Specific gut microbiome members are associated with distinct immune markers in allogeneic hematopoietic stem cell transplantation

Anna Cäcilia Masche[1], Katrine Kielsen[2,3], Malene Skovsted Cilieborg[4], Ole Lund[5], Susan Holmes[6], Frank M. Aarestrup[1], Klaus Gottlob Müller[2,3] , Sünje Johanna Pamp[1*]

[1] Research Group for Genomic Epidemiology, Technical University of Denmark, Kongens Lyngby, Denmark.
[2] Institute for Inflammation Research, Department of Rheumatology and Spine Disease, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark.
[3] Department of Pediatrics and Adolescent Medicine, Copenhagen University Hospital Rigshospitalet, Copenhagen, Denmark.
[4] Comparative Pediatrics and Nutrition, Department of Clinical Veterinary and Animal Science, University of Copenhagen, Frederiksberg, Denmark.
[5] Department of Bio and Health Informatics, Technical University of Denmark, Kongens Lyngby, Denmark.
[6] Department of Statistics, Stanford University, Stanford, USA.

*Correspondence: sjpa@food.dtu.dk
Research Group for Genomic Epidemiology, Technical University of Denmark, Kemitorvet Building 204 Room 102, 2800 Kongens Lyngby, Denmark.

Abstract

Background:
Increasing evidence reveals the importance of the microbiome in health and disease and inseparable host-microbial dependencies. Host-microbe interactions are highly relevant in patients receiving allogeneic hematopoietic stem cell transplantation, (HSCT), i.e. a replacement of the cellular components of the patients' immune system with that of a foreign donor. HSCT is employed as curative immunotherapy for a number of non-malignant and malignant hematologic conditions, including acute lymphoblastic leukemia. The procedure can be accompanied by severe side effects such as infections, acute graft-versus-host disease (aGvHD), and death. Here, we performed a longitudinal analysis of immunological markers, immune reconstitution and gut microbiota composition in relation to clinical outcomes in children undergoing HSCT. Such an analysis could reveal biomarkers, e.g. at the time point prior to HSCT, that in the future could be used to predict which patients are of high risk in relation to side effects and clinical outcomes and guide treatment strategies accordingly.

Results:
In two multivariate analyses (sparse partial least squares regression and canonical correspondence analysis), we identified three consistent clusters: (1) High concentrations of the antimicrobial peptide human beta-defensin 2 (hBD2) prior to the transplantation in patients with high abundances of *Lactobacillaceae*, who later developed moderate or severe aGvHD and exhibited high mortality. (2) Rapid reconstitution of NK and B cells in patients with high abundances of obligate anaerobes such as *Ruminococcaceae*, who developed no or mild aGvHD and exhibited low mortality. (3) High inflammation, indicated by high levels of C-reactive protein, in patients with high abundances of facultative anaerobic bacteria such as *Enterobacteriaceae.* Furthermore, we observed that antibiotic treatment influenced the bacterial community state.

Conclusions:
We identify multivariate associations between specific microbial taxa, host immune markers, immune cell reconstitution and clinical outcomes in relation to HSCT. Our findings encourage further investigations into establishing longitudinal surveillance of the intestinal microbiome and relevant immune markers, such as hBD2, in HSCT patients. Profiling of the microbiome may prove useful as a prognostic tool that could help identify patients at risk of poor immune reconstitution and adverse outcomes, such as aGvHD and death, upon HSCT, providing actionable information in guiding precision medicine.

Keywords:
Gut microbiota – pediatric cancer – HSCT – 16S rRNA gene profiling – data integration - immune reconstitution - *Ruminococcaceae* - Human beta-defensin 2 – acute GvHD - B cells and NK cells

Background

The microbiome has gained increasing attention as a crucial contributor in the course of various diseases, and as target of treatment [1–3]. A number of complications in allogeneic hematopoietic stem cell transplantation (HSCT) have recently been associated with the gut microbiota [4, 5]. HSCT is a curative treatment for various hematologic diseases including malignancies, such as acute lymphoblastic leukemia (ALL), as well as non-malignant diseases, such as metabolic disorders and immune deficiency syndromes [6]. One goal of HSCT is achieving a beneficial graft-versus-leukemia (GvL) effect where donor-derived T lymphocytes and natural killer cells target leukemic cells in the recipient [7]. Prior to allogeneic HSCT, the patients undergo a preparative conditioning regimen involving combinations of chemotherapeutic agents and total body irradiation (TBI) [8] to eradicate leukemic cells and induce immunosuppression. Immunosuppressive treatment (both prior to and post HSCT) in the stem cell recipient prevents a graft-versus-host reaction caused by cytotoxic donor T lymphocytes that attack healthy cells in the recipient [8]. To limit infectious diseases due to immunosuppression, the patients are administered broad-spectrum antibacterial and antifungal compounds. Subsequently, the patients receive a stem cell graft originating from the bone marrow, peripheral blood or umbilical cord blood of a human leukocyte antigen (HLA)-matched sibling donor or an unrelated donor (i.e., allogeneic HSCT) [9, 10].

In patients undergoing HSCT, it has been previously observed that there are associations between the microbiome and clinical outcomes such as acute graft-versus-host disease (aGvHD) [5, 11, 12] and survival [4, 13]. GvHD after HSCT has been related to an expansion of the order *Lactobacillales*, especially *Enterococcus* spp. [11] and *Lactobacillus* spp. [14] and a loss of *Clostridiales* [14]. However, so far few studies have monitored the microbiome longitudinally [11, 14, 15]. This indicates the need for more detailed investigations that take temporal monitoring of both the host immune system and the microbiome into account.

Several markers of the host immune system, including inflammatory markers (such as C-reactive protein (CRP) and interleukin 6 (IL-6)) and markers of intestinal toxicity (such as plasma citrulline), have been studied in HSCT [16, 17]. Potential novel markers, such as antimicrobial peptides (AMPs), have also been proposed to be involved in outcomes after HSCT, for example in immunomodulation and regulation of microbial homeostasis [18]. AMPs, especially defensins such as human beta-defensin 2 and 3 (hBD2 and hBD3), have previously been found to play a role in some inflammatory diseases [19, 20]. However, to our knowledge, AMPs have not yet been employed as markers in the context of HSCT. An interesting research question remains: How are known and novel markers within the host immune system

associated with each other or with changes in the microbiome? A better understanding about these associations would provide a more holistic insight into the underlying mechanisms affecting clinical outcomes after HSCT.

Another crucial factor impacting on complications and clinical outcomes after HSCT is immune reconstitution. Immune reconstitution involves the essential cellular components of the adaptive immune system, namely T and B cells, as well as key cellular components of the innate immune defense, namely natural killer (NK) cells, monocytes and neutrophils [21]. A microbial influence on immune cell differentiation has been observed previously, e.g., commensal *Clostridiales* were found to regulate Treg cell differentiation in the colon [22]. Therefore, an influence of the intestinal microbiome on immune reconstitution following HSCT is likely, but had not been investigated prior to this study.

Here, we monitor both host factors and the intestinal microbiome longitudinally. We included markers of inflammation (CRP and IL-6) and intestinal toxicity (plasma citrulline) as well as the antimicrobial peptides (hBD2 and 3), which in the following sections are collectively referred to as immune markers. To assess the prognostic potential of immune markers associated with gut microbial dynamics for immune reconstitution and clinical outcomes after HSCT (aGvHD and survival) in a holistic way, we implemented multivariate multi-table (also referred to as multi-way) approaches. This facilitated the integration of a variety of different factors that could influence the patients' convalescence. We reveal distinct clusters of bacteria associated with sets of immune markers and clinical outcomes. Patients with rapid NK and B cell reconstitution that had no or mild aGvHD and low mortality exhibited high abundances of members belonging to the family of *Ruminococcaceae*. In contrast, patients with moderate to severe aGvHD and high mortality showed high plasma concentrations of the antimicrobial peptide hBD2 already prior to HSCT. In these patients, we observed increased *Lactobacillaceae* abundances.

Results

To assess associations between immune markers and gut microbial dynamics in the context of immune reconstitution and clinical outcomes after HSCT, we monitored 37 children over time undergoing allogeneic HSCT (Figure 1A, Additional file 1: Table S1). We gained insight into the patients' immune reconstitution by determining T, B, NK, monocyte, and neutrophil cell counts in peripheral blood (Figure 1A, Additional file 1: Table S1). We measured C-reactive protein (CRP) and plasma interleukin 6 (IL-6) as markers of inflammation, and human beta-defensin 2 and 3 (hBD2 and 3) as markers for potential innate immune activation and systemic infection (Figure 1A, Additional

file 1: Table S1). Plasma citrulline levels were measured as a marker of intestinal toxicity, being produced selectively by functioning enterocytes. To relate immune marker levels to members of the intestinal microbiota in patients undergoing HSCT, we characterized the longitudinal dynamics of the human intestinal microbiome in a subset of 30 patients by utilizing 16S rRNA gene profiling (Figure 1A).

Patient cohort and outcomes

At the time of HSCT, the 37 patients were on average 8.2 years old (age range 1.1 - 18.0 years). Twenty-five patients (68%) were diagnosed with at least one bacterial infection at median 75 days post HSCT (range: day -19 to +668). Twenty-six patients (70%) had no or mild aGvHD (grade 0-I) (Additional file 1: Table S1). Eleven patients (30%) developed moderate to severe aGvHD (grade II, III or IV) at median 18 days (range: day +9 to +45) after transplantation (Additional file 1: Table S1). Seven patients (19%) died during the follow-up period at median 266 days post HSCT (range: day +9 to +784) (5 relapse-related and 2 treatment-related deaths) (Additional file 1: Table S1). In total, six patients (16%) relapsed, four of which underwent a re-transplantation. All patients received antibiotics pre- and post-transplantation (Additional file 1: Table S1). Prophylactic trimethoprim-sulfamethoxazole was administered to all patients from day -7 until transplantation. During the period of neutropenia or latest from day -1, patients received prophylactic intravenous ceftazidime. In case of infections indicated by fever or microbial culture, ceftazidime was substituted by intravenous meropenem, vancomycin or other antibiotics, according to culture-based results.

Temporal dynamics of immune markers and the intestinal microbiota in HSCT patients

Prior to assessing the interplay between clinical variables (i.e., immune markers, immune reconstitution, clinical outcomes) and the intestinal microbiota, we characterized these components separately. In order to provide an overview of changes in immune markers and immune cell counts after HSCT in our cohort (*n=37*), we assessed their temporal patterns (Figures 1B, 1C, and Additional file 2: Figure S1). Of note, these supplemental univariate analyses mainly serve the purpose of visually aiding our subsequent multivariate analyses approaches (Additional file 3: Figure S2). We characterized hBD2 for the first time in the context of HSCT by assessing plasma hBD2 concentrations over time from pre-HSCT to month +3 post HSCT in patients compared to healthy controls. The hBD2 concentration differed significantly between time points ($P < 0.001$, Kendall's $W = 0.6$). It increased from pre-HSCT to the day of HSCT ($P < 0.001$), then slightly decreased in week +1 ($P = 0.038$) before increasing again in week +3 ($P = 0.006$). HBD2 decreased again in month +2 ($P = 0.014$) (Figure 1B). CRP levels differed significantly between time points ($P < 0.001$, Kendall's $W = 0.33$). They were high pre-HSCT and until week +2, then decreasing significantly in week +3 ($P < 0.001$) with the lowest levels in weeks +4 to +6 ($P < 0.001$) (Additional file 2: Figure

S1A). Median plasma citrulline levels were significantly different between time points ($P < 0.001$, Kendall's $W = 0.32$). They decreased from pre-HSCT to week +1 ($P < 0.001$) and increased again in week +3 ($P < 0.001$) (Additional file 2: Figure S1A). B cell counts (Kendall's $W = 0.5$) as well as CD4+ T cell counts (Kendall's $W = 0.46$) increased steadily from month +1 to month +6 ($P < 0.001$) (Additional file 2: Figure S1B).

To gain insight into intestinal microbial dynamics before, at the time of, and after HSCT, we obtained a total of 97 fecal samples from a subcohort of 30 patients. Using 16S rRNA gene sequence analysis, we identified 239 operational taxonomic units (OTUs) (see Methods). Microbial alpha-diversity was lower at all time points post-HSCT compared to pre-HSCT (Figure 1B). The median inverse Simpson index decreased from 3.27 (range: 1.02 – 7.4) before HSCT to 2.89 (range: 1.04 - 10.77) on the day of transplantation and further to 2.03 (range: 1.0 – 16.51) post-HSCT (median of week +1 to +5). *Enterococcaceae* (*Firmicutes*) was the most abundant bacterial family observed, followed by other *Firmicutes*, such as three families from the class of *Bacilli* (*Lactobacillaceae*, *Streptococcaceae*, *Staphylococcaceae*), two families from the class of *Clostridia* (*Lachnospiraceae*, *Ruminococcaceae*), a family within the class of *Erysipelotrichia* (*Erysipelotrichaceae*) and a family within the class of gamma-*Proteobacteria* (*Enterobacteriaceae*) (Figure 1C).

