefit from immune checkpoint inhibitors is a hot topic (*Ma et al. 2016*), since it could lead to resistance avoidance and enhancement of the efficacy of such treatment. A prominent example for immunotherapeutic strategies are the immune checkpoint inhibitors, CTLA–4 and PD–1. In both cases, microorganisms seem to influence their efficacies. PD–1 is a programmed cell death–1 protein that binds to its ligand PDL–1. An example of such a drug would be Nivolumab (Opdivo®) which blocks PD–1 a negative regulator of T–cell activation and response, thus allowing the immune system to attack the tumour. Analysis of OTUs from patients treated with PD–1 revealed that patients enriched in *Clostridiales/Ruminococcaceae* were prone to respond to the treatment effectively, contrary to ones enriched in *Bacteroidales* (*Gopalakrishnan et al. 2018*). The responders seemed to have an overrepresentation of effector T–cells, while patients with the overrepresented phylum *Bacteroidales* had more Tregs and myeloid–derived suppressor cells. This study acknowledges that bacteria profiling might be an important tool to validate if patients are suited for this kind of treatment. Ipilimumab (Yervoy®) is a monoclonal human antibody against CTLA–4 and was the first immune checkpoint blockade drug to be approved by the FDA in the mid–1990s for the treatment of metastatic melanoma. It blocks CTLA–4 therefore, potentiating T–cells in response to tumour–associated antigens. It was found that its efficacy depends on gut microbiota. This was verified by germ–free mice treated with CTLA–4 antibodies or mouse cancer models treated with CTLA–4 antibodies and an antibiotic cocktail, where the drug–response was abolished (*Vétizou et al. 2015*). Specifically *Bacteroidetes thetaiotaomicron* and *Bacteroidetes fragilis* were identified as key players for blocking CTLA–4 and PD–1 efficacy through the induction of Th1 immune response (*Vétizou et al. 2015*).

Table 3. List of chemotherapeutic treatments and their microbiota profile

Chemotherapeutic treatment	Microbiota modifications	Reference
5–Fluorouracil	Increased Gram –negative anaerobes	(*Bültzingslöwen et al. 2003*)
	Increased translocation to mesenteric lymph nodes	(*Stringer et al. 2009*)
Irinotecan	Increased *Clostridium cluster XI* (including *Peptoclostridium difficile*) and *Enterobacteriaceae*	(*Lin et al. 2012*)
Gemcitabine	Increased *Proteobacteria, Verrucomicrobia, Akkermansia muciniphila, Escherichia coli* and *Peptoclostridium difficile;* decreased *Firmicutes, Bacteroidetes, Bacteroidales, Lachnospiraceae, Ruminococcaceae.*	(*Panebianco et al. 2018*)

The aim of scientific approaches is mostly to reshape the microbiome to create an adequate environment for anticancer therapy hence, keeping side effects to a minimum and reaching a higher drug efficacy. The approaches are very distinctive and suggest different concomitant therapies like the prescription of probiotics, prebiotics, synbiotics, postbiotics, antibiotics and inhibitors for specific bacterial genes and metabolites which might shift the microbiome back to its eubiosis or prevent drug alteration, reaching the optimum level of success for anticancer therapy (*Zitvogel et al. 2015*). Due to the high influence of the microbiome towards the health state of human beings, there is a great interest to develop drugs that impact the microbiome. Many of the recently developed drugs are part of a new field called the microbiota restoration therapy. The industrial interest began in 2011 with two microbiome modulators being in an active state of development. Since then the number of drugs in development has increased to 70 (*Sally et al. 2018)*. However, most of these drugs are still at early stages and only few have reached clinical trials (**Figure 3.**). So far, no microbiome based drug has been approved for use as human therapeutics.

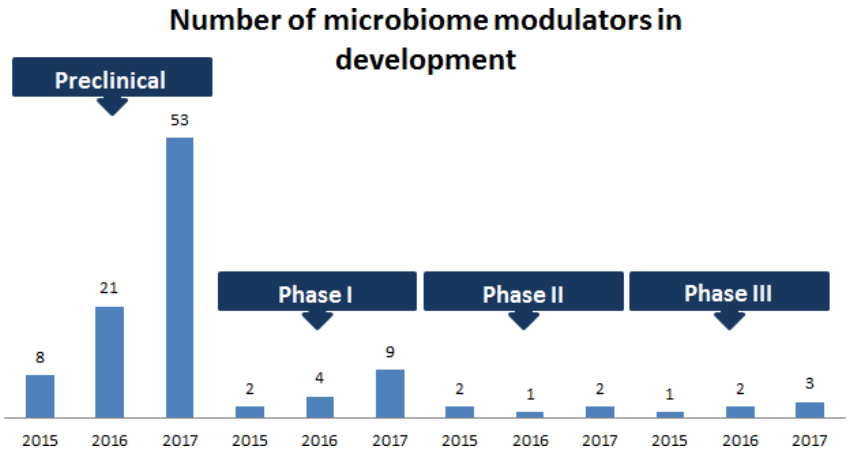

Figure 3. Number of microbiome modulators at each development phase from 2015 – 2017 (Taken from: Pharma Intelligence©Informa UK Ltd 2018)

The probably better known companies which are currently working on the development of microbiome based drugs are the American biotech's Rebiotix and Seres Therapeutics as well as the UK company Vedanta Biosciences, Inc. and the French based companies Enterome and Eligo Bioscience. However, each one of these companies has different approaches and targets. No big pharma company has yet originated any microbiome drug development technologies, instead they collaborate with some of the companies mentioned above. Janssen (Johnson & Johnson GmbH) cooperated with Vedanta to in–license their preclinical asset, VE–202 in 2015, a drug that aims on treatment of inflammatory bowel disease (IBD), aside from Vedanta, Janssen (Johnson & Johnson GmbH) has also tied knots with Enterome in 2016. This collaboration aims on discovering bioactive molecules as well as novel targets of the gut microbiome, in order to include them in the treatment of Crohn's disease. The biggest interest of microbiome based biotech's is in the fields of infections and gastrointestinal disorders. Nevertheless, with more research being executed new fields are being explored like metabolic disorders, dermatological indication, neurological disorder and cancer. In the field of microbiome therapeutics for the treatment of cancer Enterome and Seres Therapeutics have entered the leading candidates E02315 and SER–401. E02315 an immune–oncology candidate by Enterome is currently in preclinical development for the treatment of glioblastoma multiforme (GBM). This candidate is a microbiome–derived cancer vaccine and exposes a patient's immune system to antigens specific to GBM via the gut, thought to boost the immune response and allow cancer cells to be more visible to T–cells. The other candidate SER–401 by Seres Therapeutics is a combination treatment, where live bacteria is supposed to be delivered alongside immunotherapy medication (anti–PD–1) (*Gopalakrishnan et al. 2018*), to improve its efficacy and safety. These recent developments display the increased interest and potential of the microbiome as a tool for anticancer treatment.

c. The Oncobiome

The prior chapter states bacteria from the different microbiomes within or on the human body which influence carcinogenesis or anticancer therapy. However, a new interesting finding has surfaced that changed the approach of microbiome analysis once more; bacteria within the tumour tissue. Till date, there are only two publications which use the term "**oncobiome**" itself (*Banerjee et al. 2017; Thomas & Jobin, 2015*). In the review of Thomas and Jobin the oncobiome is defined as the relationship between cancer and the host microbiota. In our study we define the oncobiome as *the host microbiota that translocate into the tumour tissue*, Therefore, building up a "tumourbiome", or as we call it oncobiome. An example for an oncobiome research would be the one conducted by Banerjee who attempted to define the ovarian cancer oncobiome. In their study they identified besides bacteria, also viruses, fungi and parasites within the ovarian cancer tissue (*Banerjee et al. 2017*). They identified the tumour's specific ovarian tumour microbiome signature compared to the surrounding tissue. This outcome may enable new insights for the development of targeted therapeutics for this type of cancer.

Another approach was the finding of intratumoural bacteria in colon cancer and its effect on anticancer therapy. In a colon cancer mouse model where mice exhibited gemcitabine resistance, the circumstance could be tracked back to *Gammaproteobacteria* (*Geller et al. 2017*). Intratumoural *Gammaproteobacteria* which express the bacterial enzyme cytidine deaminase (CDD_L) were able to metabolize gemcitabine (2',2'–difluorodeoxycytidine) into its inactive form 2',2'–difluorodeoxyuridine. Thereby, making Gemcitabine into a possible pro-

drug. Moreover, was the resistance mechanism reversible when eradicating the CDD_L expressing bacteria by antibiotic treatment. Another study conducted by Bullman and her lab revealed that intratumoural bacteria are able to migrate with the cancer cells (*Bullman et al. 2017)*. Previous studies have identified *Fusobacterium nucleatum* to be notably enriched in human colon cancers (*Castellarin et al. 2012*). Bullman and her team uncovered that when colonizing human colorectal cancers with *Fusobacterium,* the bacterium could also be detected in colon derived liver metastases. Therefore, they demonstrated that the "oncobiome" is stable even between paired primary–metastatic tumours (*Bullman et al. 2017*). The "*presence of bacteria within the tumours could be due to infection via the vasculature, and their ability to survive and grow, due to the presence of nutrients within the hypoxic region of the tumour at a later stage in tumour growth"* (*Lehouritis et al. 2015*). Those special abilities would nominate bacteria as an ideal delivery system for therapeutic genes to, for example inhibit tumour growth (*Osswald et al.* 2015). However, the question whether and how bacteria are able to translocate into the tumour has not been answered so far.
There is still a lot of research to be done to find the mechanisms behind the drive of bacteria to translocate into cancerous tissue. Nevertheless, do all these findings indicate that something like an own oncobiome exist and might be as much of a crucial player in anticancer treatment as the gut microbiome or others.

d. KRAS Mutation

Different causes for the bacteria–uptake into the tumour have been investigated. Recent studies suggest that bacteria can translocate into cancer cells through activation of macropinocytosis (*G. Redelman-Sidi et al. 2013*). In a study, Bacille Calmette–Guerin (BCG) were taken up by bladder cancer cell lines through macropinocytosis which was based on KRAS expression (*Gil Redelman-Sidi et al. 2013*). They discovered, that activated Ras activates macropinocytosis which is the pathway for BCG uptake. Other than Ras, the PTEN/PI3K and Cdc42–Rac1–Pak1 which are known to be interconnected also play a crucial role in this macropinocytosis. These pathways also lead oncogenesis, therefore, making it a good target for therapies. A similar investigation was done, looking at pancreatic cancer cell lines. Here the pancreatic adenocarcinoma–derived human Mia_PaCa-2 cells which are homozygous for the $KRAS^{G12C}$ (*Lopez-Crapez et al. 1997*) allele, displayed appreciably higher levels of TMR–dextran uptake compared to BxPC-3 cells, which express wild–type KRAS (*C. Commisso et al. 2013*). The TMR–dextran labelling in the oncogenic Ras–expressing cells reflects uptake via macropinocytosis. When knocking down KRAS the macropinocytosis was attenuated therefore, confirming once more the dependence of this mechanism on oncogenic Ras expression. The same experiment was repeated with bladder carcinoma–derived T24 cells, which are homozygous for the $HRAS^{G12V}$ (*Capon et al. 1983*) allele and 5637 cells which express wild–type HRAS yielding the same results.

Therefore, it might be interesting to see whether similar phenomena can be observed in mouse models.

2. Materials & Methods

I. Animal Models

"*BomTac: NMRI-Foxn1^{nu} female mice were provided by Taconic Denmark. The mice were transferred to the local AAALAC accredited animal facility at Boehringer Ingelheim RCV GmbH & Co KG with an age of 6–8 weeks. After arrival at the animal facility, mice were allowed to adjust to conditions at least for 5 days before they were used for experiments. Standardized diet (PROVIMI KLIBA) and autoclaved tap water were provided ad libitum. Mice were group–housed (10 mice per cage) under pathogen-free and controlled environmental conditions (21 ± 1.5°C temperature, 55 ± 10% humidity) and handled according to the institutional, governmental and European Union guidelines (Austrian Animal Protection Laws, GV-SOLAS and FELASA guidelines). Animal studies were approved by the internal ethics committee and the local governmental committee*" (GZ: 1029536/2017/19).

II. Human Cancer Cell Derived Xenografts

10 NMRI-*Foxn1*^{nu} female mice per cell line were injected s.c. Prior injection, the injection site on the skin of each mouse was disinfected with Betaisodona® to provide a sterile site of injection. The injection volume of the cells was 0.1 mL per mouse and cell numbers are listed in **Table 4.** The cells were either suspended with PBS x 1 (gibco® by Life Technologies, USA) with 5% FCS, detailed information is listed in **Table 4** (gibco® by Life Technologies, USA) only or mixed in a 1:1 ratio with Corning®Matrigel®Matrix (Matrigel) (Discovery Labware Inc., USA). Cell numbers for injection as well as suspension media were taken from previous in–house data. The CDXs were measured two to three times a week (depending on their predicted growth rate), once they have reached a size of approximately 150mm³ (Tumour size was calculated as following; $L \times W^2 \times \frac{\pi}{6}$, where L is the length of the tumour and W the width). Additionally, the mice were chipped for identification. Mice were euthanized in case of necrotic skin legions of such that the integrity of the skin was not given anymore thus, those CDXs were disqualified for further analysis. Mice without tumour growth were euthanized as well as mice with a tumour size > 1500 mm³. Subsequently, tumour, skin, colon and tongue were taken out under a sterile workbench. Surgical instruments were autoclaved and opened exclusively under the sterile lamina. The instruments were changed in between cell lines, tumour, skin, colon and tongue tissue to avoid contamination. The samples were then fresh frozen on top of dry ice followed by immediate DNA extraction.

Table 4. Injection pattern of human cancer cell lines in mice

Cell line	Cells injected sc. (Mantissa/Exponent)	Suspended in
AsPC-1	5.0x10^6.0	PBS (5% FCS)
BxPC-3	5.0x10^6.0	PBS (5% FCS)/Matrigel (1:1)
CAPAN-1	5.0x10^6.0	PBS (5% FCS)
HCC1806	5.0x10^6.0	PBS (5% FCS)
HT-1197	5.0x10^6.0	PBS (5% FCS)/Matrigel (1:1)
HuP-T4	1.0x10^6.0	PBS (5% FCS)/Matrigel (1:1)
MDA-MB-231	5.0x10^6.0	PBS (5% FCS)/Matrigel (1:1)
MDA-MB-468	5.0x10^6.0	PBS (5% FCS)
MIA_PaCa-2	1.0x10^7.0	PBS (5% FCS)/Matrigel (1:1)
SUM190PT	5.0x10^6.0	PBS (5% FCS)/Matrigel (1:1)
UM-UC-3	5.0x10^6.0	PBS (5% FCS)/Matrigel (1:1)

III. Cell Lines

Human cancer cell lines stated in **Table 5** were either obtained from the American Type Culture Collection (ATCC, USA), Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (Leibniz-Institut DSMZ, Germany) or Asternad (BioIVT, USA). The cell lines which were selected according to following criteria:

- Literature associated with bacteria translocating into the tumour via micropinocytosis facilitated by KRAS Mutation (*Commisso et al. 2013; G. Redelman-Sidi et al. 2013*)
- KRAS Mutation
- KRAS WT

Furthermore, only cell lines which were already established in–house *in vivo* in NMRI-*Foxn1*^{nu} female mice were used.

The cell lines AsPC-1, BxPC-3, CAPAN-1, HCC1806, HuP-T4, HT-1197, MIA_Paca-2, SUM190PT and UM-UC-3 were maintained at 37°C under 5.0% CO_2 and 90% RH in a Forma™ Steri-Cult™ CO_2 Incubator (Thermo Scientific™, USA). MDA-MB-231 and MDA-MB- 468 were maintained at 37 °C under 0.0 % CO_2 and 90 % RH in an incubator INCO153med Memmert GmbH + Co. KG. AsPC-1, Capan-1, BxPC-3, HCC1806, MDA-MB-231, MDA-MB-468 were cultured in appropriate conditions as recommended by ATCC. HuP-T4 was cultured in appropriate conditions as suggested by DSMZ. HT-1197 and MIA_Paca-2 were cultured as proposed by ATCC however, additionally 1% MEM NEAA 100X (Gibco® by Life Technologies ,USA) and 1% NaPy 100mM (100X) (Gibco® by Life Technologies, USA) was added to the media. UM-UC-3 was also cultured as recommended by ATCC, nevertheless, 1% NaPy 100mM (100X) (Gibco® by Life Technologies, USA) was added to the media.

Table 5. Cell line list choice according to criteria: tumour type and KRAS Mutation status

Cell line	Tumour type	KRAS Mutation	Reference
AsPC-1 **(ATCC® No. CRL-1682™)**	Pancreas carcinoma	C35G>A	(*Geller et al. 2017*)
Capan-1 **(ATCC® No. HTB-79™)**	Pancreas carcinoma	C35G>T	-
HuP-T4 **(DSMZ No. ACC 223)**	Pancreas carcinoma	C35G>T	-
MIA_Paca-2 **(ATCC No. CRL-1420)**	Pancreas carcinoma	C34G>T	(*Commisso et al. 2013*)
BxPC-3 **(ATCC® No. CRL-1687™)**	Pancreas carcinoma	WT	(*Commisso et al. 2013*) (*Geller et al. 2017*)
HCC1806 **(ATCC® No. CRL-2335™)**	Breast carcinoma	WT	-
SUM190PT **(Asternad SUM-190PT)**	Breast carcinoma	WT	-
MDA-MB-468® **(ATCC® HTB-132™)**	Breast carcinoma	WT	-
MDA-MB-231 **(ATCC® No. CRM-HTB-26™)**	Breast carcinoma	C38G>A	-
UM-UC-3 **(ATCC No. CRL-1749)**	Bladder carcinoma	C34G>T	(*Redelman-Sidi et al. 2013*)
HT-1197 **(ATCC No. CRL-1473)**	Bladder carcinoma	WT	(*Redelman-Sidi et al. 2013*)

IV. DNA Extraction

Once the samples were collected, the genomic DNA was extracted and isolated under sterile conditions. The cell line samples were extracted using the QIAamp® DNA Mini kit (QIAGEN GmbH, Germany) the procedure was done according to the manufactor's instructions. The tissue samples were extracted using QIAamp® Fast DNA Tissue Kit (QIAGEN GmbH, Germany) also following the protocol provided by the manufactor. The homogenization step was done differently, after adding 200 µl AVE, 40 µl VXL, 1 µl DX Reagent, 20 µl Proteinase K, 4 µl RNase A (100 mg/ml) to each sample, Stainless Steel Beads 5 mm (200) (QIAGEN GmbH, Germany) were added, the samples were then homogenized with TissueLyser II (QIAGEN GmbH, Germany) 1 x 24Hz for 30s. Thenceforth, the protocol was followed as suggested by the manufactor. Once, the gDNA was extracted, the amount and quality was verified by NanoDrop™ 2000 Spectrophotometer (Thermo Scientific™, USA). The probes were stored at -80°C until further usage.

