haematologica
Journal of the Ferrata Storti Foundation

# The diversity of the microbiome impacts chronic lymphocytic leukemia development in mice and humans

by Tereza Faitová, Mariana Coelho, Caspar da Cunha-Bang, Selcen Öztürk, Ece Kartal, Peer Bork, Martina Seiffert, and Carsten U. Niemann

*Received: November 17, 2024.*
*Accepted: April 30, 2024.*

*Citation: Tereza Faitová, Mariana Coelho, Caspar da Cunha-Bang, Selcen Öztürk, Ece Kartal, Peer Bork, Martina Seiffert, and Carsten U. Niemann. The diversity of the microbiome impacts chronic lymphocytic leukemia development in mice and humans.*
*Haematologica. 2024 May 9. doi: 10.3324/haematol.2023.284693 [Epub ahead of print]*

# The diversity of the microbiome impacts chronic lymphocytic leukemia development in mice and humans

*Tereza Faitová[1], *Mariana Coelho[2,3], Caspar da Cunha-Bang[1], Selcen Öztürk[2], Ece Kartal[4], Peer Bork[4,5,6,7], #Martina Seiffert[2], #Carsten U. Niemann[1,8]

**Affiliations**

[1] Department of Hematology, Rigshospitalet, Copenhagen, Denmark

[2] Department of Molecular Genetics, German Cancer Research Center (DKFZ), Heidelberg, Germany

[3] Faculty of Biosciences of the University of Heidelberg, Heidelberg, Germany

[4] Structural and Computational Biology Unit, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany

[5] Department of Bioinformatics, Biocenter, University of Würzburg, Würzburg, Germany

[6] Yonsei Frontier Lab (YFL), Yonsei University, Seoul, South Korea

[7] Max Delbrück Center for Molecular Medicine, Berlin, Germany

[8] Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark

**Author Contributions**

CUN, MS, TF and MC designed the study, CdCB was responsible for sample collection and inclusion of patients, SÖ planned and performed the mouse study, EK and PB performed the sequencing of the murine microbiome samples, TF and MC performed the bioinformatics and statistical analyses, TF performed visualization of the results, TF and MC wrote the first version of the manuscript, CUN and MS reviewed and edited further versions of the manuscript, all authors read, contributed and approved the final version of the manuscript.

TF and MC contributed equally as co-first authors. CUN and MS contributed equally to this work as last co-authors.

**Running head**

Microbiome Impact on CLL Development in Mice and Humans

**Corresponding author**

Carsten U. Niemann: carsten.utoft.niemann@regionh.dk and Martina Seiffert: m.seiffert@dkfz-heidelberg.de (for human and mouse data respectively)

**Data Availability Statement**

The datasets generated and analyzed during this study are derived from patients treated in Denmark. The datasets contain sensitive patient data governed by GDPR and Danish law. Due to Danish legislation (Act No. 502 of 23 May 2018) and approvals granted by the Danish Data Protection Agency, it is not possible to upload raw data to a publicly available database. However, access to these data can be made available from the corresponding author on reasonable request, provided a data transfer agreement is entered into according to current regulations. Taxonomical profiling data of the murine cohort as well as bioinformatics analysis implemented in R are available at:
https://github.com/PERSIMUNE/PAC2023Faitova_Human_Mice_CLL

**Acknowledgments**

The authors would like to express sincere thanks to the patients providing samples for the study and making this kind of research possible. This research would not have been possible without our collaborator, PERSIMUNE Centre of Excellence, which provided the infrastructure as well as financial support and exceptional expertise. Special thank you to Mette Jørgensen for providing access and training for an in-house taxonomical profiling pipeline. The staff at the Hematology Department and at the PERSIMUNE biobank at Rigshospitalet were essential for this study as they organized and collected the samples for our research. Computerome team provided computational platform and technical support. The authors would like to thank Dr. Emma Philips for editing the manuscript for English grammar and punctuation.

**Funding**

The study was funded by the Danish Cancer Society (grant R269-A15924) and the Alfred Benzon Foundation. The infrastructure for sampling and analyses was within the Danish National Research Foundation funded PERSIMUNE project (grant 126).

**Competing Interests**

CUN received research funding and/or consultancy fees from Abbvie, Janssen, AstraZeneca, Genmab, Beigene, CSL Behring, Octapharma, Eli & Lily and Takeda. MS received research funding from Bayer AG. CdCB received consultancy fees / advisory board fees / honoraria fees from Abbvie, AstraZeneca, Beigene, Janssen and Octapharma

# Abstract

The gut microbiota play a critical role in maintaining a healthy human body and their dysregulation is associated with various diseases. In this study, we investigated the influence of the gut microbiome diversity on chronic lymphocytic leukemia (CLL) development. Stool sample analysis of 59 CLL patients revealed individual and heterogeneous microbiome compositions, but allowed for grouping of patients according to their microbiome diversity. Interestingly, CLL patients with a lower microbiome diversity and an enrichment of bacteria linked to poor health suffered from a more advanced or aggressive form of CLL. In the Eμ-TCL1 mouse model of CLL, we observed a faster course of disease when mice were housed in high hygiene conditions. Shotgun DNA sequencing of fecal samples showed that this was associated with a lower microbiome diversity which was dominated by *Mucispirillum* and *Parabacteroides* genera in comparison to mice kept under lower hygiene conditions. In conclusion, we applied taxonomic microbiome analyses to demonstrate a link between the gut microbiome diversity and the clinical course of CLL in humans, as well as the development of CLL in mice. Our novel data serve as a basis for further investigations to decipher the pathological and mechanistic role of intestinal microbiota in CLL development.

# Introduction

The gut microbiome, an ecosystem formed by commensal, symbiotic, and pathogenic microorganisms colonizing the gastrointestinal tract, is recognized as an important, life-long partner of the host (1, 2). The significance of the gut microbiome in both health and disease is a rapidly growing field of research. Recently, the role and direction of the crosstalk between the gut microbiome and immune cells, and its impact on treatment and disease development, has come into focus.

Homeostasis in the host microbiome is constantly influenced by factors such as diet, medication and stress levels (3), and the reciprocal interactions between gut microbiome and the immune system are being constantly challenged. Dysbiosis, an imbalance of the gut microbiota often associated with loss of beneficial microbes and blooms of pathogens, may lead to a disruption of the physical integrity of the intestinal barrier and/or function of the immune system. Dysbiosis can disrupt the development and distribution of immune cells, which may affect the immune response to pathogens and the ability to mount an appropriate immune defense (4).

Gut dysbiosis has also been associated with cancer susceptibility (5, 6). Within the area of hematological malignancies, the gut microbiome was suggested to play an important role in cancer microenvironment alterations and disease progression (7) as well as in treatment outcomes such as CAR-T cell and targeted therapy efficacies (8, 9).

Very little is known about the diversity of the gut microbiome and its interaction with the immune system in patients with chronic lymphocytic leukemia (CLL). In a previous study, we have demonstrated that the microbiome in CLL patients is less diverse when compared to healthy individuals, with half of the patients demonstrating severe dysbiosis caused by dominance of *Bacteroides* (10). We also showed that patients with CLL presented a lower abundance of bacterial species belonging to *Lachnospiraceae* and *Ruminococcaceae* families, including some of the main producers of short chain fatty acids (SCFA).

Given the antigen-driven nature and inherent immune dysfunction of CLL (11-13), we hypothesized based on our previous pilot study that the gut microbiome could contribute to CLL development. This might

happen by affecting the immune system through various mechanisms (14), including the production of cytokines triggered by certain bacterial species (15, 16). The gut microbiome might also be itself impacted by the immune dysfunction observed in CLL patients, and/or reflect the increased prescription of antimicrobials for this patient group (17). Here, we investigated a potential association between the fecal microbiome composition and CLL development in humans and mice. For this, we used a cohort of patients with CLL, as well as the immunocompetent Eμ-TCL1 transplantation mouse model kept under low or high hygiene conditions.

# Methods

## Patient cohort, data, sample collection, sequencing, and profiling

Fecal samples were collected from 61 patients enrolled in the CLL biobank and the PERSIMUNE biobank during regular out-patient visits at Rigshospitalet, Copenhagen, Denmark. The participants in our study are new patients, excluding the N=12 from our previous study (10), with each patient contributing one sample. One stool sample had to be excluded from further analysis due to low quality. Patient data and antibiotic records were retrieved from the Danish National CLL registry and from manual review of medical health records (18). The project was approved by the national ethics committee (approval no. 1804410) and written informed consent was obtained from all patients prior to sampling. As previously described (19), fecal samples were collected using OMNIgene.GUT (DNA Genotek) stabilization tubes according to the manufacturer's instructions. Briefly, samples were immediately fixated and subsequently frozen within 72 h. All samples were stored at −80°C until shipment for sequencing. Samples underwent shotgun metagenomic sequencing on the Illumina Hi-Seq platform. Reads preprocessing and taxonomical profiling were done using an in-house pipeline (Supplemental Methods).

## Animal models

6-8 week old female C57BL/6 J mice born and maintained in two different animal facilities of the German Cancer Research Center (DKFZ) were transferred to an experimental animal facility and subjected to adoptive transfer (AT) with leukemia cells from the TCL1 mouse model (TCL1 AT), kindly provided by Dr. Carlo Croce (The Ohio State University), as previously described (20-22). Leukemic development was

monitored through weekly blood withdrawals starting at week 2 post-TCL1 AT. All animal experiments were carried out according to governmental and institutional guidelines and authorized by the local authorities (permit numbers: DKFZ337, G-16/15).

## Murine samples collection and sequencing

Fecal samples were collected at week 0, one day before TCL1 AT, and at week 3 post-transplantation. Fecal samples were snap-frozen immediately after collection. The QIAmp DNA Stool Mini Kit (Qiagen) was used to extract DNA from fecal samples according to manufacturer instructions. Shotgun sequencing on the Illumina Hi-Seq platform was conducted at the European Molecular Biology Laboratory (EMBL, Heidelberg). Reads preprocessing and taxonomical profiling were done using an in-house pipeline (Supplemental Methods). Peripheral blood (PB) was drawn from the submandibular vein for weekly flow cytometric analysis (Supplemental Methods).

## Metabolic Potential Profiles

As previously described (23), gut metabolic modules (GMMs) were used to profile the functional potential of the bacterial community present in stool samples. In short, GMM profiling was performed by length normalizing the IGC count profiles and summing the values for each KEGG (24) gene ontology term, which were taken from IGC_catalog-v1.0.0.emapper.annotations-v2.tsv. The values were then 16sRNA normalized and turned into GMM profiles using omixer-RPM (25).

## Bioinformatics and Statistical analysis

Descriptive analyses were performed for both the human and mouse CLL cohorts with relative bacterial abundance as input data. Unless stated otherwise, Wilcoxon rank-sum testing was used to identify significant differences between subgroups and the Benjamini-Hochberg (BH) method was used for multiple-testing correction; a BH adjusted p-value of $<0.05$ was considered significant. PERMANOVA was used to test variations in the microbial composition among groups of patient samples (such as gender or antibiotic usage) or different groups of murine samples (such as cage effect). Alpha diversity measures (richness, Shannon index) were calculated at species level using the *vegan* R package (26).

The inter-individual dissimilarities in human gut microbiota composition (β-diversity) were assessed by calculating a dissimilarity matrix. Hierarchical clustering was applied on the distance matrix, dissimilarities were explored using Principal Coordinate Analysis (PCoA, *gl.pcoa*) (27), and the first 3 components of PCoA were visualized using 3D plot. Differential abundance of bacterial species in the fecal microbiome between CLL patient clusters was assessed using R implementation of SIAMCAT (28). Further details on bioinformatics analyses are provided in Supplemental Methods.

A generalized linear model (GLM) was used to visualize the relationship between two independent binary variables - TCL1 AT and hygiene of the animal facilities, and one dependent variable - bacterial relative abundance.

# Results

## Microbial composition and diversity are heterogeneous in patients with CLL

Fifty-nine patients diagnosed with CLL and one patient diagnosed with T-cell prolymphocytic leukemia (T-PLL, Pt ID=16) delivered stool samples between June 2017 and July 2020. Forty-four stool samples were collected prior to any treatment; 16 stool samples were collected from patients who received treatment before microbiome sampling. Characteristics of patients are provided in Table 1.

We evaluated the heterogeneity of microbiome composition in our patient cohort by β-diversity estimates, clustering, and statistical testing (PERMANOVA). We observed the microbial composition to be heterogeneous within the CLL patient cohort. Unsupervised hierarchical clustering revealed 3 distinct clusters based on the microbial composition of all patient samples (Figure 1A). The separation is visualized in a 3D plot using the results of a PCoA (Figure 1B). The dissimilarity between microbial communities estimated by β-diversity was evaluated and significant differences were observed between the 3 clusters ($R^2$= 0.09, $P^2$=0.001). The differences remained significant after adjusting for gender, age at sampling, and BMI ($R^2$= 0.086, $P^2$=0.001). No significant differences in the microbiome structure were observed in terms of age at microbiome sampling alone (R = 0.017, P=0.444), nor after adjustment for gender and BMI (R = 0.017, P=0.409). To evaluate differences in the microbial communities between the clusters, diversity was

assessed. The Shannon diversity for cluster 1 (C1) was lower than for clusters 2 and 3 (C2 and C3) (C1 vs C2: median, 1.80 vs. 3.50, P = 2.4e−05; C1 vs C3: median, 1.80 vs. 2.80, P = 8.1e-05; Figure 1C).

Besides the observed differences in diversity between clusters, a high variability of the microbiome composition was observed within and between clusters. At the genus level, *Bacteroides* was the most abundant genus across the 60 samples. Additionally, there was a trend of *Bacteroides* acquiring bacterial dominance (>30 % relative abundance) in 7 out of 12 samples in C1, and *Prevotella* dominating the composition of 9 samples in C2 and C3 while being completely depleted in all samples in C1. A detailed visualization of the microbiome composition of all 60 patients at the genus level can be explored through an interactive web application at: https://terezafait.shinyapps.io/microbiome_composition/. Instructions and examples of how to navigate in the app can be found in Supplemental Methods. An overview of the bacterial classification into six major taxonomic levels is provided as Supplemental Table 1.

## CLL is associated with low microbiome diversity

Given that dysbiosis, often interpreted as loss of diversity, has been documented to play a role in the development and progression of hematological diseases (29), we explored the individual course of the disease for all patients and focused on comparing patients from C1 and C2, representing those with lowest and highest gut microbiome diversity. In contrast to C2 patients, patients in C1 exhibited a more advanced and progressive CLL. This was evidenced by an extended duration from CLL diagnosis to microbiome sampling (C1: median 5.3 years, IQR: 0.3-9.2 y, C2: median 0.3y, IQR: 0.3-9.2 y; p-value: 0.47), a higher proportion of patients who needed treatment for CLL before and/or after microbiome sampling (C1: 92%; C2: 50%; p-value based on Kaplan-Meier analysis from diagnosis to first line treatment: 0.21), and a higher occurrence of patients who underwent hematopoietic stem cell transplantation (HSCT) or developed Richter's transformation (RT) (1x RT, 1x HSCT, 1x RT+HSCT) as illustrated in Figures 2 and 3. Patients in C3 (intermediate microbiome diversity) demonstrated a greater similarity to C2 than to C1 in terms of time from CLL diagnosis to microbiome sampling (median: 0.4y, IQR: 0.2-5.9y) and proportion of patients in need of CLL treatment (47%). We further monitored the most recent antimicrobial prescriptions for all patients, finding that 58% of C1 patients 25% of C2 patients, and 29% of C3 patients received antimicrobial

treatment within 6 months prior to microbiome sampling (Figure 3, Supplemental Table 2). In regard to IGHV mutational status, all CLL patients, irrespective of diversity cluster, exhibited a comparable percentage of mutated CLL (C1: 33%, C2: 37.5%, C3: 32%).