Associations between immune markers and immune cell reconstitution in HSCT patients

In order to identify patient baseline parameters and clinical outcomes (e.g. aGvHD, relapse, overall survival), as well as immune markers and immune cell types that might be important determinants in HSCT in relation to the microbiome, we performed variable assessment by permutational multivariate analysis of variance (adonis). The variables that were found to be significant ($P \leq 0.05$), i.e., those that explained most variation in the microbial community distance matrix, were selected for subsequent analyses (Additional file 3: Figure S2, Additional file 4: Table S2). Of note, the occurrence of relapse as an indication of transplantation outcome was assessed but not found to be significant in adonis and was therefore not included in follow-up analyses. We then assessed associations of the selected immune markers and immune cells in the data set comprising 37 patients by determining Spearman's rank correlations. The hBD2 concentrations pre-HSCT, on the day of HSCT and in weeks +1 and +2 post-HSCT were positively correlated with each other ($\rho = 0.73$ - 1, $P < 0.001$) (Figure 2A). NK cell counts in month +1 exhibited a positive correlation with total B cell counts ($\rho = 0.64$, $P = 0.0046$) and mature B cells counts ($\rho = 0.62$, $P = 0.0114$) in month +2. When we related immune cell reconstitution to outcomes, we observed significantly higher NK cell counts and total B cell counts in month +2 in patients with no or mild aGvHD (grade 0-I) compared with patients with moderate to severe aGvHD (grade II-IV) (Wilcoxon rank sum test, NK cells, $P = 0.011$; B cells, $P < 0.001$) (Figure 2B).

High plasma hBD2 and monocytes prior to HSCT in patients with high *Lactobacillaceae*

To gain insight into how the selected immune markers and immune cell counts co-vary with gut microbial abundances in patients with distinct outcomes, we implemented two multivariate multi-table approaches for the subcohort (*n=30*), namely sparse partial least squares (sPLS) regression and canonical correspondence analysis (CCpnA). The sPLS regression models OTU abundances as predictors and clinical variables as response variables and explains the latter in an asymmetric (i.e. unidirectional) way. In contrast, the CCpnA assesses relationships between parameters of the immune system and microbiota bidirectionally. In the following paragraphs, the results of these two analyses are reported for one observed cluster at a time, respectively.

First, we performed sPLS regression to reveal multivariate correlation structures between immune markers, immune cell counts, and OTU abundances, modeling the latter as explanatory variables. The sPLS regression and subsequent hierarchical clustering suggested that the data separated into three clusters (Figure 3A). High monocyte counts and high plasma hBD2 concentration prior to HSCT, high patient age at the time of transplantation, and high abundances of *Lactobacillaceae* independent of time point contributed the most to the formation of cluster 1. Of note, monocyte counts and hBD2 were positively correlated with each other, in agreement with the correlation analysis above (Figure 2A). The *Lactobacillaceae* were represented by microaerophilic *Lactobacillus* sp. OTUs (e.g., AF413523.1, GU451064.1, KF029502.1) (Figure 3B, Additional file 5: Table S3). These OTUs exhibited high loading weights in sPLS dimension 2 (Figure 3C), indicating that they contributed strongly to the separation of clusters in dimension 2 (Figure 3A and 3B).

Second, we applied CCpnA to model the canonical relationships between OTU abundances and clinical variables through the construction of common “latent” variables. The CCpnA confirmed the separation of the data into three clusters as observed in the sPLS regression (Figures 3A, 4, and Additional file 6: Figure S3A), including the clustering of OTUs. In addition, the CCpnA facilitated the inclusion of categorical variables, such as the patients’ baseline parameters (e.g., recipient sex, donor type) and clinical endpoints (aGvHD grade, overall survival). Because CCpnA is an unsupervised method and upheld the results of the sPLS regression, it provides confidence in the cluster findings.

Cluster 1 in the CCpnA seemed to include patients who developed moderate to severe aGvHD (grade II - IV) and who died. As suggested by both, sPLS and CCpnA, these

patients exhibited high levels of plasma hBD2 before HSCT (and in week +1 and +2) and high monocyte counts before HSCT. OTUs within this CCpnA cluster predominantly included members of the family *Lactobacillaceae* and were most abundant in fecal samples of these patients (Figure 4 and Additional file 7: Figure S4). OTUs that were assigned to cluster 1 by the sPLS-based hierarchical clustering were congruently associated with the same clinical variables in the CCpnA (Figures 3B and 4).

Temporal patterns of the *Lactobacillaceae*-dominated community state type

Upon revealing an association between high plasma hBD2 concentrations and monocyte counts prior to HSCT, moderate to severe aGvHD, and high mortality with high abundances of *Lactobacillaceae* in multivariate analyses, we assessed in more detail the importance of longitudinal changes in these components. We implemented an additional approach to identify distinct bacterial community patterns by employing partitioning around medoid (PAM) clustering (see Methods). In this analysis, we detected four community state types (CSTs). Dominating taxa, similar to those identified by the sPLS-based hierarchical clustering, were revealed in the CSTs, e.g., *Lactobacillaceae* members dominated CST 1 (Additional file 5: Table S3 and Additional file 8: Figure S5). We then used this information to examine temporal community state changes in individual patients, i.e., their transitions between CSTs over the course of time (Figure 5). Based on clinical outcomes, the patients can be divided into four groups: 1. Patients who developed no or mild aGvHD (aGvHD grade 0-I) and survived compared to 2. Patients who died, and 3. Patients who developed moderate to severe aGvHD (grade II-IV) and survived compared to 4. Patients who died (Figure 5). We observed that 6 out of 8 patients (75%) with moderate to severe aGvHD (groups 3 and 4) harboured the *Lactobacillaceae*–dominated CST 1 at least once during the monitored period (1x: *n=2*, 2x: *n=2*, 3x: *n=2*) (Figure 5). In comparison, only 5 out of 22 patients (23%) with aGvHD grade 0-I (groups 1 and 2) carried CST 1 one or two times (Figure 5). Interestingly, high abundances of *Lactobacillaceae* in patients with aGvHD grade II-IV (groups 3 and 4) occurred predominantly at late time points (week +1 and later) (Figure 5 and Additional file 7: Figure S4).

Temporal association of *Lactobacillaceae* with aGvHD and immune markers

To relate temporal changes in immune markers to those in the bacterial community composition in patients with aGvHD grade II-IV who died (group 4), we assessed their individual longitudinal profiles (Figure 6). In all three patients (P24, P26, P30), a large expansion of *Lactobacillaceae* abundances after the onset of aGvHD was observed. The average relative abundance of *Lactobacillaceae* after aGvHD onset was 72.92% (range 0.22 – 97.04%), as compared to before aGvHD (average 9.88%, range 0.25 – 37.83%) (Figure 6). Furthermore, bacterial alpha diversity was lower at the time point after aGvHD onset compared to the time point before (Figure 6). All three patients

were treated with antibiotics for different durations between these two time points and prior to the time point before aGvHD onset.

In agreement with the results of the multivariate analyses, two of the patients (P24 and P26) exhibited between 1.95 and 14.56 times higher plasma hBD2 concentrations for all measured time points compared with the average of the whole data set (average plasma hBD2 concentration: 10,983.9 pg/ml, range: 0-177,400.28 pg/ml). The plasma hBD2 concentration was the highest at the time of HSCT in patient P26 and before HSCT in patient P24 (Figure 6). Of note, patient P24 received corticosteroid-based GvHD prophylaxis prior to HSCT and both patients received anthymocyte globulin as part of their conditioning regimen. These two patients had underlying malignant diseases and their pre-HSCT CRP levels, providing insight into underlying inflammation, were 1.6 times lower and 1.29 times higher compared with the average of the whole data set on that time point (average plasma CRP concentration before HSCT: 19.21 mg/L: , range: 1 – 60.36 mg/L), respectively. Monocyte counts before HSCT, i.e., recipient-derived cells, were 6.25 times higher in patient P24 compared with the average of the whole data set (average monocyte count before HSCT: 0.45 x 10^9/L, range: 0.08 – 2.83 x 10^9/L) and close to average in patient P26. In both patients, monocyte counts were higher before (i.e., recipient-derived cells) than after HSCT (i.e., donor-derived cells) (Figure 6).

In contrast to the *Lactobacillaceae* expansion after aGvHD onset in group 4, we observed high *Lactobacillaceae* abundances already before aGvHD onset in three survivors (P16, P22 and P29) who developed aGvHD grade II-IV (group 3) (Additional file 9: Figure S6). The average relative abundance of *Lactobacillaceae* before aGvHD onset was 53.18% (range 0.35 – 98.86%), as compared to after aGvHD (average 18.25%, range 0 – 54.78%). However, these observations were limited by a small number of patients per outcome group (groups 3 and 4) (Figure 5) and therefore cannot serve to provide statistical evidence.

High NK and B cells and no or mild aGvHD in patients with high obligate anaerobes such as *Ruminococcaceae*

The association between NK and total B cell counts after HSCT observed in the Spearman's correlation analysis (Figure 2) was supported by the multivariate analyses (Figures 3 and 4). In the sPLS regression, cluster 2 included high NK cell counts in months +1 and +2, as well as high immature, mature, and total B cell counts in month +2 (Figure 3A). Furthermore, NK and B cell counts were positively correlated with OTUs found in cluster 2 (Figure 3A and B), in particular with members of the family *Ruminococcacea* (Order: *Clostridiales*) (Figure 3B, Additional file 5: Table S3). Strong positive correlations between mature and total B cell counts in month +2 and NK cell counts in month +1 were observed, for example, with *Faecalibacterium sp.*

(DQ804549.1) (Figures 3B and 3C, Additional file 5: Table S3). Additionally, we observed positive associations of these immune cells with *Lachnospiraceae* (Order: *Clostridiales*), among those two *Blautia* spp. (DQ800353.1 and DQ802363.1) (Figures 3B and 3C, Additional file 5: Table S3).

In support of the sPLS regression, the CCpnA also revealed high NK and B cell counts in cluster 2 in months +1 and +2. Based on the CCpnA, we found that NK and B cell reconstitution was associated with certain disease outcomes. For example, patients who had no or mild aGvHD and survived were predominantly represented in cluster 2. Cluster 2 included OTUs mainly belonging to *Ruminococcaceae* and *Lachnospiraceae*, exhibiting their highest abundances in patient samples associated with this cluster. OTUs with the highest scores in dimension 2 predominantly belonged to the family of *Ruminococcacea* and matched the sPLS-based hierarchical cluster assignment (Figure 4).

As all patients received antibiotic treatment prior to and post transplantation, we examined the potential effect of antibiotics on the intestinal bacterial community composition. A trending influence of the vancomycin treatment was revealed by adonis analysis ($P = 0.055$). Using a CCpnA model that also included information about antibiotic treatments, we found that patients exhibited higher abundances of *Ruminococcaceae* and *Lachnospiraceae* and lower abundances of *Enterobacteriaceae* at time points without the vancomycin treatment compared to with vancomycin (Additional file 6: Figure S3B). The same pattern was observed for treatment with ciprofloxacin (Additional file 6: Figure S3B).

Community state typing also revealed a state type dominated by *Ruminococcaceae* and *Lachnospiraceae* family members (CST 2) (Additional file 8: Figure S5). A total of 10 out of 22 patients (45%) with no or mild aGvHD were assigned to CST 2 at least at one time point (1x: *n=6*, 2x: *n=3*, 3x: *n=1*) (Figure 5). Half of the patients (4 out of 8) with moderate or severe aGvHD were also assigned to CST 2 once, but only at the time point before transplantation (Figure 5). That is, a *Ruminococcaceae*- and *Lachnospiraceae*-dominated community only persisted in patients with no or mild aGvHD.

Persistence of *Enterococcaceae*-dominated community state type

A subcluster of cluster 2 comprised two facultative anaerobe *Enterococcus* spp. (GQ1330038.1 and AJ272200.1), exhibiting positive correlations with high NK and B cell counts (Figure 3B) and contributing to the separation of the clusters (Figure 3C). *Enterococcaceae* was the most abundant family in the overall study population (Figure 1D), and community state typing revealed a state type predominantly characterized by *Enterococcaceae* (CST 4, Additional file 8: Figure S5). A total of 9 out of 22 patients

(41%) with no or mild aGvHD were assigned to CST 4 at least at one time point (1x: *n=2*, 2x: *n=2*, ≥3x: *n=5*) and this *Enterococcaceae*-dominated CST often persisted in the patient over time (Figure 5). A quarter of the patients (2 out of 8, both survivors) with moderate or severe aGvHD were assigned to CST 4 at least at one time point on the day of or post-HSCT (Figure 5).

High inflammation in patients with high facultative anaerobic bacteria

In multivariate analyses, OTUs belonging to facultative anaerobic *Enterobacteriaceae* and *Staphylococcaceae* were characteristic for sPLS cluster 3 (Figures 3A and 3B, Additional file 5: Table S3). Cluster 3 was further comprised of high plasma citrulline concentrations pre-HSCT (before conditioning) and week +1, high monocyte counts in week +3, high CD4+ T cell counts in month +2, high CRP levels, particularly in weeks +1, +5, and +6 (Figures 3A and 3B). The projection of the variables in the sPLS suggested a weak positive correlation between these clinical variables in dimension 1 (Figure 3A). However, a weak negative association between CRP levels and monocytes (week +3) as well as citrulline and CD4+ T cell counts is indicated in dimension 2 (Figure 3A), which is in agreement with some results of the Spearman's rank correlation tests (Figure 2).