V. 16S rRNA

Posterior to DNA extraction, the 16S rRNA Library Preparation was performed using the 16S Metagenomic Sequencing Library Preparation protocol provided by Illumina® (USA). The Protocol contains three parts where during the first part the hypervariable region V3–V4 of the 16S gene is amplified. In a second PCR each sample is barcoded using the Nextera® XT indices kit (Illumina®, USA). Once the desired amplicon is generated and barcoded, the library is ready for sequencing. For the purpose of this study, the protocol was slightly adjusted as stated below.

Three libraries were generated in total making up to 239 samples (**Table 10.**). The libraries resulted from the different samples taken from the generated CDX models plus each cell line used, on the day of the injection. Additionally, tumour samples which were cultured under hypoxic and anoxic conditions called "*Hypoxia*" and "*Anoxia*" were collected. *Hypoxia* refers to the tumour samples which were cultured at 37°C under 5.0% CO_2 and 1.0% O_2 in an HERACELL 150i ThermoFisher Scientific incubator for up to six days. *Anoxia* refers to the tumour samples which were cultured under the same conditions as the cell culture (at 37 °C under 5.0% CO_2 and 90% RH in an incubator INCO153med Memmert GmbH + Co. KG). Those conditions were supposed to trigger a type of selection, Therefore, resulting in differences in bacteria presence due to the strictly anaerobic and aerobic conditions. Additionally to our normal samples we used two probes DSM–7942 (*Helicobacter pylori*) and DSM–8245 (*Clostridium spp.*) obtained by Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (Leibniz-Institut DSMZ, Germany) as positive control. The selection of those controls was random and based on their easy accessibility towards those samples.

(1) Amplicon PCR

The gDNA was first subjected to 16S amplification using a mix of 4 x FWD and 4 x REV primers (Merck KGaA, Germany) (**Table 6.**) which target the hypervariable regions V3–V4 as the 16S Metagenomic Sequencing Library Preparation protocol provided by Illumina® suggests. The amplification PCR reaction contained the mix stated in **Table 7**. The cycling conditions were as following, the PCR reaction started with an initial denaturation step at 95°C for 3 minute, followed by 35 cycles and another denaturation step at 95°C for 30 sec, an annealing step at 55°C for 30 sec and the extension at 72°C for 30 sec. A final extension step was executed at 72°C for 5 min lastly; the samples were cooled down to 4°C.

Table 6. Primer list used for the amplification of the hypervariable regions V3–V4

Primer	Sequence (5'-3')
FWD 1	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG
FWD 2	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNCCTACGGGNGGCWGCAG
FWD 3	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNCCTACGGGNGGCWGCAG
FWD 4	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNCCTACGGGNGGCWGCAG
REV 1	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC
REV 2	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNGACTACHVGGGTATCTAATCC
REV 3	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNGACTACHVGGGTATCTAATCC
REV 4	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNGACTACHVGGGTATCTAATCC

Once the amplicon was generated, it was purified according to protocol. AMPure® XP beads (Beckman Coulter Holdings GmbH, USA) were vortexed for 30 sec, 20 µL beads were added to each sample, mixed and incubated at RT for 5 mins. Afterwards, they were placed on a magnetic stand for 2 mins and the supernatant was discarded. The beads were then washed twice with 80% ethanol and left to dry. Samples were removed from the magnetic stand, 52.5µL of DNA Suspension Buffer (TEKNOVA, USA) was added and mixed. The samples were put back on the magnetic rack and the supernatant was collected into fresh tubes. The size and purification of the ~ 550 bp amplicon was verified via Agilent 2200 TapeStation System (Agilent Technologies, USA) using the High Sensitivity D1000 ScreenTape Kit (Agilent Technologies, USA). The samples were then stored at -28°C until further usage.

Table 7. Amplification PCR mix

Item	Volume
gDNA (25–50 ng/µL)	2.5 µL
Forward Primer Mix (1 µM)	5 µL
Reverse Primer Mix (1 µM)	5 µL
2 x KAPA HiFi HotStart Ready Mix (Roche, Switzerland)	12.5 µL

(2) Index PCR

In the second part, Illumina sequencing adapters and dual–index barcodes are added to the amplicon target. The index PCR contained the reaction mix stated in **Table 8**. The index PCR reaction started with an initial denaturation step at 95°C for 3mins, 8 cycles of another denaturation step at 95°C for 30sec, the annealing at 55°C for 30sec and the extension at 72°C for 30sec followed by an extension step at 72°C for 5mins and a cool down to 4°C. The amplicon is purified by adding 56 µL of AMPure® XP beads (Beckman Coulter Holdings GmbH, USA). The suspension was incubated at RT for 5 mins and placed on a magnetic stand for 2 mins, where the supernatant was discarded afterwards. Posterior the beads were washed twice with 80% Ethanol and let to dry. Afterwards 27.5 µL DNA Suspension Buffer was added and incubated for 2mins at RT. Clear supernatant was collected and its purity and size of ~ 600 bp was verified via Agilent 2200 TapeStation System (Agilent Technologies, USA) using the High Sensitivity D1000 ScreenTape Kit (Agilent Technologies, USA). Samples were stored at -28°C until further usage. Once the full amplicon was generated samples were pooled in equimolar concentrations followed by a size selection (500–700 bp) step on the PippinPrep™ (Sage Science, USA) to reduce non–specific amplification products from the host DNA.

Table 8. Index PCR mix

Item	Volume
16S DNA	5 µL
Nextera XT Index 1 Primers (N7XX) from the Nextera XT Index kit	5 µL
Nextera XT Index 2 Primers (S5XX) from the Nextera XT Index kit	5 µL
2 x KAPA HiFi HotStart Ready Mix (Roche, Switzerland)	25 µL
PCR Grade Water	10 µL

(3) Sequencing

The complete library was submitted to the Vienna Biocenter Core Facilities GmbH (VBCF). Prior sequencing the library was quantified according to the 16S Metagenomic Sequencing Library Preparation protocol provided by Illumina®. Sequencing was performed on the Illumina MiSeq using the MiSeq Reagent Kit v3 (600 cycles, 5% PhiX) following the 2 x 300- bp paired–end sequencing protocol.

VI. Bioinformatics And Multivariant Statistics

A dual-indices protocol was used for de–multiplexing to generate fastq files. In total three libraries were submitted for sequencing. The first two libraries were analysed by Ardigen (Poland), the third and last one was analysed inhouse. The upcoming process was executed by Ardigen with the first two libraries. To generate a good database, the two databases RDP DB (3 196 041 sequences) and NCBI 16s (20 470 sequences) were merged, after removal of duplicated sequences and definition of a unified taxonomy the used database contained 2 836 773 sequences. For the definition of the taxonomy the standard taxonomy in the NCBI taxonomy database was used. The obtained reads were submitted to a quality control originally by FastQC. In additional non–bacterial fractions were filtered by mapping against a hybrid mouse/human genome using Burrow Wheeler Aligner. Moreover, the read length distribution was reported by calculating the overlap of reads and continuous sequence fragment. Once the QC is executed, the fragments were mapped by vsearch, where reads were merged and primers trimmed. Short reads were removed and merged reads were filtered for duplicates. The reads were then assigned against the mentioned database (Max 400 hits, id > 70%). An elastic threshold was applied to enable adjustments; taxonomy was assigned when 75% of hits agree. Lastly, normalized counts took the bacteria 16S copy– number into account thereby, reporting the bacteria gene copy number. The inhouse pipeline established by Dr. Markus Bauer followed a similar approach as Ardigen's proprietary. The widely used GreenGenes database was used as a reference. Taxonomic assignments were done using a Bayesian approach.

VII. *In Situ* Hybridization

Fresh frozen CDX–tissue samples from the cell line MIA_PaCa-2 were first formalin–fixed and paraffin embedded. The samples were fixed in 10% neutral buffered formalin (Merck KGaA, Germany**)** and let there for several hours (< 16 hours). The samples were then washed afterwards with 1 X PBS and dehydrated using a standard ethanol series, followed by xylene. Then the samples were embedded in paraffin and the blocks were cut into 4 µm sections using a microtome.

Due to the already high background of FFPE slides, we decided to do *in situ* hybridization without fluorescence. The ISH procedure was executed by following the protocol RNAscope®2.5 LS Reagent Kit–RED User Manual for BDZ 11 according to the manufacture (Advanced Cell Diagnostics, Inc.). As target probe we used RNAscope®Probe–EB–16S rRNA (Advanced Cell Diagnostics, Inc.) which is the probe provided by Advanced Cell Diagnostics, Inc. to generally stain eubacteria. The manufacture confirmed, that this probe covers the same region of the 16S rRNA as the widely used EU–338 probe (5'- GCTGCCTCCGTAGGAGT-3') (*Amann et al. 1990*) . As general probe for the xenograft we used RNAscope®Positive Control Probe–Hs–PPIB which binds to the human peptidylprolyl isomerase B. Our positive FFPE slides were the tissue samples tongue, skin and colon as well as a Helicobacter TISSUE-TROL™ Control Slide (Merck KGaA, Germany**)** purchased slide. The purchased slide is a mouse intestine tissue containing *Helicobacter* pylori. Our negative control was provided by our self–generated muscle tissue FFPE slide.

3. Aim

The main intention of this project was to establish mouse models that could be used to investigate the oncobiome. Additionally, we also intended to characterize the oncobiome in the different cancer mouse models by applying 16S rRNA sequencing. Thereby, we aspired to find answers to questions like whether bacteria are even present in cell–derived xenografts in pathogen–free mice like NMRI-*Foxn1*^{nu}, if the bacteria present within the oncobiome are more of beneficial or pathogenic nature and comparing tumour models harbouring a certain KRAS or HRAS mutation and KRAS or HRAS WT. Finding patterns between bacteria living under hypoxic versus bacteria living under anoxic conditions and eventually, discovering a linkage between the age of the mice with their gut microbiome. 16S rRNA sequencing proofs to be the state of the art method when investigating the microbiome and is the number one tool used to analyse microbiota in low biomass samples such as CDX. Therefore, it was the method chosen to conduct our research.

To achieve this aim we established a protocol for 16S sequencing and created the bioinformatics pipeline in order to process our data correctly. However, due to the high prevalence of contamination in our data we conducted an *in silico* analysis to evaluate 16S sequencing as a method for low bacteria biomass samples to round up our 16S sequencing analysis we added a troubleshooting chapter to pinpoint how the method can be executed correctly. Due to the challenges that we faced with 16S sequencing we also decided to perform one simple ISH experiment with a CDX mouse model to set up a method by which the oncobiome could be investigated in a reliable, less contamination–prone matter.

4. Results & Discussion

The tissue and cell line samples from the different CDX mouse-models were categorized in libraries; one library meant one sequencing job. Three libraries were generated in total making up to 239 samples (**Table 10.**).
The Illumina® protocol used, allows to process and sequence up to 96 samples simultaneously, since the kit includes barcodes, which are attached to the 16S amplicon during the Index PCR (**2.Materials & Methods**). The barcodes can be uniquely combined maximum 96 times, thus, one library consists of maximum 96 samples.
The results are organized by library in a chronological matter. Since our in-house bioinformatical pipeline did not seem to be sophisticated enough, we decided to have the first two libraries analysed by the polish biotechnical company Ardigen, a company specialized in microbiome analysis.
In the third library a highly bioinformatical output was not needed, which is why, the third library was analysed by our in-house pipeline to get a rough estimate about the presence of bacteria in the chosen samples.
More detailed information about the samples of each library is stated in **Table 10**.

I. Ardigen Results-First Two Libraries

As already mentioned, Ardigen processed the sequencing data of the first two libraries (**Table 10.**).

In **Figure 4** the relative abundance of taxonomy composition across all groups from samples of the 1st and 2nd library is stated. The positive controls (*Helicobacter pylori*: DSM–7942 and *Clostridium sp.*: DSM–8245) in the graph (**Figure 4.A**) are clearly visible and have been plotted together.

In our study skin and tongue samples are the two cohorts that resemble each other the most throughout all CDX models (**Figure 4.A**). They are also the tissue types with the highest abundance of *Actinobacteria,* which correlates with publications (*Morgan et al. 2013*). Furthermore, *Cyanobacteria* appear in low numbers but are exclusive for skin samples (**Figure 4.A**). In microbiome analysis the prevalence of this phylum is regularly mentioned, but its presence is not further explained (*C. Wun et al. 2013, Morgan et al. 2013*).

Colon samples display the characteristic family *Deferribacteres* which has been determined as part of the gut microbiome preferentially in mice, in humans only sporadic cases are described by previous publications (*F. Hugenholtz & de Vos, 2018*). The most representative genus in our study is *Mucispirillum* out of the *Deferribacteres,* which is a coincidence with previous studies (*F. Hugenholtz & de Vos, 2018*) (**Figure 4.B**). The finding on phyla level is consistent with previous publications (*Xiao et al. 2015*) where *Proteobacteria, Firmicutes* and *Bacteroidetes* make up more than 70% of the gut microbiota (**Figure 4.–Figure 7.**). This event also includes most of the tissue samples as well (**Figure 4.–Figure 7.**). Moreover, *Proteobacteria* is the most prevalent phylum with a relative abundance of <50% among all untreated tissue samples (excluding *Anoxia* and *Hypoxia* samples) (**Figure 4.**). Here are our findings are also in line with the literature (*F. Zhang et al. 2018*) where those three phyla are determined as the most prevalent in mouse. *Proteobacteria* includes the classes *Alphaproteobacteria, Betaproteobacteria, Epsilonproteobacteria* and *Gammaproteobacteria* which could be detected as well. In our analysis *Alphaproteobacteria* and *Betaproteobacteria* occur in a very low abundance number making up less than 5 % of the prevalence of *Proteobacteria* while the rest is made up by *Epsilon*–and *Gammaproteobacteria* the most abundant classes within that phylum (**Figure 4.–Figure 7.**).

The most prevalent class among the cell samples is *Epsilonproteobacteria* while among the CDX tumour samples it appears to be *Gammaproteobacteria* indicating that the environment of the tumour might interact thus, causing the shift from one bacteria class to a another bacteria class (**Figure 4.A.**).

There is another shift from one most relative abundant class to a least abundant class within a phylum when comparing the cell samples with the tumour tissue. Within the phylum of *Firmicutes* a shift occurs from cell line sample to tumour tissue (**Figure 4.**). This shift includes the classes *Bacilli* and *Clostridia,* where *Clostridia*, a high abundant class in the cell line sample becomes a low abundant class in the tumour tissue. On the other hand, *Bacilli* which is low abundant in the cell line sample suddenly becomes the most abundant class within the *Firmicutes* in the tumour samples. This matches with the other cohort's colon, tongue and skin where *Bacilli* are the most prevalent of the *Firmicutes* phylum which might be another hint that bacteria potentially migrate from other cohorts into tumour tissue changing the relative abundances from the original cell line sample.

Clostridiaceae is the most prevalent family of the *Clostridia* class in cell samples while *Enterococcus* seems to be the most prevalent genus of the *Bacilli* class in tumour tissue,

tongue, skin and colon which is suitable with previous studies (*F. Hugenholtz & de Vos, 2018*) (**Figure 4.A.**).
To further investigate the changes in relative abundances especially between the cell lines and the tumour samples, a deeper analysis on genus–level has been conducted (**Figure 4.B**). The analysis on genus–level (**Figure 4.B**) identifies *Escherichia/Shigella* being the most prevalent bacteria of the *Gammaproteobacteria* phylum. In our case, *Escherichia* and *Shigella* could not be differentiated, which is a common phenomenon in 16S rRNA sequencing due to their close genetic relationship (*Khot & Fisher, 2013*). *Shigellae* are phylogenetically *Escherichia coli* which were later classified as separate species on the bases of biochemical characteristics and clinical relevance. However, since *Shigella spp.* are enteric pathogens and our mice have been bred under pathogen free conditions and tested against a variety of pathogens (FELASA recommendation tested), the most probable assumption is that the identified *Escherichia/Shigella* are actually *Escherichia*. In publications, both are listed as common genera for mice and humans, probably due to the differentiation problem (*Xiao et al. 2015; F. Zhang et al. 2018*).

When looking at the cell lines per se from the different models, it can be observed that there is a high prevalence of bacteria overall where *Helicobacter, Escherichia/Shigella, Clostridium sensu stricto* and *Anaerobacillus* are the genera which were most abundant. This is a result which was quite unexpected. As cell lines are cultured under extremely sterile conditions and treated with intense care, we would have anticipated that they would be the samples with the least bacteria prevalence. However, the results presented the opposite (**Figure 4.–Figure 7**). Furthermore, did the cell lines display a comparable read count to high bacteria prevalent samples like colon. Bacteria are about 0.5µm–5.0µm therefore, it might be possible that some bacteria are not visible under the microscope, especially if those bacteria are intracellular. Nevertheless, bacteria present in cell lines are more likely to proliferate rapidly resulting in some kind of visible phenotype in the cell culture. Especially since cells were cultured without antibiotics like in our case. Therefore, it is more likely, that the bacteria observed in the sequencing data is an artefact subsequent to low bacterial biomass 16S rRNA sequencing, meaning that this method might be too sensitive to investigate low bacterial biomass cohorts like cells or that our experimental environment is too contaminated to perform an adequate pure PCR processes.
Finally, we want to discuss the high abundance of *Helicobacter pylori* throughout all the samples (**Figure 4.–Figure 7.**), which was quite alarming, the mice have been tested negatively towards *Helicobacter pylori* prior delivery by Taconic Bioscience, Inc. and all internal health checks were negative for that species. In the case of *Helicobacter* we were able to identify the species *Helicobacter pylori*. This is particularly outstanding since that species was used as a positive control DSM-7492 (*Helicobacter pylori* pure DNA - isolate*)*. Normally our pipeline and library–preparation is not powerful enough to identify bacteria down to the species level. This strongly suggests that the species might be a result of cross– contamination with the positive control *Helicobacter pylori*. The assumption is reinforced by the fact that *Helicobacter pylori* infection happens orally or through stool not skin contact, meaning that the species either already had to be present in a sample and contaminate the rest throughout the experimental procedure, or the reagents got contaminated and when reused, contaminated the rest of the samples. Ardigen made sure to eradicate overrepresentation of certain bacteria in the database to prevent false–positive hits. When *Helicobacter pylori* was still present in such high numbers, even after Ardigen's processing it proved to not be a bioinformatics artifact but moreover, a result of the experimental

procedure, proving the sensitivity of the method. This incident will be discussed throughout this work.