## Differential abundance of bacterial taxa illustrates heterogeneity among CLL patients

Having observed a correlation between the clinical course of CLL and the gut microbiome composition, we further aimed to identify groups of bacterial taxa differing significantly between the patient clusters. . In total, the abundance of 30 bacterial genera was significantly different between C1 and C2 (log2 fold change >1) as determined by SIAMCAT (Supplemental Table 3). Of these, *Hungatella*, *Anaerotruncus*, *Dialister*, *Erysipelatoclostridium/Clostridiales*, *Lachnoclostridium* and *Flavonifractor* were more abundant among C1 (low diversity) patients, while *Parabacteroides*, *Barnesiella*, *Odoribacter* and *Bilophila,* amongst others, were noticeably enriched among C2 (high diversity) patients (Figure 4). Interestingly, patient no. 2 from the C2 cluster was clinically similar to patients in C1 (diagnosis 3 years prior to microbiome sampling and antimicrobial treatment prior to sampling), and also showed a similar microbiome composition as C1 patients. Along the same line, patient no. 4 from C1, clinically similar to patients in C2 with regard to time from diagnosis to microbiome sampling (<1 month), showed an enrichment of bacteria with higher abundance in C2 patients. Bacterial genera differentially abundant between C1 and C3 partly overlapped with those identified as differentially abundant between C1 and C2 (Supplemental Figure 4), whereas no taxa were identified to be significantly different between C2 and C3 (Supplemental Table 3).

At the species level, 110 bacterial species were identified to be differentially abundant between C1 and C2 patient samples (Supplemental Table 3). All detected bacterial species were enriched in C2 and depleted in C1 (Supplemental Figure 5), which is likely due to strong differences in abundance of individual species and not groups of species as demonstrated at the genus level above. Intriguingly, bacteria such as *Prevotella copri, Dorea longicatena,* or *Bifidobacterium adolescentis* that belong to a healthy microbiome signature (30) were enriched among C2 (high diversity) patients.

## Clustering of patients does not reflect shared metabolic functions

It has been shown that different bacterial species can have similar metabolic function (31). Hence, an assessment of functional bacterial groups might be more informative than the bacterial composition itself. Thus, we used the bacterial genes identified in a stool sample that were annotated to metabolic functions by omixer-RPM (25) as estimates of a potential function of the bacterial community, i.e. gut metabolic modules (GMMs). The most abundant GMM in samples from CLL patients was lactose degradation (Table 2). Despite the differences in diversity and composition of the microbiome, no clear pattern in GMM between C1 (low diversity) and C2 (high diversity) could be detected. The potentially clinically relevant GMMs related to production of overall beneficial SCFAs (32), and a variety of indole-derivatives promoting fortification of the gut epithelial barrier (33) were detected in many samples, but not different between patient clusters (Table 2). A detailed visualization of GMMs grouped according to Vieira-Silva, et al. (34) (Supplemental Table 4) in all 60 patient samples can be explored through an interactive web application accessible via this link: https://terezafait.shinyapps.io/gmm_modules/.

## Hygiene level influences CLL progression in mice

Adoptive transfer of Eμ-TCL1 leukemia (TCL1 AT) in C57BL/6 mice housed in two animal facilities at the German Cancer Research Center with a different hygiene status revealed differences in the development of CLL. In order to elucidate whether the gut microbiome is causally contributing to this observation, we performed TCL1 AT with C57BL/6 mice that were born and kept in either, a closed breeding facility with altered Schaedler flora (high hygiene, HH) (35) or an experimental barrier with individually-ventilated cages (low hygiene, LH). One day before TCL1 AT, mice were brought into a common experimental facility with LH conditions and kept there for the rest of the experiment (Figure 5A). Mice originating from the HH facility developed CLL more rapidly compared to LH mice, reaching higher percentages and absolute numbers of CLL cells in the blood over time (week 2: LH = 398.5 cells/μL vs HH = 1,390 cells/μL, p-value = 0.0037; and week 4: LH = 17,489 cells/μL vs HH = 31,918 cells/μL, p-value = 0.0274; Figure 5B, C). To assess potential differences in the immune system at a similar stage of leukemia development, HH mice were euthanized at

4 weeks post-TCL1 AT, and mice from the LH group were euthanized 5 days later (Figure 5A). At these time points, similar tumor burden in the spleen was achieved in both groups (Supplemental Figure 1A). Immunophenotyping of splenic immune cell populations, specifically of the T-cell compartment, revealed no differences between the two groups (Supplemental Figure 1,2).

## Hygiene level and microbiome diversity in mice are inversely proportional

We hypothesized that mice kept in the HH environment would develop a less diverse microbiome than mice kept in the LH facility (36). To confirm this in the setup described above, we analyzed the gut microbiome by shotgun DNA sequencing of fecal samples. One day before TCL1 AT (TP1), corresponding to the untouched microbiome status of the mice maintained in the two different facilities, lower diversity was seen for the HH mice as shown by richness and diversity index (Figure 6A). Three weeks after TCL1 AT (TP2) and co-housing of mice in the same LH facility, which was necessary to allow for experimental interventions to the mice, the microbiome diversity of the two groups became more similar, with a massive increase in diversity for the HH group, and only a minor change towards higher diversity in the LH group (Figure 6A).

Focusing on TP1 as the baseline condition for CLL onset in the mice, we explored differences of the gut microbiome between the two groups (Figure 6B). Major differences were observed between the composition of HH and LH microbiomes, with *Mucispirillum* and *Muribaculaceae* dominating (i.e., constituting more than 30 % of sequencing reads) the microbiomes of HH and LH, respectively. We further explored relative abundances of specific bacterial species in relation to two predictor variables, high hygiene condition and TCL1 AT, by running a generalized linear model (GLM). Each bacterial species is represented as a dot weighted according to the predictor variable values (Figure 6C). Beta (β) values, coefficients obtained from GLM analysis represent the potential influence of high hygiene condition and TCL1 AT on the relative abundance of each bacterial species. For instance, we observed *Clostridium* to have a β-value of 1.9 on the x-axis (representing high hygiene condition as correlated with high abundance) and a β-value close to zero on the y-axis (representing no clear correlation with TCL1 AT).

In summary, the results of our study clearly link, for the first time, clinical course of CLL in patients and development of CLL mouse models with the diversity of the gut microbiome, where higher microbiome

diversity is associated with slower disease progression. Our in-depth characterization of bacterial species in groups of patients with a difference in outcome provides relevant data to study the role and pathological function of these microorganisms, with implications for stratification and therapy of CLL patients.

## Discussion

There is increasing evidence for an important role of gut microbiota in human physiology, arguing for a critical role of the microbiome to maintain a healthy state (37). A recent update of Hanahan's hallmarks of cancer has included the microbiota as an important player in carcinogenesis (38). This is based on results from several studies showing that the microbiome contributes to the development of several cancer entities, such as colorectal, gastric, or biliary cancer, and studying the underlying mechanisms will help to develop novel therapies (39-41).

In this study, lower diversity of the gut microbiome was linked to more aggressive and/or more progressive disease development in patients with CLL and TCL1 AT mice. The study of human stool samples showed that upon unsupervised clustering of patients with CLL based on gut bacterial distribution, a group of patients with lower microbiome diversity showed more severe clinical course. The severe disease course was characterized by longer time from diagnosis to microbiome sampling signaling more advanced disease, higher frequency of CLL treatment and disease progressions implying more aggressive disease, as well as increased antimicrobial usage either implying pre-existing immune system impairment or being a cause of the identified microbiota disruption. By applying TCL1 AT in immunocompetent C57BL/6 mice with basal differences in microbiome diversity, co-housed during the development of CLL, we provide evidence for a causal link between a lower gut microbiome diversity at onset of disease and a faster development of CLL. Our novel data identify the microbiome as a driver of disease progression and therefore, as a potential target to impact the course of CLL development.

In line with our findings suggesting more advanced or more aggressive CLL correlating with a less diverse, dysbiotic microbiome, a study of B-cell lymphomas demonstrated an association between gut microbiome composition and disease severity, where patients with indolent lymphomas presented higher microbiome

diversity and enrichment of certain bacterial genera when compared to diffuse large B-cell lymphoma patients (42). Similarly, the majority of patients with low diversity were either treated for CLL prior to or 1 month after microbiome sampling, illustrating the link between a dysbiotic microbiome and more advanced CLL. Several studies, mainly focusing on chemotherapy regimens, reported changes in the gut microbiota after treatment, some of which persist for years, which could also be part of the mechanism for dysbiosis in patients having received CLL treatment prior to collection of the microbiome sample (43-45). However, further studies and randomized clinical trials are needed to elucidate the influence of combination and targeted therapies on CLL microbiomes.

As signs of dysbiosis among CLL patients we documented the loss of diversity (Shannon diversity < 2.0) as well as blooms of bacteria associated with poor health. Several clinically important bacterial taxa enriched in low diversity patients (C1) included *Flavonifractor*, *Anaerotruncus* and *Dialister* genera, whose members were among the top 40 microbial species associated with disease by Gacesa et al. (30). *Flavonifractor plautii* was recently shown to be associated with young-onset colorectal cancer (46), and together with *Anaerotruncus colihominis* was strongly associated with disease and smoking (30); *Dialister invisus* was a common bacterium in individuals with poor diet habits (30, 47). Patients with higher diversity (C2) showed significant abundance of bacterial species such as *Prevotella copri, Dorea longicatena* or *Bifidobacterium adolescentis,* which are known to produce SCFAs through fermentation of dietary fibers (32, 48), also overlapping with the pattern of healthy-like microbiome described by Gacesa et al (30). Thus, we speculated that the presented bacterial composition in patients with higher microbiome diversity might lead to beneficial outcomes during their course of CLL.

While results regarding the function of gut bacteria are sometimes contradictory and isolated effects of specific bacteria are difficult to prove due to complex interactions (49), we hypothesized that exploring taxonomic distributions reflected in the gut metabolic potential profiles, representing metabolic function of individual gut microbiomes (25), might be more informative. A study using gut metabolic potential profiles in patients receiving HSCT revealed that the conditioning regimen is associated with the degree of changes in metabolic potential of their gut microbiomes (23). We focused on production of SCFAs such as

propionate, and butyrate, the crucial gut microbiome metabolites with known ability to exhibit immunomodulatory effects, and on starch degradation metabolism, which shows largely consistent health-promoting effects. However, the abundances of metabolic pathways directly involved in the production of SCFA and other compounds did not show substantial differences between patient subgroups. It may be that the gut metabolic potential profiles will still reveal differences between patients with CLL as compared to patients with other diseases and healthy volunteers. We are currently undertaking such studies to extend on our previous exploration of a CLL gut microbiome signature (10).

Studying the impact of the gut microbiome on cancer development in mouse models remains a challenging but essential task. Most studies of tumor mouse models are performed in facilities with various and often unknown hygiene levels and microbiome status. In our study, we used mice that were born and maintained in either a high or low hygiene facility which assured that the two groups were distinct in terms of their gut microbiome. The clear difference in microbiome diversity that we observed in these mice impacted CLL development. This approach comes, however, with the limitation of not using littermates in our study and the risk of a slightly different genetic background in the two groups. To overcome this limitation, future experiments should include animals from germ-free facilities that fully block the exposure of mice to any microorganisms (50). Exposure of these mice to defined gut bacteria of interest will help to clarify their impact on tumor development. A crucial and unequivocal takeaway from the findings in this study is that when conducting tumor development studies, it is imperative to use animals that are co-housed and possess identical microbial compositions.

Among the bacteria that were detected in the LH but not HH group, *Muribaculaceae* has been described as an immune-protective bacterial family in a CT26 melanoma mouse model (51). *Helicobacter*, also highly present in LH mice, is widely known for its correlation with the occurrence of gastric cancer. Its metabolites are known to drive macrophages into an anti-inflammatory state (52, 53). In our study, the presence of these bacteria in the LH mice could however have beneficial effects, perhaps by shaping myeloid cells into a phenotype that is less supportive of CLL growth. Lastly, *Bacteroidales*, also upregulated in this group and considered to be beneficial for gut health, but also correlated with worse disease outcome in lung cancer

patients (54, 55), could be priming the gut-associated immune system in the LH mice and contributing to its immune-protective action against leukemia development.

In HH mice, we detected an enrichment of *Mucispirillum*, which has been described as cancer-promoting due to its induction of lipopolysaccharide production, which enhances inflammation (56). *Parabacteroides*, also highly enriched among HH mice, is a bacterial genus generally considered as anti-inflammatory (57). Such species present in the HH mice could contribute to enhancing immune suppression and thereby promote CLL development. Importantly, linking the presence of specific bacteria in the gut of these mice with specific functions according to the literature is not straightforward, partly because of the multiple effects that bacteria can have in different settings.

Our study is limited by only assessing the microbiome on DNA level, whereas a more precise way would be the inclusion of metatranscriptomics and/or metabolomics focused on the microbiome. Other limitations of our study are the lack of consecutive samples per patient, which would allow us to describe the microbiome changes during disease progression more precisely, and small numbers in the extreme clusters (C1 and C2), which most likely prevent us from achieving statistical significance. A study set-up where stool samples are collected before and after treatment initiation has been applied as translational studies adjoined to several clinical trials (NCT04008706; NCT04639362; NCT04608318); thus, elucidation of microbiome dynamics throughout treatment will be the focus of upcoming studies. The observed association between microbiome low diversity and advanced CLL may be influenced by sampling bias with potential overrepresentation of patients starting treatment in planned clinical trials. Additionally, antibiotic exposures accompanying CLL treatment could confound our findings, influencing microbial composition and diversity. Given the purpose of this study, while a descriptive overview of the patient and mouse data itself is insightful, a translational and functional comparison of our findings in humans and mice would be ideal. However, exploration of the microbial overlap between human and mouse showed that 85% of bacterial genera found in the murine microbiome are not present in humans (58). These impressive differences might be caused by the obvious dissimilarity between the mouse and human systems, as well as

by external factors (59). Therefore, translating conclusions from murine to human data remains challenging.

In conclusion, taxonomic analyses of gut microbiota provide evidence for a link between microbiome diversity and CLL aggressiveness and development in patients with CLL and mouse models, respectively. In the patient study, we grapple with a classic chicken-and-egg dilemma as it remains unclear whether the microbiome dysbiosis is a result of the CLL, its treatment, and antibiotic use, or whether it represents an underlying condition driving the disease's development. However, in the mouse study, we provide evidence through the TCL 1 AT which elucidates that the microbiome alterations are not just a consequence but indeed play a significant role in the progression of the disease. Further, we provide a complete overview of the taxonomical and functional composition identified in patient samples. Lastly, we attempt to apply metabolic potential analysis to provide a superior understanding of the biological processes underlying gut dysbiosis in this patient cohort. However, profound taxonomical changes were not reflected in changes in the gut metabolic potential. It is thus appealing to further investigate whether the intestinal microbial composition and function could serve as a potential predictor for CLL development.