The clinical variables in cluster 3, in particular high CRP levels post HSCT, exhibited positive correlations with the OTUs predominantly affiliated with *Proteobacteria* (e.g., *Enterobacteriaceae*), *Bacteroidetes*, and *Staphylococcus* spp. (Figure 3B, Additional file 5: Table S3). The strongest positive correlations occurred with several facultative anaerobe bacteria, e.g., *Enterobacter* sp. LCR81 (FJ976590.1), *Escherichia coli* (FJ950694.1) and *Staphylococcus* sp. (JF109069.1). The CCpnA supported observations from the sPLS regression regarding cluster 3, but indicated that they were only represented by a few patient samples (Figure 4 and Additional file 6: Figure S3A). These samples were characterized by high CRP levels, especially in week +1, +5, and +6. High CRP levels at these time points were not associated with aGvHD grade, i.e. they were not higher in patients with either aGvHD grade 0-I, or grade II-IV.OTUs of cluster 3, i.e., members of *Enterobacteriaceae and Staphylococcaceae,* exhibited their highest abundances in samples of this cluster (Figure 4 and Additional file 6: Figure S3A). Of note, patients represented in cluster 3 were younger compared to those in the other two clusters (Figure 4).

Discussion

The human immune system and host-associated microorganisms are closely interlinked and play central roles in health and disease. The underlying components and mechanisms facilitating interactions between the immune system and microorganisms are however not completely understood. Understanding these

associations is particularly relevant in patients that receive components of a "foreign" immune system, such as in patients undergoing allogeneic HSCT. Because both the patients' immune system and microbiome appears severely affected, they potentially jointly impact on clinical outcomes. Here, we perform an integrated analysis of immune markers, immune reconstitution data, clinical outcomes, and microbiota, and we provide evidence for the association between specific microbial taxa, host immune markers, immune cells and clinical outcomes.

We observed that a predominance of *Clostridiales*, represented by *Ruminococcaceae* and *Lachnospiraceae,* in the intestine did not persist after transplantation in patients who either developed aGvHD grade II-IV, died, or both. *Clostridiales* are common colonizers of the healthy distal gut [23], and their loss is associated with microbial community disruption, reduced diversity, and GvHD [14]. A low diversity before conditioning and at the time of engraftment is associated with increased mortality [4, 5], in line with our findings.

In addition, in our multivariate analyses we observed high plasma hBD2 levels before HSCT and in weeks +1 and +2 post HSCT in patients who developed aGvHD grade II-IV, died, or both. The reason for the high hBD2 levels before HSCT in patients with increased mortality and GvHD is unknown at present. One could speculate that it relates to a higher burden of inflammation before and during the first two weeks post-transplantation. This was also reflected in higher counts of (recipient-derived) monocytes pre-HSCT that, by secreting hBD2, might contribute to an inflammatory reaction and thereby a potentially higher risk of aGvHD after HSCT. High hBD2 concentrations in weeks +1 and +2 could be an indicator for an innate immune reaction involving donor-derived cells, for example, against opportunistic pathogens that may have translocated to the bloodstream [24]. This might have been promoted by the decrease in *Ruminococcaceae* and *Lachnospiraceae* abundances in these patients after HSCT, indicating a microbial community disruption. Plasma hBD2 concentrations were characterized for the first time here in HSCT patients, and our findings emphasize the importance for further investigations to exploit the potential of hBD2 as a novel candidate marker for outcomes after HSCT. Importantly, our data suggested that differences in hBD2 secretion levels between patients were highly dependent on both, the microbial community composition and the time point relative to HSCT. We suggest that further investigations take into account variations of microbial community patterns and temporal changes when assessing hBD2 in the HSCT context. Moreover, these findings may be refined with a sampling time point homogeneity that is higher than what we provide in this study.

An interesting observation was that high pre-HSCT hBD2 levels and monocyte counts were associated with high abundance of *Lactobacillaceae* independent of time point. Probiotic *Lactobacillus* spp. have previously been shown to enhance hBD2-secretion

in immune cells, thereby contributing to the innate immune defense [25, 26]. This mechanism could play a protective role during blood stream infections by opportunistic pathogens in HSCT patients. The increase of *Lactobacillaceae* abundances, particularly after the onset of aGvHD in patients who died, may be explained by a previously proposed compensatory mechanism to reduce aGvHD severity after onset in mice and humans [14]. For instance, high abundances of *Lactobacillaceae* could indicate homeostasis in the gut microbiome of children and thereby prevent inflammation caused by opportunistic pathogen expansion [27, 28]. Reduced intestinal inflammation might then benefit the outcome of aGvHD. This might however also lead to a less effective graft-versus-leukemia (GvL) effect as suggested by the finding that all patients, for which we observed moderate to severe aGvHD and subsequent increased *Lactobacillaceae* post HSCT, overcame aGvHD, but died following a relapse. Only a few patients in our study represented this combination of outcomes (moderate to severe aGvHD with subsequent increase of *Lactobacillaceae* and death), therefore no significant conclusion can be drawn at this point and further studies are needed to address these observations in more detail.
We provide a discussion on survival following high *Lactobacillaceae* abundances prior to the onset of aGvHD in Additional file 10: Supplementary discussion.

In contrast to patients with moderate to severe aGvHD, we observed a higher overall survival in patients with no or mild aGvHD in multivariate analyses. The latter patients also had increased numbers of NK and B cells. In agreement with our study, previous studies have shown slower NK cell reconstitution after HSCT in patients with moderate to severe aGvHD compared to those with no or mild aGvHD [29]. Moreover, our results are in agreement with the previously described association of low NK cell numbers after HSCT and reduced overall survival [30], in line with NK cells' crucial role in the GvL effect [7]. A poor recovery of B cells has been found to pose an increased risk of late infections [31], which might be an explanation for the association of high B cell counts and lower survival. The lower B cell numbers we observed in patients with more severe aGvHD could partially be a consequence of aGvHD or the treatment with corticosteroids, which is known to reduce the number of B cell precursors [21, 29]. However, based on our findings, the intestinal microbiota could also play a contributing role, because faster NK and B cell reconstitution as well as lower aGvHD severity, and higher overall survival were associated with high abundances of obligate anaerobes belonging to *Ruminococcaceae* and *Lachnospiraceae*. Indeed, decreased abundances of these bacterial families have previously been associated with an overall microbial disruption and decreased diversity [5, 15, 32]. In line with our observations, reduced microbial diversity was shown to contribute to lower survival [4] and a reduction of *Clostridiales* was observed in patients presenting aGvHD [33]. Furthermore, in melanoma patients, a low diversity and decreased *Ruminococcaceae* abundance were associated with a poor response to immunotherapy [34]. A potential

explanation for this association could be that members of the order *Clostridiales* can downregulate inflammation and might thereby prevent aGvHD. Anti-inflammatory components produced by *Clostridiales* include, for example, urinary 3-indoxyl sulfate (3-IS) [35] and butyrate [36]. Acute GvHD might also reinforce microbial community disruption, as it is known to be accompanied by a reduction of Paneth cell numbers in the intestine. Paneth cells are secretors of α-defensins, important modulators of gut microbial homeostasis [37].

In contrast to the patient group with high abundances of obligate anaerobic bacteria (e.g., *Ruminococcaceae* and *Lachnospiraceae*), patients with high abundances of facultative anaerobic bacteria (e.g., *Enterobacteriaceae*, *Staphylococcus* spp., and *Streptococcus* spp.), showed slow NK and B cell reconstitution. An increase in facultative anaerobic bacteria has previously been observed in the gut microbiome of HSCT recipients before the start of conditioning compared with donors [5]. Here, we additionally observed high levels of C-reactive protein (CRP), indicating high inflammation in these patients. A possible explanation for this association might involve the shift to microbial growth conditions that favor facultative anaerobic bacteria during intestinal inflammation, such as an increased availability of oxygen caused by inflammatory products [38]. Our multivariate analyses indicated that the patients in which we observed these associations were younger compared to the rest of the cohort. Therefore, one could speculate that an immature intestinal microbiome might exhibit a higher susceptibility to opportunistic growth of facultative anaerobic bacteria. Further investigations will have to elucidate this relation and its potential consequences for adjustments of monitoring and treatment by age.
We provide a discussion on our findings regarding associations of adverse outcomes with *Enterococcus* compared with previous studies in Additional file 10: Supplementary discussion.

Antimicrobial treatment of patients can significantly affect the gut microbiota, especially in children [39]. Early use of antibiotics in general has been found to reduce *Clostridiales* in the intestine of HSCT patients [40]. Furthermore, a link of high total amounts of antibiotic in GvHD development has been observed in pediatric stem cell recipients [41]. However, to our knowledge, the effects of specific antibiotics on the intestinal microbiome especially in pediatric HSCT patients have not yet been elucidated in detail. Here, we identified a number of specific antibiotics, including vancomycin and ciprofloxacin, associated with simultaneous reduction of *Clostridiales* (in particular *Ruminococcaceae* and *Lachnospiraceae*), similar to what has previously been observed in adult patients [42]. Most interestingly, we found that treatment with vancomycin and ciprofloxacin was not only associated with reduced *Clostridiales*, but also with increased abundances of facultative anaerobic bacteria, e.g., *Enterobacteriaceae* (gamma-*Proteobacteria*). Vancomycin-treatment in a cohort of

563 rheumatoid arthritis patients was previously shown to be associated with an
564 expansion of *Proteobacteria* [43]. The use of a prophylactic ciprofloxacin treatment to
565 prevent chemotherapy- and transplantation-related bloodstream infections is an
566 established method [44], and it was proposed that fluoroquinolones (the antibiotic
567 class comprising ciprofloxacin) can prevent intestinal domination of *Proteobacteria* in
568 HSCT patients [15]. However, ciprofloxacin treatment in healthy subjects has been
569 associated with decreased microbial diversity and decreased *Ruminococcaceae* and
570 *Lachnospiraceae* abundances [45], indicating microbial community disruption.
571 Therefore, our findings further challenge the choice of antibiotics, such as vancomycin
572 and ciprofloxacin, in patients undergoing chemotherapy and HSCT. Interestingly, in
573 the present study a number of less frequently used antibiotic agents, e.g. ceftazidime
574 (a cephalosporin), showed positive associations with high *Clostridiales* abundances. In
575 agreement, another cephalosporin (cefepime) has previously been attributed
576 clostridial sparing effects [46]. However, ceftazidime was associated with reduced
577 bacterial alpha diversity to a similar degree as vancomycin and ciprofloxacin in a
578 previous study [42]. Elucidating the effects of antibiotic agents potentially
579 contributing to maintaining gut microbial homeostasis in HSCT therefore require
580 further investigation.Of note, we did not take the mode of application of the
581 antibiotics into account here, which might limit the strength of our conclusion. It
582 remains to be determined whether certain antibiotics modulate the patients'
583 microbiome and how this might lead to either positive or adverse clinical outcomes.
584

585 ## Conclusions

586 Our findings support the increasing evidence of microbial involvement in the context
587 of HSCT in cancer patients. We provide evidence for the association between specific
588 microbial taxa and host immune markers. In particular, we examined the prognostic
589 potential of immune markers and gut microbial community dynamics for immune
590 reconstitution and outcomes after HSCT by revealing multivariate associations. We
591 observed increased human beta-defensin 2 in patients with moderate to severe
592 aGvHD and high mortality. In those patients, NK and B cell reconstitution was slow
593 compared to patients with low mortality. These associations only applied when
594 distinct gut microbial abundance patterns were observed, namely low abundances of
595 *Ruminococcaceae* and high abundances of *Lactobacillaceae*. Therefore, hBD2, in
596 connection with longitudinal microbial community pattern surveillance, could be
597 further evaluated as a potential novel candidate marker to identify patients at risk of
598 adverse outcomes (e.g. aGvHD) and slow immune cell reconstitution after HSCT,
599 contributing to improved clinical outcomes. Of note, our cohort comprised a relatively
600 small number of patients with different primary diseases and conditioning regimens,
601 and our findings would therefore benefit from being assessed in larger, more
602 homogenous patient groups. Importantly, microbial abundances also depended on

603 antibiotic treatment. Our findings suggest that certain antimicrobial agents might
604 contribute to a shift from obligate to facultative anaerobes. This highlights the need
605 to assess the usage of specific antibiotics in more detail and to take antibiotic
606 treatment into consideration when describing microbial communities in HSCT
607 recipients.

608

609 ## Methods

610 **Patient recruitment and sample collection**

611 We recruited 37 children (age range 1.1 - 18.0 years) undergoing their first
612 myeloablative allogeneic hematopoietic stem cell transplantation at Copenhagen
613 University Hospital Rigshospitalet, Denmark, from June 2010 to September 2012.
614 Patients' clinical characteristics are listed in Additional file 1: Table S1. Further
615 information can also be found in previous studies where the cohort has been
616 examined in relation to other questions [17, 47–49]. All patients received
617 pretreatment with a myeloablative conditioning regimen, starting on day -7
618 (Additional file 1: Table S1). Four patients were re-transplanted at day +157, +518,
619 +712 and +1360 after the first transplantation, respectively.
620 Sampling time points were defined according to the following intervals: pre-HSCT
621 (collected between day -33 and day -3), at the time of HSCT, preferably before graft
622 infusion (collected between day -2 day +2) and weekly during the first six weeks after
623 transplantation (week +1: day +3 to day +10, week +2: day +11 to day +17, week +3:
624 day +18 to day +24, week +4: day +25 to day +31, week +5: day +32 to day +38, week
625 +6: day +39 to day +45) (Figure 1A). Broader intervals applied to follow-up time points:
626 Month +1 (between days +21 and +44), month +2 (between days +45 and +70), month
627 +3 (between days +77 and +105), month +6 (between days +161 and +197) and 1 year
628 post transplantation (between days +346 and +375).

629

630 **Infections and antibiotics**

631 Bacterial infections from before transplantation until 1 year post-HSCT were taken
632 into consideration for downstream analysis. For each time point, it was recorded
633 whether any bacterial infection occurred within the respective specified interval or
634 not (1/0). Antibiotic treatment from before HSCT (from day -90) until month +2 was
635 taken into consideration. We included only those time points corresponding to the
636 time points of microbiota profiling into downstream analyses.