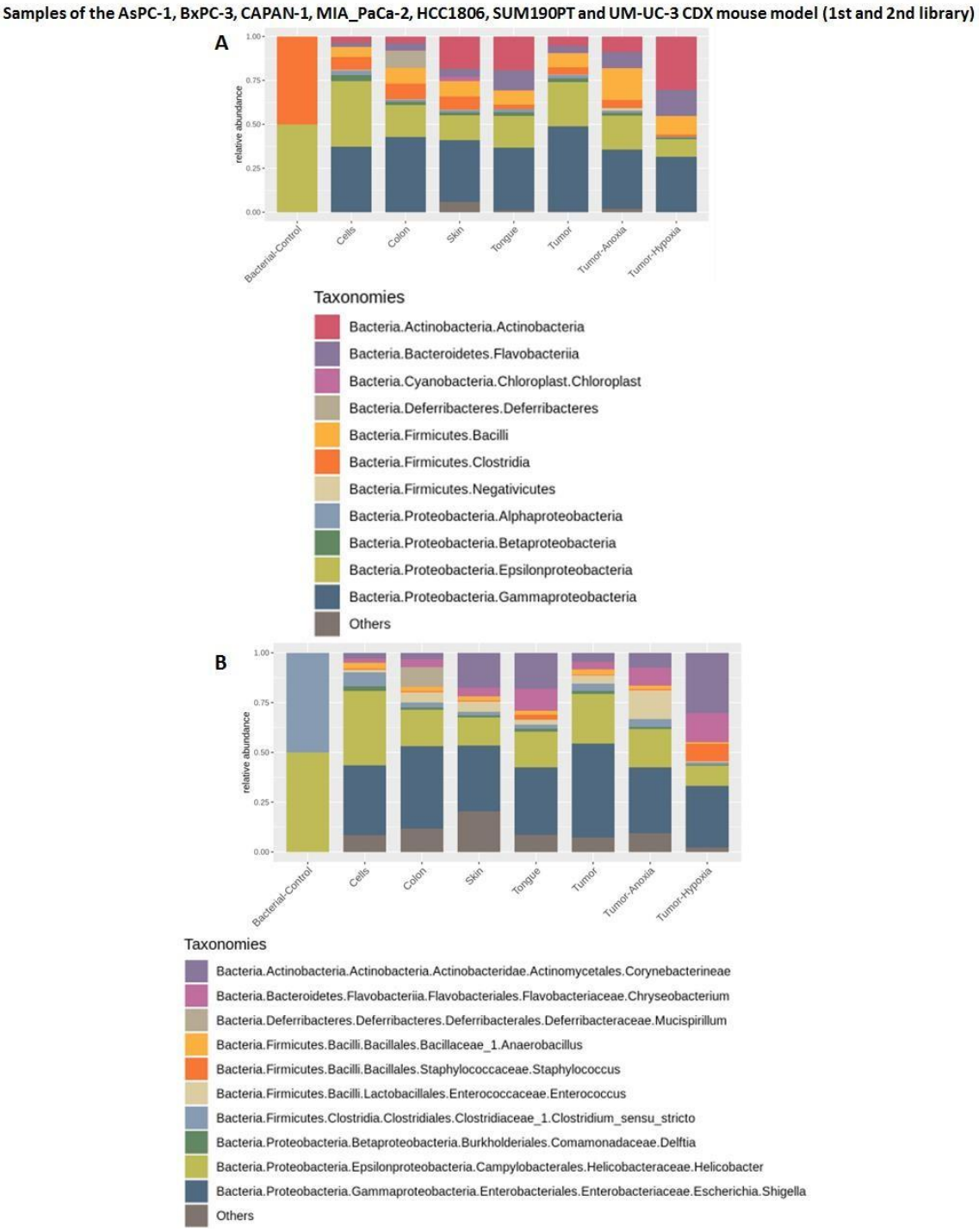

Figure 4. Survey of relative abundances of the microbial communities across all groups and cohorts at A. class level and B. genus level from 1st and 2nd library(for detailed information see Table 10.). Samples were assayed by 16S sequencing while the bioinformatical pipeline was processed by Ardigen. The plots show the mean relative abundances of the taxa with the relative abundance greater than 1% in at least one sample.

Other genera that we have identified that strongly suggest to be result of DNA–contaminant are *Escherichia/Shigella* which is also a high prevalent phylum in the cell samples and is most likely rather the genus *Escherichia* which has been identified as a common contamination in 16S sequencing, same applies for the genus *Anaerobacillus* (*Salter et al. 2014)* which is also significantly prevalent amongst the cells.
Among all samples with abundances of *Clostridium* we identified it as a new genus called *Clostridium sensu stricto*, which exhibited a high prevalence especially in the cell line samples. Since as a second positive control we used *Clostridium sp.* we cannot exclude that this finding might be the result of cross–contamination as well. However, in this case we were able to identify it as *Clostridium sensu stricto.* Nevertheless, most bacteria could be tracked back to contamination, which is the reason why all cell lines and other samples were again sequenced in a third library without using any positive controls (**II.3rd Library**).

Figure 5. Survey of relative abundances at class level of microbial communities residing in pancreas carcinoma CDX mouse model samples assayed by 16S sequencing (1st and 2nd library; for detailed information see Table 10). The bioinformatical pipeline was processed by Ardigen. The numbers on the x-axis represent the experimental replicates of each cohort (n=mouse). The plots show the mean relative abundances of the taxa with the relative abundance greater than 1% in at least one sample.

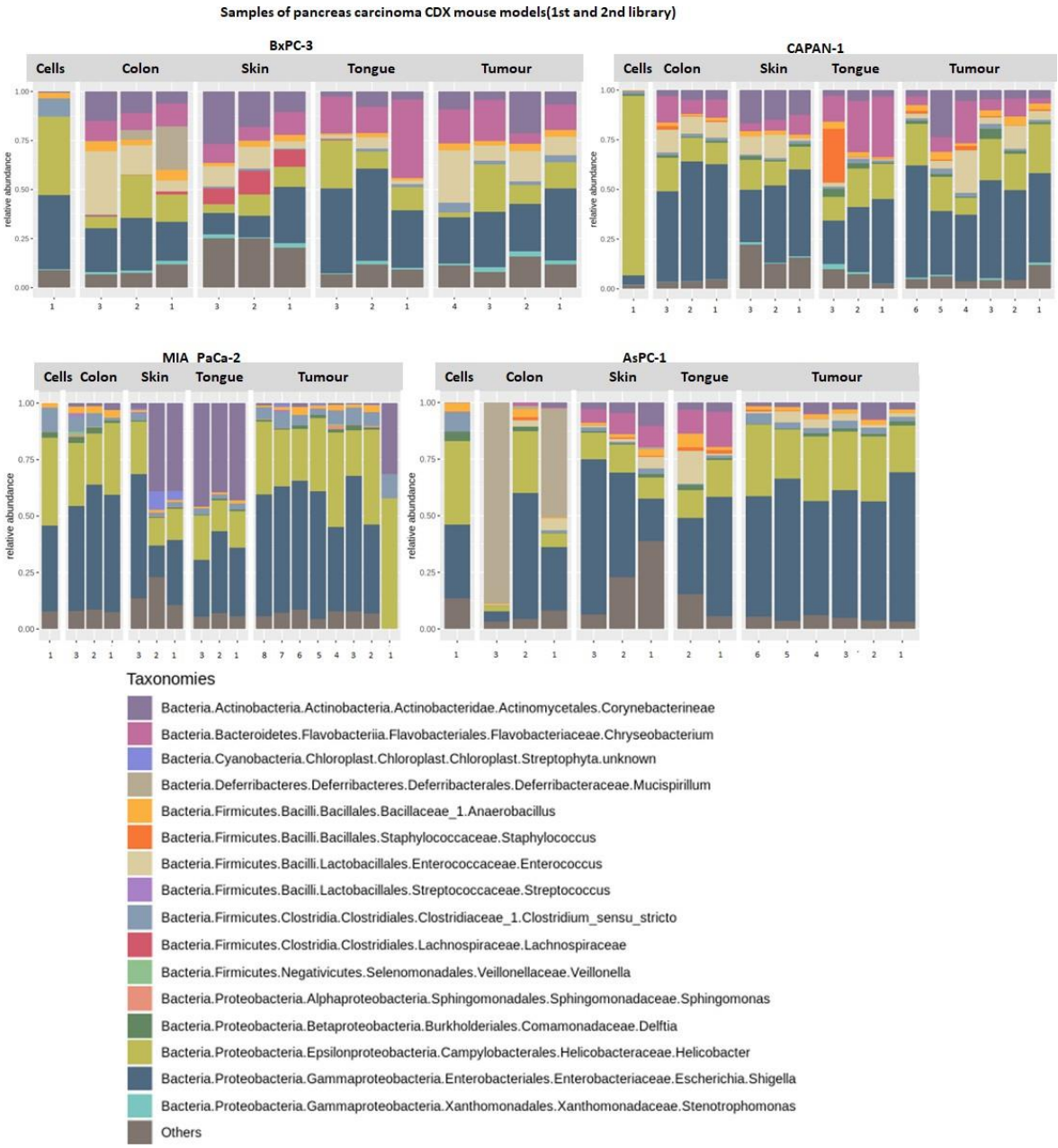

Figure 6. Survey of relative abundances at <u>genus level</u> of microbial communities residing in pancreas carcinoma CDX mouse model samples assayed by 16S sequencing (1st and 2nd library; for detailed information see Table 10). The bioinformatical pipeline was processed by Ardigen. The numbers on the x-axis represent the experimental replicates of each cohort (n=mouse). The plots show the mean relative abundances of the taxa with the relative abundance greater than 1% in at least one sample.

The main difference between the *<u>Hypoxia</u>* and *<u>Anoxia</u>* samples consists in the relative abundance of the *Actinobacteria* class, which is more prevalent in the *<u>Hypoxia</u>* samples. However, when looking at the bacteria on genus–level one can also observe differences especially within the *Firmicutes* family. While *Staphylococcus* is the most abundant genus in *<u>Hypoxia</u>*, *Enterococcus* is more abundant in *<u>Anoxia</u>* which fits better with the native tissues. Generally, it can be said that the *<u>Anoxia</u>* samples resemble more the native tissue than the

Hypoxia samples. However, due to the high inconsistency amongst the single *Hypoxia* samples among all models (**Figure 4.–Figure 9.**) we decided to not investigate the samples further, since they were not reproducible and therefore, not representative.

In **Figure 5** the relative abundances of the bacteria among the pancreas carcinoma CDX mouse model are presented on class and genus level. One of our aims was to find out whether or not there are differences between CDX mouse models and consequently, finding characteristics between the KRAS mutated CDX models (MIA_PaCa-2, AsPC-1 and CAPAN-1) and the KRAS WT CDX (BxPC-3, SUM190PT and HCC1806) models. However, from our data we could not identify any characteristic differences between those two groups neither on phyla level nor on genus level (**Figure 5.–Figure 7.**). Since our library preparation and sequencing data is not quantifiable we could not proof the theory that KRAS mutated cell lines harbor more bacteria than WT cell lines, as it is stated in some publications (*Commisso et al. 2013*). Nevertheless, do the CDX models possess different bacteria patterns (**Figure 5.–Figure 6.**). Among the pancreas carcinoma cell line samples it is noticeable that all of them look alike except for CAPAN-1. CAPAN-1 is the cell line with the least diversity of phyla where almost 95% compromise of exclusively *Helicobacteraceae*. The reason for that would have to be further investigated. Apart from that we can also observe that while the tumour tissue samples from the cell lines BxPC-3 and CAPAN-1 harbor *Flavobacteria*, the tumours deriving from the cell lines MIA_PaCa-2 and AsPC-1 do not display that phylum at all (**Figure 5.**).
Lastly, the most striking occurrence within that group is the extreme difference of phyla abundance between MIA_PaCa-2 and the other three pancreas carcinoma cell lines. While BxPC-3, AsPC-1 and CAPAN-1 mostly exhibit the same pattern, the samples derived from MIA_PaCa-2 mouse model fall completely out of line which could have multiple reasons. It is hard to make any hypothesis why the cell line is so distinctive. Further investigations would have to be executed. When looking at the breast carcinoma group (**Figure 7.A. –B.**), drastic differences can be seen. The HCC1806 derived tissue samples own the same bacteria patterns as the MIA_PaCa-2 samples (**Figure 5.** and **Figure 7.A.**). This is particularly interesting since they are carcinomas from different cohorts, Therefore, would be expected to have little in common. The difference is made up by their high abundance (<50%) of *Actinobacteria* which is unique in those two groups, since normally this phylum does not surpass the 40% threshold in the other models. On the other hand do SUM190PT derived samples resemble BxPC-3 and CAPAN-1 just like the bladder cancer CDX model UM-UC-3, except for the fact that the UM-UC-3 cell line which, presents an abnormal amount of *Gammaproteobacteria* (**Figure 5.** and **Figure 7.C.).**

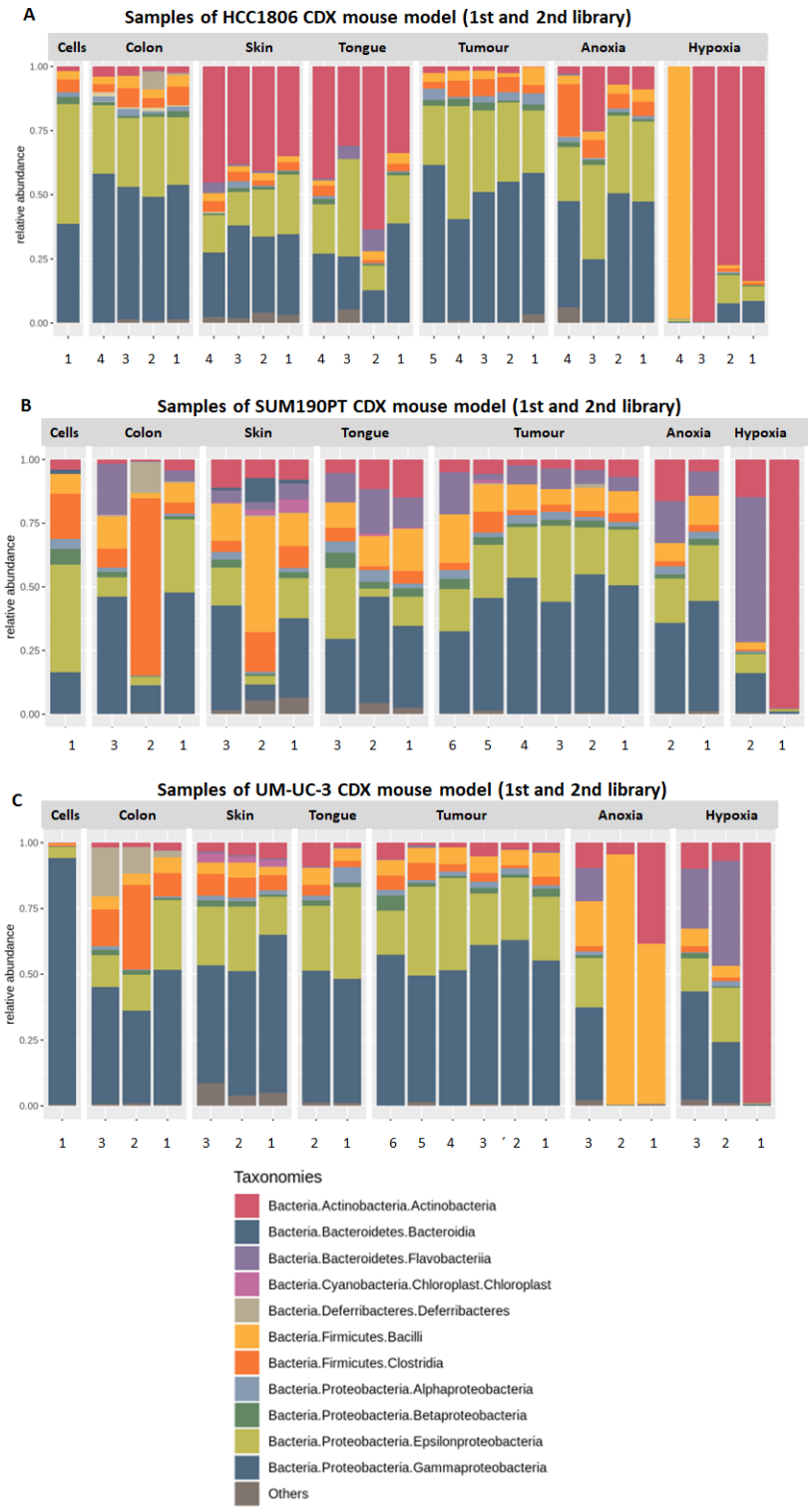

Figure 7. Survey of relative abundances at <u>class level</u> of microbial communities residing in A. HCC1806 B. SUM190PT and C. UM-UC-3 CDX mouse model samples assayed by 16S sequencing (1st and 2nd library; for detailed information see Table 10). The bioinformatical pipeline was processed by Ardigen. The numbers on the x-axis represent the experimental replicates of each cohort (n=mouse) (Hypoxia and Anoxia refer to the tumour samples cultured under either hypoxic or anoxic conditions). The plots show the mean relative abundances of the taxa with the relative abundance greater than 1% in at least one sample.

Ardigen also analyzed the homogeneity of the different samples within a model. This, analysis is quite crucial since microbiome studies need to overcome the bottom - line issue of reproducibility due to inconsistency. Therefore, a correlation analysis (**Figure 8.–Figure 9.**) was executed to illustrate that the organ cohorts (colon, tongue and skin) are the most homogeneous as expected since the mice for each CDX model were housed in the same cage and came all (except for one) from the same barrier and vender; thereby, providing a consistency within the models. Moreover, mice housed in the same cage come to share a similar microbiome due to mixing by coprophagia (*Campbell et al. 2012*) which is verified by our results.
The tumour samples within a model also demonstrated a relatively high homogeneity, while the *Hypoxia* samples are the least homogenous. The tumour samples were expected to be comparable with each other, since they were handled equally. The *Hypoxia* samples on the other hand display the importance of consistency. The handling of those samples was quite inconsistent. Since different conditions varied, from the incubation time of the samples to the time point when the samples were further processed. This inconsistency is reflected in the data and is illustrated in the correlation analysis (**Figure 8.–Figure 9.**).

Different publications indicated that the microbiome changes with age (*Langille et al. 2014*). This linkage might be of high importance since the microbiota influence different processes in our body (immunsystem, permeability of intestinal layers, metabolism) therefore, by identifying the link and the differences, therapies could be adapted to elderly people such that for example drugs are more easily tolerated.
This is why we included mice of different age, which is why all though all mice were delivered at an age of 6–8 weeks, they reached an age from 35–147 days thereby, causing an age difference of so much as 112 days which we expected to be enough to see changes in the microbiome. Unfortunately, we were unable to spot any significant shift or difference age dependent in the microbiome of the mice. However, Langille and his team, who have already executed such study defined three age–stages in mice; “young” (174 days old), “middle” (589 days old) and “old” (857 days old). A microbiome difference could only be identified significantly between the “young” and “old” mice. Therefore, it is of no surprise that we could not identify any linkage between the mouse age and the microbiome in our study since it seems like that our chosen age–gap was too little to make a difference.