# References


1. Valdes AM, Walter J, Segal E, Spector TD. Role of the gut microbiota in nutrition and health. BMJ. 2018;361:k2179.
2. Gorkiewicz G, Moschen A. Gut microbiome: a new player in gastrointestinal disease. Virchows Arch. 2018;472(1):159-172.
3. Rodríguez JM, Murphy K, Stanton C, et al. The composition of the gut microbiota throughout life, with an emphasis on early life. Microb Ecol Health Dis. 2015;26:26050.
4. Yoo JY, Groer M, Dutra SV, Sarkar A, McSkimming DI. Gut Microbiota and Immune System Interactions. Microorganisms. 2020;8(10):1587.
5. Brüssow H. Problems with the concept of gut microbiota dysbiosis. Microb Biotechnol. 2020;13(2):423-434.
6. Bonnet M, Buc E, Sauvanet P, et al. Colonization of the human gut by E. coli and colorectal cancer risk. Clin Cancer Res. 2014;20(4):859-867.
7. Wang R, Yang X, Liu J, et al. Gut microbiota regulates acute myeloid leukaemia via alteration of intestinal barrier function mediated by butyrate. Nat Commun. 2022;13(1):2522.
8. Vicente-Dueñas C, Janssen S, Oldenburg M, et al. An intact gut microbiome protects genetically predisposed mice against leukemia. Blood. 2020;136(18):2003-2017.
9. Park JS, Gazzaniga FS, Wu M, et al. Targeting PD-L2–RGMb overcomes microbiome-related immunotherapy resistance. Nature. 2023;617(7960):377-385.
10. Faitová T, Svanberg R, da Cunha-Bang C, et al. The gut microbiome in patients with chronic lymphocytic leukemia. Haematologica. 2022;107(9):2238-2243.
11. Caligaris-Cappio F. Inflammation, the microenvironment and chronic lymphocytic leukemia. Haematologica. 2011;96(3):353-355.
12. Niemann CU, Wiestner A. B-cell receptor signaling as a driver of lymphoma development and evolution. Semin Cancer Biol. 2013;23(6):410-421.
13. Svanberg R, Janum S, Patten PEM, Ramsay AG, Niemann CU. Targeting the tumor microenvironment in chronic lymphocytic leukemia. Haematologica. 2021;106(9):2312-2324.
14. Aghamajidi A, Maleki Vareki S. The Effect of the Gut Microbiota on Systemic and Anti-Tumor Immunity and Response to Systemic Therapy against Cancer. Cancers (Basel). 2022;14(15):3563.
15. Schulthess J, Pandey S, Capitani M, et al. The Short Chain Fatty Acid Butyrate Imprints an Antimicrobial Program in Macrophages. Immunity. 2019;50(2):432-445.
16. Modoux M, Rolhion N, Lefevre JH, et al. Butyrate acts through HDAC inhibition to enhance aryl hydrocarbon receptor activation by gut microbiota-derived ligands. Gut Microbes. 2022;14(1):2105637.
17. Andersen MA, Rostgaard K, Niemann CU, Hjalgrim H. Antimicrobial use before chronic lymphocytic leukemia: a retrospective cohort study. Leukemia. 2021;35(3):747-751.
18. da Cunha-Bang C, Geisler CH, Enggaard L, et al. The Danish National Chronic Lymphocytic Leukemia Registry. Clin Epidemiol. 2016;8:561-565.
19. Ilett EE, Jørgensen M, Noguera-Julian M, et al. Associations of the gut microbiome and clinical factors with acute GVHD in allogeneic HSCT recipients. Blood Adv. 2020;4(22):5797-5809.
20. Hanna BS, McClanahan F, Yazdanparast H, et al. Depletion of CLL-associated patrolling monocytes and macrophages controls disease development and repairs immune dysfunction in vivo. Leukemia. 2016;30(3):570-579.
21. Bichi R, Shinton SA, Martin ES, et al. Human chronic lymphocytic leukemia modeled in mouse by targeted TCL1 expression. Proc Natl Acad Sci U S A. 2002;99(10):6955-6960.
22. McClanahan F, Hanna B, Miller S, et al. PD-L1 checkpoint blockade prevents immune dysfunction and leukemia development in a mouse model of chronic lymphocytic leukemia. Blood. 2015;126(2):203-211.
23. Jørgensen M, Nørgaard JC, Ilett EE, et al. Metabolic Potential of the Gut Microbiome Is Significantly Impacted by Conditioning Regimen in Allogeneic Hematopoietic Stem Cell Transplantation Recipients. Int J Mol Sci. 2022; 23(19):11115.
24. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27-30.

25. Darzi Y, Falony G, Vieira-Silva S, Raes J. Towards biome-specific analysis of meta-omics data. ISME J. 2016;10(5):1025-1028.
26. Oksanen J, Simpson GL, Blanchet FG, et al. vegan: Community Ecology Package. R package version 2.6-2. Retrieved from: https://github.com/vegandevs/vegan. Accessed on: 23.09.2022.
27. Gruber B, Unmack PJ, Berry OF, Georges A. dartr: An r package to facilitate analysis of SNP data generated from reduced representation genome sequencing. Mol Ecol Resour. 2018;18(3):691-699.
28. Wirbel J, Zych K, Essex M, et al. Microbiome meta-analysis and cross-disease comparison enabled by the SIAMCAT machine learning toolbox. Genome Biol. 2021;22(1):93.
29. Uribe-Herranz M, Klein-González N, Rodríguez-Lobato LG, Juan M, de Larrea CF. Gut Microbiota Influence in Hematological Malignancies: From Genesis to Cure. Int J Mol Sci. 2021; 22(3):1026.
30. Gacesa R, Kurilshikov A, Vich Vila A, et al. Environmental factors shaping the gut microbiome in a Dutch population. Nature. 2022;604(7907):732-739.
31. Deleu S, Machiels K, Raes J, Verbeke K, Vermeire S. Short chain fatty acids and its producing organisms: An overlooked therapy for IBD? EBioMedicine. 2021;66:103293.
32. Fusco W, Lorenzo MB, Cintoni M, et al. Short-Chain Fatty-Acid-Producing Bacteria: Key Components of the Human Gut Microbiota. Nutrients. 2023;15(9):2211.
33. Roager HM, Licht TR. Microbial tryptophan catabolites in health and disease. Nat Commun. 2018;9(1):3294.
34. Vieira-Silva S, Falony G, Darzi Y, et al. Species–function relationships shape ecological properties of the human gut microbiome. Nat Microbiol. 2016;1(8):16088.
35. Dewhirst FE, Chien CC, Paster BJ, et al. Phylogeny of the defined murine microbiota: altered Schaedler flora. Appl Environ Microbiol. 1999;65(8):3287-3292.
36. Ege MJ. The Hygiene Hypothesis in the Age of the Microbiome. Ann Am Thorac Soc. 2017;14(Supplement_5):S348-S353.
37. Mohajeri MH, Brummer RJM, Rastall RA, et al. The role of the microbiome for human health: from basic science to clinical applications. Eur J Nutr. 2018;57(Suppl 1):1-14.
38. Hanahan D. Hallmarks of Cancer: New Dimensions. Cancer Discov. 2022;12(1):31-46.
39. Saus E, Iraola-Guzmán S, Willis JR, Brunet-Vega A, Gabaldón T. Microbiome and colorectal cancer: Roles in carcinogenesis and clinical potential. Mol Aspects Med. 2019;69:93-106.
40. Liu Y, Baba Y, Ishimoto T, et al. Gut microbiome in gastrointestinal cancer: a friend or foe? Int J Biol Sci. 2022;18(10):4101-4117.
41. Wheatley RC, Kilgour E, Jacobs T, et al. Potential influence of the microbiome environment in patients with biliary tract cancer and implications for therapy. Br J Cancer. 2022;126(5):693-705.
42. Diefenbach CS, Peters BA, Li H, et al. Microbial dysbiosis is associated with aggressive histology and adverse clinical outcome in B-cell non-Hodgkin lymphoma. Blood Adv. 2021;5(5):1194-1198.
43. Montassier E, Gastinne T, Vangay P, et al. Chemotherapy-driven dysbiosis in the intestinal microbiome. Aliment Pharmacol Ther. 2015;42(5):515-528.
44. Galloway-Peña JR, Shi Y, Peterson CB, et al. Gut Microbiome Signatures Are Predictive of Infectious Risk Following Induction Therapy for Acute Myeloid Leukemia. Clin Infect Dis. 2020;71(1):63-71.
45. Rajagopala SV, Singh H, Yu Y, et al. Persistent Gut Microbial Dysbiosis in Children with Acute Lymphoblastic Leukemia (ALL) During Chemotherapy. Microb Ecol. 2020;79(4):1034-1043.
46. Yang Y, Du L, Shi D, et al. Dysbiosis of human gut microbiome in young-onset colorectal cancer. Nat Commun. 2021;12(1):6757.
47. Vals-Delgado C, Alcala-Diaz JF, Molina-Abril H, et al. An altered microbiota pattern precedes Type 2 diabetes mellitus development: From the CORDIOPREV study. J Adv Res. 2021;35:99-108.
48. Louis P, Flint HJ. Formation of propionate and butyrate by the human colonic microbiota. Environ Microbiol. 2017;19(1):29-41.
49. Shetty SA, Kuipers B, Atashgahi S, Aalvink S, Smidt H, de Vos WM. Inter-species Metabolic Interactions in an In-vitro Minimal Human Gut Microbiome of Core Bacteria. NPJ Biofilms Microbiomes. 2022;8(1):21.
50. Rodriguez-Palacios A, Aladyshkina N, Ezeji JC, et al. 'Cyclical Bias' in Microbiome Research Revealed by A Portable Germ-Free Housing System Using Nested Isolation. Sci Rep. 2018;8(1):3801.

51. Tomasi M, Dalsass M, Beghini F, et al. Commensal Bifidobacterium Strains Enhance the Efficacy of Neo-Epitope Based Cancer Vaccines. Vaccines (Basel). 2021; 9(11):1356.
52. Danne C, Ryzhakov G, Martínez-López M, et al. A Large Polysaccharide Produced by Helicobacter hepaticus Induces an Anti-inflammatory Gene Signature in Macrophages. Cell Host Microbe. 2017;22(6):733-745.
53. Wang F, Meng W, Wang B, Qiao L. Helicobacter pylori-induced gastric inflammation and gastric cancer. Cancer Lett. 2014;345(2):196-202.
54. Peters BA, Pass HI, Burk RD, et al. The lung microbiome, peripheral gene expression, and recurrence-free survival after resection of stage II non-small cell lung cancer. Genome Med. 2022;14(1):121.
55. Giri S, Mangalam A. Chapter 34 - The Gut Microbiome and Metabolome in Multiple Sclerosis. In: Faintuch J, Faintuch S, editors. Microbiome and Metabolome in Diagnosis, Therapy, and other Strategic Applications: Academic Press; 2019. p. 333-340.
56. Daniel SG, Ball CL, Besselsen DG, Doetschman T, Hurwitz BL. Functional Changes in the Gut Microbiome Contribute to Transforming Growth Factor β-Deficient Colon Cancer. mSystems. 2017;2(5):e00065-17.
57. Koh GY, Kane A, Lee K, et al. Parabacteroides distasonis attenuates toll-like receptor 4 signaling and Akt activation and blocks colon tumor formation in high-fat diet-fed azoxymethane-treated mice. Int J Cancer. 2018;143(7):1797-1805.
58. Ley RE, Bäckhed F, Turnbaugh P, Lozupone CA, Knight RD, Gordon JI. Obesity alters gut microbial ecology. Proc Natl Acad Sci U S A. 2005;102(31):11070-11075.
59. Nguyen TL, Vieira-Silva S, Liston A, Raes J. How informative is the mouse for human gut microbiota research? Dis Model Mech. 2015;8(1):1-16.

# Tables

| Feature | | All (n=60) | Cluster 1 (n=12) | Cluster 2 (n=8) | Cluster 3 (n=40) |
|---|---|---|---|---|---|
| **Gender** | Female | 21 (35%) | 6 (50%) | 5 (62.5%) | 10 (25%) |
| | Male | 39 (65%) | 6 (50%) | 3 (37.5%) | 30 (75%) |
| **Median age** | at diagnosis (yr) | 63.7 | 64.4 | 67.5 | 63.6 |
| | at microbiome sampling (yr) | 68.6 | 69.3 | 69.6 | 66.4 |
| **BMI** | kg/m2 | 26.2 | 24.6 | 25.2 | 26.9 |
| **IGHV** | M-CLL | 38 (63%) | 8 (66.6%) | 5 (62.5%) | 25 (62.5%) |
| | U-CLL | 19 (32%) | 4 (33.3%) | 3 (37.5%) | 12 (30%) |
| | NA | 3 (5%) | 0 | 0 | 3 (7.5%) |
| **FISH** | Del17p | 4 (6.6%) | 0 | 1 (12.5%) | 3 (7.5%) |
| | Del11q | 5 (8.3%) | 2 (16.6%) | 0 | 3 (7.5%) |
| | Tri12 | 11 (18.3%) | 1 (8.3%) | 2 (25%) | 8 (20%) |
| | Normal | 14 (23.3%) | 2 (16.6%) | 2 (25%) | 10 (25%) |
| | Del13q | 24 (40%) | 6 (50%) | 3 (37.5%) | 15 (37.5%) |
| | NA | 2 (3.3%) | 1 (8.3%) | 0 | 1 (2.5%) |

**Table 1. Patient characteristics.**

N, number; yr, years; BMI, Body mass index (body mass divided by the square of the body height, units: $kg/m^2$); IGHV, Immunoglobulin heavy-chain variable region gene, U-CLL, CLL with unmutated IGHV; M-CLL, CLL with mutated IGHV.

Kruskal-Wallis H test of difference between the 3 clusters: Age: p = 0.74, Gender: p = 0.06, BMI: p = 0.45, IGHV status: p = 0.75, FISH status: p = 0.16.

| | GMM | Mean (IQR) | | | |
|---|---|---|---|---|---|
| | | All (N=60) | Cluster1 (N=12) | Cluster 2 (N=8) | Cluster 3 (N=40) |
| 7 most abundant GMMs | Lactose degradation | 68.5 (53.8-83.1) | 77.9 (66.2-99.8) | 64.0 (58.1-75.8) | 66.6 (51.8-78.0) |
| | Melibiose degradation | 18.5 (15.5-22.0) | 19.0 (15.1-25.2) | 18.7 (17.3-20.2) | 18.4(15.2-21.2) |
| | Mannose degradation | 17.8 (14.8-19.9) | 16.6 (12.2-19.9) | 16.1 (13.8-19.6) | 18.5 (15.6-20.2) |
| | Glycolysis (prep. phase) | 17.4 (14.9-19.9) | 15.8 (14.6-17.6) | 19.0 (16.2-21.2) | 17.6 (15.1-19.9) |
| | Arabinoxylan degradation | 17.4 (13.9-21.4) | 15.7 (13.0-19.7) | 17.7 (15.1-20.0) | 17.8 (13.9-22.2) |
| | Starch degradation | 15.7 (10.8-20.0) | 12.9 (9.0-16.0) | 17.7 (14.3-20.4) | 16.1 (11.3-20.1) |
| | Mucin degradation | 14.7 (9.5-18.3) | 18.2 (14.4-23.4) | 11.5 (9.4-14.3) | 14.2 (9.8-18.8) |
| 7 apriori selected GMMs | Propionate Production I | 0.15 (0-0.05) | 0.13 (0.03-0.13) | 0.06 (0-0.01) | 0.12 (0-0.05) |
| | Propionate Production II | 0.7 (0.4-0.8) | 1.2 (0.6-1.8) | 0.7 (0.6-0.9) | 0.6 (0.4-0.7) |
| | Butyrate Production I | 5.6 (4.5-6.9) | 5.5 (4.6-6.6) | 5.3 (3.3-6.8) | 5.7 (4.6-7.0) |
| | Butyrate Production II | 5.7 (5.0-6.4) | 5.2 (4.5-5.9) | 6.4 (5.4-7.0) | 5.7 (5.0-6.3) |
| | Acetyl-Coa to Acetate | 10.4 (9.4-11.6) | 9.5 (8.9-11.5) | 10.4 (8.9-11.1) | 10.6 (9.5-11.6) |
| | Tryptophan Degradation | 3.1 (1.8-4.0) | 3.3 (1.2-5.2) | 2.9 (1.4-4.4) | 3.1 (1.9-3.8) |
| | Tyrosine degradation I | 6.4 (5.8-7.2) | 6.8 (5.7-8.4) | 6.8 (6.4-7.3) | 6.2 (5.7-6.8) |
| GMM abundance = normalized proxy for the portion of bacteria in a sample that can perform a specific function | | | | | |

**Table 2. Abundance of the seven most abundant and seven a priori selected GMMs.**

GMM, gut metabolic modules. Values, mean GMM values across samples belonging to the indicated clusters. IQR, interquartile range.

# Figure legends

**Figure 1.** Assessment of microbiome (dis)similarity and diversity in CLL patients. The (dis)similarity was measured by distance matrix constructed using Robust Aitchison distance. A) Hierarchical clustering over bacterial taxa. Hierarchical clustering (*hclust* function in R) with Ward's minimum variance method was run on the distance matrix calculated based on robust Aitchison distances. The clustering approach used was purely data-driven and the number of resulting clusters was not specified in advance. Cutting a hierarchical clustering tree at the point of the largest distance ("cut"), resulted in three clusters (Clusters 1 (C1): purple, C2: blue, and C3: orange). The patient diagnosed with T-PLL showed average microbiome values and has been marked by a black circle. B) Principal Coordinate Analysis (PCoA) representation of the CLL cohort (dis)similarity. Each dot in the PCoA plot represents one sample. Samples ordinated closer to one another are more similar than those ordinated further away. C) Shannon alpha diversity in CLL samples grouped according to clusters from A). Alpha diversity measures include richness - representing observed number of genera, and Shannon index - representing evenness of species in a community. In the box plots, box edges represent the 25th and 75th percentiles, the center line shows the median and whiskers extend from the box edges to the most extreme data point. The p-values (adjusted for multiple testing with the Benjamini-Hochberg (BH)) obtained upon Wilcoxon rank-sum tests are indicated, values < 0.05 were considered significant. Ns, $P > 0.05$; *, $P < 0.05$; **, $P < 0.01$.

