

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

Tereza Faitová^{*1}, Mariana Coelho^{*2,3}, Caspar da Cunha Bang¹, Selcen Öztürk², Ece Kartal⁴, 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. Supplemental Methods
2. R shiny applications
 - a. Methods
 - b. Tutorial

1. Supplemental Methods

Animal models

E μ -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. 2×10^7 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 μ L extraction buffer (methanol/ddH₂O at a ratio of 4/1). Sample mixture was grinded for 6 minutes and sonicated at 5 $^{\circ}$ C for 30 minutes. Mixture was then kept at -20 $^{\circ}$ C for 30 minutes and centrifuged at 13,000 g and 4 $^{\circ}$ C for 15 minutes. Supernatant was collected and snap-frozen. Snap-frozen sample was re-dissolved in 100 μ L 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 °C 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 4°C 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 °C. after two washes with 1X permeabilization buffer, cells were resuspended in the same buffer and stored at 4 °C 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 °C 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: $\text{absolute count (cells/}\mu\text{L)} = (\text{cell count} \times \text{bead volume} \times \text{bead concentration}) / (\text{bead count} \times \text{cell volume})$.

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 °C 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 $\pm$ 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 2C 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 3B (*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).

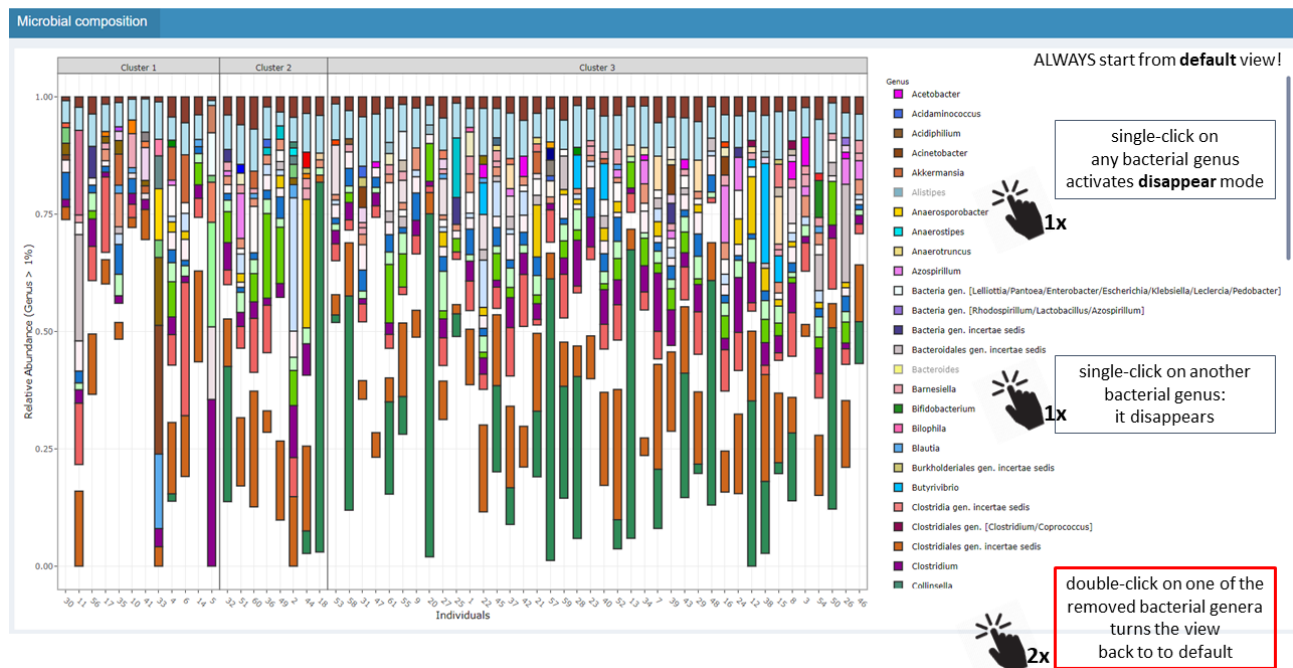
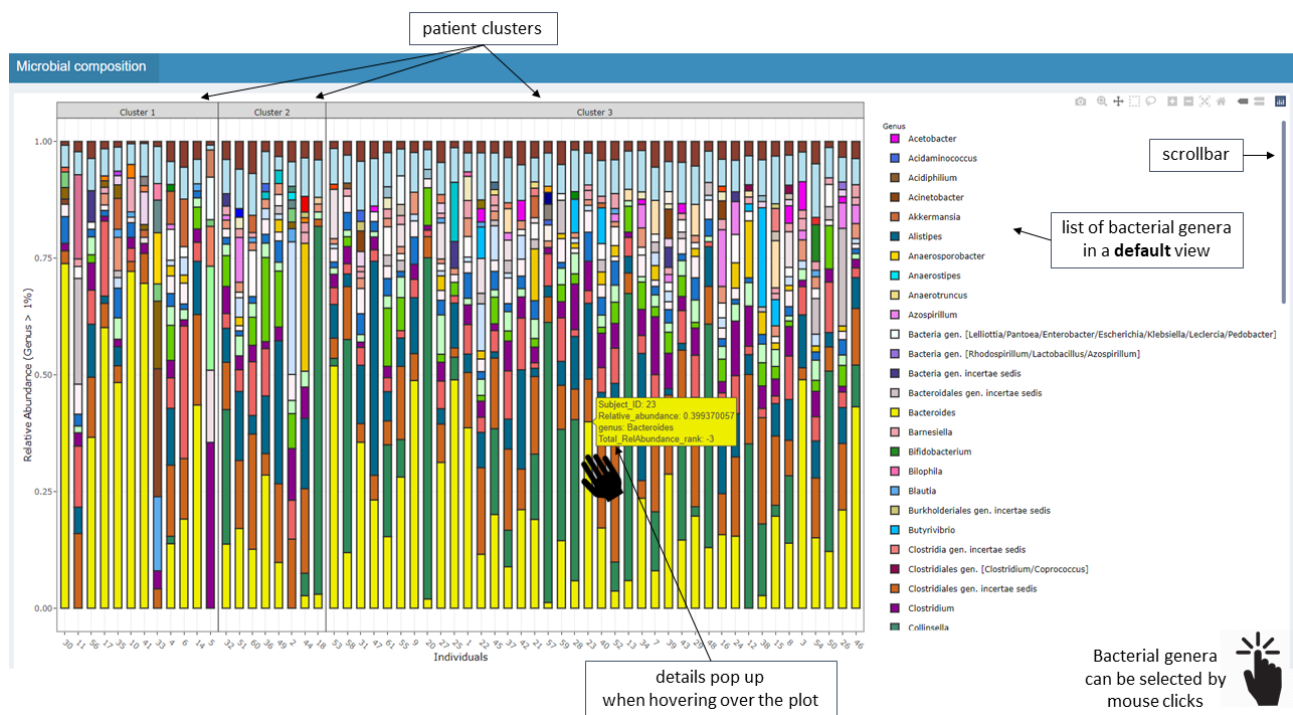
R script as well as taxonomical profiling data of the mice cohort are available at https://github.com/PERSIMUNE/PAC2023Faitova_Human_Mice_CLL.