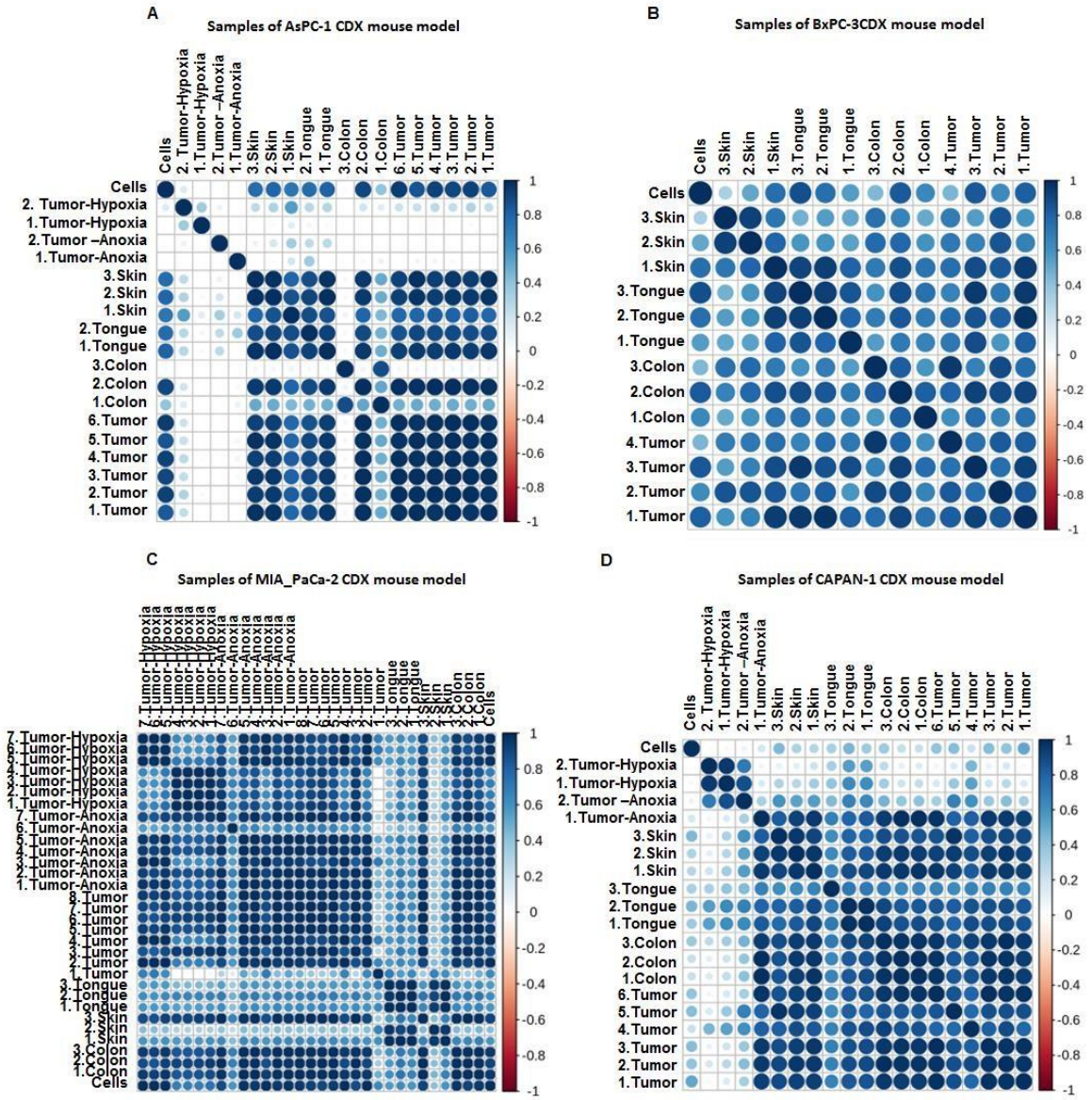

Figure 8. Survey of correlation of taxonomies of relative abundance results at <u>class</u> <u>level</u> from the first two libraries (for detailed information see Table 10.) processed by Ardigen. Similarity of composition of samples from **A.** samples of AsPC-1 CDX mouse model **B.** samples of BxPC-3 CDX mouse model **C.** samples of MIA_PaCa-2 CDX mouse model **D.** samples of CAPAN-1 CDX mouse model, the Pearson coefficient was calculated on the profiles of relative abundances on the <u>genus level</u>.

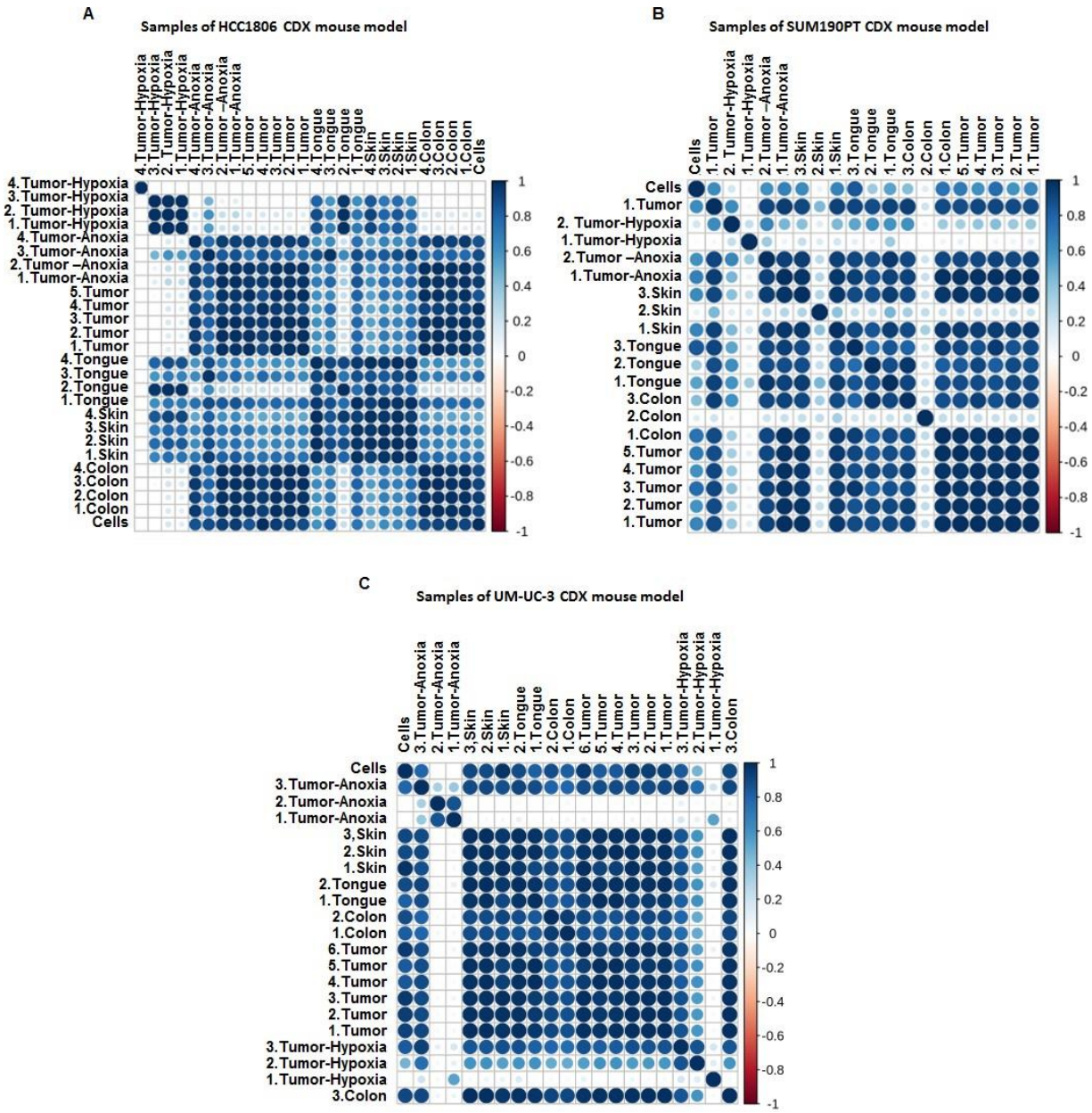

Figure 9. Survey of correlation of taxonomies of relative abundance results at <u>class</u> <u>level</u> from the first two libraries (for detailed information see Table 10). Similarity of composition of samples from **A.** samples of HCC1806 CDX mouse model **B.** samples of SUM190PT CDX mouse model **C.** samples of UM-UC-3 CDX mouse model, the Pearson coefficient was calculated on the profiles of relative abundances on the <u>genus level</u>.

II. 3rd Library

In the previous chapter it was only discussed if the data yielded from our analysis, corresponds with current and previous literature as well as comparing the different cohorts to each other. However, due to the problems we have faced with possible cross–contamination or oversensitivity we decided to process a third library. The third and last library contained the samples stated in **Table 10**. Previous positive controls (*Helicobacter pylori* and *Clostridium spp.*) were not used in this library, instead we decided to generate "PCR-controls" (**Table 11.**) used during the procedure to evaluate contamination during the experiment.

In **Figure 10.A** the presence of the taxa of the three libraries are represented as a heatmap. In the first two libraries the *Epsilonproteobacteria* is a sign of *Helicobacter pylori* which is present in almost every sample but in the last library *Epsilonproteobacteria* are almost eradicated. Moreover, in the two cases where it is actually present, when going into the genus level we could identify the *Camplyobacterales* genus but not the *Helicobacter pylori* species. Meaning that in the 3rd library the *Helicobacter pylori* was completely absent thus, supporting our previous assumption that this species had to be a consequence of cross– contamination with the positive control.

These results show that it is crucial to separate every experimental procedure physically from each other to prevent contamination as good as possible and to use positive and negative controls as validation of the experimental procedure.

The 3rd library was not processed by Ardigen, the processing was performed by our in–house bioinformatic pipeline. Here, we should mentioned that the first two libraries were also run by our in–house bioinformatics pipeline once, demonstrating that even though, Ardigen's pipeline is more sensitive, both pipelines are comparable on qualitylevel.

It is easily visible that throughout all 3 libraries (**Figure 10.A.**) the classes which are present in most samples stay the same, being; *Gammaproteobacteria, Actinobacteria, Clostridia, Betaproteobacteria, Epsilonproteobacteria* and *Bacilli*. While in the first two libraries *Deferribacteres* were exclusively present in the colon samples, in the third library *Deferribacteres* also appeared in the tumour tissue, tongue and skin, nevertheless, it stayed persisted as one of the most prevalent phylum in the colon samples.

The PCR controls should have functioned as negative controls, since they were reagents which were sterilizes and autoclaved prior usage and supposed to be "DNA–and RNA–free", however, we could detect low but significant prevalence of different bacteria (**Figure 10.B.**). This reminds of the occurrence observed in the cell line samples (**Figure 4.–Figure 7.**), where despite of their sterilized culture–conditions, bacteria in high numbers were detected, which was also unexpected. This could be another hint that the sensitivity of the sequencing method is too high for close–to–zero bacteria biomass samples.

Furthermore, when comparing the bacteria present in the PCR controls and all other samples (**Figure 10.**) it is nicely exemplified that a lot of phyla present in the PCR controls are also present in all samples. Therefore, we cannot guarantee the credibility of the generated data. The origin for the bacteria in the PCR controls cannot be determined, since they might be a cause of the experimental procedure like contamination during the amplification PCR or already present bacteria from the DNA Extraction kits or other reagents used. The problem that we faced when conducting this study is not original (*Eisenhofer et al. 2019; D. Kim et al. 2017; Salter et al. 2014*) even though, most investigations and laboratories still rely strongly on the identification of bacteria in low biomass samples by 16S rRNA sequencing. The reason for having found that much bacteria even in our negative controls could have more than just one reason. Either our results suggest that 16S rRNA sequencing is too sensitive

for low bacteria biomass samples or our conditions under which we are performing our experiment are simply not sterile enough to execute such a pure non–contaminant PCR process and thus not being capable of performing not contaminated sequencing. To find out which of the two are true, further investigation was needed. Therefore, we decided to dig deeper into literature and suggest possible improvement and furthermore, possible alternatives in the following chapters.

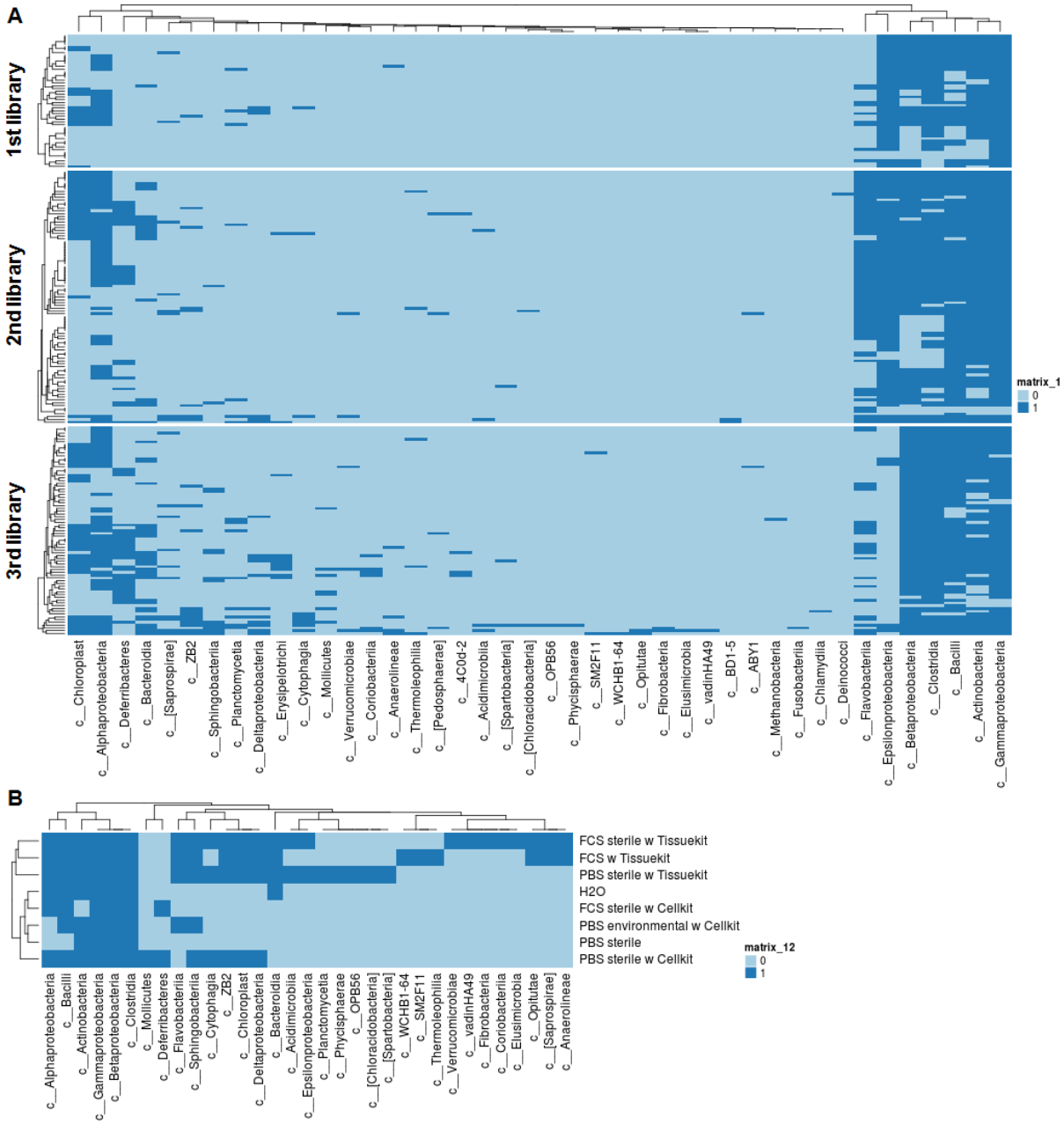

Figure 10. Heatmap of the presence of taxonomies at phylum level among A. all three libraries and B. PCR controls.

III. In Silico

Due to the findings of our obtained data, we decided to further compare our 16S rRNA sequencing method procedure with other publications. Six publications were chosen which have executed a similar research to this project, meaning that the objective was to investigate the bacterial profile from two different cohorts in low bacterial biomass samples. In most cases it would be cancerous versus healthy tissue or different cancerous tissues compared to each other (**Table 9**.). The points that we compare include the conditions of the experimental procedures as well as the used tools such as how the DNA of the samples was extracted, how the library was prepared for sequencing and what method was used for sequencing. It is crucial to note that the results of the analysis are alike in terms that each paper states the relative abundance on at least phylum level and sometimes even genus level between the two investigated cohorts. However, the execution of the 16S rRNA sequencing method differs drastically. Each step in the experimental pipeline introduces variations influencing the final output and probably leading to the appearing inconsistency among microbiome analysis.

First, samples are taken under sterile conditions. Here, differences or contamination could occur which is why it would be advisable to always add an environmental probe at each step of the experiment to exclude contamination by the environment or to then incorporate it in the bioinformatics analysis. Such an inclusion of an environmental probe was not mentioned in any of the six publications. However, literature exists, where an environmental probe was used (*Urbaniak, 2014*) and the data which was obtained from the environmental probe was incorporated in the contamination predictor tool SourceTracker. The SourceTracker compared the microbial population in the tissue samples to that of environmental control that were processed alongside the tissue samples, demonstrating that the contamination present made up an average of 10%. Nevertheless, even a low percentage weighs heavily on low biomass samples; Therefore, we would suggest that results from the environmental probe should also be published alongside and not just incorporated in statistics.
Optimal storage conditions of the isolated DNA have been investigated for different samples (*Lauber et al. 2010*). However, even though, Lauber et al. clarified that the relative abundances of most bacterial taxa in the investigated samples (soil, human skin and stool) did not change with different storage conditions, it is advisable to store samples promptly after collection at -80°C, which has been the case among all the publications.
Second, the DNA is extracted from samples with a DNA isolation kit. When looking at **Table 9** hat amongst the stated publications, different DNA extraction kits were used. The MO BIO commercial kits are very widely used, different variations were used in 3/6 (*Geller et al. 2017 Pushalkar et al. 2018 and Hieken et al. 2016*) publications (**Table 9**). The popularity of these kits is probably due the fact that the MO BIO PowerSoil DNA Isolation Kit (by MO BIO Laboratories, Carlsbad, CA, USA) was used in the Human Microbiome Project (*Wagner et al. 2015*) and is described to provide high yields. However, originally it was designed to isolate DNA from soil, stool and environmental samples, thus samples which are high in microbial DNA (*Kim et al. 2017*). It is recommended to use kits which were designed to minimize kit contamination for low microbial biomass samples, such as tumour tissue. The QIAGEN commercial kits e.g. QIAmp UCP (UltraClean production) Pathogen Mini Kit (QIAGEN) like used by 1/6 publication (*Xuan et al.)* (**Table 9**.) are examples for such kits. Yield and quality of the extracted DNA differs between kits which has been just recently proven (*Panek et al. 2018*). Therefore, it can be said that investigations using different DNA isolation kits are not

directly comparable with each other. Additionally, just recently an event known as the “kitome”, which refers to microbiota already present in the extraction kit prior use has been elucidated (*Kim et al. 2017*) thus, it is suggestable that one does not change kits within a study nor kit–batch since that might influence the output and introduce different contamination resulting in different analysis outcome. Also, due to the fact that in low biomass samples, even the kit possesses a microbiome, one needs to identify the “kitome” first. Thus, with every new used kit–batch it would be necessary to run a blank sample parallel to the other samples to ensure that there is no contamination from the samples due to the used kit or define a standard contamination factor for the kit which can be subtracted out during data fitting.
After the DNA was isolated the next step is to prepare the library. Here, most investigators use the 16S Metagenomic Sequencing Library Preparation protocol provided by Illumina®. However, only 2/6 papers (*Pushalkar et al. 2018, Bullman et al. 2017*) used primers covering the hypervariable regions V3–V4 as suggested by Illumina®, 3/6 (*Xuan et al. 2014, Geller et al. 2017 and Zhang et al*. *2018*) used primers covering V4 and only one paper used primers covering V3–V5 (*Hieken et al*. *2016*). Different studies have been executed, where the sensitivity and correlation of the different hypervariable regions in the 16S rRNA gene for phylogenetic analysis purpose was investigated (*Chakravorty et al. 2016*). Yang et al. determined V4 as the region with the highest sensitivity at phylum level for bacterial phylogenetic analysis, which is consistent with previous taxonomic results (*Wang et al. 2007*). This is of no surprise since V4 is the most commonly used hypervariable region in 16S rRNA sequencing analysis when identifying taxonomic communities. Nevertheless, does Yang and his team emphasiz that the output date is also influenced by the bioinformatic pipeline used. On the other hand their study demonstrated a poor performance for V2 and V3 in terms of the phylogenetic analysis. In the end they claim that they have a combination of the hypervariable regions V4–V6 to represent the optimal sub–regions for the bacterial phylogenetic study of new phyla. A more detailed research was executed by Chakravorty and his team who stated the purpose of the different regions. They determined V1 as the best region to differentiate among *Staphylococcs sp.*, while V2, V3 and V6 were most suitable to distinguish all bacteria at genus level, except for closely related *Enterobacteriaceae*. V4, V5, V7 and V8 were less useful for genus identification within his study (*Chakravorty et al. 2007*). In general the most used regions found in the literature overall are V3, V4 and V6 (*Caporaso et al. 2011; Finkel et al. 2011; Shepherd et al. 2012*). For identifying in a database existing taxonomy the common used regions should be sufficient and by using a combination of more than one hypervariable region one enables deeper sequencing into the taxonomy and identification not only at phylum–level but also at genus–level. The inconsistency of the usage of different hypervariable regions might not be a decision based on science alone but is most probably based on personal preferences. Nevertheless, changing the target region, also changes the primers that are used for amplifying the 16S rRNA gene which does have an influence on the output and its accuracy (*Klindworth et al. 2013*).