637

638 **Analysis of T, B and NK cells in peripheral blood**

639 T, B and NK cell counts were determined in months +1, +2, +3, and +6 post
640 transplantation. Lymphocyte subsets in peripheral blood were quantified using
641 Trucount Tubes (Becton Dickinson, Albertslund, Denmark) together with the following
642 panel of conjugated monoclonal antibodies and analysed on a FC500 flow cytometer

(Beckman Coulter, Copenhagen, Denmark): CD3-PerCP, CD3- FITC, CD4-FITC, CD8-PE, CD45-PerCP, CD16/56-PE, CD20-FITC and CD19-PE (Becton Dickinson). CD3+ T cells, CD3+CD4+ T cells and CD3+CD8+ T cells were determined. NK cells were differentiated by CD3-CD45+CD16+CD56+ phenotype. The following B cell phenotypes were distinguished: total B cells (CD45+CD19+), mature B cells (CD45+CD19+CD20+) and immature B cells (CD45+CD19+CD20-). Data of these immune cell populations have been published previously in a different context [47].

Analysis of monocytes and neutrophils

Leukocyte numbers and subsets were monitored daily during hospitalization and subsequently every week in the outpatient clinic using flow cytometry (Sysmex XN) or, in case of very low leucocytes, counted by microscopy (CellaVision DM96 microscope). Mean monocyte and neutrophil counts were calculated for further analysis per time point according to the intervals specified above.

Quantification of inflammatory markers

EDTA-anticoagulated and heparinized blood was sampled and then centrifuged within 2 hours after collection. The plasma was isolated and cryopreserved in 0.5-ml aliquots at −80°C. IL-6 levels on day +7 were determined in EDTA-anticoagulated plasma using the Human Th1/Th2/Th17 Cytometric Bead Array kit (Becton, Dickinson and Co., Denmark) and a FACSCalibur flowcytometer (Becton, Dickinson and Co), according to the manufacturer's instructions with a detection limit of 2.5 pg/mL. IL-6 data have been published previously in another context for a larger cohort than the patients included here [17, 48]. CRP levels were measured daily by Modular P Modular (normal range, 0 to 10 mg/L) at the Department of Clinical Biochemistry, Copenhagen University Hospital Rigshospitalet, Denmark. Mean CRP levels were calculated for further analysis per time point according to the intervals specified above. Mean CRP levels pre-HSCT include measurements from day -7 to day -3, i.e. the days after the start of conditioning (except for day -7).

Quantification of Citrulline

Plasma citrulline concentrations pre-HSCT (before the start of conditioning (day -7)) and at days +7 and +21 were measured by reverse-phase high-performance liquid chromatography of their phenylisothiocyanate derivatives from heparinized plasma, as described previously [17, 48]. Citrulline levels have previously been described for patients of this cohort in a different context [17, 48].

Enzyme-linked immunosorbent assay (ELISA) of human beta defensins

Human beta defensin 2 and 3 (hBD2 and hBD3) concentrations in heparinized plasma samples of 37 patients at eight time points (pre-HSCT (before start of conditioning, except for 4 patients sampled at day -6 or -5), on the day of transplantation, weeks +1

to +4, month +2 and +3) and 10 healthy controls (sampled once each) were quantified by two-step sandwich ELISA following the manufacturer's instructions (Peprotech Human BD-2 and BD-3, Standard ABTS ELISA Development Kit, cat.no. 900-K172 and 900-K210, respectively). Samples of 3 out of 10 healthy individuals were additionally spiked to a peptide concentration of 1000 pg/ml. Two replicates were measured per sample and their mean was used for further analysis. Samples were measured undiluted as well as in 1:4 and 1:16 dilutions to also cover concentrations potentially exceeding the upper detection limit. Samples with very high concentrations were additionally measured in 1:32 and 1:128 dilutions. Detection limits were 16 – 2000 pg/ml for hBD2 and 31– 4000 pg/ml for hBD3. Absorbance was measured on a VICTOR™ X3 Multilabel Plate Reader (Perkin Elmer, Inc., USA) at 405 nm. Wavelength correction at 540 nm was used to prevent optical interference caused by the material of the microtiter plate. Concentrations of hBD3 were mostly below the limit of detection, except for a few exceptionally high measurements (average: 1450.69 pg/ml, median: 0 pg/ml, range: 0 – 279038.71 pg/ml).

DNA isolation from fecal samples and 16S rRNA gene sequencing

Fecal samples for analysis of the intestinal microbiome were collected from a subset of 30 patients at 7 time points: pre-HSCT (5 patients were sampled after the start of conditioning (between day -6 and day -4)), at the time of HSCT and once weekly during the first five weeks after transplantation. The intestinal microbiome was characterized at 1-2 time points in 8 patients (27%), at 3-4 time points in 15 patients (50%) and at 5-6 time points in 8 patients (27%) (Additional file 1: Table S1). In patients who underwent re-transplantation, no feces samples collected after the second transplantation were included in this study. In total, 97 fecal samples were obtained.

DNA from fecal samples and a blank control were isolated with the use of the Maxwell 16 Instrument (Promega Corporation) following the manufacturer's instructions for the low elution volume blood DNA system. Alterations to the protocol included additional lysozyme treatment and bead beating with stainless steel beads for 2 min/20 Hz in a tissue lyser (Qiagen). In each sample including the blank control, the V4-V5 region of the 16S ribosomal RNA gene were amplified in PCR using the following barcoded primers: 519F (5#-CAGCAGCCGCGGTAATAC-3#) and 926R (5#-CCGTCAATTCCTTTGAGTTT-3#). Amplicons were then analyzed for quantity and quality in an Agilent 2100 Bioanalyzer (Agilent Technologies) with the use of an Agilent RNA 1000 Nano Kit. For library preparation, 50 ng of DNA from each sample was pooled with multiplex identifiers for 2-region 454 sequencing on GS FLX Titanium PicoTiterPlates (70675) with the use of a GS FLX Titanium Sequencing Kit XLR70 (Roche Diagnostics). Library construction and 454 pyrosequencing were performed at the National High-Throughput DNA Sequencing Centre, University of Copenhagen.

16S rRNA gene sequence pre-processing

Raw 454 sequence reads stored in standard flowgram format (SFF) were extracted, converted to and stored in FASTA format with associated quality files (containing sequence quality scores) using the *sffinfo* command of the bioinformatics software tool mothur [50]. Trimming according to the clipQualLeft and clipQualRight values provided by the sequence provider was disabled because cut-off values are opaque and not customizable.

Analysis was continued in the Quantitative Insights Into Microbial Ecology (QIIME; version 1.9.0) bioinformatics pipeline [51]. FASTA files were demultiplexed and quality filtered using the script *split_libraries.py* (Mapping files are available from figshare (https://dx.doi.org/10.6084/m9.figshare.6508250). As the samples were sequenced bidirectional, each FASTA file was demultiplexed in two steps. Firstly, based on a mapping file containing the 519F primer as the “LinkerPrimerSequence” and the 926R primer as the "ReversePrimer", both in 5' to 3' orientation. Secondly, based on a mapping file containing the 926R primer as the “LinkerPrimerSequence” and the 519F primer as the "ReversePrimer", again both in 5' to 3' orientation.

Reads between 200 and 1000 bp length and a minimum quality score of 25 were retained (default). Sequences with homopolymers longer than 200 bp were removed from the data set. Removal of reverse primer sequences (*-z truncate_only* option) was disabled during demultiplexing. Subsequently, the demultiplexed FASTA files that were not yet primer-truncated were then used to denoise flowgrams (.sff.txt files also generated by mothur’s *sffinfo*) with QIIME’s *denoise_wrapper.py* script. Reverse primer-truncation had not been done yet to ensure compatibility between FASTA and .sff.txt files. The denoised FASTA output files were then inflated, i.e., flowgram similarity between cluster centroids was translated to sequence similarity, to be used for OTU picking. Reverse primers and subsequent sequences in the demultiplexed and denoised FASTA files were then truncated using the *truncate_reverse_primer.py* script. In the following step, the orientation of the primer-truncated reads that started with the 926R primer as the “LinkerPrimerSequence” was adjusted by reverse complementation (with the script *adjust_seq_orientation.py*). All trimmed reads were then concatenated to a single file for further analysis.

Chimeras were identified using the script *identify_chimeric_seqs.py* and method *usearch61,* which performs both de novo (abundance-based) and reference-based detection (by comparing the dataset to the chimera-free reference database Ribosomal Database Project (RDP; training database version 15)). Only those sequences that were flagged as non-chimeras from both detection methods were retained (option *–non_chimeras_retention = intersection*). Operational taxonomic unit (OTU) clustering was performed, using the script *pick_otus.py* (based on the SILVA

766 database, *Silva_119_rep_set97)*. OTU tables in BIOM format were created with
767 *make_otu_table.py* (and subsequently converted to JSON BIOM format to be
768 compatible with analysis in R [52] with the package *phyloseq* [53]). The OTU table and
769 the taxonomy table are available from figshare
770 (https://dx.doi.org/10.6084/m9.figshare.6508187).
771
772 **Statistical analyses**
773 Statistical analyses and creation of graphs were performed with the program R
774 (Version 3.4.0, R Foundation for Statistical Computing, Vienna, Austria) [52]. All R
775 scripts documenting our statistical analyses are available from figshare
776 (https://dx.doi.org/10.6084/m9.figshare.6508238). Sequencing data and all related
777 experimental and clinical data (data sets available from figshare,
778 https://dx.doi.org/10.6084/m9.figshare.6508232) were integrated for analysis with
779 the R package *phyloseq* and its dependencies [53] (Additional file 11). The resulting
780 phyloseq objects are provided through figshare
781 (https://dx.doi.org/10.6084/m9.figshare.6508235). Plots were generated with the
782 packages *ggplot2* [54], *plotly* [55], and *mixOmics* [56]. Dose-response analysis of the
783 ELISA data was performed with four-parameter log-logistic models in the R package
784 *drc* [57].
785
786 Alpha diversity (measured by inverse Simpson index), levels of human beta-defensin
787 2 (hBD2) concentration, citrulline and C-reactive protein (CRP), as well as monocyte
788 counts, NK cell counts, total B cell counts, and CD4+ T cell counts at different time
789 points were compared using Friedman tests with Benjamini-Hochberg correction for
790 multiple testing (Additional files 3: Figure S2, Additional files 11 and 12). In addition,
791 Kendall's coefficient of concordance (Kendall's *W*) was calculated on ranked data for
792 each marker (Additional file 13). Like the Friedman test, the test for Kendall's *W*
793 allowed the comparison of marker levels between time points. In addition, the
794 coefficient of concordance informs about the level of agreement between patients.
795 Therefore, Kendall's *W* can be interpreted as a measure of effect size for the Friedman
796 tests. A Kendall's $W < 0.1$ was considered as indicating a small effect, 0.1 – 0.5 as a
797 moderate effect, and > 0.5 as a strong effect. As an exception, hBD2 in healthy controls
798 was compared with hBD2 in patients at individual time points by using Wilcoxon rank-
799 sum tests, because hBD2 was only measured once in the healthy control individuals
800 and can therefore not be analyzed in a Friedman test designed for repeated
801 measurements. Monocyte counts in Additional file 2: Figure S1B are depicted at more
802 time points than indicated in Figure 1A because not all time points were included into
803 further analyses. The day of HSCT, week +1 and week +2 were excluded as data were
804 missing for ≥40% of the patients. Months +2, +3, and +6, as well as 1 year were chosen
805 as representative follow-up time points for further analyses as indicated in Figure 1A.
806

A core set of OTUs was obtained by retaining 256 OTUs (out of 756 OTUs) with ≥5 reads in ≥ 2 samples using the function *kOverA()* from R package *genefilter* [58]. Subsequently, 17 OTUs that were more abundant in the blank control than in the majority of samples were removed as potential contaminants prior to downstream analyses. The resulting count data set of 239 OTUs was transformed for subsequent analyses using the function *varianceStabilizingTransformation()* in the package *DESeq2* [59] (Additional file 11). The function implements a Gamma-Poisson mixture model [60] to account for both library size differences and biological variability.
Median imputations were performed for continuous clinical and immune marker data with less than 20% missing values. Variables with more than 20% missing values were excluded from the analysis (Additional files 13 and 14). Central tendencies of immune cell counts at single time points in relation to clinical outcomes (maximum aGvHD grade 0-I vs. grade II-IV) were assessed in univariate analyses by Wilcoxon rank-sum tests and displayed in boxplots (Additional file 14).