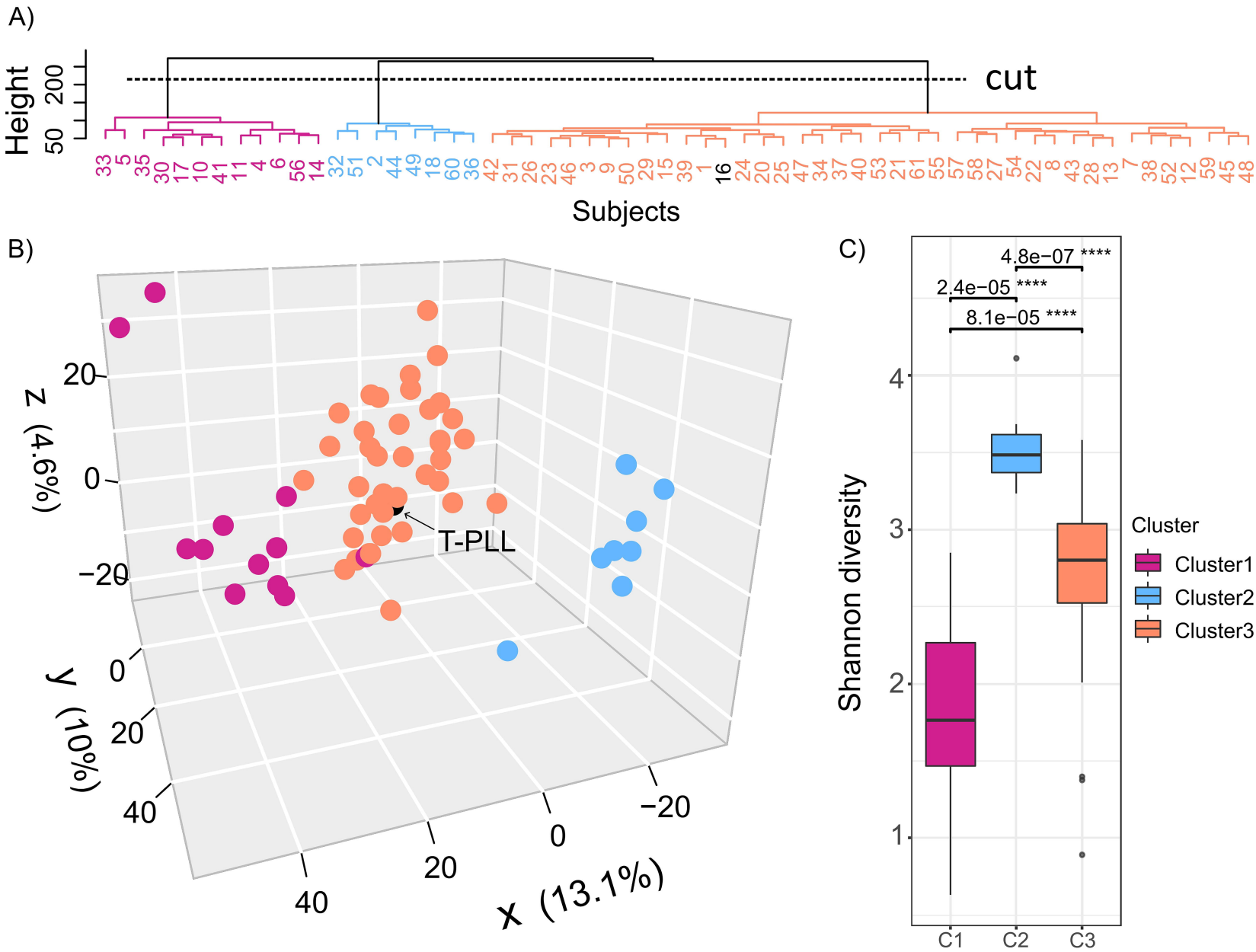
**Figure 2.** Swimmer plot illustrating the clinical course of disease of CLL patients. Patients from clusters C1 and C2 in Figure 1 are included in the swimmer plot. All included patients were alive by 15/09/2022, the end of the follow-up period. The time points of CLL diagnosis, CLL treatment, antimicrobial treatment, and microbiome sampling are shown in the swimmer plot. X-axis: Time before and after microbiome sampling, non-continuous time scale is marked by dashed lines. Y-axis: Subject numbers colored according to patient clusters (purple: C1, low diversity; blue: C2, high diversity). HSCT: hematopoietic stem cell transplantation; m: months; y: years.

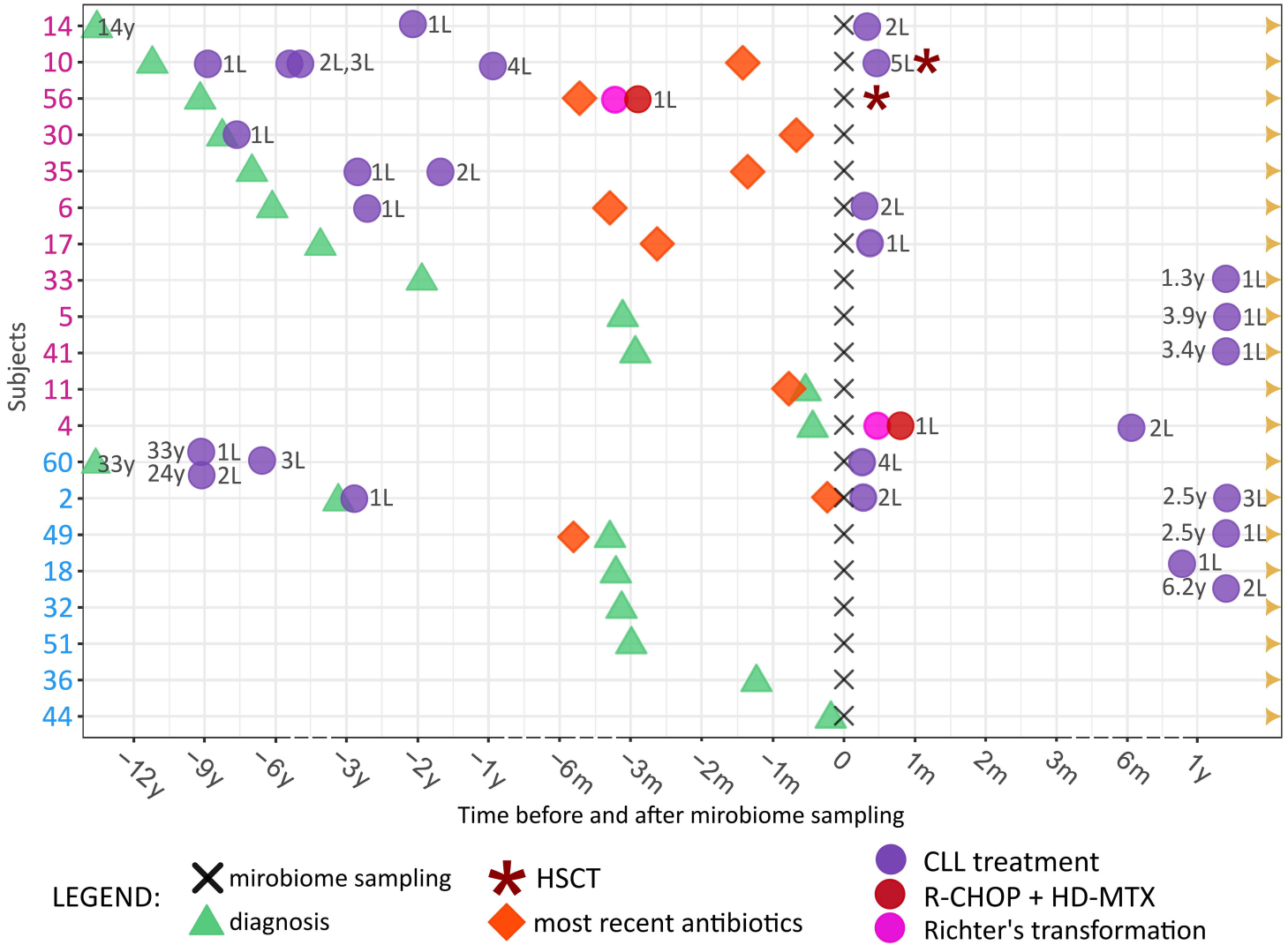
**Figure 3.** Heatmap representation of clinical outcomes over time for all patients. The color-coded cells in the heatmap depict different temporal intervals: time from diagnosis to sampling, time from diagnosis to initiation of first-line treatment (1L), time from 1L to sampling, time from sampling to 1L for patients without prior 1L treatment before microbiome sampling, and time from sampling to progression for patients who received 1L treatment before microbiome sampling. For instance, the time from diagnosis to 1L: in C1, 92% of patients required 1L within a median period of 3.7 years; in C2, 50% of patients required 1L within a median period of 0.8 years; in C3, 45% of patients required 1L within a median timeframe of 3.7 years. Annotation of each patient sample: cluster affiliation based on results from Figure 1, microbiome sample obtained before receiving 1L treatment (yes/no), IGHV mutation status (M-CLL/U-CLL), antibiotics treatment within the 6 months preceding microbiome sampling (yes/no), hematopoietic stem cell transplantation (HSCT), and/or Richters transformation. White color represents missing values. IGHV: Immunoglobulin heavy-chain variable region gene; U-CLL: CLL with unmutated IGHV; M-CLL: CLL with mutated IGHV; HSCT: Hematopoietic Stem Cell Transplantation; 1L: First-line treatment; AB: Antibiotics.

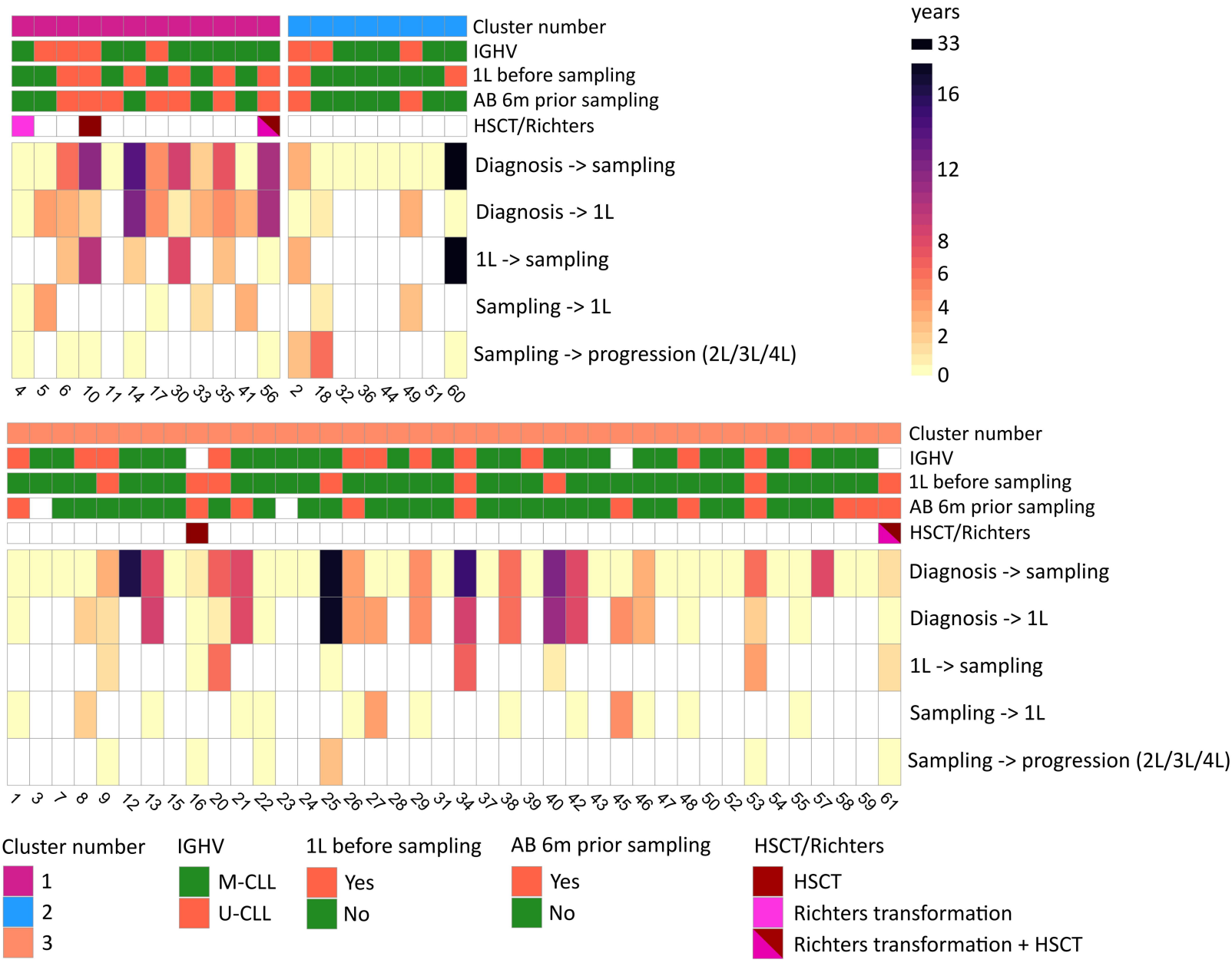
**Figure 4.** Heatmap of differential abundance of bacteria in C1 and C2 patient samples. Relative abundances of all genera with logFC >1 in C1 and twelve genera with highest logFC in C2 patient samples are visualized. Color scale: Centered log ratio transformed relative abundance of bacterial genera, scaled by columns in *pheatmap* function in R. C1: patient cluster 1; C2: patient cluster 2 according to Figure 1. Subject 5 (from C1) omitted from visualization due to extremely low relative abundances across all genera. Abbreviations: gen.i.s., genus incertae sedis; for instance, Ruminococcaceae gen.i.s.: reads could not be certainly classified as Ruminococcus (genus), but were classified as Ruminococcaceae (family).

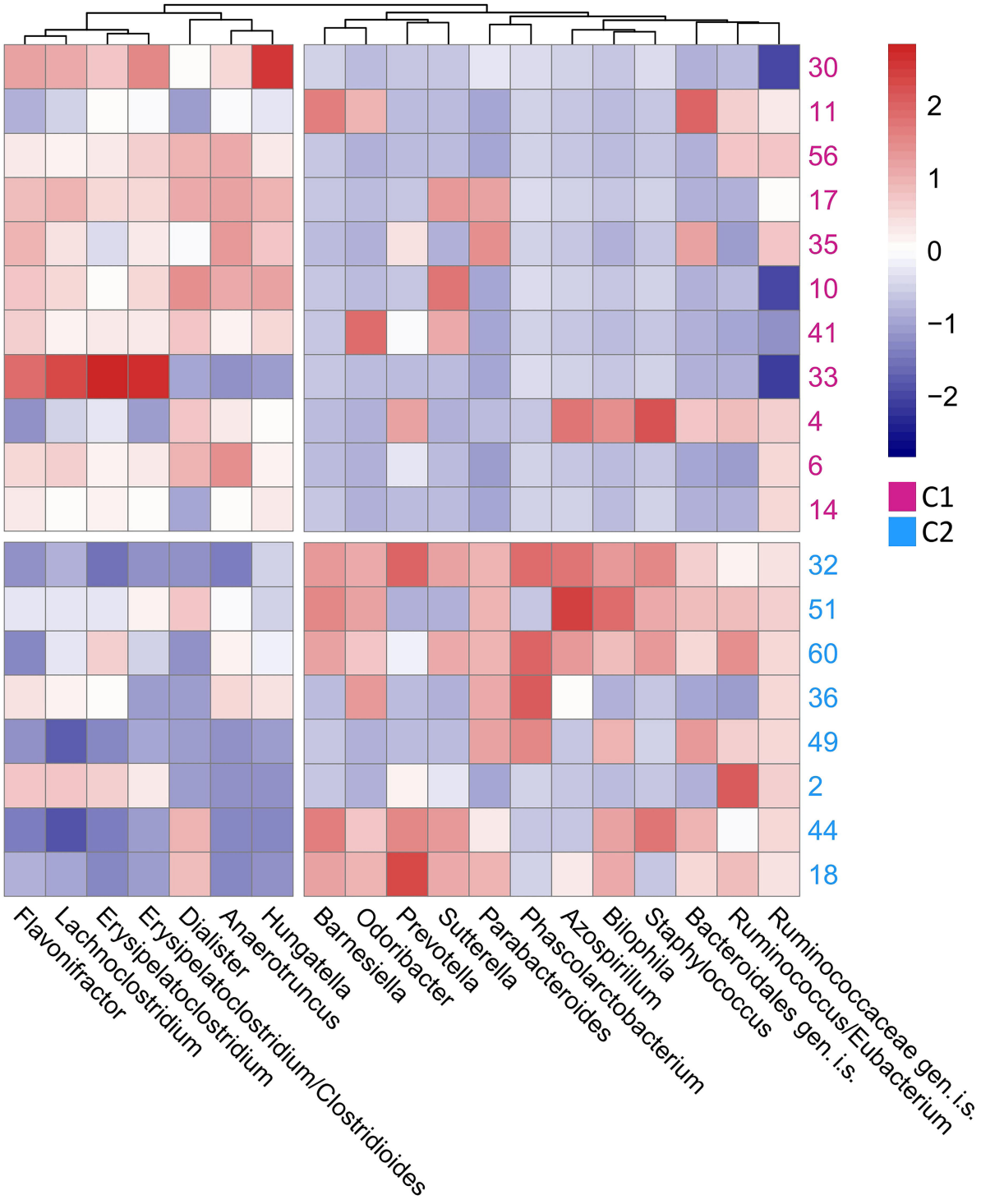
**Figure 5.** Leukemia development after TCL1 AT in mice derived from facilities with low (LH) or high hygiene (HH) levels. A) Schematics of experimental design of adoptive transfer of TCL1 leukemia (TCL1 AT) of LH mice (n = 9) and HH mice (n = 8). Mice without engraftment of TCL1 cells (n = 1 in LH group, n = 2 in HH groups) were removed from the study. Figure was created with BioRender.com B) Percentage of CLL cells in peripheral blood (PB) out of CD45+ viable cells 2, 3 and 4 weeks after TCL1 AT. C) Number of CLL cells per µL of PB at the same time points. Statistics: One independent study including 2 groups of 10 mice. Mann-Whitney non-parametric test for each time point (p-value: * < 0.05, ** < 0.01).