Apart from the variation in the hypervariable region, the bigger challenge when reading literature is most probably that the primers which are used in the different microbiome analysis are often not stated, even though, most microbiome analyses are based on the 16S Metagenomic Sequencing Library Preparation protocol by Illumina®. However, that protocol does not come with a kit or any reagents and the primers stated in the protocol of Illumina® are merely a suggestion. 1/6 papers (*Xuan et al. 2014)* (**Table 9**.) did not state any reference or primers with which the protocol was executed. 2/6 papers (*Geller et al. 2017*

and *Zhang et al. 2018*) (**Table 9**.) used the primer set 515F and 806R which was used in the Earth Microbiome Project (*Caporaso et al. 2011; Gilbert et al. 2014*). The primer set 515F/806R has been continuously optimized (*Apprill et al. 2015; Parada et al. 2016*) and has been developed to target bacteria and archea in soil ecology studies, which is why it is surprising that it is still widely used for human tissue samples. Moreover 2/6 papers (*Pushlakar et al. 2018* and *Bullman et al. 2017*) (**Table 9**.) used the primer set 341F and 805R which has been reported to be a set with a very high coverage for interrogating bacterial communities (*Klindworth et al. 2013; Mizrahi-Man et al. 2013*). Klindworth and her team who executed an evaluation of primers used for 16S rRNA gene sequencing, noted that some phyla (*Armatimonadetes*, *Chlamydiae*, *Caldiserica*, Hyd24-12, GOUTA4, Kazan-3B-28, SM2F11, as well as Candidate divisions WS6, OP11, TM7 and OD1) are not detected by that primer set (*Klindworth et al. 2013*). Only one paper (*Hieken et al. 2016*) used the primer set 357F and 926R which has been little used or reported (*Sim et al. 2012*). Klindworth concluded that even commonly used single primers exhibit significant differences in overall coverage and phylum spectrum and that the choice of database for the bioinformatic pipeline is crucial for the data output. Therefore, it is noteworthy that primers should be thoroughly checked and validated prior usage to avoid biases within the analysis. Also, the reason for the primer set choice should be added in the *Methods and Materials* section of a publication to better comprehend the angle of the analysis.

After the primers have been chosen, the next step would be to establish the conditions for the first 16S rRNA gene amplification PCR. This procedure strongly influences the data output and are often poorly described in publications. Here, two main points need to be considered; the DNA template that is inputted and the amount of PCR cycles executed. When looking at **Table 9**. one can observe a wild discrepancy. 1/6 papers (*Zhang et al. 2018*) gives no information about the PCR conditions at all and none of the papers give full information about the DNA amount and volume inputted and PCR cycles used. This is a common deviation in microbiome analysis especially in studies which include samples of low bacterial biomass. The challenge lies in finding the balance of being able to amplify the low bacterial biome while simultaneously losing possible contamination and doubtful low level sequences. We also faced the problem to find the balance in our study. When establishing the protocol, we determined that 25 cycles of 5 ng/µl 2.5 µl DNA amount input as suggested by Illumina®, is too less to amplify the samples with low bacterial biomass such as tumour and cell lines. However, when inputting 40 cycles of 50 ng/µl 2.5 µl DNA amount, the signal of high bacterial biomass samples such as colon were too high however, the tumour signal was exquisite. The reason why the suggested template by Illumina® is not applicable is probably due to the fact that the protocol refers to microbial DNA and most researchers input genomic DNA, since microbial DNA is tedious and expensive to isolate. "*Salter and colleagues investigated effects of the number of PCR cycles for amplification, stating that for low bacterial biomass samples 20 cycles was too low to fully represent the PCR product"* while 40 cycles recovered both contaminating and authentic low level sequences (*Salter et al. 2014*). Other studies claim that the concentration of the starting template inputted is the major factor behind variability in sequencing results (*Kennedy et al. 2014*). When establishing the protocol we run the amplification PCR with different DNA template dilution as well as different cycle numbers with colon and tumour the samples therefore, representing the highest and lowest biomass within our library. Finally, we decided to increase the input DNA template concentration to 50 ng/µL and the PCR cycles to 32. It is probable that PCR conditions vary among studies Therefore, causing fluctuation in sequencing results. It is not assured that standardizing PCR conditions would be possible, since sample biomass varies

too much and the conditions subsequently need to be adjusted. Nevertheless, the amplification PCR conditions have to be prior established properly, such that the method yields representable and reproducible results.

After the 16S rRNA gene is amplified, the PCR product is indexed. This step is pretty much standardized and mostly executed the same way. After the library has been completed, sequencing rounds up the experimental procedure. As for the NGS platform there are two options described either Illumina MiSeq or Ion Torrent PGM. Surprisingly, researchers seem to agree to use the Illumina MiSeq as sequencing platform (all papers listed used Illumina MiSeq) but the reason for choosing MiSeq stays unclear. The difference in sequencing method can be traced back on the primers used at the 16S rRNA gene amplification.

Even though, the whole bioinformatic pipeline to analyze the data resulting from sequencing influences the output completely, only one aspect will be discussed during this study; the choice of database. In all except one publication (*Zhang et al. 2018*) which are stated in this study, Greengenes (http://greengenes.secondgenome.com/) was used as database, while Zhang et al. used the RDP database (http://rdp.cme.msu.edu/). The reason for the high agreement of Greengenes as database is simple; Greengenes is a free and easy–to–use database. However, when Ardigen analyzed our data they chose not using it, stating that this database relies on quantity instead of quality meaning, an overrepresentation of many taxonomic units, Therefore, yielding to a bias of results. Thus, they chose to establish their own database by merging two databases together (RDP DB and NCBI 16s) and modifying it as such that the quality and quantity of the database reach an optimum. The generation of such a high quality database is quite tedious and cost and time consuming which is why most researches choose the easier and cheaper way, which is in this case Greengenes. In our study Greengenes was used in–house, exactly for the reasons named above and it resulted in decent results. The high agreement of use of Greengenes might suggest using Greengenes as well such that the results become better comparable with literature. However, one should always validate the chosen database on its quality more than its accessibility.

Based on our findings, we have developed a set of best practices to guide co-occurrence network analysis.

Table 9. Comparison of selected papers in terms of 16S rRNA sequencing methodology stating their Res. (Research), DNA Ext. (DNA Extraction), Sequ. (Sequencing), DB (Database) and Result

Paper	Xuan et al. PLOSONE 2014	Geller et al. Science 2017	Pushalkar et al. Cancer Discovery 2018	Hieken et al. Nature 2016	Bullman et al. Science 2017	Zhang et al. Mol. Medicine Reports 2018
Res.	Different breast cancer tissue	Gemcitabine resistance	Cancerous pancreas vs. healthy	Benign vs. Malignant breast tissue	Colorectal cancer vs. liver metastisis	Human vs. Mouse tumours
DNA Ext.	Breast Tissue QIAmp DNA FFPE Tissue kit	PDAC tissue Mo-Bio UltraClean Tissue DNA Kit	PDX & fecas Lysis & MoBio Power fecal kit	Breast tissue PowerSoil DNA Isolation Kit & Microbiome Enrichment Kit	Tissue ZymoBIOMICS™ Service-(Metagenomic Sequencing)	Tissue TIANamp Genomic DNA kit
Lib. Prep	V4 16S rRNA Not stated 2 µL DNA template 25 cycles	V4 16S rRNA 515F & 806R 10 µL DNA template 30 cycles	V3-V4 16S rRNa 341F & 805R 10 ng/µL DNA template 25cycles	V3-V5 16S rRNA 357F & 926R 50 ng/µL DNA template 34cycles	V3-V4 16S rRNA 341F & 805R Not stated 42 cycles	V4 16S rRNA 515F & 806R Not stated
Sequ.	Ilumina MiSeq 150bp PE	Illumina MiSeq 175bp PE	Illumina MiSeq 300bp PE, 600cycles	Illumina MiSeq	Illumina MiSeq 600cycles	Illumina MiSeq
DB	Greengenes	Greengenes	Greengenes	Greengenes	Greengenes	RDP
Result	Rel. abundance on phylum level– different breast cancer tissues	Rel. abundance on phylum level-PDAC tissue	Rel. abundance on genus level-pancreatic cancer tissues	Rel. abundances on genus level-Breast vs. Skin	Rel. abundances on genus level-Colorectal cancer vs. Liver metastasis	Rel. abundances on genus level-human vs. Mouse tumours

At this point we are identifying contamination per se, as the most troublesome obstacle in our kind. Microbiome analysis is challenging to standardize since every study needs readjustment, establishment and revalidation of previously used methods. Therefore, we continued with troubleshooting for future research.

IV. Troubleshooting

In the previous chapter we stated different protocols that were used and their influences they can have on every single step along the 16S sequencing process. In this chapter we decided to suggest a possible troubleshooting of this method. It has to be noted, that 16S rRNA sequencing for low bacterial biomass samples has been widely discussed (*Salter et al. 2014*) (*Bender et al. 2018; Eisenhofer et al. 2019*). Studies are being executed exhibiting the difficulty of the methods high sensitivity by concluding that in low biomass samples like analysis of peripheral blood or placenta samples, they were unable to differentiate samples from blank controls (*Glassing et al. 2016; Lauder et al. 2016*). Moreover, in the investigation of Glassing where his team re-evaluated some low bacterial biomass analysis, 99 % of the identified sequences corresponded to contaminant taxa.

There are two crucial contamination types appearing in microbiome studies.

One is contaminant DNA which is the kind of contamination originated from the experimental procedure despite utmost care. This type has been discussed in the *in–silico* analysis. In that

case the contaminant taxa arise from laboratory environments, reagents, DNA extraction kits and so on.

The second occurring contamination type is cross–contamination which was mentioned when talking about our *Helicobacter pylori* contaminant taxa across all samples. In this case, a present taxa cross contaminants other samples, in our case it was the positive control *Helicobacter pylori* which cross-contaminated all samples. These contaminant taxa are extremely persistent and are shared and identified across multiple studies (*Eisenhofer et al. 2019; Salter et al. 2014*). The reason of why cross–contamination might happen despite absolute care, is because different events might take place like aerosolization during pipetting, plate cover removal or barcode cross–contamination where incorrect barcodes jump into neighbouring tubes, a phenomenon named “tag switching” (*Carlsen et al. 2012*). But even in the sequencing instrument contamination can occur like barcode sequencing errors, residual amplicons from past sequencing runs or “index hopping”, where the platform mismatches indexing reads to sequencing reads (*Larsson et al. 2018*). At this point we made clear that 16S sequencing is very contamination prone, Therefore, one has to take different measurements to either avoid or supervise contamination.

In the paper by Eisenhofer et al. they state a so–called “RIDE” checklist which consists of guidance’s and measurements one can take to keep the risk of contamination at a minimum (*Eisenhofer et al. 2019*). The most important points being “using sampling blank control”, “report taxa found in negative controls” and lastly “comparison of taxa found in controls to those identified in biological samples”. This is especially striking since when going through our *in silico* analysis we painfully had to experience that none of the papers that we chose mentioned neither positive nor negative controls.

It is crucial to always include controls in the preparation of the libraries, especially since that enables another step of the “RIDE” checklist; “Use subtractive filtering to remove contaminants from biological samples” (*Eisenhofer et al. 2019*).

Positive controls are crucial in any analysis and especially in microbiome analysis they serve to verify the sample preparation and sequencing procedure, particularly when samples are prepared side by side. Positive controls are also important to track cross–contamination. Nevertheless, the choice of controls is not that trivial since, in our case it led to cross–contamination and was not easily traced back. Therefore, it is suggestable to generate either a mixture of cultured organisms (“mock communities”) (*Salter et al. 2014*) or known mixtures of free DNA (“mock DNA” samples) (*Wu et al. 2010*). Kim et al gave an example for a positive control where selected DNA was synthesized “*using regions of the 16S gene in archaeal species which would not be detected in experimental data since, the amplification primer binding sites in the archaeal region does not match the bacterial primer binding site”*. However, to allow amplification, the bacterial primer binding sites were synthetically added. This procedure enables to distinguish the positive control from the experimental samples while still being processed through the same pipeline equally. In general it is advisable to use artificially altered DNA sequences that can be traced while running through the whole method. Thus, being able to test the level of cross–contamination between experimental samples during wet–lab preparation and sequence acquisition.

Besides the positive control, there is also a need for an adequate negative control to enable assessment of the contamination background, enabling to monitor contaminant DNA, which is especially crucial when profiling low bacterial biomass samples. It is suggestable to have at least three different negative controls, to watch over each step thus, a “sampling blank control”, a “DNA extraction blank control” and a “no–template amplification control”. All of the three samples are DNA–free water going through different points of the sequencing protocol.

The “blank–control” is a DNA–free water sample with which the whole protocol is run through, it serves as environmental control and should be present at each step of sample preparation. The “sampling blank control” is a DNA–free water samples with which the extraction step is done as well as all subsequent steps without any additional input material. The “no–template amplification control” is DNA–free water with which only the sequencing protocol is executed. It would be advisable to add another two “blank–control” samples which run only through either the amplification or the indexing process, to monitor the contamination at those steps separately. It has to be mentioned that even DNA/RNA–free purchased samples can harbour bacteria, meaning that those samples do not necessarily have to stay bacteria free until the end of the analysis but moreover, enable the identification and monitoring of contaminant DNA.
Both positive and negative controls should be placed randomly, especially when handling the whole library at once. This would provide better detection of possible cross–contamination.
After having discussed the assessment of contamination and how it can be partly avoided or at least supervised, we want to go further into detail about the problems of our analysis and where we would see improvement. Another thing is that for 16S sequencing it would be recommendable to use technical triplets additionally to experimental triplets thus, supervising the reliability and reproducibility of the process.
Looking at our process we can say that in terms of tissue dissection and library preparation we took all measurements possible for the 3rd library. The tissue was dissected under a lamina flow, wearing gloves, a lab coat and a face mask. The DNA–extraction was done under a different lamina flow, always sterilizing the work bench prior usage with UV–light. The PCR–Mastermix preparation for the amplification PCR was done under a PCR bench, while the pipetting of the PCR was executed under a different lamina flow at 4°C. After the amplification PCR, the amplicon was purified which was again done in a different lamina flow. We applied the same procedure for our index PCR more or less fulfilling the “RIDE” checklist (*Eisenhofer et al. 2019*). Nevertheless, there is still room for improvement to make our PCR process even purer or better traceable.
First of all, we should have validated the sequencing process before starting the actual experiment. The validation could be done by performing an easy qPCR with the 16S primers and titrate different concentrations of bacteria such that we would be able to estimate our detection limit and sensitivity. Afterwards the same qPCR should have been performed for the expected highest bacterial biomass sample (e.g. colon tissue) and the expected lowest bacterial biomass sample (e.g. blank). Thus, we would monitor the bacteria appearance and estimate better the number of cycles needed to get a significant result while simultaneously finding out which cycle number would cause recovery of contamination in the low bacteria biomass samples.
Another aspect is that we did not quantify each single sample, but the pooled library. This has different disadvantages. One, being that if “tag switching” really did occur we would not be able to see it anymore in our output data. Also, by not quantifying the single samples we probably caused the effect that samples with high biomass like the colon and tongue tissue samples dominate the sequencing effort.
Lastly, we wanted to emphasize on the fact, that Ardigen found a significant amount of reads partially mapped to Hsa and Mmu, which were trimmed in the pipeline. However, the event that we generated a high amount of mapped Hsa and Mmu reads, suggests that our primers might not have been optimal in our case. Therefore, we would suggest using, the probably most used 16S rRNA primers 515F and 806R (*Geller et al. 2017*).

Nevertheless, due to our high contamination prevalence in our samples, the probability that we would not be able to perform a pure library and moreover, the fact that we do not own a MiSeq platform therefore, making it unpractical to sequence many samples in different libraries, we decided to not pursue any of those experimental adjustments.
When building up the pipeline it is probably suggestable to filter out infrequent species, especially when handling low bacterial biomass samples. Infrequent species should be removed from the "*dataset until the mean site similarity (Jaccard similarity) is at least 20%. Therefore, communities with very low similarity produce less specific co–occurrence networks. Even though, one might lose diversity, it would increase the sensitivity of the network"*.
Despite the technical biases that we have listed, biological processes also have to be taken in account which can lead to biased results. "*Microorganisms have variable copy number of targeted genes such as rRNA genes* (*Farrelly et al. 1995*) *that can influence abundance in sequencing libraries. Additionally, the targeted gene or gene region may have insufficient resolution to resolve different species, as in the case for the 16S rRNA gene in the Escherichia coli* (*Khot & Fisher, 2013*). *In order to minimize the biological events possible, we recommend using the highest level of sequence resolution possible with the targeted gene and sequencing approach".*

At last, Eisenhofer et al. suggest the use of a non–DNA sequencing approach (Fluorescence *in situ* hybridization) to verify the sequencing output data, which is the approach that we chose to pursue to validate our sequencing method.