A model selection procedure was implemented to find the relevant variables to be included in subsequent multivariate analyses of how microbiome patterns are associated with clinical outcomes, baseline parameters and immune parameters during the course of transplantation: A Manhattan distance matrix of the variance-stabilized bacterial community data was calculated using the *distance()* function in *phyloseq* [53]. Subsequently, permutational multivariate analysis of variance using distance matrices (adonis) for model selection was performed by applying the *adonis2()* function in the package *vegan* [61] (Additional files 3: Figure S2, Additional file 15). Permutation design was set up with respect to repeated measurements within the same patients and the intact chronological order of time points. Besides immune marker levels and immune cell counts (pre- and post-HSCT, i.e. recipient- and donor-derived cells, respectively) at the time points described above, we included clinical outcomes (i.e. overall survival, aGvHD (grade 0-I vs II-IV), and relapse) after transplantations, antibiotic treatment during the course of transplantation and clinical patient characteristics in the model to account for possible effects of recipient age at the time of transplantation, recipient sex, donor type (sibling vs. unrelated), malignant vs. benign diagnosis, graft type (stem cell source: bone marrow, umbilical cord blood or peripheral blood) and application of irradiation therapy (yes/no). Variables that were found to be significant ($P \leq 0.05$) in the adonis analysis were included in subsequent multivariate multi-table analyses, i.e., sparse partial least squares (sPLS) regression and canonical correspondence analysis (CCpnA) (Additional file 15). Choosing variables with significant effects in adonis for follow-up statistical testing has been performed previously [62]. Even though validation of the set of selected variables through a data-splitting approach might be preferable, this was not feasible due to our relatively small data set. To account for this and to avoid post-selective

847 inference, we renounce calculation of p-values from the two analyses that directly
848 depend on the pre-selection (sPLS and CCpnA).
849
850 Correlations among the selected clinical variables were assessed in correlation
851 matrices based on Spearman's rank correlation tests (Additional file 3: Figure S2,
852 Additional file 14). Matrices were calculated using the *rcorr()* function of the R package
853 *Hmisc* [63] and displayed with the package *corrplot* [64]. *P*-values were calculated with
854 the *rcorr.adjust()* function with correction for multiple testing (method "Holm"). The
855 Spearman's rank correlation tests were performed on the set of variables selected
856 from the adonis analysis. However, here we assess correlations among those
857 variables, and not between variables and microbial abundances.
858
859 Sparse PLS regression was performed by applying the *spls()* function in the package
860 *mixOmics* [56] (Additional file 3: Figure S2, Additional file 15). The sPLS regression
861 allows the integration of the microbial community data matrix and the clinical variable
862 matrix for multiple regression. It is robust enough to handle collinearity and noise in
863 the data and is suitable to model multiple response variables [65]. The number of
864 clinical variables to be kept in the model for each component (*keepY*) was set to 23,
865 corresponding to the number of variables pre-selected with adonis. We ran the sPLS
866 regression with a range of numbers (20-40) of OTUs to be kept for each component
867 (*keepX*). As the results were robust to this choice, *keepX* was set to 30. The number of
868 components to choose was estimated with the *perf*() function and set to *ncomp* = 2.
869 The sPLS model was run in regression mode. Thereafter hierarchical clustering was
870 performed within the mixOmics *cim()* function based on the sPLS regression model
871 with the clustering method "complete linkage" and the distance method "Pearson's
872 correlation". Coefficients of pairwise correlations between OTU abundances and
873 clinical variables were thereby obtained. Furthermore, loading plots were generated
874 with the function *plotLoadings()* (method = "mean") to visualize loading vectors of
875 specific OTUs that contribute most to the separation of variables in components 1 and
876 2.
877
878 Canonical (i.e., bidirectional) correspondence analysis (CCpnA), a multivariate
879 constrained ordination method, was performed by using the *cca()* function in the
880 package *vegan* [61] (Additional file 3: Figure S2, Additional file 15). In this method, the
881 microbial community data matrix is Chi-square transformed and weighted linear
882 regression on pre-selected constraining variables is performed. The resulting fitted
883 values are used for correspondence analysis by singular value decomposition. CCpnA
884 is a constrained method in the sense that it does not aim at depicting all variation in
885 the data, but only the variation directly explained by the constraints (i.e., the provided
886 set of pre-selected variables). The resulting triplot is not displayed as a square
887 representation, but rather corresponds to the percentage of variance explained by

axis 1 and 2, respectively, as previously suggested [66]. In contrast to the sPLS analysis, the CCpnA was performed in canonical mode, i.e., modeling bidirectional relations between OTU abundances and clinical variables. OTUs with a correlation of >0.2/<-0.2 in the sPLS analysis were included in the CCpnA model.

As another approach to distinguish between microbial community states of the intestinal microbiome, we assigned samples to community state types (CSTs) by partitioning around medoid (PAM) clustering (function *pam()* in package *cluster* [67]) based on a Jensen-Shannon distance of the variance stabilized microbial count data (R code modified after [68]) (Additional file 3: Figure S2, Additional file 16). The number of clusters was determined by gap statistic evaluation and silhouette width quality validation. We further assessed patients' transitions between CSTs over time. OTUs were assigned to CST - based clusters (Additional file 4: Table S2) based on in which CST they exhibited the highest average abundance over all samples (within each CST). Furthermore, we showed detailed longitudinal profiles of the microbial community on family-level, and selected immune markers for individual patients with aGvHD (Additional file 17).

List of abbreviations

3-IS: 3-indoxyl sulfate
aGvHD: Acute graft-versus-host disease
ALL: Acute lymphoblastic leukemia
AML: Acute myeloid leukemia
AMP: Antimicrobial peptide CCpnA
Canonical correspondence analysis
CRP: C-reactive protein
CST: Community state typeELISA
Enzyme-linked immunosorbent assay
GvL effect: Graft-versus-leukemia effect
hBD2/hBD3: Human beta-defensin 2/3
HLA: Human leukocyte antigen
HSCT: Hematopoietic stem cell transplantation
IL-6: Interleukin-6
NK cell: Natural killer cell
OTU: Operational taxonomic unit
P: Patient
PAM clustering: Partitioning around medoid clustering
QIIME: Quantitative Insights Into Microbial Ecology
SFF: Standard flowgram format
sPLS regression: sparse partial least squares regression
TBI: Total body irradiation

Declarations

Acknowledgements

We thank the patients and their families for participating in the study, and the National High-Throughput DNA Sequencing Centre at the University of Copenhagen for library construction and sequencing. Furthermore, we are grateful for helpful discussions about the statistical analyses with Pratheepa Jeganathan and Kris Sankaran from the Department of Statistics at Stanford University. We thank Jeffrey Edward Skiby for reviewing the manuscript.

Funding

This work was supported by the European Union's Framework program for Research and Innovation, Horizon2020 (643476), and by the National Food Institute, Technical University of Denmark.

Availability of data and material

The 16S rRNA gene sequences are available through the European Nucleotide Archive (ENA) at the European Bioinformatics Institute (EBI) under accession number PRJEB25221. The datasets generated and/or analysed during the current study as well as the R code used to analyze the data are available from the figshare repository at https://figshare.com/projects/Specific_gut_microbiome_members_are_associated_with_distinct_immune_markers_in_allogeneic_hematopoietic_stem_cell_transplantation/35201.

Author's contributions

A.C.M., K.K., K.G.M., and S.J.P. designed the research; A.C.M., K.K., M.S.C., and S.J.P. performed the research; A.C.M., S.H., and S.J.P. contributed analytic tools; A.C.M., and S.J.P. analysed the data; A.C.M. and S.J.P. wrote the manuscript; and K.K., S.H., F.M.A., O.L., and K.G.M. edited the manuscript. All authors have read and approved the manuscript as submitted.

Ethics approval and consent to participate

Written informed consent was obtained from the patients and/or their legal guardians. The study protocol was approved by the local ethics committee (H-1-2010-009) and the Danish Data Protection Agency.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Additional files

Additional file 1: **Table S1. Clinical patient characteristics.** General patient characteristics, conditioning regimens, complications, and outcomes for the pediatric cohort (*n=37*) and the subcohort (*n=30*) for which the intestinal microbiome was characterized. Abbreviations: HLA, human leukocyte antigen; TBI, total body irradiation; CY, Cyclophosphamide; VP16, Etoposide; BU, Busulfan; MEL, Melphalan; GvHD, graft-versus-host disease. (PDF)

Additional file 2: **Figure S1. Temporal patterns of immune markers and immune cells in HSCT patients.** (A) C-reactive protein (CRP) and plasma citrulline levels in HSCT patients (*n = 37*) over time. CRP levels were significantly higher prior to HSCT and until week +2 compared to all following time points, e.g. at the day of HSCT (median: 16.93 mg/L, range: 1.22 - 85.28 mg/L) compared to week +3 (median: 3.92 mg/L, range: 1.22 - 55.89 mg/L) ($P < 0.001$). Plasma citrulline levels were significantly lower in week +1 compared to pre-HSCT ($P < 0.001$) and week +3 ($P < 0.001$). (B) Immune cell counts in HSCT patients over time. Monocyte counts are depicted at more time points than indicated in Figure 1A, because not all time points were included into further analyses (see Methods). NK cell counts were higher in months +2 to +6 compared to in month +1 ($P < 0.001$). B cell counts as well as CD4+ T cell counts increased steadily from month +1 to month +6 ($P < 0.001$). Y -axes in all plots, except for citrulline, were log10-transformed for better visualization. Zeros were replaced with 1 to avoid undefined values on the log-transformed axes. Asterisks indicate whether the component at each respective time point was significantly different from any of the other time points (showing the maximum significance level). * $P < 0.05$, ** $P < 0.01$ and *** $P < 0.001$. (PDF)

Additional file 3: **Figure S2. Workflow of the statistical analysis approach.** The diagram displays the major steps of the statistical analyses and their dependencies. Multivariate analyses (blue box) constitute the main approach, especially the multi-table analyses and clustering analyses (green box). To unravel the complexity of the multivariate analyses, these were supplemented with univariate analyses (upper grey box). (PDF)

Additional file 4: **Table S2. Results of Permutational Multivariate Analysis of Variance Using Distance Matrices (adonis).** Adonis was employed for model selection to identify relevant immune markers and immune cell types to be included in downstream analyses (See Methods for details). Significant variables (P<0.05) are marked in bold. Abbreviations: hBD2_sim, plasma human beta-defensin 2 levels at time points simultaneous to microbiome characterization; CRP_sim, C-reactive protein levels at time points simultaneous to microbiome characterization;

Lymphocyte_count_sim, total lymphocyte counts at time points simultaneous to microbiome characterization; pIL6, plasma interleukin 6 concentration; Citr, plasma citrulline concentration; CD3+, CD3+ T cell counts; CD4+, CD3+CD4+ T cell counts; CD8+, CD3+CD8+ T cell counts; B, total B cell (CD45+CD19+) counts; mat_B, mature B cell (CD45+CD19+CD20+) counts. immat_B, immature B cell (CD45+CD19+CD20-) counts; NK, natural killer cell counts; mean_mono, mean monocyte counts at indicated time point; mean_neutro, mean neutrophil counts at indicated time point; Timepoints: pre, prior to transplantation; w0, on the day of transplantation; w1, w2, w3, w4, w5: one, two, three, four and five weeks after transplantation, respectively; m1, m2, m3, m4, m6: one, two, three, four and six months after transplantation, respectively; 1y, 1 year post transplantation. (PDF)

Additional file 5: **Table S3: Taxonomy and cluster affiliation of OTUs strongly associated with host-related variables based on sPLS analysis and community state typing (CST).** List of the 57 OTUs correlated strongest with variables in the sPLS analysis (>0.2/<-0.2) . SPLS-based clusters were determined by applying the mixOmics *cim()* function to the sPLS regression model (hierarchical clustering method: complete linkage, distance method: Pearson's correlation) (see Methods). Four community state types (CSTs) were defined by clustering of fecal samples with similar microbial community compositions by partitioning around medoid (PAM) clustering (see Methods). OTUs were then assigned to the CST-based clusters in which they exhibited the highest average abundance over all samples. The same taxonomic families dominated in sPLS- and CST-based clusters, respectively. Cluster 1 was dominated by *Lactobacillaceae*. Cluster 2 was characterized mainly by *Ruminococcaceae* and *Lachnospiraceae*. Cluster 3 harbored Proteobacteria (P), e. g. *Enterobacteriaceae*. CS-typing revealed one additional cluster (4), characterized by a high abundance of *Enterococcaceae* and *Staphylococcaceae.* OTU numbers refer to the SILVA database (*silva_119_rep_set97).* Phyla abbreviations: F, Firmicutes; B, Bacteroidetes; A, Actinobacteria; P, Proteobacteria; FU, Fusobacteria. (PDF)

Additional file 6: **Figure S3. Canonical correspondence analysis (CCpnA) of immune markers and intestinal bacterial taxa in patients undergoing HSCT.** Triplots showing dimension 1 and 2 of the CCpnA that includes continuous clinical variables (arrows), categorical variables (+), and OTUs (circles). Samples are depicted as triangles. OTUs with a correlation of >0.2/<-0.2 in the sPLS analysis were included in the CCpnA model. Only the variables and OTUs with a score >0.2/<-0.2 in at least one CCpnA dimension are shown. The OTUs in the CCpnA plots are colored according to the cluster they were affiliated with in the sPLS-based hierarchical clustering analysis, and the ellipses present an 80% confidence interval, assuming normal distribution. (A) Full size visualization corresponding to the CCpnA model shown in Figure 4. Plot dimensions correspond to the explained variances of each component. (B) CCpnA including

antibiotic treatment at time points simultaneous to microbiome characterization. Antibiotics were added as categorical variables. Depiction of the antibiotic's name (in red) indicates administration of the particular antibiotic, and the extension "_0" indicates no administration of the respective antibiotic. Abbreviations of variables are the same as in Figure 2. Further abbreviations: graft_BM: stem cell source bone marrow; graft_UC: stem cell source umbilical cord blood. (PDF)

Additional file 7: **Figure S4. Clustered image map (CIM) of OTU abundances by patient in the first two sPLS dimensions.** Hierarchical clustering of OTU abundances (bottom) and patients' fecal samples (right) (clustering method: complete linkage, distance method: Pearson's correlation) was performed within the mixOmics *cim()* function based on the sPLS regression model. High abundance of an OTU in a sample is represented as positive correlation in the map (red) and low abundance as negative correlation (blue). The sampling time points of the fecal samples are displayed in the side bar on the left (blue gradient from pre-HSCT time point (light blue) to week +5 post HSCT (dark blue)). The top side bar shows taxonomic information on family level. Sample names on the right indicate patient (P) and time point (pre: pre-HSCT, d0: day of HSCT, w: week). An "a" or "b" indicates that two samples were collected from the respective patient at the same time point, but on two different days. (PDF)