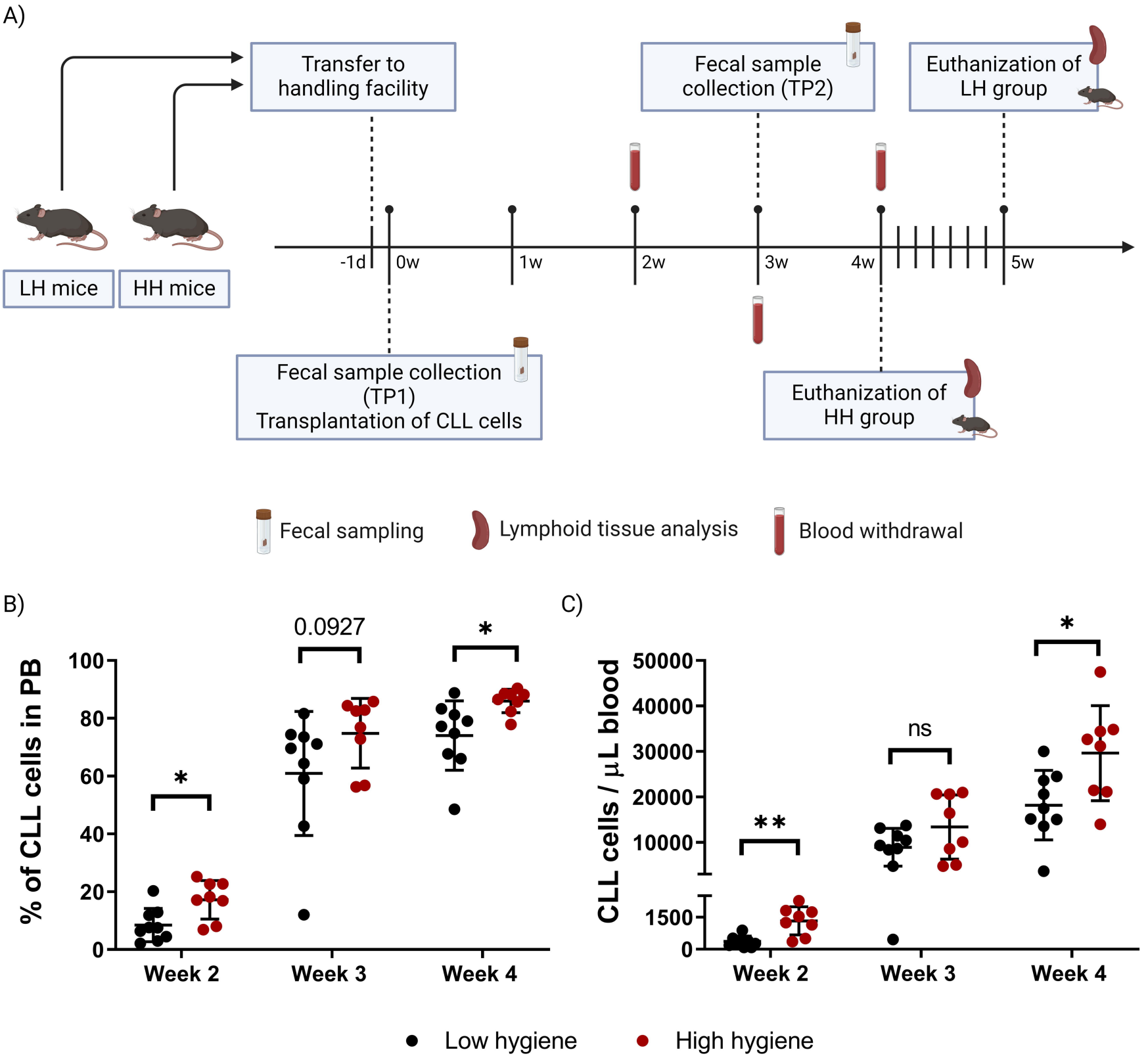
**Figure 6.** Microbiome composition and diversity in the TCL1 AT mouse model. A) Fecal alpha diversity in murine samples from the study described in Figure 5 collected at two different timepoints (before and after TCL1 AT; TP1 and TP2 respectively). Mice with low quality of their microbiome sequencing upon quality control were excluded from the analysis (n = 3 from LH group, n = 5 from HH group). Boxplots were constructed as described in Figure 1C. B) The relative abundance of bacterial genera in murine samples taken at TP1. Bacterial genera with abundance < 1% in a sample were omitted form plotting. Sequences that could not be assigned to a genus were grouped as *Unclassified*. C) Generalized linear model (GLM) for every bacterial genus to representing its abundance based on two predictor variables: hygiene and TCL1 AT. Center log ratio (clr)-transformed relative abundance data at genus level were used as input. The position of a point is given by coefficients (betas), where betas represent weights assigned to the predictor variables. In other words, each of the points illustrate to what degree is the bacterial genus relative abundance influenced by the two predictor variables. Positive (negative) value at the x-axis indicates that mice initially housed in a high hygiene barrier will have higher (lower) relative abundance of a bacterial genera compared to mice initially housed in low hygiene barrier. Positive (negative) value at the y-axis indicates that mice transplanted with CLL cells will have higher (lower) relative abundance of a bacterial genera than mice before transplantation of CLL cells. Concrete example: *Helicobacter* is positioned at coordinates x: -0.95 and y: -0.25, which can be interpreted as *Helicobacter*'s relative abundance is more influenced by the hygiene of the barrier than by CLL cells transplantation. Also, based on this model, relative abundance of *Helicobacter* will be lower in mice kept in high hygiene barrier and slightly lower in mice transplanted with CLL cells.

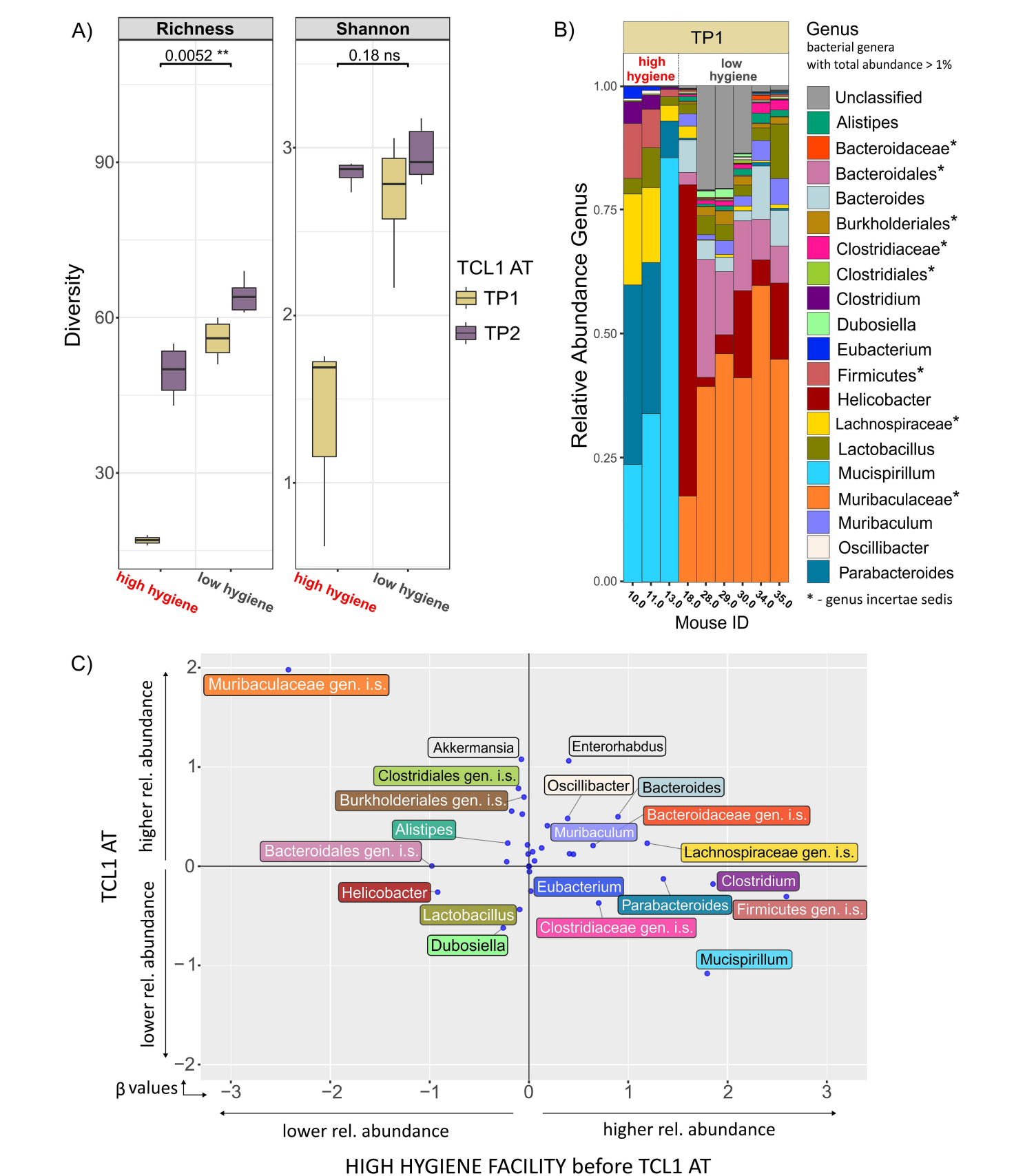

A)
Height
50
200
cut
33 5 35 30 17 10 41 11 4 6 56 14 32 51 2 44 49 18 60 36 42 31 26 23 46 3 9 50 29 15 39 1 16 24 20 25 47 34 37 40 53 21 61 55 57 58 27 54 22 8 43 28 13 7 38 52 12 59 45 48
Subjects
B)
z (4.6%)
20
0
−20
T-PLL
y (10%)
0
20
40
x (13.1%)
40
20
0
−20
C)
Shannon diversity
4
3
2
1
4.8e−07 ****
2.4e−05 ****
8.1e−05 ****
C1
C2
C3
Cluster
Cluster1
Cluster2
Cluster3


Subjects
14
10
56
30
35
6
17
33
5
41
11
4
60
2
49
18
32
51
36
44
14y
1L
2L,3L
4L
1L
1L
1L
2L
1L
2L
1L
1L
2L
1L
1.3y 1L
3.9y 1L
3.4y 1L
1L
2L
33y
33y 1L
24y 2L
3L
1L
4L
2L
2.5y 3L
2.5y 1L
1L
6.2y 2L
−12y
−9y
−6y
−3y
−2y
−1y
−6m
−3m
−2m
−1m
0
1m
2m
3m
6m
1y
Time before and after mirobiome sampling
LEGEND:
mirobiome sampling
diagnosis
HSCT
most recent antibiotics
CLL treatment
R-CHOP + HD-MTX
Richter's transformation


Cluster number
IGHV
1L before sampling
AB 6m prior sampling
HSCT/Richters
Diagnosis -> sampling
Diagnosis -> 1L
1L -> sampling
Sampling -> 1L
Sampling -> progression (2L/3L/4L)
4 5 6 10 11 14 17 30 33 35 41 56 2 18 32 36 44 49 51 60
years
33
16
12
8
6
4
2
0
Cluster number
IGHV
1L before sampling
AB 6m prior sampling
HSCT/Richters
Diagnosis -> sampling
Diagnosis -> 1L
1L -> sampling
Sampling -> 1L
Sampling -> progression (2L/3L/4L)
1 3 7 8 9 12 13 15 16 20 21 22 23 24 25 26 27 28 29 31 34 37 38 39 40 42 43 45 46 47 48 50 52 53 54 55 57 58 59 61
Cluster number
1
2
3
IGHV
M-CLL
U-CLL
1L before sampling
Yes
No
AB 6m prior sampling
Yes
No
HSCT/Richters
HSCT
Richters transformation
Richters transformation + HSCT


30
11
56
17
35
10
41
33
4
6
14
32
51
60
36
49
2
44
18
2
1
0
−1
−2
C1
C2
Flavonifractor
Lachnoclostridium
Erysipelatoclostridium
Erysipelatoclostridium/Clostridioides
Dialister
Anaerotruncus
Hungatella
Barnesiella
Odoribacter
Prevotella
Sutterella
Parabacteroides
Phascolarctobacterium
Azospirillum
Bilophila
Staphylococcus
Bacteroidales gen. i.s.
Ruminococcus/Eubacterium
Ruminococcaceae gen. i.s.