V. *In Situ* Hybridization (ISH)

We decided to perform *in situ* hybridization since this method does not include any PCR and Therefore, is less contamination prone. The aim of this experiment was to determine whether or not ISH would be a possible alternative to 16S sequencing and whether or not we find bacteria present within the tumour tissue. So far ISH is not as widely used compared to 16S. It is mostly performed when trying to find out the exact location in or on the certain cell or tissue type, after already having identified the taxa by 16S sequencing. The reason for the unpopularity of ISH is simple explained.
First, when comparing ISH and 16S sequencing, the work effort increases, while 96 samples can be processed simultaneously with 16S sequencing, in ISH each sample needs to be processed one by one. This is quite a disadvantage since microbiome analysis normally takes a tremendous amount of samples to obtain reliable and representable results. Second, while ISH gives information about the location of the bacteria, it is not able to identify taxa levels with a general probe, something that 16S sequencing is able to do. Lastly, even though, ISH is less contamination prone, the background–noise problem is stronger prevalent in this method. According to literature we would expect one bacteria cell per 150 tumour cells (*Geller et al. 2017*), meaning that the signal would be already very weak by default therefore, it might become difficult to differentiate between background and signal in the case of ISH.
The most prominent probe to target the 16S rRNA region of bacteria is the EUB 338 probe, here most people refer to a variant of the probe cited in a paper by Amann et al. (*Amann et al. 1990*). This probe is a rRNA probe that hybridizes with the covering 16S rRNA region of eubacteria thus, generating a signal. However, it is not commercially available since, like the primers used in 16S sequencing, they are normally adapted individually. Due to the fact that we wanted to keep the experiment as simple as possible and use it as a proof of method, we

decided to use the commercially available probe RNAscope®Probe–EB–16S rRNA (Advanced Cell Diagnostics, Inc.), covering the same region as the EUB 338 probe. As samples to analyse we chose the MIA_PaCa-2 model where we analysed tongue, muscle, colon and tumour tissue (**Figure 11**).

The *technical positive control*, where the human tumour tissue was hybridized with a general probe, is as expected positive (**Figure 11.E**.) thus, demonstrating that the hybridization protocol works and there are no technical issues.

The *negative control*, where *muscle tissue* from the MIA_PaCa-2 CDX model was stained with our EB probe (**Figure 11.F**.) is as expected negative towards the probe. There are no bacteria visible in that sample.

However, when looking at the *positive controls* the picture is ambiguous (**Figure 11.A.–C**.). The *Helicobacter pylori–positive control* slide should have been clearly positive however, only few bacteria are seen (**Figure 11.A**.). Despite of the event that there is very little bacteria signal, the signal present is very sharp and clear therefore, we determined it as a successful positive control.

The *tongue–positive control* slide display exactly the pattern we would have anticipated, where the bacteria lies on the papillary dermis (dorsum of tongue), the part of the tongue which is exposed to the environment while there is no signal in the connective tissue or epithelium of the tongue.

The *colon–positive control* slide is poorly colonized by bacteria, which is especially surprising since the colon should harbour most of the bacteria. Nevertheless, bacteria was detectable but not in the dimension that we have anticipated.

This experiment perfectly illustrates the importance of technical replicates in ISH. Due to the fact that we processed our analysis without technical replicates we cannot determine our results as outliners or not. Another reason for the poor signal might be that the tissues were not processed directly after dissection but were fresh frozen and stored at -80 °C for. Generally, it is better to process tissue directly after dissection for staining to obtain the “realest” picture therefore, immediate 10% neutral buffered formalin fixing subsequently to dissection.

Nevertheless, overall we can say that the *positive–controls* are sufficiently sophisticated to proof the technical function of the method as valid. Still, the EB probe that we have chosen, might just not be sensitive enough, such that not every bacteria is detected which results in little signal. But when comparing the *positive–controls* with the *tumour tissue* we can clearly see that the *tumour tissue* does not harbour many bacteria. Only single dots like presented in **Figure 11.D** are visible, which makes it hard to interpret the result.

Currently, it functions as a proof of method. ISH visualizes successfully bacteria and is less contamination prone than sequencing. The challenges are to score a good noise–signal ratio by testing out different probes and their sensitivity.

Lastly, we would like to emphasize that we used experimental and technical singlets even though, when using ISH one should always use technical and experimental triplets to get conclusive statistics and results.

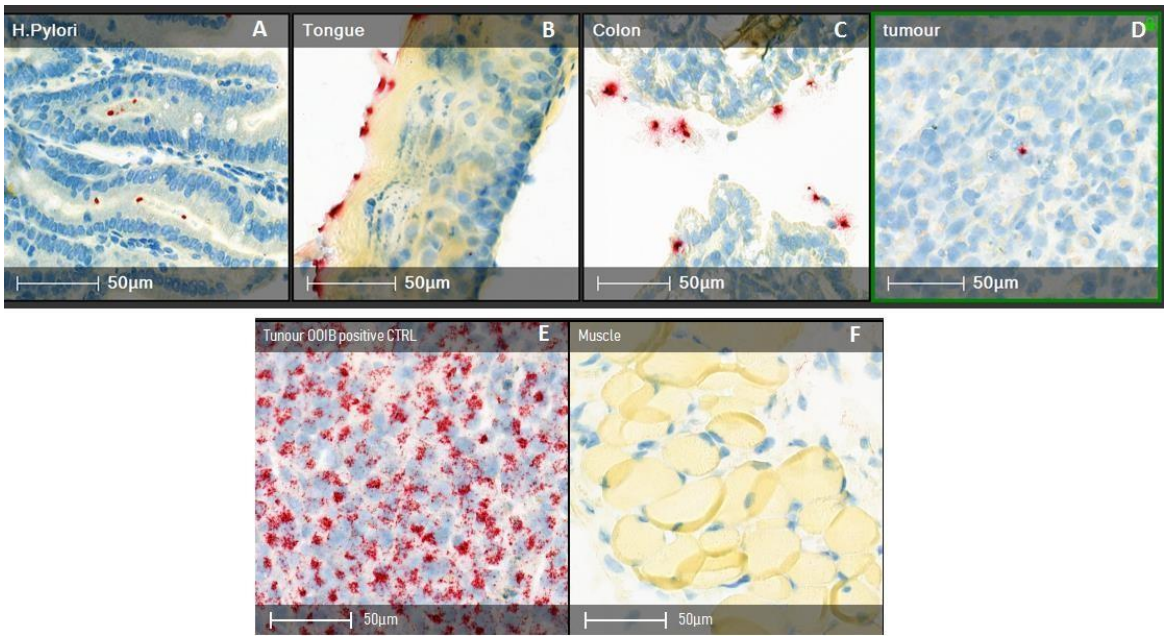

Figure 11. *in situ* hybridization staining x40 with RNAscope®Probe–EB–16S rRNA (red) of tissue samples from MIA_PaCa-2 CDX model
A *Positive control*–slide with *Helicobacter pylori* from mouse intestine **B** *Positive control*–tongue tissue **C** *Positive control*–colon tissue **D** *tumour tissue* **E** *Technical positive control*–tumour tissue stained with general probe (RNAscope®Positive Control Probe-Hs-PPIB).
F *Negative control*–Muscle tissue

5. Conclusion

In conclusion, we can wrap up by saying that 16S rRNA sequencing is still not a fully established method and even though, the protocol is quite simple and low cost as well as little time consuming, the data yielded from that analysis is hard to validate if the method has not been verified thoroughly before.
Each step of the analysis needs establishment, validation and detailed logging, such that the steps can be traced back and are comprehensible. According to our results, 16S rRNA sequencing is sufficient to make conclusions about relative abundances on phyla level with a satisfying confidence, when comparing two cohorts with each other. However, it would not be advisable to rely on this method to identify single genera, investigate the significant abundance of classes or quantification of the output data.
The technical differences aside, if we look at the relative abundances of our CDX models, we suspect that bacteria are able to interact with the tumour tissue. We saw a significant difference between the cell line samples and the tumour tissue. Therefore, the xenograft is most likely to be able to communicate and alter, influenced by its environment. The mechanisms of how or why the bacteria enter or exit the tumour tissue are yet to be investigated.
In our study we could not verify the previous hypothesis that KRAS mutations play a role in that matter, since we did not find any significant differences between the KRAS mutated and WT CDX models.
Nevertheless, we cannot ensure at this point that our results from 16S sequencing are real or just contamination. We have observed along our analysis that microbiome studies in general are not standardized and too laxly performed without any usage of positive or negative controls. Therefore, publishing results with a method that has not been prior validated. The probability is high that in the past, contamination rather than descriptive results have been reported which is one of the reasons why microbiome analysis is widely criticized (*Eisenhofer et al. 2019; Dorothy Kim et al. 2017; Klindworth et al. 2013; Salter et al. 2014*).
In our case, we would have to re–evaluate and re–establish our method for future experiments (including more blanks, which then can be incorporated in the pipeline as a contamination factor, validation of the PCR amplification by performing a qPCR, using fresh reagents etc.). **Figure 12** exhibits an agenda that should be followed when performing DNA- based sequencing to analyse microbiota in low biomass samples. It should serve as a guideline to improve the sequencing procedure and consequently the results.

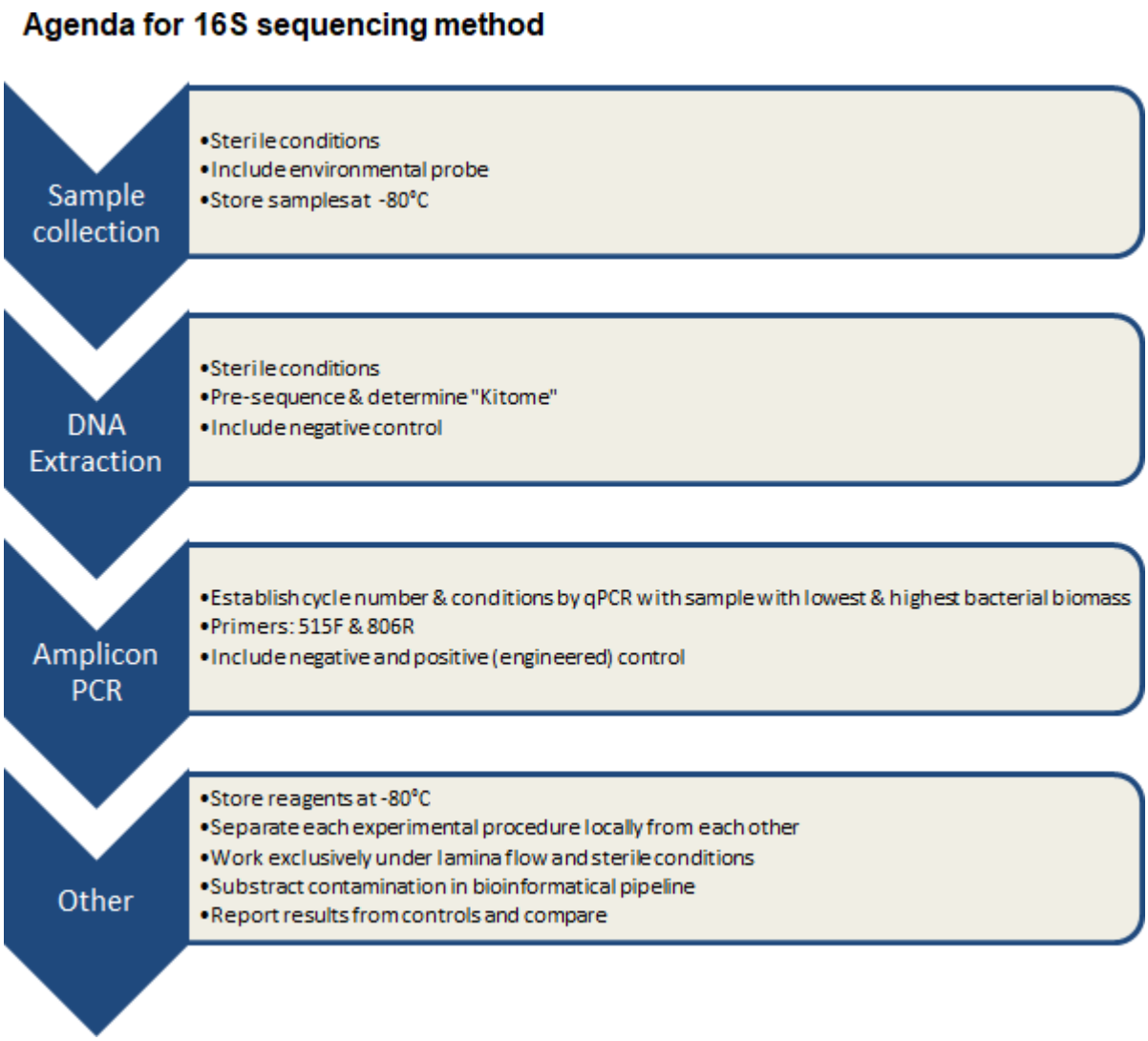

Figure 12 Collection of recommendations to overcome challenges of low microbial biomass microbiome assessed by 16S rRNA sequencing.

As an alternative, we would suggest to always perform FISH or ISH alongside sequencing to include a non DNA sequencing approach and verify sequencing results.
In our case ISH did work from the technical side, however, our simple experiment displayed that technical and experimental triplets of each sample are necessary to receive conclusive results. Additionally, one would have to look for a probe, which is more sensitive, since in our case we have to overcome the problem of the high noise to signal ratio.
Microbiome analysis is still a shaky field of research especially, since even though much data has been collected in the past, the causality between the different microbial patterns is yet to be determined. Additionally, due to the arising contamination dispute already published data might just be distorted. Microbiome analysis is not straightforward and many investigations are aiming on finding better approaches. In the future different techniques would have to be combined to achieve the most reliable and representable results. A forthcoming approach is FISH and ISH analysis, which is combined with other techniques to improve its sophistication. One of the many possibilities would be to combine FISH or ISH with fluorescence–activated cell sorting (FACS) which would enable quantification (*Vandeputte et al. 2017*) or laser capture microdissection (LCM) (*Pédron et al. 2012*) which would permit detailed characterization and identification of bacterial taxa. Nevertheless, since those kind of

protocols need much ground and preliminary work, the focus still relies on 16S sequencing due to its convenience. **Figure 13** presents a guideline of how microbiome analysis based on FISH or ISH could be exhibited in order to obtain sophisticated results.
Conclusively, we would suggest to distant ourselves from CDX models in general, since their bacterial load might be under a threshold such that no reliable results can be obtained by sequencing or ISH. We would propose to repeat the same experiment with PDX (Patient Derived Xenograft) samples to test whether the bacterial load of PDX samples are in the same dimension of CDX samples or harbour significantly more bacteria. Additionally, we would recommend adding colorectal cancer to the panel, due to the reason that colorectal cancer alongside gastric cancer are the two cancer types most associated with microbiome dysbiosis (*Pushalkar et al. 2018*). Therefore, both cancer types would be suitable since they most probably harbour bacteria even as CDX–models.

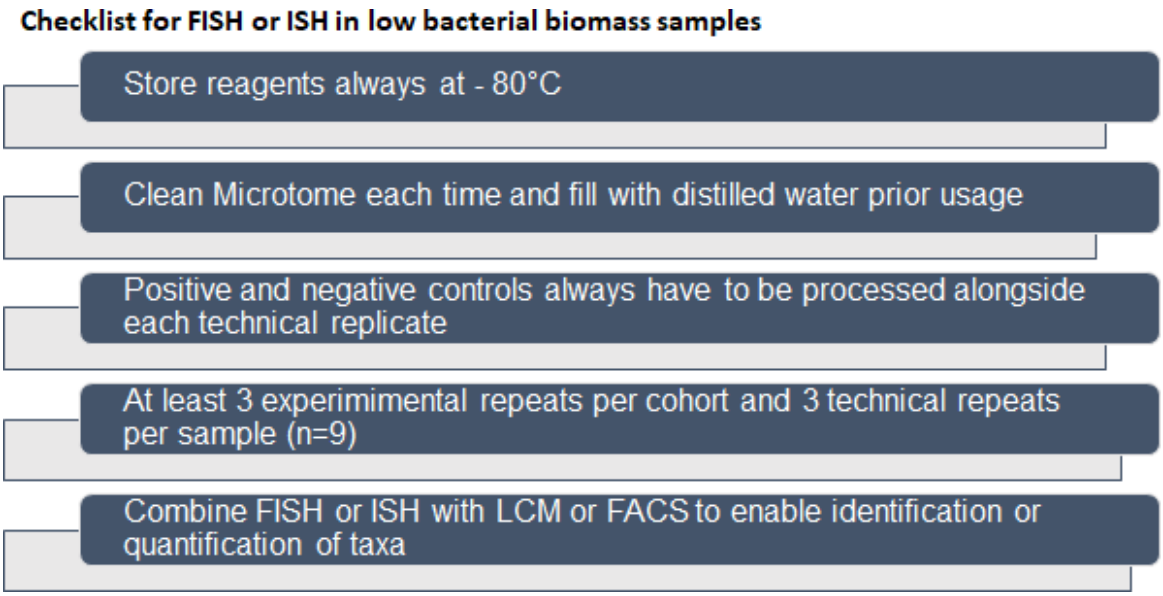

Figure 13 Guidance for performing sophisticated FISH or ISH on low bacterial biomass samples, including further analysis to obtain more detailed information about taxa.

6. Appendix

Table 10. Samples of each library

Cell Line	Sample Type	Library N°
AsPC-1, BxPC-3, CAPAN-1, HCC1806, HuP-T4, HT-1197, MDA-MB-231, MDA-MB-468, MIA_PaCa-2, UM-UC-3, SUM190PT	Cells	1 & 3
DSM-7492	*Helicobacter pylori*	1
DSM-8245	*Clostridium sp.*	1
AsPC-1	CDX–Skin, Tongue, Tumour, Colon, *Hypoxia*, *Anoxia*	2
BxPC-3	CDX–Tongue, Tumour, Colon	2
CAPAN-1	CDX–Skin, Tongue, Tumour, Colon, *Hypoxia*, *Anoxia*	2
HCC1806	CDX–Skin, Tongue, Tumour, Colon, *Hypoxia*, *Anoxia*	1
HuP-T4	CDX–Tumour, Colon	3
HT-1197	CDX–Tumour, Colon	3
MDA-MB-231	CDX–Tumour, Colon	3
MDA-MB-468	CDX–Tumour, Colon	3
MIA_PaCa-2	CDX–Skin, Tongue, Tumour, Colon, *Hypoxia*, *Anoxia*	1
UM-UC-3	CDX–Skin, Tongue, Tumour, Colon, *Hypoxia*, *Anoxia*	2
SUM190PT	CDX–Skin, Tongue, Tumour, Colon, *Hypoxia*, *Anoxia*	2
HCC1806	CDX–Tumour, Colon	3
MIA_PaCa-2	CDX–Tumour, Colon	3
PCR Control	PBS sterile PBS sterile with Cellkit PBS environmental with Cellkit FCS sterile w Cellkit FCS with Cellkit PBS sterile with Tissuekit FCS sterile with Tissuekit FCS with Tissuekit H2O	3

Table 11. PCR Controls used in the 3rd Library

Control	Extraction Kit
FCS sterile filtered QIAamp® Fast DNA kit (QIAGEN GmbH, Germany)	Tissue Kit
FCS QIAamp® Fast DNA kit (QIAGEN GmbH, Germany)	Tissue Kit
FCS sterile filtered QIAamp® DNA Mini kit (QIAGEN GmbH, Germany)	Cell Kit
PBS sterile QIAamp® Fast DNA kit (QIAGEN GmbH, Germany)	Tissue Kit
PBS sterile	No Kit
PBS sterile QIAamp® DNA Mini kit (QIAGEN GmbH, Germany)	Cell Kit
PBS environmental QIAamp® DNA Mini kit (QIAGEN GmbH, Germany)	Cell Kit
H2O	No Kit

7. References

Abdel.Aziz, M.I., Vijverberg, S.J.H., Neerincx, A.H., Aletta, D. K., Maitland-van der Zee, A.H. (2019). The cross-talk between microbiome and asthma: exploring associations and challenges. *Clin. Exo. Allergy,* doi: 10.1111/cea.13444. [Epub ahead of print]

Alexander, J. L., Wilson, I. D., Teare, J., Marchesi, J. R., Nicholson, J. K., & Kinross, J. M. (2017). Gut microbiota modulation of chemotherapy efficacy and toxicity. *Nature Reviews Gastroenterology &Amp; Hepatology, 14*, 356. doi:10.1038/nrgastro.2017.20

Amann, R. I., Binder, B. J., Olson, R. J., Chisholm, S. W., Devereux, R., & Stahl, D. A. (1990). Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl Environ Microbiol, 56*(6), 1919-1925.