2a. R Shiny applications – Methods

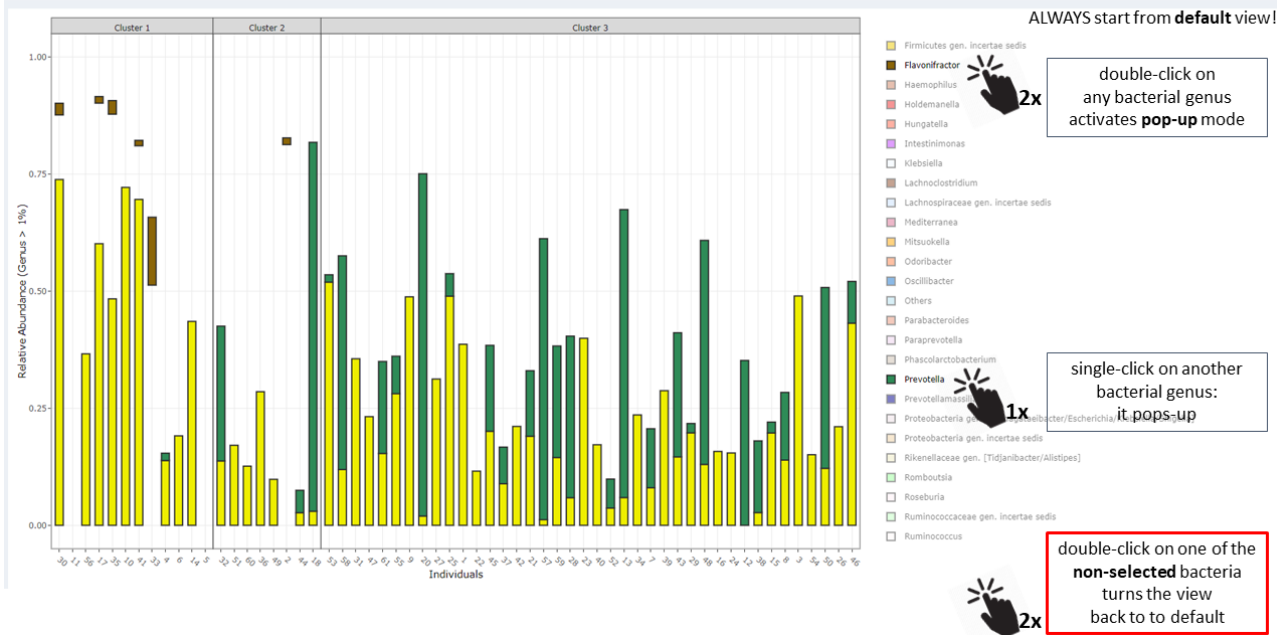
Relative bacterial abundance and GMM abundance data from individual patient samples were retrieved as described in main Methods section. Relative bacterial abundance data refers to information about the proportional representation of different bacterial species or taxa within a given sample. GMM abundance 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



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.