A)
Transfer to handling facility
Fecal sample collection (TP2)
Euthanization of LH group
LH mice
HH mice
-1d
0w
1w
2w
3w
4w
5w
Fecal sample collection (TP1) Transplantation of CLL cells
Euthanization of HH group
Fecal sampling
Lymphoid tissue analysis
Blood withdrawal
B)
% of CLL cells in PB
0.0927
*
*
Week 2
Week 3
Week 4
C)
CLL cells / µL blood
**
ns
*
Week 2
Week 3
Week 4
Low hygiene
High hygiene


A)
Richness
0.0052 **
Shannon
0.18 ns
Diversity
90
60
30
3
2
1
high hygiene
low hygiene
high hygiene
low hygiene
TCL1 AT
TP1
TP2
B)
TP1
high hygiene
low hygiene
Relative Abundance Genus
1.00
0.75
0.50
0.25
0.00
10.0
11.0
13.0
18.0
28.0
29.0
30.0
34.0
35.0
Mouse ID
Genus
bacterial genera
with total abundance > 1%
Unclassified
Alistipes
Bacteroidaceae*
Bacteroidales*
Bacteroides
Burkholderiales*
Clostridiaceae*
Clostridiales*
Clostridium
Dubosiella
Eubacterium
Firmicutes*
Helicobacter
Lachnospiraceae*
Lactobacillus
Mucispirillum
Muribaculaceae*
Muribaculum
Oscillibacter
Parabacteroides
* - genus incertae sedis
C)
TCL1 AT
higher rel. abundance
lower rel. abundance
2
1
0
−1
−2
Muribaculaceae gen. i.s.
Akkermansia
Enterorhabdus
Clostridiales gen. i.s.
Oscillibacter
Bacteroides
Burkholderiales gen. i.s.
Bacteroidaceae gen. i.s.
Alistipes
Muribaculum
Bacteroidales gen. i.s.
Lachnospiraceae gen. i.s.
Helicobacter
Eubacterium
Clostridium
Parabacteroides
Lactobacillus
Firmicutes gen. i.s.
Clostridiaceae gen. i.s.
Dubosiella
Mucispirillum
β values
−3
−2
−1
0
1
2
3
lower rel. abundance
higher rel. abundance
HIGH HYGIENE FACILITY before TCL1 AT

# The Diversity of the Microbiome Impacts Chronic Lymphocytic Leukemia Development in Mice and Humans - Supplemental Methods and Supplemental Figures

*Tereza Faitová[1], *Mariana Coelho[2,3], Caspar da Cunha Bang[1], Selcen Öztürk[2], Ece Kartal[4], Peer Bork[4,5,6,7], #Martina Seiffert[2], #Carsten U. Niemann[1,8]

**Affiliations**

[1] Department of Hematology, Rigshospitalet, Copenhagen, Denmark

[2] Department of Molecular Genetics, German Cancer Research Center (DKFZ), Heidelberg, Germany

[3] Faculty of Biosciences of the University of Heidelberg, Heidelberg, Germany

[4] Structural and Computational Biology Unit, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany

[5] Department of Bioinformatics, Biocenter, University of Würzburg, Würzburg, Germany

[6] Yonsei Frontier Lab (YFL), Yonsei University, Seoul, South Korea

[7] Max Delbrück Center for Molecular Medicine, Berlin, Germany

[8] Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark

**Co-first/co-senior authorship**

*/# These authors contributed equally to this work

**Content**

# 1. Supplementary Methods

## Animal models

Eμ-TCL1 (TCL1) mice on a C57BL/6 J background were kindly provided by Carlo M. Croce (The Ohio State University, Columbus, Ohio, USA), and crossed at least 10 times to ensure C57BL/6 J background (1). 6-8 week old female C57BL/6 J mice for adoptive transfer (AT) of TCL1 tumors were bred and maintained at the central animal facility of the German Cancer Research Center (DKFZ, Heidelberg).

Two colonies of C57BL/6 mice were maintained in two different animal facilities of the German Cancer Research Center (DKFZ) – a low hygiene facility, without user-entry restrictions, and a high hygiene facility (where altered Schaedler flora was ensured). Mice were kept on the same water and food restrictions on both facilities. Ten 7-week-old mice from each colony were submitted to adoptive transfer (AT) with leukemia cells from the TCL1 mouse model (TCL1 AT), as previously described (2, 3). In short, splenocyte suspensions obtained from 40-60 week-old female TCL1 mice underwent B cell purification with the EasySep Mouse Pan-B Cell Isolation Kit (StemCell Technologies Inc., 19844) according to manufacturer instructions. The tumor content (percentage of CD5+CD19+ cells) post-purification reached 95% or higher, according to flow cytometry (FC) measurement. 2x107 cells were injected intraperitoneously, and animals were monitored for the development of leukemia, through weekly blood withdrawals starting at week 2 post-transplantation. Leukemic animals were identified given the presence of a growing CD5+CD19+ tumor population in the blood (as measured by FC), as well as palpable splenomegaly. 3 mice were removed from the study upon failing in tumor engraftment.

## Murine samples collection, DNA extraction and sequencing

Fecal samples were collected at week 0, one day before TCL1 AT, and at week 3 post-transplantation (after tumor was established), into Stool Nucleic Acid Collection and Preservation System tubes (Norgen Biotek, 63700). Fecal samples were snap-frozen with liquid N2 immediately after collection. QIAmp DNA Stool Mini Kit (Qiagen, Hilden, Germany) was used to extract DNA from fecal samples according to manufacturer instructions. The concentration of DNA was determined by NanoDrop spectrophotometry (NanoDrop, Germany). In short, 50 mg of fecal sample was lyophilized in 400 µL extraction buffer (methanol/ddH2O at a ratio of 4/1). Sample mixture was grinded for 6 minutes and sonicated at 5 ºC for 30 minutes. Mixture was then kept at -20 ºC for 30 minutes and centrifuged at 13,000 g and 4 ºC for 15 minutes. Supernatant was collected and snap-frozen. Snap-frozen sample was re-dissolved in 100 µL 90% methanol (in ddH2O). Quality control sample was prepared for NanoDrop measurement. Shotgun sequencing was conducted at the European Molecular Biology Laboratory (EMBL, Heidelberg) on Illumina HiSeq 4000 platform in a 2x150bp paired-end setup at the Genomics Core Facility, European Molecular Biology Laboratory in Heidelberg, Germany.

Peripheral blood (PB) was drawn from the submandibular vein for weekly measurements, and collected in ethylenediaminetetraacetic acid (EDTA)-coated tubes (Sarstedt) for immunostaining and FC measurement. Mice were euthanized by increasing concentrations of carbon dioxide (CO2). Single-cell suspensions from spleen, bone marrow (BM) and inguinal lymph nodes (LN) were prepared as described previously (2, 3). Briefly, splenocyte suspensions were generated with the GentleMACS tissue dissociator (Miltenyi Biotec), followed by red blood cell (RBC) lysis, and passing the cells by 70-µm strainers (BD Biosciences) to exclude fat and cell clumps. BM cells were flushed from femurs with PBS/2% fetal calf serum (FCS). LN cell suspensions were prepared through grinding the lymph nodes through 70-µm strainers (BD Biosciences).

## Flow cytometry

All antibodies were purchased from BD, Biolegend or ThermoFisher Scientific. For surface staining, single-cell suspensions previously obtained were washed with PBS/2% FCS, and incubated with recommended antibodies against cell surface markers, using the respective recommended antibody dilutions. Incubation went on for 30 minutes, at 4 ºC in the dark. After washing twice with PBS/2% FCS, cells were fixed using IC fixation buffer (ThermoFisher Scientific, 00-8222-49), washed and stored in PBS/2% FCS at 4ºC in the dark until analyzed by FC.

For intracellular staining with FoxP3, cells were fixed after surface staining using FoxP3 fixation/permeabilization buffer (ThermoFisher Scientific, 00-552300) for 30 minutes at RT (in the dark). Then, permeabilization was performed with 1X permeabilization buffer (ThermoFisher Scientific, 00-5523-00) and staining with antibody against FoxP3 transcription factor in 1X permeabilization buffer was started for 30 minutes at 4 ºC. after two washes with 1X permeabilization buffer, cells were resuspended in the same buffer and stored at 4 ºC in dark conditions, until analyzed by FC.

For PB staining, 25 µL of blood were stained with surface molecule antibodies for 30 minutes at 4 ºC in the dark. Then, blood samples were incubated for 10 minutes in 2 mL of 1X 1-step Fix/Lyse Solution (ThermoFisher Scientific, 00-5333-57) for RBC removal. After centrifugation and discarding of the supernatant, cell pellets were resuspended in PBS. Less than 1 hour before FC measurement, 25 µL of 123count eBeads (ThermoFisher Scientific, 01-1234-42) were added, to allow calculation of absolute number of cells in the blood according to the formula: absolute count (cells/µL) = (cell count x bead volume x bead concentration)/(bead count x cell volume).

Flow cytometry data acquisition was done using BD LSRFortessa flow cytometer (BD Biosciences). For analysis, dead cells and doublets were excluded. Median Fluorescence Intensity (MFI) was calculated and normalized by subtracting the MFI of the respective fluorescence minus one (FMO) control, for markers without a clear negative vs positive population distinction. Data analysis was performed using FlowJo X 10.0.7 software (FlowJo LLC, BD Biosciences).

## Patient cohort and patient data

Fecal samples were collected from 60 patients enrolled in the CLL biobank and the PERSIMUNE biobank during regular out-patient visits at Rigshospitalet, Copenhagen, Denmark, between June 2017, and July 2020. The project was approved by the national ethics committee (approval no. 1804410) and written informed consent was obtained from all patients prior to sampling.

## Patient sample collection and sequencing

Fecal samples were collected by the patient or nursing staff using the OMNIgene.GUT (DNA Stabilized-frozen Inc., Ottawa, ON, Canada) stabilization tube according to the manufacturer's instructions and refrigerated for a maximum of 7 days before freezing at -80 C. All samples were processed and stored at the PERSIMUNE biobank located at Rigshospitalet. Samples were extracted from the biobank and transported on dry ice to the IrsiCaixa AIDS Research Institute, Spain. Here samples were thawed on ice and DNA was extracted from ~200 mg of each fecal sample using the Power Soil DNA Extraction Kit (MO BIO Laboratories, Carlsbad, CA, US). 4 Extracted DNA was then stored at -80 ºC until sequencing. For sequencing, whole fecal DNA was chemically fragmented using the Nextera-XT® Illumina kit. Total fecal DNA was sequenced on an Illumina Hi-Seq® platform in a 2x150bp paired-end setup.

## Preprocessing and taxonomical profiling

Both, reads preprocessing and taxonomical profiling were done for the mouse and CLL samples using an in-house pipeline implemented in ngless (4). After sequencing, reads underwent quality control and preprocessing steps, including trimming of reads using a quality score cutoff of 20, and removal of reads below 100 base pairs. Reads were mapped to the mouse and human genomes (mm39 and hg19, respectively) using BWA-MEM2 (5) and reads with a minimum identity of 80 across 90 bases were discarded. Gene profiling was carried out by mapping to the integrated gene catalogue (IGC) (6) with BWA-MEM2 and counting the number of reads mapping to each gene with a minimum match size of 60 and minimum identity of 90. Taxonomic profiling and estimation of the relative abundances at all taxonomical levels in mice and humans were done using mOTUs2 (7) and reference-independent method MetaPhlAn3 (8), respectively. The naming for mOTUs gives all possible species that the specific mOTU potentially represents. The format is species1/species2/species3/etc., meaning that all the species are possible annotations. MetaPhlAn3 utilizes the marker gene database to align the metagenomic sequencing data. By identifying the most suitable marker gene sequences, MetaPhlAn3 ensures a unique species assignment during the alignment of metagenomic sequencing data to the marker gene database. Taxa failing to be classified into any taxonomic ranks were marked as *Unclassified* by both mOTUs2 and MetaPhlAn3.

## Bioinformatics and Statistical analysis

Statistical analyses of FC mouse data were performed using GraphPad Prism software version 9. All graphs show means ± SEM, unless otherwise indicated. Comparisons of two different sample groups (n = 8, n = 9) at each blood withdrawal timepoint were performed using a non-parametric Mann Whitney t-test. All other statistical analyses were performed using R (version 4.2.0).

Descriptive analyses were performed for both the mouse and human CLL cohorts with relative bacterial abundance as input data. Unless stated otherwise, Mann Whitney U Test (Wilcoxon Rank Sum Test) was used to identify significant differences between subgroups; Benjamini-Hochberg (BH) method was used for multiple-testing correction; BH adjusted p-value of <0.05 was considered significant. PERMANOVA (*adonis2* from *vegan* R package version 2.6-2) was used to test statistical differences in microbial composition among different groups of murine samples (such as cage effect) or groups of patient samples (such as gender, groups patients with different antibiotic usage, or resulting patient clusters).

Alpha diversity measures (richness, Shannon index) were calculated at species level using *vegan* R package version 2.6-2 (9). Generalized linear model (GLM) was used to visualize relationship between two independent binary variables - TCL1 AT and hygiene of the animal facilities, and one dependent variable - bacterial relative abundance. Interpretation of the GLM statistical analysis is detailed in Figure 6C legend.

The interindividual dissimilarities in gut microbiota composition (β-diversity) in patients with CLL were assessed by robust aitchison dissimilarity index (*vegdist* function from *vegan* R package version 2.6-2). Hierarchical clustering (*hclust* from base R package *stats*) with Ward's minimum variance method was run on the distance matrix. The interindividual dissimilarities represented by the distance matrix and the formed clusters were visualized using Principal Coordinate Analysis (PCoA, *gl.pcoa* function from *dartR* R package version 2.7.2) (10). First 3 components of PCoA were visualized using 3D plot as Figure 1B (*scatter3D* function from *plot3D* graphics library in R, version 1.4). Statistical differences in microbial composition given by β-diversity estimates between the resulting cluster were tested using PERMANOVA.

Differential abundance of bacterial species in fecal microbiome between CLL patient clusters was assessed using the R implementation of SIAMCAT (*SIAMCAT* R package version 2.0.0) (11). The cutoff value for bacterial taxa to be considered differentially abundant was log fold change > 1 (logFC).

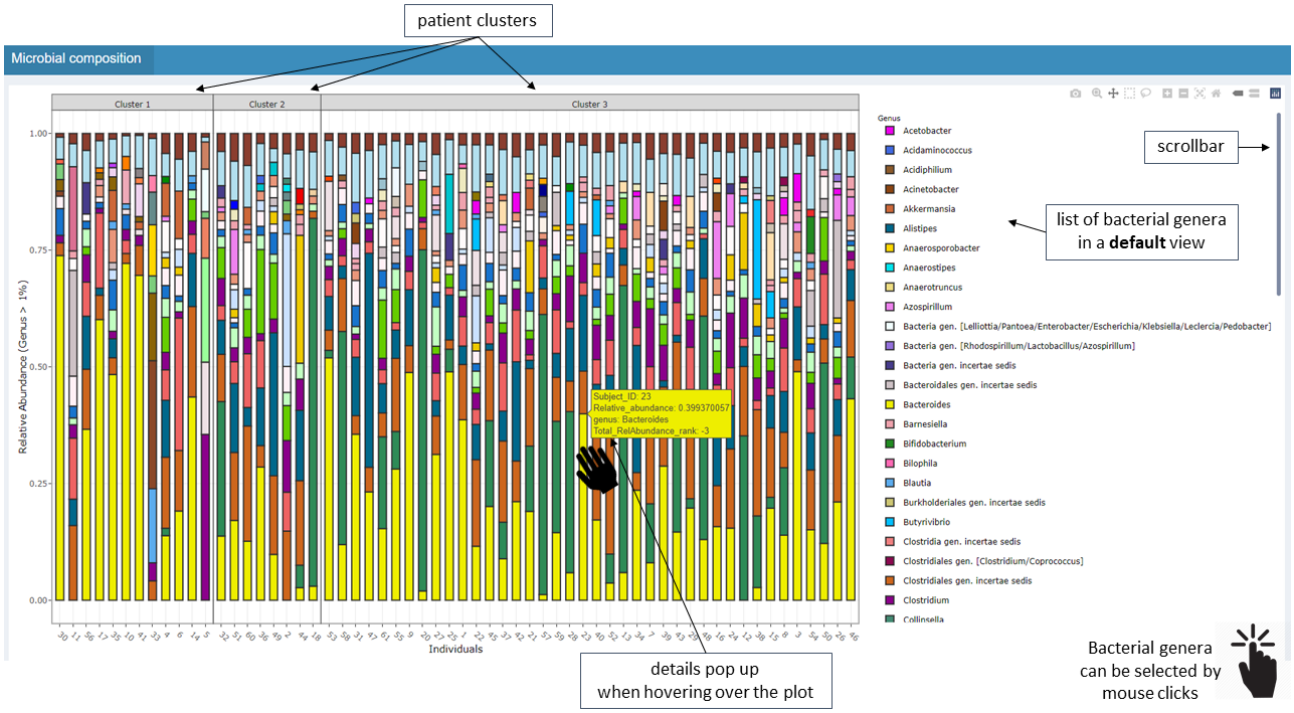
R script as well as taxonomical profiling data of the mice cohort are available at https://github.com/PERSIMUNE/PAC2023Faitova_Human_Mice_CLL.