Apostolou, P., Tsantsaridou, A., Papasotiriou, I., Toloudi, M., Chatziioannou, M., & Giamouzis, G. (2011). Bacterial and fungal microflora in surgically removed lung cancer samples. *Journal of Cardiothoracic Surgery, 6*, 137-137. doi:10.1186/1749-8090-6-137

Apprill, A., McNally, S., Parsons, R., & Weber, L. (2015). Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquatic Microbial Ecology, 75*(2), 129-137.

Baldridge, M. T., Nice, T. J., McCune, B. T., Yokoyama, C. C., Kambal, A., Wheadon, M., . . . Virgin, H. W. (2015). Commensal microbes and interferon-λ determine persistence of enteric murine norovirus infection. *Science, 347*(6219), 266-269. doi:10.1126/science.1258025

Banerjee, S., Tian, T., Wei, Z., Shih, N., Feldman, M. D., Alwine, J. C., . . . Robertson, E. S. (2017). The ovarian cancer oncobiome. *Oncotarget, 8*(22), 36225-36245. doi:10.18632/oncotarget.16717

Bender, J. M., Li, F., Adisetiyo, H., Lee, D., Zabih, S., Hung, L., . . . Aldrovandi, G. M. (2018). Quantification of variation and the impact of biomass in targeted 16S rRNA gene sequencing studies. *Microbiome, 6*(1), 155. doi:10.1186/s40168-018-0543-z

Bergey, D. H., & Etris, S. (1936). Active Immunization Against Tetanus Infection with Refined Tetanus Toxoid. *The Journal of Immunology, 31*(5), 363-371.

Bessa, L. J., Fazii, P., Di Giulio, M., & Cellini, L. (2015). Bacterial isolates from infected wounds and their antibiotic susceptibility pattern: some remarks about wound infection. *International Wound Journal, 12*(1), 47-52. doi:doi:10.1111/iwj.12049

Bik, E. M., Eckburg, P. B., Gill, S. R., Nelson, K. E., Purdom, E. A., Francois, F., . . . Relman, D. A. (2006). Molecular analysis of the bacterial microbiota in the human stomach. *Proceedings of the National Academy of Sciences of the United States of America, 103*(3), 732-737. doi:10.1073/pnas.0506655103

Bullman, S., Pedamallu, C. S., Sicinska, E., Clancy, T. E., Zhang, X., Cai, D., . . . Meyerson, M. (2017). Analysis of Fusobacterium persistence and antibiotic response in colorectal cancer. *Science, 358*(6369), 1443-1448. doi:10.1126/science.aal5240

Bültzingslöwen, I., Adlerberth, I., Wold, A. E., Dahlén, G., & Jontell, M. (2003). Oral and intestinal microflora in 5-fluorouracil treated rats, translocation to cervical and mesenteric lymph nodes and effects of probiotic bacteria. *Oral Microbiology and Immunology, 18*(5), 278-284. doi:doi:10.1034/j.1399-302X.2003.00075.x

Campbell, J. H., Foster, C. M., Vishnivetskaya, T., Campbell, A. G., Yang, Z. K., Wymore, A., . . . Podar, M. (2012). Host genetic and environmental effects on mouse intestinal microbiota. *The Isme Journal, 6*, 2033. doi:10.1038/ismej.2012.54
https://www.nature.com/articles/ismej201254#supplementary-information

Cantwell, A. R., Jr., & Kelso, D. W. (1981). Microbial findings in cancers of the breast and in their metastases to the skin. Implications for etiology. *J Dermatol Surg Oncol, 7*(6), 483-491.

Capon, D. J., Chen, E. Y., Levinson, A. D., Seeburg, P. H., & Goeddel, D. V. (1983). Complete nucleotide sequences of the T24 human bladder carcinoma oncogene and its normal homologue. *Nature, 302*(5903), 33-37.

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., . . . Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences of the United States of America, 108 Suppl 1*(Suppl 1), 4516-4522. doi:10.1073/pnas.1000080107

Carlsen, T., Aas, A. B., Lindner, D., Vrålstad, T., Schumacher, T., & Kauserud, H. (2012). Don't make a mista(g)ke: is tag switching an overlooked source of error in amplicon pyrosequencing studies? *Fungal Ecology, 5*(6), 747-749. doi:https://doi.org/10.1016/j.funeco.2012.06.003

Castellarin, M., Warren, R. L., Freeman, J. D., Dreolini, L., Krzywinski, M., Strauss, J., . . . Holt, R. A. (2012). Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. *Genome Res, 22*(2), 299-306. doi:10.1101/gr.126516.111

Chakravorty, S., Helb, D., Burday, M., Connell, N., & Alland, D. (2007). A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *Journal of microbiological methods, 69*(2), 330-339. doi:10.1016/j.mimet.2007.02.005

Collins, D., Hogan, A. M., & Winter, D. C. (2011). Microbial and viral pathogens in colorectal cancer. *The Lancet Oncology, 12*(5), 504-512. doi:10.1016/S1470-2045(10)70186-8

Commisso, C., Davidson, S. M., Soydaner-Azeloglu, R. G., Parker, S. J., Kamphorst, J. J., Hackett, S., . . . Bar-Sagi, D. (2013). Macropinocytosis of protein is an amino acid supply route in Ras-transformed cells. *Nature, 497*(7451), 633-637. doi:10.1038/nature12138

Commisso, C., Davidson, S. M., Soydaner-Azeloglu, R. G., Parker, S. J., Kamphorst, J. J., Hackett, S., . . . Bar-Sagi, D. (2013). Macropinocytosis of protein is an amino acid supply route in Ras-transformed cells. *Nature, 497*(7451), 633-637. doi:10.1038/nature12138

de Martel, C., Plummer, M., Parsonnet, J., van Doorn, L. J., & Franceschi, S. (2009). Helicobacter species in cancers of the gallbladder and extrahepatic biliary tract. *Br J Cancer, 100*(1), 194-199. doi:10.1038/sj.bjc.6604780

Ding, H. T., Taur, Y., & Walkup, J. T. (2017). Gut Microbiota and Autism: Key Concepts and Findings. *Journal of Autism and Developmental Disorders, 47*(2), 480-489. doi:10.1007/s10803-016-2960-9

Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., . . . Relman, D. A. (2005). Diversity of the Human Intestinal Microbial Flora. *Science (New York, N.Y.), 308*(5728), 1635-1638. doi:10.1126/science.1110591

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

Elderman, M., Hugenholtz, F., Belzer, C., Boekschoten, M., van Beek, A., de Haan, Bart., Savelkoul, Huub., de Vos, P. and Faas, Marijke F. (2018). Sex and strain dependent differences in mucosal immunology and microbiota composition in mice. *Biology of Sex Differences*, 9:26. doi: 10.1186/s13293-018-0186-6

Eppig, J. T., Blake, J. A., Bult, C. J., Kadin, J. A., & Richardson, J. E. (2015). The Mouse Genome Database (MGD): facilitating mouse as a model for human biology and disease. *Nucleic Acids Res, 43*(Database issue), D726-736. doi:10.1093/nar/gku967

Faith, J. J., Rey, F. E., O'Donnell, D., Karlsson, M., McNulty, N. P., Kallstrom, G., . . . Gordon, J. I. (2010). Creating and characterizing communities of human gut microbes in gnotobiotic mice. *The Isme Journal, 4*, 1094. doi:10.1038/ismej.2010.110

Farrelly, V., Rainey, F., & Stackebrandt, E. (1995). *Effect of genome size and rrn gene copy number on PCR amplification of 16S rRNA genes from a mixture of bacterial species* (Vol. 61).

Fierer, N., Lauber, C. L., Zhou, N., McDonald, D., Costello, E. K., & Knight, R. (2010). Forensic identification using skin bacterial communities. *Proceedings of the National Academy of Sciences, 107*(14), 6477-6481. doi:10.1073/pnas.1000162107

Filyk, H. A., & Osborne, L. C. (2016). The Multibiome: The Intestinal Ecosystem's Influence on Immune Homeostasis, Health, and Disease. *EBioMedicine, 13*, 46-54. doi:https://doi.org/10.1016/j.ebiom.2016.10.007

Finkel, O. M., Burch, A. Y., Lindow, S. E., Post, A. F., & Belkin, S. (2011). Geographical location determines the population structure in phyllosphere microbial communities of a salt-excreting desert tree. *Applied and environmental microbiology, 77*(21), 7647-7655. doi:10.1128/AEM.05565-11

Fitz-Gibbon, S., Tomida, S., Chiu, B.-H., Nguyen, L., Du, C., Liu, M., . . . Li, H. (2013). *Propionibacterium Acnes Strain Populations in the Human Skin Microbiome Associated with Acne* (Vol. 133).

Franzosa, E. A., Huang, K., Meadow, J. F., Gevers, D., Lemon, K. P., Bohannan, B. J. M., & Huttenhower, C. (2015). Identifying personal microbiomes using metagenomic codes. *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.1423854112

Fuccio, L., Zagari, R., Eusebi, L., & et al. (2009). Meta-analysis: Can helicobacter pylori eradication treatment reduce the risk for gastric cancer? *Annals of Internal Medicine, 151*(2), 121-128. doi:10.7326/0003-4819-151-2-200907210-00009

Fujimori, M. (2008). Anaerobic bacteria as a gene delivery system for cancer therapy. *Drug Delivery System, 23*(5), 560-566. doi:10.2745/dds.23.560

Geller, L. T., Barzily-Rokni, M., Danino, T., Jonas, O. H., Shental, N., Nejman, D., . . . Straussman, R. (2017). Potential role of intratumor bacteria in mediating tumor resistance to the chemotherapeutic drug gemcitabine. *Science, 357*(6356), 1156-1160. doi:10.1126/science.aah5043

Gilbert, J. A., Jansson, J. K., & Knight, R. (2014). The Earth Microbiome project: successes and aspirations. *BMC Biol, 12*, 69. doi:10.1186/s12915-014-0069-1

Gill, S. R., Pop, M., DeBoy, R. T., Eckburg, P. B., Turnbaugh, P. J., Samuel, B. S., . . . Nelson, K. E. (2006). Metagenomic Analysis of the Human Distal Gut Microbiome. *Science, 312*(5778), 1355-1359. doi:10.1126/science.1124234

Glassing, A., Dowd, S. E., Galandiuk, S., Davis, B., & Chiodini, R. J. (2016). Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples. *Gut Pathog, 8*. doi:10.1186/s13099-016-0103-7

Goodman, B., & Gardner, H. (2018). The microbiome and cancer. *The Journal of Pathology, 244*(5), 667-676. doi:doi:10.1002/path.5047

Gopalakrishnan, V., Spencer, C. N., Nezi, L., Reuben, A., Andrews, M. C., Karpinets, T. V., . . . Wargo, J. A. (2018). Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. *Science, 359*(6371), 97-103. doi:10.1126/science.aan4236

Hashimoto, T., Perlot, T., Rehman, A., Trichereau, J., Ishiguro, H., Paolino, M., . . . Penninger, J. M. (2012). ACE2 links amino acid malnutrition to microbial ecology and intestinal inflammation. *Nature, 487*, 477. doi:10.1038/nature11228
https://www.nature.com/articles/nature11228#supplementary-information

Hehemann, J.-H., Correc, G., Barbeyron, T., Helbert, W., Czjzek, M., & Michel, G. (2010). Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature, 464*, 908. doi:10.1038/nature08937
https://www.nature.com/articles/nature08937#supplementary-information

Hernandez, C. J., Guss, J. D., Luna, M., & Goldring, S. R. (2016). Links Between the Microbiome and Bone. *Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research, 31*(9), 1638-1646. doi:10.1002/jbmr.2887

Hooper, S. J., Crean, S. J., Fardy, M. J., Lewis, M. A., Spratt, D. A., Wade, W. G., & Wilson, M. J. (2007). A molecular analysis of the bacteria present within oral squamous cell carcinoma. *J Med Microbiol, 56*(Pt 12), 1651-1659. doi:10.1099/jmm.0.46918-0

Hubbard, A. L., Harrison, D. J., Moyes, C., & McOrist, S. (1998). Direct detection of eae-positive bacteria in human and veterinary colorectal specimens by PCR. *J Clin Microbiol, 36*(8), 2326-2330.

Hugenholtz, F., & De Vos, W. (2017). *Mouse models for human intestinal microbiota research: a critical evaluation* (Vol. 75).

Hugenholtz, F., & de Vos, W. M. (2018). Mouse models for human intestinal microbiota research: a critical evaluation. *Cell Mol Life Sci, 75*(1), 149-160. doi:10.1007/s00018-017-2693-8

Iida, N., Dzutsev, A., Stewart, C. A., Smith, L., Bouladoux, N., Weingarten, R. A., . . . Goldszmid, R. S. (2013). Commensal Bacteria Control Cancer Response to Therapy by Modulating the Tumor Microenvironment. *Science, 342*(6161), 967-970. doi:10.1126/science.1240527

Initial sequence of the chimpanzee genome and comparison with the human genome. (2005). *Nature, 437*(7055), 69-87. doi:10.1038/nature04072

Jangi, S., Gandhi, R., Cox, L. M., Li, N., von Glehn, F., Yan, R., . . . Weiner, H. L. (2016). Alterations of the human gut microbiome in multiple sclerosis. *Nature Communications, 7*, 12015. doi:10.1038/ncomms12015

Jonsson, A. L., & Bäckhed, F. (2016). Role of gut microbiota in atherosclerosis. *Nature Reviews Cardiology, 14*, 79. doi:10.1038/nrcardio.2016.183

Kane, A., M Dinh, D., & Ward, H. (2014). *Childhood Malnutrition and the Intestinal Microbiome* (Vol. 77).

Kennedy, K., Hall, M. W., Lynch, M. D. J., Moreno-Hagelsieb, G., & Neufeld, J. D. (2014). Evaluating Bias of Illumina-Based Bacterial 16S rRNA Gene Profiles. *Applied and environmental microbiology, 80*(18), 5717-5722. doi:10.1128/aem.01451-14

Khot, P. D., & Fisher, M. A. (2013). Novel approach for differentiating Shigella species and Escherichia coli by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J Clin Microbiol, 51*(11), 3711-3716. doi:10.1128/jcm.01526-13

Kim, D., Hofstaedter, C. E., Zhao, C., Mattei, L., Tanes, C., Clarke, E., . . . Kelsen, J. (2017). Optimizing methods and dodging pitfalls in microbiome research. *Microbiome, 5*. doi:10.1186/s40168-017-0267-5

Kim, D., Hofstaedter, C. E., Zhao, C., Mattei, L., Tanes, C., Clarke, E., . . . Bittinger, K. (2017). Optimizing methods and dodging pitfalls in microbiome research. *Microbiome, 5*(1), 52. doi:10.1186/s40168-017-0267-5

Kiraly, D. D., Walker, D. M., Calipari, E. S., Labonte, B., Issler, O., Pena, C. J., . . . Nestler, E. J. (2016). Alterations of the Host Microbiome Affect Behavioral Responses to Cocaine. *Scientific Reports, 6*, 35455. doi:10.1038/srep35455

Kleiman, S. C., Watson, H. J., Bulik-Sullivan, E. C., Huh, E. Y., Tarantino, L. M., Bulik, C. M., & Carroll, I. M. (2015). The Intestinal Microbiota in Acute Anorexia Nervosa and During Renourishment: Relationship to Depression, Anxiety, and Eating Disorder Psychopathology. *Psychosomatic Medicine, 77*(9), 969-981. doi:10.1097/psy.0000000000000247

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., & Glöckner, F. O. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res, 41*(1), e1-e1. doi:10.1093/nar/gks808

Knight, R., Callewaert, C., Marotz, C., Hyde, E. R., Debelius, J. W., McDonald, D., & Sogin, M. L. (2017). The Microbiome and Human Biology. *Annual Review of Genomics and Human Genetics, 18*(1), 65-86. doi:10.1146/annurev-genom-083115-022438

Kobayashi, T., Glatz, M., Horiuchi, K., Kawasaki, H., Akiyama, H., Kaplan, D. H., . . . Nagao, K. (2015). Dysbiosis and Staphylococcus aureus colonization drives inflammation in atopic dermatitis. *Immunity, 42*(4), 756-766. doi:10.1016/j.immuni.2015.03.014

Kostic, A. D., Xavier, R. J., & Gevers, D. (2014). The Microbiome in Inflammatory Bowel Diseases: Current Status and the Future Ahead. *Gastroenterology, 146*(6), 1489-1499. doi:10.1053/j.gastro.2014.02.009

Langille, M.G., Meehane, C.J., Koenig, J.E., Dhanani, A.S., Rose R.A., Howlett, S.E., Beiko, R.G. (2014). Microbial shifts in the aging mouse gut. *Microbiome,* 2(1).50. doi: 10.1186/s40168-014-0050-9

Larsson, A. J. M., Stanley, G., Sinha, R., Weissman, I. L., & Sandberg, R. (2018). Computational correction of index switching in multiplexed sequencing libraries. *Nature Methods, 15*, 305. doi:10.1038/nmeth.4666
https://www.nature.com/articles/nmeth.4666#supplementary-information

Lauber, C. L., Zhou, N., Gordon, J. I., Knight, R., & Fierer, N. (2010). Effect of storage conditions on the assessment of bacterial community structure in soil and human-associated samples. *FEMS Microbiology Letters, 307*(1), 80-86. doi:doi:10.1111/j.1574-6968.2010.01965.x

Lauder, A. P., Roche, A. M., Sherrill-Mix, S., Bailey, A., Laughlin, A. L., Bittinger, K., . . . Bushman, F. D. (2016). Comparison of placenta samples with contamination controls does not provide evidence for a distinct placenta microbiota. *Microbiome, 4*. doi:10.1186/s40168-016-0172-3

Lehouritis, P., Cummins, J., Stanton, M., Murphy, C. T., McCarthy, F. O., Reid, G., . . . Tangney, M. (2015). Local bacteria affect the efficacy of chemotherapeutic drugs. *Scientific Reports, 5*, 14554. doi:10.1038/srep14554
https://www.nature.com/articles/srep14554#supplementary-information

Ley, R. E., Turnbaugh, P. J., Klein, S., & Gordon, J. I. (2006). Human gut microbes associated with obesity. *Nature, 444*, 1022. doi:10.1038/4441022a
https://www.nature.com/articles/4441022a#supplementary-information

Lin, X. B., Dieleman, L. A., Ketabi, A., Bibova, I., Sawyer, M. B., Xue, H., . . . Gänzle, M. G. (2012). Irinotecan (CPT-11) Chemotherapy Alters Intestinal Microbiota in Tumour Bearing Rats. *PLoS ONE, 7*(7), e39764. doi:10.1371/journal.pone.0039764

Lopez-Crapez, E., Chypre, C., Saavedra, J., Marchand, J., & Grenier, J. (1997). Rapid and large-scale method to detect K-ras gene mutations in tumor samples. *Clin Chem, 43*(6 Pt 1), 936-942.