Additional file 8: **Figure S5. Community state types and gut microbial patterns.** Heat map of variance stabilized counts of the 50 most abundant OTUs of the intestinal microbiome over all samples, grouped into community state types (CSTs). Based on their OTU-composition, samples were assigned to community state types (CSTs) by partitioning around medoid (PAM) clustering using Bray-Curtis distance. The optimal number of clusters (k = 4) was estimated from the gap statistic and Silhouette width validation. Members of the *Lactobacillaceae* family dominated the abundance profiles within CST 1. CST 2 exhibited domination by *Lachnospiraceae*, *Erysipelotrichaceae* and *Ruminococcaceae* members. *Enterobacteriaceae*, *Streptococcaceae* and *Staphylococcacea* were characteristic for CST 3. CST 4 was characterized by a high abundance of *Enterococcaceae.* Average Silhouette width was s(*i*) = 0.16 (range: -0.02 – 0.36), with CST 1 and CST 4 being the best defined clusters (s(*i*) = 0.23 and 0.36, respectively). A Silhouette coefficient s(*i*) close to 1 indicates appropriate clustering of the respective samples. Sample names at the bottom indicate patient (P) and time point (pre: pre-HSCT, d0: day of HSCT, w: week). An "a" or "b" indicates that two samples were collected from the respective patient at the same time point, but on two different days. (PDF)

Additional file 9: **Figure S6. Longitudinal profiles of microbial community composition and immune markers in patients with aGvHD who survived**. In three representative patients with moderate to severe aGvHD who survived, high

abundance of *Lactobacillaceae* was observed already before aGvHD onset. None of the depicted patients had a bacterial infection recorded during the monitored period. InvSimpson, inverse Simpson diversity index; hBD2, human beta-defensin 2. (PDF)

Additional file 10: **Supplementary Discussion.** Discussion concerning survival following high *Lactobacillaceae* abundances prior to the onset of aGvHD, and associations of adverse outcomes with *Enterococcus* compared with previous studies. (PDF)

Additional file 11: **R data analysis report 1**. Data preparation, filtering and transformation. (HTML)

Additional file 12: **R data analysis report 2**. Bacterial alpha-diversity over time and rank abundance curve for the gut microbiome of HSCT patients. (HTML)

Additional file 13: **R data analysis report 3**. Temporal patterns of immune markers and immune cells in HSCT patients. (HTML)

Additional file 14: **R data analysis report 4**. Correlations between immune markers, immune cell counts, and outcomes in HSCT patients. (HTML)

Additional file 15: **R data analysis report 5**. Variable selection and multivariate analyses of immune parameters and intestinal bacterial taxa in HSCT patients. (HTML)

Additional file 16: **R data analysis report 6**. Clustering of samples into Community State Types (CSTs) based on Jenson-Shannon divergence. (HTML)

Additional file 17: **R data analysis report 7**. Longitudinal profiles of microbial community composition and immune markers. (HTML)

References

1. Maukonen J, Kolho K-L, Paasela M, Honkanen J, Klemetti P, Vaarala O, et al. Altered Fecal Microbiota in Paediatric Inflammatory Bowel Disease. J Crohn's Colitis. 2015;9:1088–95. doi:10.1093/ecco-jcc/jjv147.

2. Lynch S V., Pedersen O. The Human Intestinal Microbiome in Health and Disease. N Engl J Med. 2016;375:2369–79. doi:10.1056/NEJMra1600266.

3. Lemon KP, Armitage GC, Relman DA, Fischbach MA. Microbiota-Targeted Therapies: An Ecological Perspective. Sci Transl Med. 2012;4:137rv5-137rv5. doi:10.1126/scitranslmed.3004183.

4. Taur Y, Jenq RR, Perales M, Littmann ER, Morjaria S, Ling L, et al. The effects of intestinal tract bacterial diversity on mortality following allogeneic hematopoietic stem cell transplantation. Transplantation. 2014;124:1174–82. doi:10.1182/blood-2014-02-554725.The.

5. Liu C, Frank DN, Horch M, Chau S, Ir D, Horch EA, et al. Associations between acute gastrointestinal GvHD and the baseline gut microbiota of allogeneic hematopoietic stem cell transplant recipients and donors. Bone Marrow Transplant Adv online Publ. 2017;doi August:1–8. doi:10.1038/bmt.2017.200.

6. Appelbaum FR. Haematopoietic cell transplantation as immunotherapy. Nature. 2001;411:385–9. doi:10.1038/35077251.

7. Barrett AJ. Understanding and harnessing the graft-versus-leukaemia effect. Br J Haematol. 2008;142:877–88. doi:10.1111/j.1365-2141.2008.07260.x.

8. Gyurkocza B, Sandmaier BM. Conditioning regimens for hematopoietic cell transplantation: one size does not fit all. Blood. 2014;124:344–53. doi:10.1182/blood-2014-02-514778.

9. Holtick U, Albrecht M, Chemnitz JM, Theurich S, Skoetz N, Scheid C, et al. Bone marrow versus peripheral blood allogeneic haematopoietic stem cell transplantation for haematological malignancies in adults. Cochrane Database Syst Rev. 2014. doi:10.1002/14651858.CD010189.pub2.

10. Tiercy J-M. How to select the best available related or unrelated donor of hematopoietic stem cells? Haematologica. 2016;101:680–7. doi:10.3324/haematol.2015.141119.

11. Holler E, Butzhammer P, Schmid K, Hundsrucker C, Koestler J, Peter K, et al. Metagenomic Analysis of the Stool Microbiome in Patients Receiving Allogeneic Stem Cell Transplantation: Loss of Diversity Is Associated with Use of Systemic Antibiotics and More Pronounced in Gastrointestinal Graft-versus-Host Disease. Biol Blood Marrow Transplant. 2014;20:640–5. doi:10.1016/j.bbmt.2014.01.030.

12. Shono Y, Docampo MD, Peled JU, Perobelli SM, Jenq RR. Intestinal microbiota-related effects on graft-versus-host disease. Int J Hematol. 2015;101:428–37. doi:10.1007/s12185-015-1781-5.

13. Jenq RR, Taur Y, Devlin SM, Ponce DM, Goldberg JD, Ahr KF, et al. Intestinal Blautia Is Associated with Reduced Death from Graft-versus-Host Disease. Biol Blood Marrow Transplant. 2015;21:1373–83. doi:10.1016/j.bbmt.2015.04.016.

14. Jenq RR, Ubeda C, Taur Y, Menezes CC, Khanin R, Dudakov JA, et al. Regulation of

1166 intestinal inflammation by microbiota following allogeneic bone marrow
1167 transplantation. J Exp Med. 2012;209:903–11. doi:10.1084/jem.20112408.

1168 15. Taur Y, Xavier JB, Lipuma L, Ubeda C, Goldberg J, Gobourne a., et al. Intestinal
1169 Domination and the Risk of Bacteremia in Patients Undergoing Allogeneic
1170 Hematopoietic Stem Cell Transplantation. Clin Infect Dis. 2012;55:905–14.
1171 doi:10.1093/cid/cis580.

1172 16. Gosselin KB, Feldman HA, Sonis AL, Bechard LJ, Kellogg MD, Gura K, et al. Serum
1173 citrulline as a biomarker of gastrointestinal function during hematopoietic cell
1174 transplantation in children. J Pediatr Gastroenterol Nutr. 2014;58:709–14.
1175 doi:10.1097/MPG.0000000000000335.

1176 17. Pontoppidan PL, Jordan K, Carlsen AL, Uhlving HH, Kielsen K, Christensen M, et al.
1177 Associations between gastrointestinal toxicity, micro RNA and cytokine production in
1178 patients undergoing myeloablative allogeneic stem cell transplantation. Int
1179 Immunopharmacol. 2015;25:180–8. doi:10.1016/j.intimp.2014.12.038.

1180 18. Sporrer D, Gessner A, Hehlgans T, Oefner PJ, Holler E. The Microbiome and
1181 Allogeneic Stem Cell Transplantation. Curr Stem Cell Reports. 2015;1:53–9.
1182 doi:10.1007/s40778-014-0006-9.

1183 19. Liu S, He LR, Wang W, Wang GH, He ZY. Prognostic value of plasma human beta-
1184 defensin 2 level on short-term clinical outcomes in patients with community-
1185 acquired pneumonia: a preliminary study. Respir Care. 2013;58:655–61.
1186 doi:58/4/655 [pii];10.4187/respcare.01827 [doi].

1187 20. Meisch JP, Nishimura M, Vogel RM, Sung HC, Bednarchik BA, Ghosh SK, et al.
1188 Human β-Defensin 3 Peptide is Increased and Redistributed in Crohn's Ileitis.
1189 Inflamm Bowel Dis. 2013;19:942–53. doi:10.1097/MIB.0b013e318280b11a.

1190 21. de Koning C, Plantinga M, Besseling P, Boelens JJ, Nierkens S. Immune
1191 Reconstitution after Allogeneic Hematopoietic Cell Transplantation in Children. Biol
1192 Blood Marrow Transplant. doi:10.1016/j.bbmt.2015.08.028.

1193 22. Furusawa Y, Obata Y, Hase K. Commensal microbiota regulates T cell fate
1194 decision in the gut. Semin Immunopathol. 2015;37:17–25. doi:10.1007/s00281-014-
1195 0455-3.

1196 23. Donaldson GP, Lee SM, Mazmanian SK. Gut biogeography of the bacterial
1197 microbiota. Nat Rev Microbiol. 2015;14:20–32. doi:10.1038/nrmicro3552.

1198 24. Taur Y, Pamer EG. Microbiome mediation of infections in the cancer setting.
1199 Genome Med. 2016;8:40. doi:10.1186/s13073-016-0306-z.

1200 25. Schlee M, Harder J, Köten B, Stange EF, Wehkamp J, Fellermann K. Probiotic
1201 lactobacilli and VSL#3 induce enterocyte β-defensin 2. Clin Exp Immunol.
1202 2008;151:528–35. doi:10.1111/j.1365-2249.2007.03587.x.

1203 26. Paolillo R, Romano Carratelli C, Sorrentino S, Mazzola N, Rizzo A.
1204 Immunomodulatory effects of Lactobacillus plantarum on human colon cancer cells.
1205 Int Immunopharmacol. 2009;9:1265–71. doi:10.1016/j.intimp.2009.07.008.

1206 27. Chung H, Pamp SJ, Hill JA, Surana NK, Edelman SM, Troy EB, et al. Gut immune
1207 maturation depends on colonization with a host-specific microbiota. Cell.
1208 2012;149:1578–93. doi:10.1016/j.cell.2012.04.037.

1209 28. Bäumler AJ, Sperandio V. Interactions between the microbiota and pathogenic
1210 bacteria in the gut. Nature. 2016;535:85–93. doi:10.1038/nature18849.

1211 29. Fujimaki K, Maruta A, Yoshida M, Kodama F, Matsuzaki M, Fujisawa S, et al.
1212 Immune reconstitution assessed during five years after allogeneic bone marrow
1213 transplantation. Bone Marrow Transplant 2001 2712. 2001;27:1275.
1214 doi:10.1038/sj.bmt.1703056.

1215 30. Bartelink IH, Belitser S V., Knibbe CAJ, Danhof M, de Pagter AJ, Egberts TCG, et al.
1216 Immune Reconstitution Kinetics as an Early Predictor for Mortality using Various
1217 Hematopoietic Stem Cell Sources in Children. Biol Blood Marrow Transplant.
1218 2013;19:305–13. doi:10.1016/j.bbmt.2012.10.010.

1219 31. Corre E, Carmagnat M, Busson M, de Latour RP, Robin M, Ribaud P, et al. Long-
1220 term immune deficiency after allogeneic stem cell transplantation: B-cell deficiency
1221 is associated with late infections. Haematologica. 2010;95:1025–9.
1222 doi:10.3324/haematol.2009.018853.

1223 32. Montassier E, Gastinne T, Vangay P, Al-Ghalith GA, Bruley Des Varannes S,
1224 Massart S, et al. Chemotherapy-driven dysbiosis in the intestinal microbiome.
1225 Aliment Pharmacol Ther. 2015;42:515–28.

1226 33. Jenq RR, Ubeda C, Taur Y, Menezes CC, Khanin R, Dudakov JA, et al. Regulation of
1227 intestinal inflammation by microbiota following allogeneic bone marrow
1228 transplantation. J Exp Med. 2012;209:903–11. doi:10.1084/jem.20112408.

1229 34. Gopalakrishnan V, Spencer CN, Nezi L, Reuben A, Andrews MC, Karpinets T V, et
1230 al. Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma
1231 patients. Science. 2018;359:97–103. doi:10.1126/science.aan4236.

1232 35. Weber D, Oefner PJ, Hiergeist A, Koestler J, Gessner A, Weber M, et al. Low
1233 urinary indoxyl sulfate levels early after transplantation reflect a disrupted

microbiome and are associated with poor outcome. Blood. 2015;126:1723–8. doi:10.1182/blood-2015-04-638858.

36. Mathewson ND, Jenq R, Mathew A V, Koenigsknecht M, Hanash A, Toubai T, et al. Gut microbiome-derived metabolites modulate intestinal epithelial cell damage and mitigate graft-versus-host disease. Nat Immunol. 2016;17:505–13. doi:10.1038/ni.3400.