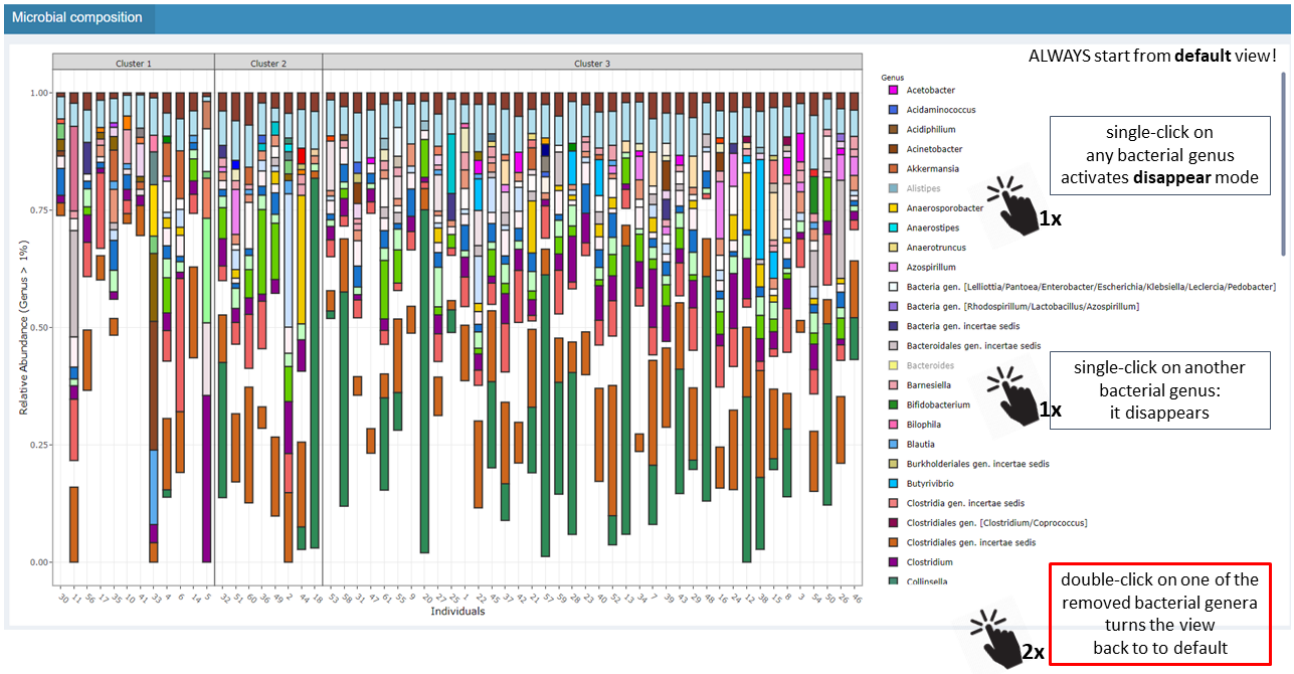
## 2a. R Shiny applications – Methods

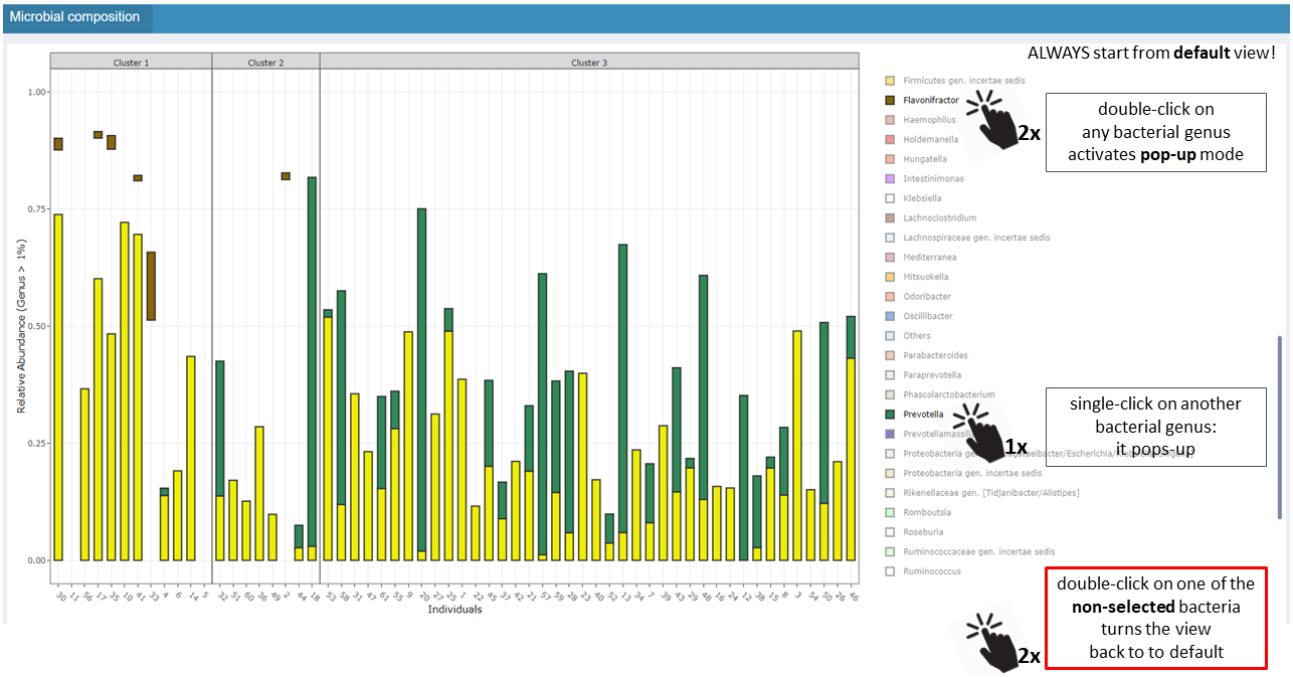
Relative bacterial abundance and GMM abundance data from individual patient samples were retrieved as described in main Methods section. Relative bacterial abundance data refers to information about the proportional representation of different bacterial species or taxa within a given sample. GMM abundance represents a normalized proxy for the portion of bacteria in a sample that have the ability to perform a specific function. Graphs representing the abundances were constructed using ggplot2 package in R (version 3.3.6). Interactive features such as tooltips, zooming, panning, and hover effects were introduced to the static ggplot2 graphs using plotly package (version 4.10.0). The *ggplotly* function was employed to convert the ggplot2 graphs into interactive plotly graphs.

The R Shiny framework is employed to develop interactive web applications which were deposited on the open source Shinyapps.io. The applications are accessible through links that can be found in the main text.

# 2b. R Shiny applications – Tutorial



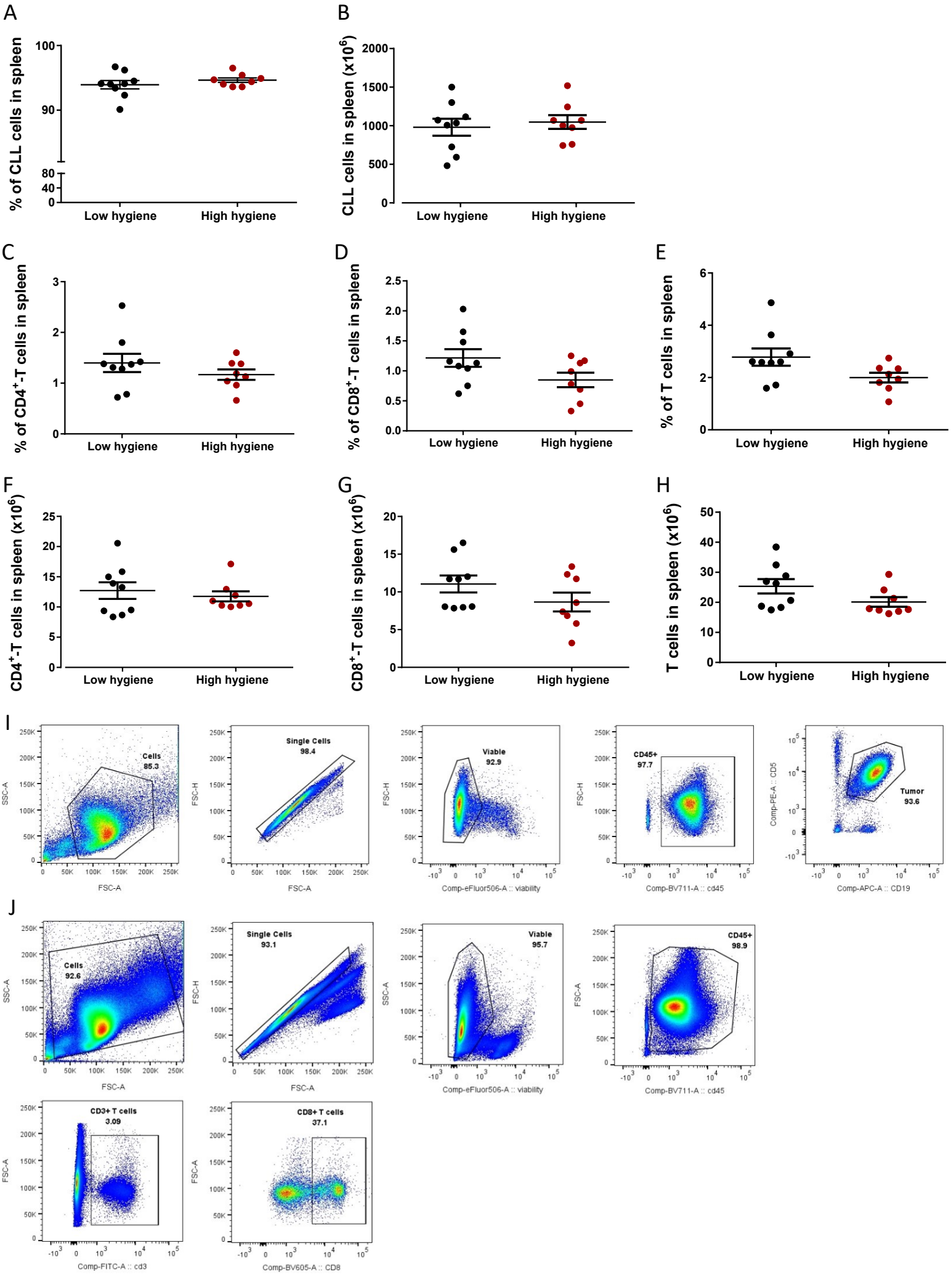




Microbial composition
Cluster 1
Cluster 2
Cluster 3
Relative Abundance (Genus > 1%)
1.00
0.75
0.50
0.25
0.00
Individuals
ALWAYS start from **default** view!
Firmicutes gen. incertae sedis
Flavonifractor
Haemophilus
Holdemanella
Hungatella
Intestinimonas
Klebsiella
Lachnoclostridium
Lachnospiraceae gen. incertae sedis
Mediterranea
Mitsuokella
Odoribacter
Oscillibacter
Others
Parabacteroides
Paraprevotella
Phascolarctobacterium
Prevotella
Proteobacteria gen. incertae sedis
Rikenellaceae gen. [Tidjanibacter/Alistipes]
Romboutsia
Roseburia
Ruminococcaceae gen. incertae sedis
Ruminococcus
2x
double-click on any bacterial genus activates **pop-up** mode
1x
single-click on another bacterial genus: it pops-up
2x
double-click on one of the **non-selected** bacteria turns the view back to to default

## References

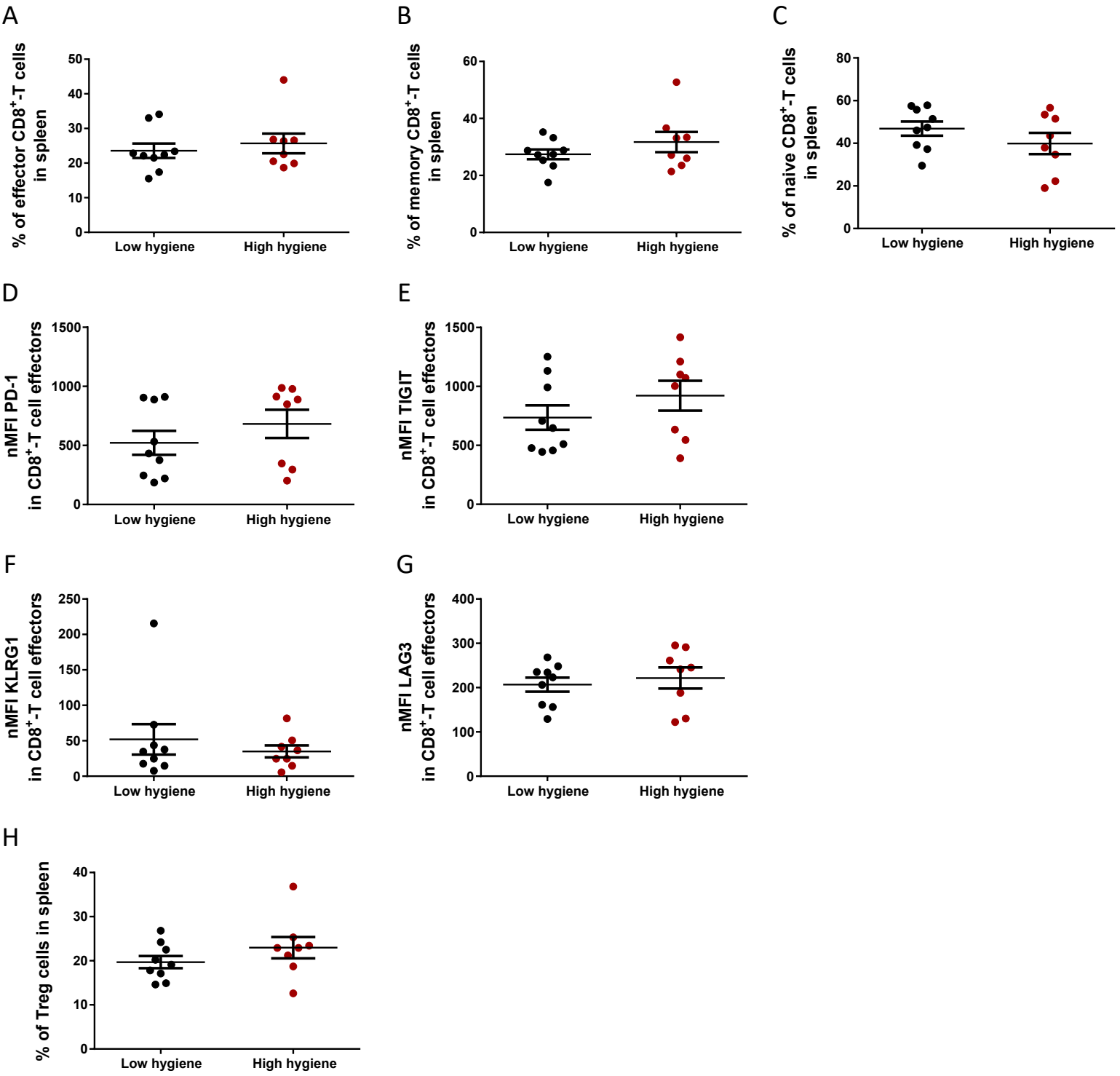

1. Bichi R, Shinton SA, Martin ES, Koval A, Calin GA, Cesari R, et al. Human chronic lymphocytic leukemia modeled in mouse by targeted TCL1 expression. Proc Natl Acad Sci U S A. 2002;99(10):6955-60.
2. Hanna BS, McClanahan F, Yazdanparast H, Zaborsky N, Kalter V, Rossner PM, et al. Depletion of CLL-associated patrolling monocytes and macrophages controls disease development and repairs immune dysfunction in vivo. Leukemia. 2016;30(3):570-9.
3. McClanahan F, Hanna B, Miller S, Clear AJ, Lichter P, Gribben JG, et al. PD-L1 checkpoint blockade prevents immune dysfunction and leukemia development in a mouse model of chronic lymphocytic leukemia. Blood. 2015;126(2):203-11.
4. Coelho LP, Alves R, Monteiro P, Huerta-Cepas J, Freitas AT, Bork P. NG-meta-profiler: fast processing of metagenomes using NGLess, a domain-specific language. Microbiome. 2019;7(1):84.
5. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-60.
6. Li J, Jia H, Cai X, Zhong H, Feng Q, Sunagawa S, et al. An integrated catalog of reference genes in the human gut microbiome. Nature Biotechnology. 2014;32(8):834-41.
7. Milanese A, Mende DR, Paoli L, Salazar G, Ruscheweyh H-J, Cuenca M, et al. Microbial abundance, activity and population genomic profiling with mOTUs2. Nature Communications. 2019;10(1):1014.
8. Beghini F, McIver LJ, Blanco-Míguez A, Dubois L, Asnicar F, Maharjan S, et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. Elife. 2021;10.
9. Oksanen J SG, Blanchet F, Kindt R, Legendre P, Minchin P, O'Hara, R SP, Stevens M, Szoecs E, Wagner H, Barbour M, Bedward M,, Bolker B BD, Carvalho G, Chirico M, De Caceres M, Durand S,, Evangelista H FR, Friendly M, Furneaux B, Hannigan G, Hill M,, Lahti L MD, Ouellette M, Ribeiro Cunha E, Smith T, Stier A, Ter, Braak C WJ. _vegan: Community Ecology Package_. In: package R, editor. 2022.
10. Gruber B, Unmack PJ, Berry OF, Georges A. dartr: An r package to facilitate analysis of SNP data generated from reduced representation genome sequencing. Mol Ecol Resour. 2018;18(3):691-9.
11. Wirbel J, Zych K, Essex M, Karcher N, Kartal E, Salazar G, et al. Microbiome meta-analysis and cross-disease comparison enabled by the SIAMCAT machine learning toolbox. Genome Biology. 2021;22(1):93.
12. Gacesa R, Kurilshikov A, Vich Vila A, Sinha T, Klaassen MAY, Bolte LA, et al. Environmental factors shaping the gut microbiome in a Dutch population. Nature. 2022;604(7907):732-9.