Low, K. B., Ittensohn, M., Le, T., Platt, J., Sodi, S., Amoss, M., . . . Bermudes*, D. (1999). Lipid A mutant Salmonella with suppressed virulence and TNFα induction retain tumor-targeting in vivo. *Nature Biotechnology, 17*, 37. doi:10.1038/5205

Ma, W., Gilligan, B. M., Yuan, J., & Li, T. (2016). Current status and perspectives in translational biomarker research for PD-1/PD-L1 immune checkpoint blockade therapy. *Journal of Hematology & Oncology, 9*(1), 47. doi:10.1186/s13045-016-0277-y

Marchesi, J. R., Dutilh, B. E., Hall, N., Peters, W. H., Roelofs, R., Boleij, A., & Tjalsma, H. (2011). Towards the human colorectal cancer microbiome. *PLoS ONE, 6*(5), e20447. doi:10.1371/journal.pone.0020447

Methé, B. A., Nelson, K. E., Pop, M., Creasy, H. H., Giglio, M. G., Huttenhower, C., . . . White, O. (2012). A framework for human microbiome research. *Nature, 486*(7402), 215-221. doi:10.1038/nature11209

Mizrahi-Man, O., Davenport, E. R., & Gilad, Y. (2013). Taxonomic classification of bacterial 16S rRNA genes using short sequencing reads: evaluation of effective study designs. *PLoS ONE, 8*(1), e53608. doi:10.1371/journal.pone.0053608

Morgan, X. C., Segata, N., & Huttenhower, C. (2013). Biodiversity and functional genomics in the human microbiome. *Trends in Genetics, 29*(1), 51-58. doi:https://doi.org/10.1016/j.tig.2012.09.005

Nath, G., Singh, H., & Shukla, V. K. (1997). Chronic typhoid carriage and carcinoma of the gallbladder. *Eur J Cancer Prev, 6*(6), 557-559.

Nemunaitis, J., Cunningham, C., Senzer, N., Kuhn, J., Cramm, J., Litz, C., . . . Sznol, M. (2003). Pilot trial of genetically modified, attenuated Salmonella expressing the E. coli cytosine deaminase gene in refractory cancer patients. *Cancer Gene Therapy, 10*, 737. doi:10.1038/sj.cgt.7700634

Nguyen, T. L. A., Vieira-Silva, S., Liston, A., & Raes, J. (2015). How informative is the mouse for human gut microbiota research? *Disease Models & Mechanisms, 8*(1), 1-16. doi:10.1242/dmm.017400

Osswald, A., Zhongke, S., Grimm, V., Ampem, G., Riegel, K., Westendorf, A.M., Somergruber, W., Otte, K., Fürre, P., Riedel, C.U. (2015). Three-dimensional tumor spheroids for in vitro analysis of bacteria as gene delivery vectors in tumor therapy. *Microbial Cell Fact,* 14:199. doi:https://doi.org/10.1186/s12934-015-0383-5

Pace, B., & Campbell, L. L. (1971). Homology of Ribosomal Ribonucleic Acid of Diverse Bacterial Species with Escherichia coli and Bacillus stearothermophilus. *Journal of Bacteriology, 107*(2), 543-547.

Panebianco, C., Adamberg, K., Jaagura, M., Copetti, M., Fontana, A., Adamberg, S., . . . Pazienza, V. (2018). Influence of gemcitabine chemotherapy on the microbiota of pancreatic cancer xenografted mice. *Cancer Chemotherapy and Pharmacology, 81*(4), 773-782. doi:10.1007/s00280-018-3549-0

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

Parada, A. E., Needham, D. M., & Fuhrman, J. A. (2016). Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ Microbiol, 18*(5), 1403-1414. doi:10.1111/1462-2920.13023

Pédron, T., Mulet, C., Dauga, C., Frangeul, L., Chervaus, C., Grompone, G., Sansonetti, P.J. (2012). A Crypt-Specific Core Microbiota Resides in the Mouse Colon. *mBio*, 3(3) doi: 10.1128/mBio.00116-12

Pope, J. L., Tomkovich, S., Yang, Y., & Jobin, C. (2017). Microbiota as a mediator of cancer progression and therapy. *Translational research : the journal of laboratory and clinical medicine, 179*, 139-154. doi:10.1016/j.trsl.2016.07.021

Pushalkar, S., Hundeyin, M., Daley, D., Zambirinis, C. P., Kurz, E., Mishra, A., . . . Miller, G. (2018). The Pancreatic Cancer Microbiome Promotes Oncogenesis by Induction of Innate and Adaptive Immune Suppression. *Cancer Discovery, 8*(4), 403-416. doi:10.1158/2159-8290.cd-17-1134

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., . . . Wang, J. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature, 464*, 59. doi:10.1038/nature08821
https://www.nature.com/articles/nature08821#supplementary-information

Raymond, F., Ahmed Ouameur, A., Deraspe, M., Iqbal, N., Gingras, H., Dridi, B., . . . Corbeil, J. (2015). *The initial state of the human gut microbiome determines its reshaping by antibiotics* (Vol. 10).

Redelman-Sidi, G., Iyer, G., Solit, D., & Glickman, M. S. (2013). Oncogenic activation of Pak1-dependent pathway of macropinocytosis determines BCG entry into bladder cancer cells. *Cancer Res, 73*(3), 1156-1167. doi:10.1158/0008-5472.can-12-1882

Redelman-Sidi, G., Iyer, G., Solit, D. B., & Glickman, M. S. (2013). Oncogenic Activation of Pak1-Dependent Pathway of Macropinocytosis Determines BCG Entry into Bladder Cancer Cells. *Cancer Research, 73*(3), 1156-1167. doi:10.1158/0008-5472.can-12-1882

Relman, D. A. (2012). The human microbiome: ecosystem resilience and health. *Nutrition Reviews, 70*(s1), S2-S9. doi:doi:10.1111/j.1753-4887.2012.00489.x

Ridaura, V. K., Faith, J. J., Rey, F. E., Cheng, J., Duncan, A. E., Kau, A. L., . . . Gordon, J. I. (2013). Cultured gut microbiota from twins discordant for obesity modulate adiposity and metabolic phenotypes in mice. *Science (New York, N.Y.), 341*(6150), 10.1126/science.1241214. doi:10.1126/science.1241214

Rogers, M. B. (2011). Mycoplasma and cancer: in search of the link. *Oncotarget, 2*(4), 271-273.

Sally, H. (2018). Microbiome Modulator Drugs - The New Generation of Therapeutic. *Pharma Intelligence* ©.

Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., . . . Walker, A. W. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biology, 12*(1), 87. doi:10.1186/s12915-014-0087-z

Sampson, T. R., Debelius, J. W., Thron, T., Janssen, S., Shastri, G. G., Ilhan, Z. E., . . . Mazmanian, S. K. (2016). Gut Microbiota Regulate Motor Deficits and Neuroinflammation in a Model of Parkinson's Disease. *Cell, 167*(6), 1469-1480.e1412. doi:10.1016/j.cell.2016.11.018

Sender, R., Fuchs, S., & Milo, R. (2016). Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biol, 14*(8), e1002533. doi:10.1371/journal.pbio.1002533

Shanmughapriya, S., Senthilkumar, G., Vinodhini, K., Das, B. C., Vasanthi, N., & Natarajaseenivasan, K. (2012). Viral and bacterial aetiologies of epithelial ovarian cancer. *Eur J Clin Microbiol Infect Dis, 31*(9), 2311-2317. doi:10.1007/s10096-012-1570-5

Shepherd, M. L., Swecker, W. S., Jr., Jensen, R. V., & Ponder, M. A. (2012). Characterization of the fecal bacteria communities of forage-fed horses by pyrosequencing of 16S rRNA V4 gene amplicons. *FEMS Microbiol Lett, 326*(1), 62-68. doi:10.1111/j.1574-6968.2011.02434.x

Sim, K., Cox, M. J., Wopereis, H., Martin, R., Knol, J., Li, M.-S., . . . Kroll, J. S. (2012). Improved Detection of Bifidobacteria with Optimised 16S rRNA-Gene Based Pyrosequencing. *PLoS ONE, 7*(3), e32543. doi:10.1371/journal.pone.0032543

Simrén, M., Barbara, G., Flint, H. J., Spiegel, B. M. R., Spiller, R. C., Vanner, S., . . . Zoetendal, E. G. (2013). Intestinal microbiota in functional bowel disorders: a Rome foundation report. *Gut, 62*(1), 159-176. doi:10.1136/gutjnl-2012-302167

SNEL, J., HEINEN, P. P., BLOK, H. J., CARMAN, R. J., DUNCAN, A. J., ALLEN, P. C., & COLLINS, M. D. (1995). Comparison of 16S rRNA Sequences of Segmented Filamentous Bacteria Isolated from Mice, Rats, and Chickens and Proposal of "Candidatus Arthromitus". *International Journal of Systematic and Evolutionary Microbiology, 45*(4), 780-782. doi:doi:10.1099/00207713-45-4-780

Sogin, S. J., Sogin, M. L., & Woese, C. R. (1971). Phylogenetic measurement in procaryotes by primary structural characterization. *J Mol Evol, 1*(1), 173-184.

Song, S. J., Lauber, C., Costello, E. K., Lozupone, C. A., Humphrey, G., Berg-Lyons, D., . . . Knight, R. (2013). Cohabiting family members share microbiota with one another and with their dogs. *eLife, 2*, e00458. doi:10.7554/eLife.00458

Stringer, A. M., Gibson, R. J., Logan, R. M., Bowen, J. M., Yeoh, A. S. J., Hamilton, J., & Keefe, D. M. K. (2009). Gastrointestinal Microflora and Mucins May Play a Critical Role in the Development of 5-Fluorouracil-Induced Gastrointestinal Mucositis. *Experimental Biology and Medicine, 234*(4), 430-441. doi:10.3181/0810-RM-301

Tang, W. H. W., & Hazen, S. L. (2014). The contributory role of gut microbiota in cardiovascular disease. *The Journal of Clinical Investigation, 124*(10), 4204-4211. doi:10.1172/JCI72331

The Human Microbiome Project, C. (2012). Structure, Function and Diversity of the Healthy Human Microbiome. *Nature, 486*(7402), 207-214. doi:10.1038/nature11234

Thomas, R. M., & Jobin, C. (2015). The Microbiome and Cancer: Is the 'Oncobiome' Mirage Real? *Trends in cancer, 1*(1), 24-35. doi:10.1016/j.trecan.2015.07.005

Toso, J. F., Gill, V. J., Hwu, P., Marincola, F. M., Restifo, N. P., Schwartzentruber, D. J., . . . Rosenberg, S. A. (2002). Phase I Study of the Intravenous Administration of Attenuated Salmonella typhimurium to Patients With Metastatic Melanoma. *Journal of Clinical Oncology, 20*(1), 142-152. doi:10.1200/jco.2002.20.1.142

Turnbaugh, P. J., & Gordon, J. I. (2009). The core gut microbiome, energy balance and obesity. *The Journal of Physiology, 587*(17), 4153-4158. doi:doi:10.1113/jphysiol.2009.174136

Turnbaugh, P. J., Ley, R. E., Mahowald, M. A., Magrini, V., Mardis, E. R., & Gordon, J. I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature, 444*, 1027. doi:10.1038/nature05414
https://www.nature.com/articles/nature05414#supplementary-information

Urbaniak, C., Cummins, J., Brackstone, M., Macklain, JM., Gloor, GB., Baban, CK., Scott, L., O'Hanlon, DM., Burton, JP., Francis, KP., Tangeny, M., Reid, G. (2014). Microbiota of human breast tissue. *Appl. Environ. Microbiol.*, 80 (10): 3007-14. doi: 10.1128/AEM.00242-14

Vandamme, T. F. (2014). Use of rodents as models of human diseases. *Journal of Pharmacy & Bioallied Sciences, 6*(1), 2-9. doi:10.4103/0975-7406.124301

Vandeputte, D., Kathagen, G., D'hoe, K., Vieira-Silva, S., Valles-Colomer, M., Sabino, J., Wang, Jun., Tito, R.Y., De Commer, L., Darzi, Y., Vermeire, S., Falony, G., Raes, J. (2017). Quantitative microbiome profiling links gut community variation to microbial load. *Nature,551,* p 507-51 doi:10.1038/nature24460

Vétizou, M., Pitt, J. M., Daillère, R., Lepage, P., Waldschmitt, N., Flament, C., . . . Zitvogel, L. (2015). Anticancer immunotherapy by CTLA-4 blockade relies on the gut microbiota. *Science (New York, N.Y.), 350*(6264), 1079-1084. doi:10.1126/science.aad1329

Wagner Mackenzie, B., Waite, D. W., & Taylor, M. W. (2015). Evaluating variation in human gut microbiota profiles due to DNA extraction method and inter-subject differences. *Frontiers in Microbiology, 6*(130). doi:10.3389/fmicb.2015.00130

Wallace, B. D., Wang, H., Lane, K. T., Scott, J. E., Orans, J., Koo, J. S., . . . Redinbo, M. R. (2010). Alleviating Cancer Drug Toxicity by Inhibiting a Bacterial Enzyme. *Science (New York, N.Y.), 330*(6005), 831-835. doi:10.1126/science.1191175

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and environmental microbiology, 73*(16), 5261-5267. doi:10.1128/AEM.00062-07

Weingarden, A., González, A., Vázquez-Baeza, Y., Weiss, S., Humphry, G., Berg-Lyons, D., . . . Sadowsky, M. J. (2015). Dynamic changes in short- and long-term bacterial composition following fecal microbiota transplantation for recurrent Clostridium difficile infection. *Microbiome, 3*(1), 10. doi:10.1186/s40168-015-0070-0

Wernitznig, D., Kiakos, K., Del Favero, G., Harrer, N., Machat, H., Osswald, A., Jakupec, M.A., Wernitznig, A., Sommergruber, W., Keppler, B.K. (2019). First-in-class ruthenium anticancer drug (KP1339/IT-139) induces an immunogenic cell death signature in colorectal spheroids *in vitro*. *Metallomics.* doi: 10.1039/c9mt00051h. [Epub ahead of print]

Woese, C. R. (1967). The Genetic Code: the Molecular basis for Genetic Expression.

Woese, C. R., & Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A, 74*(11), 5088-5090.

Wu, G. D., Lewis, J. D., Hoffmann, C., Chen, Y.-Y., Knight, R., Bittinger, K., . . . Bushman, F. D. (2010). Sampling and pyrosequencing methods for characterizing bacterial communities in the human gut using 16S sequence tags. *BMC microbiology, 10*, 206-206. doi:10.1186/1471-2180-10-206

Wylie, K. M., Truty, R. M., Sharpton, T. J., Mihindukulasuriya, K. A., Zhou, Y., Gao, H., . . . Pollard, K. S. (2012). Novel Bacterial Taxa in the Human Microbiome. *PLoS ONE, 7*(6), e35294. doi:10.1371/journal.pone.0035294

Xiao, L., Feng, Q., Liang, S., Sonne, S. B., Xia, Z., Qiu, X., . . . Kristiansen, K. (2015). A catalog of the mouse gut metagenome. *Nature Biotechnology, 33*, 1103. doi:10.1038/nbt.3353
https://www.nature.com/articles/nbt.3353#supplementary-information

Yang, B., Wang, Y., & Qian, P.-Y. (2016). Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. *BMC bioinformatics, 17*, 135-135. doi:10.1186/s12859-016-0992-y

Zeuthen, L. H., Christensen, H. R., & Frøkiær, H. (2006). Lactic Acid Bacteria Inducing a Weak Interleukin-12 and Tumor Necrosis Factor Alpha Response in Human Dendritic Cells Inhibit Strongly Stimulating Lactic Acid Bacteria but Act Synergistically with Gram-Negative Bacteria. *Clinical and Vaccine Immunology, 13*(3), 365-375. doi:10.1128/cvi.13.3.365-375.2006

Zhang, F., Zhang, M., Wang, Y., Li, C., & Chen, T. (2018). Comparison of the common bacteria in human and mouse tumours using high-throughput sequencing. *Mol Med Rep, 17*(5), 6717-6722. doi:10.3892/mmr.2018.8689

Zhang, X., Zhang, D., Jia, H., Feng, Q., Wang, D., Liang, D., . . . Wang, J. (2015). The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. *Nature Medicine, 21*, 895. doi:10.1038/nm.3914
https://www.nature.com/articles/nm.3914#supplementary-information

Zhao, M., Geller, J., Ma, H., Yang, M., Penman, S., & Hoffman, R. M. (2007). Monotherapy with a tumor-targeting mutant of Salmonella typhimurium cures orthotopic metastatic mouse models of human prostate cancer. *Proceedings of the National Academy of Sciences, 104*(24), 10170-10174. doi:10.1073/pnas.0703867104

Zheng, P., Zeng, B., Zhou, C., Liu, M., Fang, Z., Xu, X., . . . Xie, P. (2016). Gut microbiome remodeling induces depressive-like behaviors through a pathway mediated by the host's metabolism. *Molecular Psychiatry, 21*, 786. doi:10.1038/mp.2016.44
https://www.nature.com/articles/mp201644#supplementary-information

Zitvogel, L., Galluzzi, L., Viaud, S., Vétizou, M., Daillère, R., Merad, M., & Kroemer, G. (2015). Cancer and the gut microbiota: An unexpected link. *Science translational medicine, 7*(271), 271ps271-271ps271. doi:10.1126/scitranslmed.3010473

Zoetendal, E. G., Raes, J., van den Bogert, B., Arumugam, M., Booijink, C. C. G. M., Troost, F. J., . . . Kleerebezem, M. (2012). The human small intestinal microbiota is driven by rapid uptake and conversion of simple carbohydrates. *The Isme Journal, 6*, 1415. doi:10.1038/ismej.2011.212
https://www.nature.com/articles/ismej2011212#supplementary-information

Zozaya, M., Ferris, M. J., Siren, J. D., Lillis, R., Myers, L., Nsuami, M. J., . . . Martin, D. H. (2016). Bacterial communities in penile skin, male urethra, and vaginas of heterosexual couples with and without bacterial vaginosis. *Microbiome, 4*(1), 16. doi:10.1186/s40168-016-0161-6