37. Salzman NH, Hung K, Haribhai D, Chu H, Karlsson-Sjöberg J, Amir E, et al. Enteric defensins are essential regulators of intestinal microbial ecology. Nat Immunol. 2010;11:76–82. doi:10.1038/ni.1825.

38. Rivera-Chávez F, Lopez CA, Bäumler AJ. Oxygen as a driver of gut dysbiosis. Free Radic Biol Med. 2017;105:93–101. doi:10.1016/J.FREERADBIOMED.2016.09.022.

39. Fouhy F, Guinane CM, Hussey S, Wall R, Ryan CA, Dempsey EM, et al. High-throughput sequencing reveals the incomplete, short-term recovery of infant gut microbiota following parenteral antibiotic treatment with ampicillin and gentamicin. Antimicrob Agents Chemother. 2012;56:5811–20. doi:10.1128/AAC.00789-12.

40. Weber D, Jenq RR, Peled JU, Taur Y, Hiergeist A, Koestler J, et al. Microbiota Disruption Induced by Early Use of Broad Spectrum Antibiotics is an Independent Risk Factor of Outcome after Allogeneic Stem Cell Transplantation. Elsevier Inc.; 2017. doi:10.1016/j.bbmt.2017.02.006.

41. Simms-Waldrip TR, Sunkersett G, Coughlin LA, Savani MR, Arana C, Kim J, et al. Antibiotic-Induced Depletion of Anti-inflammatory Clostridia Is Associated with the Development of Graft-versus-Host Disease in Pediatric Stem Cell Transplantation Patients. Biol Blood Marrow Transplant. 2017;23:820–9. doi:10.1016/J.BBMT.2017.02.004.

42. Weber D, Hiergeist A, Weber M, Dettmer K, Wolff D, Hahn J, et al. Detrimental effect of broad-spectrum antibiotics on intestinal microbiome diversity in patients after allogeneic stem cell transplantation: Lack of commensal sparing antibiotics. Clin Infect Dis. 2018. doi:10.1093/cid/ciy711.

43. Isaac S, Scher JU, Djukovic A, Jiménez N, Littman DR, Abramson SB, et al. Short- and long-term effects of oral vancomycin on the human intestinal microbiota. J Antimicrob Chemother. 2017;72:128–36. doi:10.1093/jac/dkw383.

44. Yeh T-C, Liu H-C, Hou J-Y, Chen K-H, Huang T-H, Chang C-Y, et al. Severe infections in children with acute leukemia undergoing intensive chemotherapy can successfully be prevented by ciprofloxacin, voriconazole, or micafungin prophylaxis. Cancer. 2014;120:1255–62. doi:10.1002/cncr.28524.

45. Dethlefsen L, Relman DA. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. Proc Natl Acad Sci U S A. 2011;108 Suppl Supplement 1:4554–61. doi:10.1073/pnas.1000087107.

46. Shono Y, Docampo MD, Peled JU, Perobelli SM, Velardi E, Tsai JJ, et al. Increased GVHD-related mortality with broad-spectrum antibiotic use after allogeneic hematopoietic stem cell transplantation in human patients and mice. Sci Transl Med. 2016;8:339ra71-339ra71. doi:10.1126/scitranslmed.aaf2311.

47. Kielsen K, Jordan KK, Uhlving HH, Pontoppidan PL, Shamim Z, Ifversen M, et al. T cell Reconstitution in Allogeneic Haematopoietic Stem Cell Transplantation: Prognostic Significance of Plasma Interleukin-7. Scand J Immunol. 2015;81:72–80. doi:10.1111/sji.12244.

48. Jordan K, Pontoppidan P, Uhlving HH, Kielsen K, Burrin DG, Weischendorff S, et al. Gastrointestinal Toxicity, Systemic Inflammation, and Liver Biochemistry in Allogeneic Hematopoietic Stem Cell Transplantation. Biol Blood Marrow Transplant. 2017;23:1170–6. doi:10.1016/j.bbmt.2017.03.021.

49. Kielsen K, Ryder LP, Lennox-Hvenekilde D, Gad M, Nielsen CH, Heilmann C, et al. Reconstitution of Th17, Tc17 and Treg cells after paediatric haematopoietic stem cell transplantation: Impact of interleukin-7. Immunobiology. 2018;223:220–6. doi:10.1016/j.imbio.2017.10.023.

50. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol. 2009;75:7537–41. doi:10.1128/AEM.01541-09.

51. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7:335–6. doi:10.1038/nmeth.f.303.

52. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2017. https://www.r-project.org/.

53. McMurdie PJ, Holmes S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS One. 2013;8:e61217. doi:10.1371/journal.pone.0061217.

54. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer Verlag New York; 2016.

55. Sievert C, Parmer C, Hocking T, Chamberlain S, Ram K, Corvellec, Marianne Despouy P. plotly: Create Interactive Web Graphics via “plotly. js.” 2016. https://plot.ly/r, https://cpsievert.github.io/plotly_book/, https://github.com/ropensci/plotly#readme.

56. Cao K Le, Rohart F, Gonzalez I, Dejean S, Gautier B, Bartolo F, et al. mixOmics: Omics Data Integration Project. R package version 6.1.3. 2017. https://cran.r-project.org/package=mixOmics.

57. Ritz C, Baty F, Streibig JC, Gerhard D. Dose-Response Analysis Using R. PLoS One. 2015;10:e0146021. doi:10.1371/journal.pone.0146021.

58. Gentleman R, Carey V, Huber W, Hahne F. genefilter: methods for filtering genes from microarray experiments. R package version 1.58.1. 2017.

59. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550. doi:10.1186/s13059-014-0550-8.

60. McMurdie PJ, Holmes S. Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. PLoS Comput Biol. 2014;10:e1003531. doi:10.1371/journal.pcbi.1003531.

61. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. Vegan: Community Ecology Package. R Package Version. 2.4-3. https://cran.R-project.org/package=vegan. 2017. doi:10.1029/2006JF000545.

62. Pérez-Losada M, Freishtat RJ, Kwak C, Authelet KJ, Hoptay CE, Crandall KA. Pediatric asthma comprises different phenotypic clusters with unique nasal microbiotas. Microbiome. 2018;6:1–13.

63. Harrell Jr FE, Dupont C. Hmisc: Harrell Miscellaneous. R package version 4.0-3. 2017.

64. Wei T, Simko V. corrplot: Visualization of a Correlation Matrix. R package version 0.77. 2017. https://github.com/taiyun/corrplot.

65. Lee D, Lee W, Lee Y, Pawitan Y. Sparse partial least-squares regression and its applications to high-throughput data analysis. Chemom Intell Lab Syst. 2011;109:1–8. doi:10.1016/J.CHEMOLAB.2011.07.002.

66. Callahan BJ, Sankaran K, Fukuyama JA, McMurdie PJ, Holmes SP. Bioconductor workflow for microbiome data analysis: from raw reads to community analyses. F1000Research. 2016;5:1492. doi:10.12688/f1000research.8986.1.

67. Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K. cluster: Cluster Analysis Basics and Extensions. R package version 2.0.6. 2017.

68. DiGiulio DB, Callahan BJ, McMurdie PJ, Costello EK, Lyell DJ, Robaczewska A, et al. Temporal and spatial variation of the human microbiota during pregnancy. Proc Natl Acad Sci. 2015;112:11060–5. doi:10.1073/pnas.1502875112.

Figure legends

Figure 1. Monitoring of the host immune system and intestinal microbiome in hematopoietic stem cell transplantation (HSCT). (A) Study outline. A cohort of 37 pediatric recipients of allogeneic HSCT was monitored prior to HSCT, at the time point of HSCT, and post HSCT (median follow-up time 5.2 years). A range of patient characteristics as well as disease outcomes, immune markers, immune cell counts, and intestinal patterns of microbial community composition were recorded at the noted time points (up to 12 months post HSCT). See Table S1 for details regarding the patient characteristics. (B) Plasma hBD2 concentrations over time and in comparison to healthy controls. The Y-axis was log10-transformed for better visualization. Zeros were replaced with 1 to avoid undefined values on the log-transformed axis. Asterisks indicate whether the concentrations at each respective time point were significantly different from any of the other time points (showing the maximum significance level). (C) Bacterial alpha-diversity, measured by inverse Simpson index, of the intestinal microbiome shown with log10-transformed y-axis. (D) Rank abundance curve displaying the 8 most abundant taxonomic families in the dataset (comprising 98 fecal samples).

Figure 2. Correlations between immune markers, immune cell counts, and outcomes in patients undergoing HSCT. A) Pairwise Spearman's correlation between immune markers and immune cell counts in HSCT patients (*n=37*) that were determined to be significant in a permutational multivariate analysis of variance using (microbial) distance matrices (adonis) (See Table S2). Positive and negative correlations are represented by red or blue circles, respectively, and the size of circles and intensity of color refer to the strength of the correlation. Correlations that are significant ($P \leq 0.05$) are indicated by a black outline of the circle. (B) Natural killer (NK) and total B cell (mature and immature) reconstitution in month +2 with respect to the maximum acute GvHD (aGvHD) grade (0-I vs. II-IV) in HSCT patients (*n=37*). Abbreviations: hBD2_pre, hBD2_w0, hBD2_w1, hBD2_w2: plasma human beta-defensin 2 concentration pre-HSCT, on the day of HSCT, and in weeks +1 and +2, respectively;

mono_pre, mono_w3: monocyte counts pre-HSCT and in week +3, respectively; neutro_m3: neutrophil count in month +3; CD8+_m1: CD8+ T cell counts in month +1; Age: Recipient age at time point of transplantation; NK_m1, NK_m2: Natural killer cell counts in months +1 and +2, respectively; B_m2, mat_B_m2, immat_B_m2: all, mature, and immature B cell counts in month +2; CD4+_m2: CD4+ T cell counts in month +2; Citr_pre, Citr_w1: plasma citrulline levels pre-HSCT and in week +1, respectively; CRP, CRP_w1, CRP_w5, CRP_w6, CRP_m3, CRP_m6: C-reactive protein levels at time points simultaneous to microbiome characterization, in weeks +1, +5, and +6, and in months +3 and +6, respectively. * $P < 0.05$, ** $P < 0.01$ and *** $P < 0.001$.

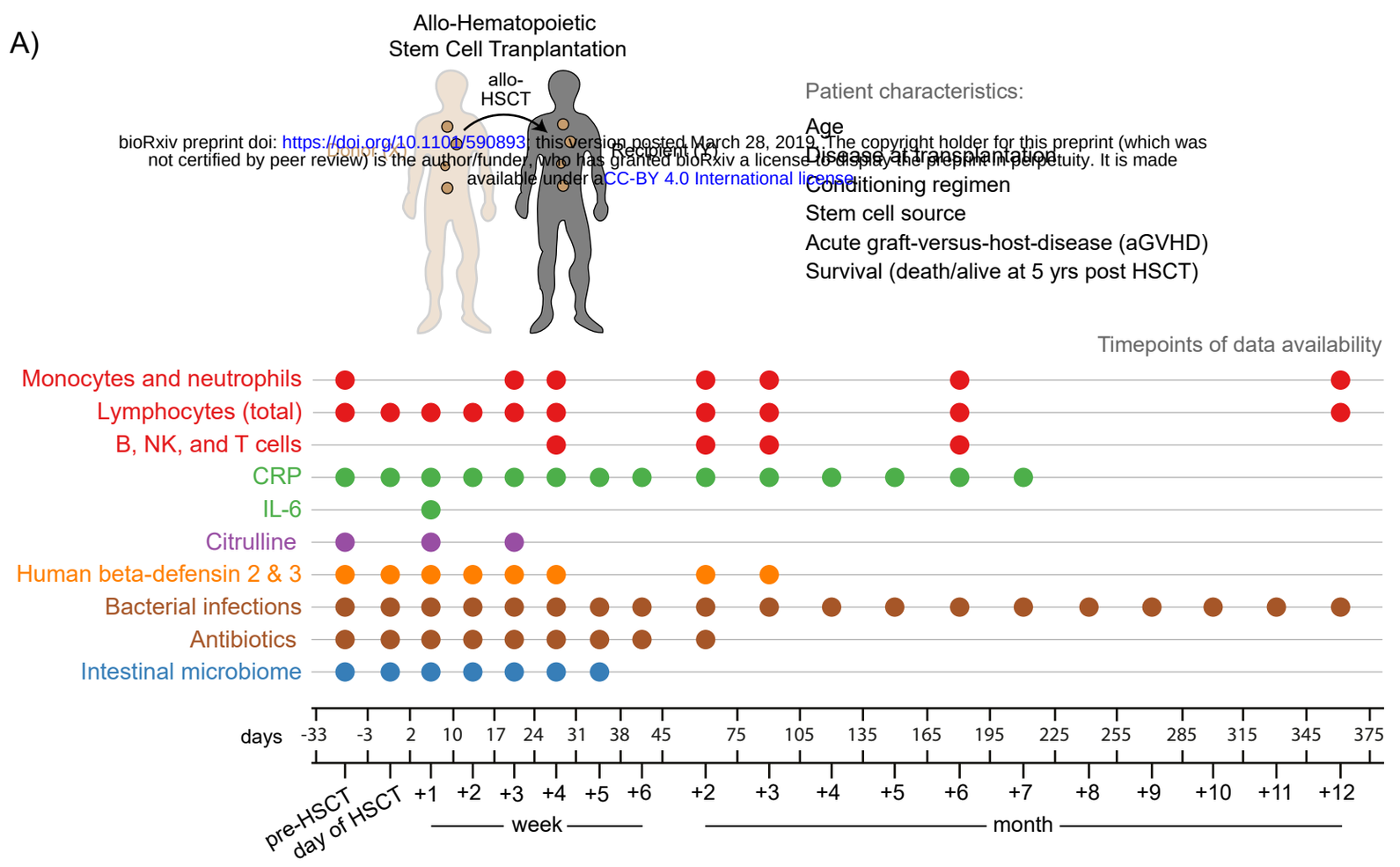
Figure 3. Sparse partial least squares (sPLS) regression of immune parameters and intestinal bacterial taxa during HSCT. (A) Correlation circle plot for the first two sPLS dimensions with correlations displayed for >0.2/<-0.2. The two grey circles indicate correlation coefficient radii at 0.5 and 1.0. Bacterial operational taxonomic units (OTUs) are displayed as circles, and are colored according to the cluster they are affiliated with (Cluster 1: blue; Cluster 2: orange; Cluster 3: grey). Variables projected closely to each other are positively correlated. Variables projected diametrically opposite from each other are negatively correlated. Variables situated perpendicularly to each other are not correlated. (B) Clustered image map (CIM) of the first two sPLS dimensions, displaying pairwise correlations between OTUs (bottom) and clinical variables (left). Red and blue indicate positive and negative correlations, respectively. Hierarchical clustering (clustering method: complete linkage, distance method: Pearson's correlation) was performed within the mixOmics *cim()* function based on the sPLS regression model. An overview of the OTU abundances in the individual samples is provided in Figure S3, and a list of the individual OTUs and their cluster-affiliation is provided in Table S3. (C) Loading plots of OTUs with maximum contributions on the first (left) and second (right) component, respectively. The bars are coloured according to the cluster they are affiliated with. The family-affiliation for each respective OTU is indicated by color (for legend see B). Abbreviations of variables are the same as in Figure 2.