# Supplemental Figure 1



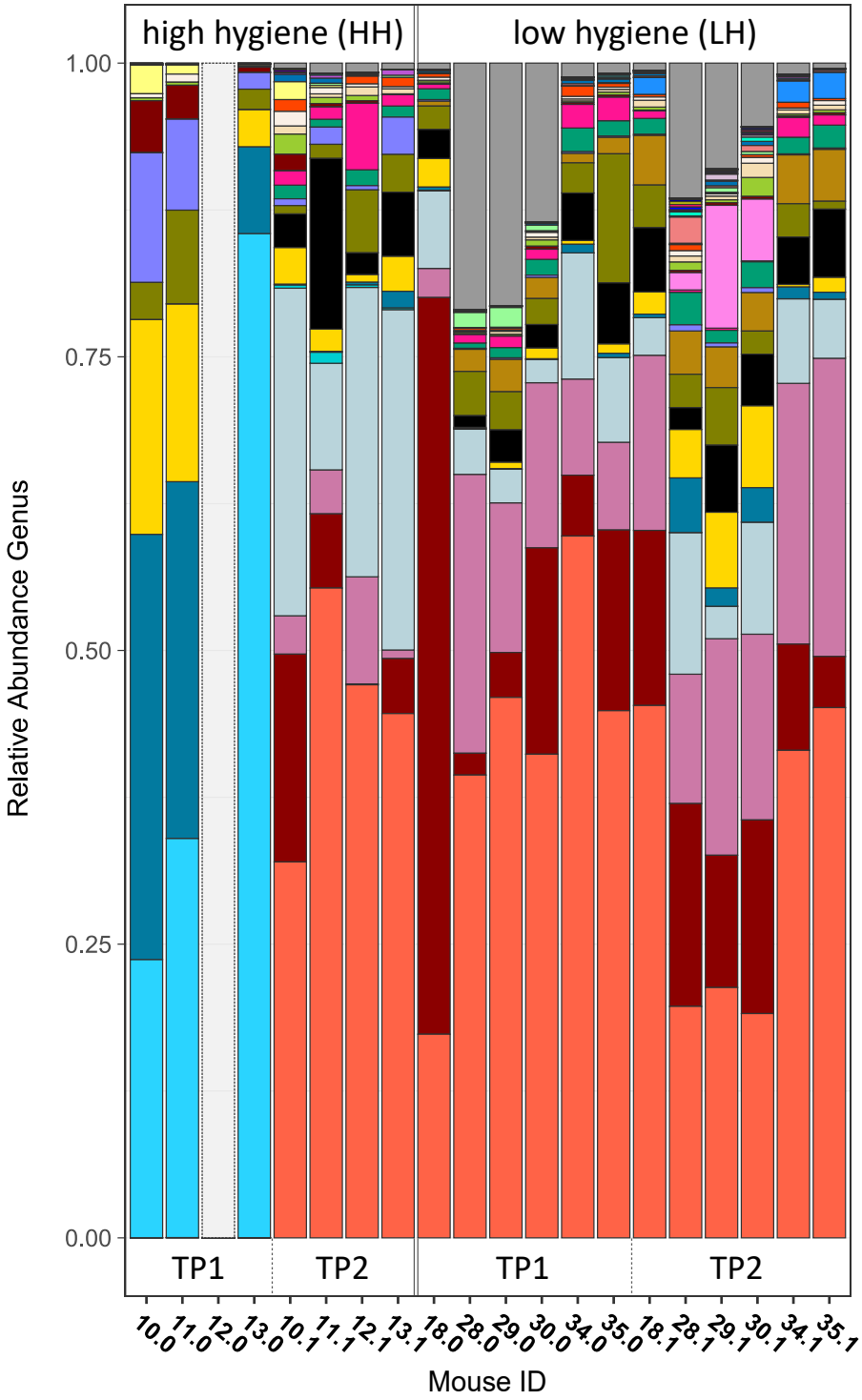
**Supplemental Figure 1.** A) Percentage of CLL cells (CD19+CD5+) out of CD45+ compartment in spleen. B) Absolute number of CLL cells in spleen (x10^6). C-E) Percentage of CD4+, CD8+ and CD3+ T cells in spleen out of viable compartment. F-H) Absolute number of CD4+, CD8+ and CD3+ T cells in spleen (x10^6). I) Exemplary gating strategy for CD19+CD5+ CLL cells ("Tumor"). J) Exemplary gating strategy for CD8+ T cells.

Supplemental Figure 2




**Supplemental Figure 2.** A-C) Percentage of effector, memory and naive CD8+ T cells out of T cell splenic compartment. D-G) Normalized median fluorescence intensity (nMFI) of PD-1, TIGIT, KLRG1 and LAG3 in CD8+ T effector cells in t he splenic compartment. H) Percentage of FoxP3+CD25+ Treg cells out of CD4+ cells in the splenic compartment.

Supplemental Figure 3



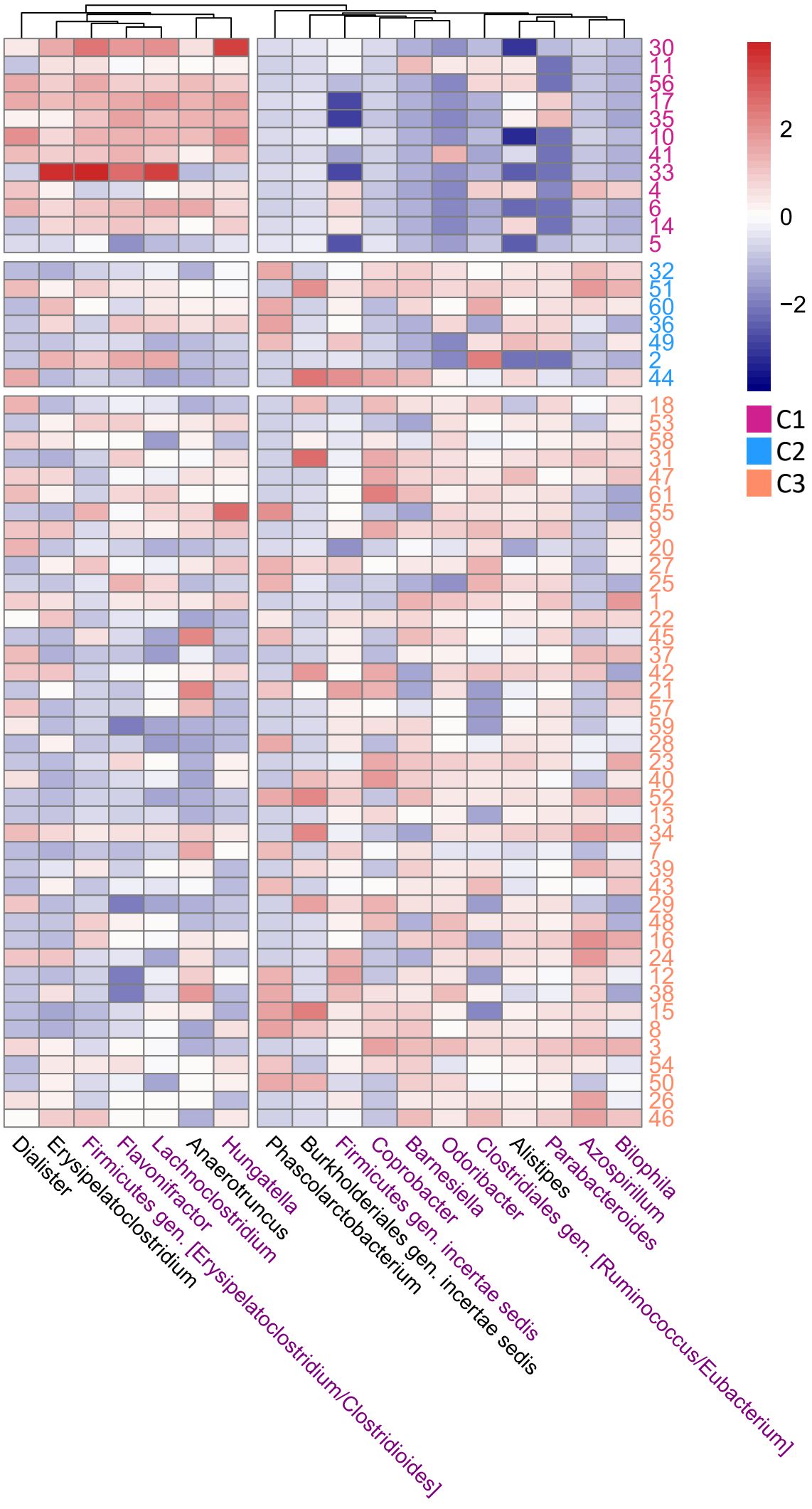

Genus

bacterial genera
with total abundance > 1%

- Unclassified
- Adlercreutzia
- Akkermansia
- 12 Alistipes
- Anaerostipes
- Anaerotruncus
- Bacteroidaceae gen. incertae sedis
- 4 Bacteroidales gen. incertae sedis
- 5 Bacteroides
- Bifidobacterium
- Bittarella
- 10 Burkholderiales gen. incertae sedis
- 13 Clostridiaceae gen. incertae sedis
- Clostridiales gen. [Clostridium/Flavonifractor]
- 15 Clostridiales gen. incertae sedis
- 14 Clostridium
- Dorea
- Dubosiella
- Eggerthella
- Enterorhabdus
- Erysipelatoclostridium
- Eubacterium
- Faecalibaculum
- Firmicutes gen. [Erysipelatoclostridium/Clostridioides]
- 11 Firmicutes gen. incertae sedis
- Flavonifractor
- 3 Helicobacter
- Hungateiclostridiaceae gen. incertae sedis
- Ileibacterium
- Intestinimonas
- 7 Lachnospiraceae gen. incertae sedis
- 9 Lactobacillus
- 2 Mucispirillum
- 1 Muribaculaceae gen. incertae sedis
- 8 Muribaculum
- Oscillibacter
- Paeniclostridium
- 6 Parabacteroides
- Proteobacteria gen. [Escherichia/Klebsiella/Shigella]
- Pseudoflavonifractor
- Romboutsia
- Ruminococcaceae gen. incertae sedis
- Staphylococcus

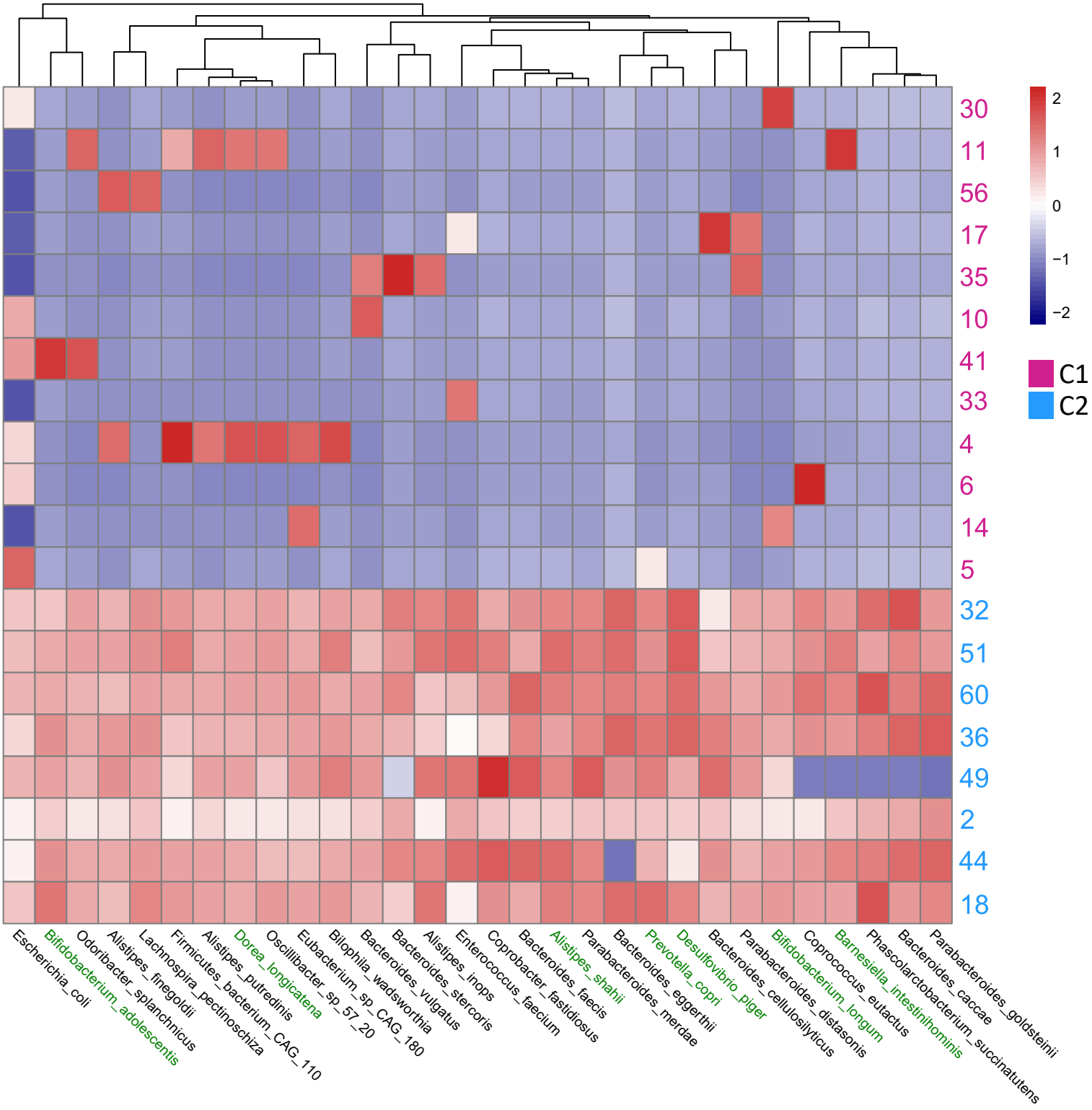
1-15: 15 bacterial genera with highest total rel. abundance, ranked

Supplemental Figure 4



**Supplemental Figure 4. Heatmap visualization of differentially abundant bacterial genera between C1 and C2, all clusters.** Relative abundances of selected genera visualized for all 3 clusters. Selected genera: genera with logFC >1 enriched in C1, and twelve genera with highest logFC enriched in C2 patient samples. Patient samples (y-axis) are labeled according to the cluster membership from Figure 1. Subject 5 (from C1) was omitted from visualization due to extremely low relative abundances across all genera. Bacterial species that were found to be differentially abundant also between C1 and C3 are marked in blue. Full list of differentially abundant bacterial genera and species between C1 and C3 can be found in Supplemental Table 3. Data used for visualization were clr-transformed relative abundances. Clustered heatmap was built and visualized using kmeans clustering method from *pheatmap* function in R.

Supplemental Figure 5




**Supplemental Figure 5. Heatmap visualization of differentially abundant bacterial species between C1 and C2.** Thirty bacterial species (x-axis) with highest fold change (analysis by SIAMCAT; Supplemental Table 3) were selected for visualization. Patient samples (y-axis) were labeled according to the cluster membership from Fig 1. Bacterial species enriched in patients from C2 corresponding to healthy signature identified by Gacesa et al. (12) are marked in green. Data used for visualization were clr-transformed relative abundances of bacterial species. Clustered heatmap was built and visualized using k-means clustering method from pheatmap function in R.