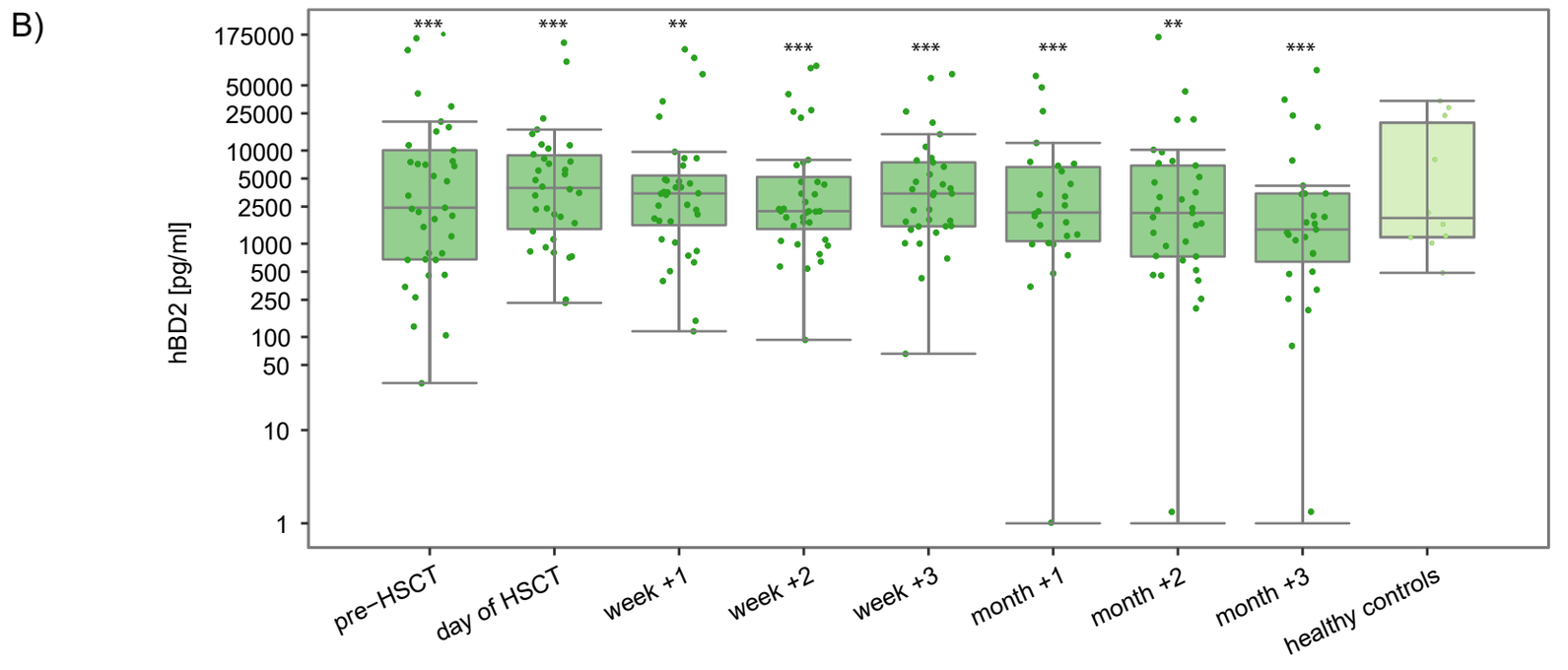
Figure 4. Canonical correspondence analysis (CCpnA) of immune parameters and intestinal bacterial taxa in patients undergoing HSCT. Triplot showing dimension 1 and 2 of the CCpnA that included continuous clinical variables (arrows), categorical variables (+), and OTUs (circles). Samples are depicted as triangles. OTUs with a correlation of >0.2/<-0.2 in the sPLS analysis were included in the CCpnA model. Only the variables and OTUs with a score >0.2/<-0.2 in at least one CCpnA dimension are shown. The OTUs in the CCpnA plot are colored according to the cluster they were affiliated with in the sPLS-based hierarchical clustering analysis, and the ellipses present an 80% confidence interval, assuming normal distribution. For visualization

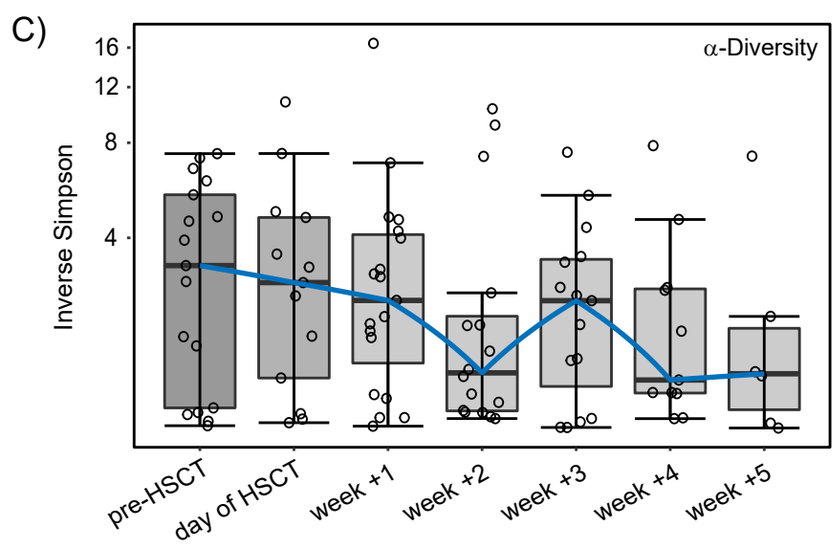
purposes, this plot is a section focussing on the categorical and continuous variables contributing to the ordination. The full size version of the CCpnA triplot, including all samples and OTUs, is presented in Figure S2A. Abbreviations of variables are the same as in Figure 2.

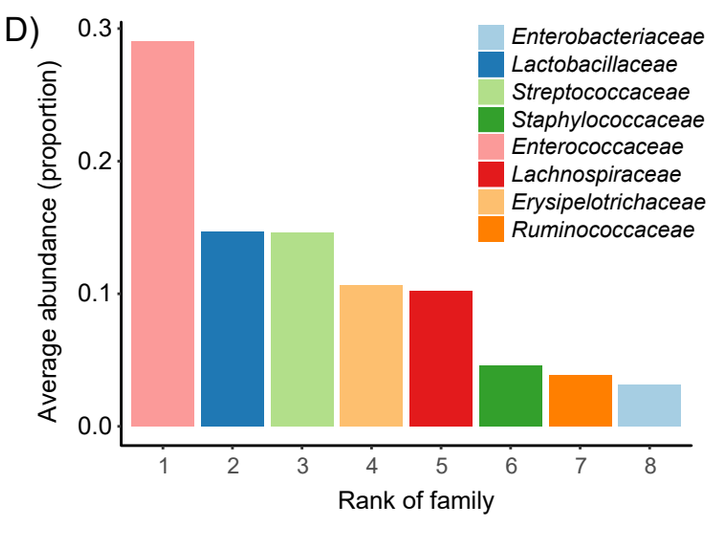
Figure 5. Bacterial community state types over time in patients undergoing allogeneic HSCT. Patients are grouped into four outcome groups: 1. Patients who developed no or mild aGvHD (grade 0-I) and survived vs. 2. Patients who did not survive; 3. Patients who developed moderate to severe aGvHD (grade II-IV) and survived vs. 4. Patients who did not survive. The day of commencement of aGVHD grade II-IV and the day of death post HSCT are displayed to the right. Patients with moderate to severe aGvHD (groups 3 and 4) most often harbored the *Lactobacillaceae*-dominated community state type 1 (CST 1), especially at late time points. CST2, dominated by *Ruminococcaceae* and *Lachnospiraceae*, did not persist after HSCT in any of the patients in groups 3 and 4. A detailed overview of the CSTs is provided in Figure S4, and information about individual OTUs and their cluster-affiliation is provided in Table S3.

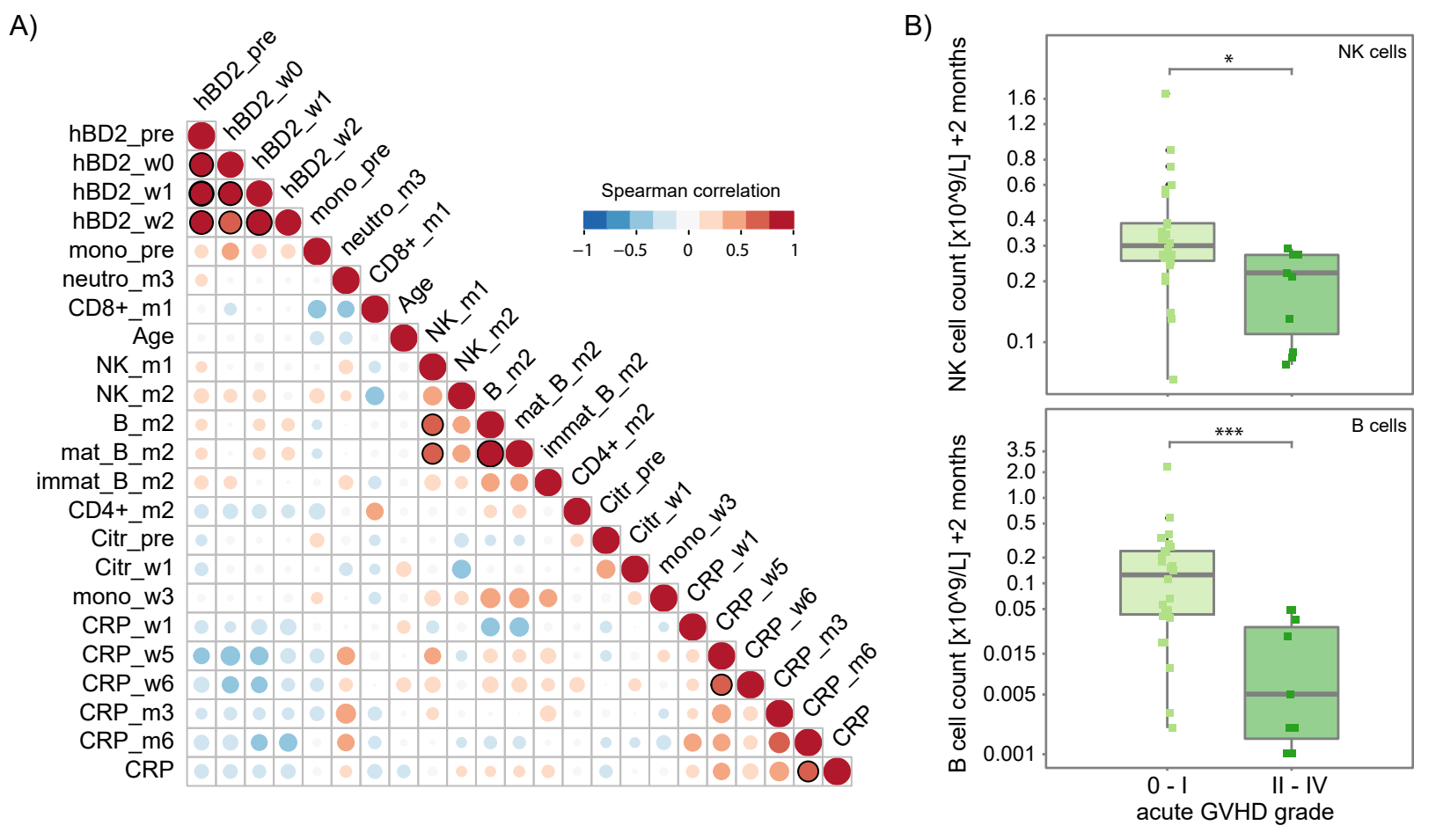
Figure 6. Longitudinal profiles of microbial community composition and immune markers in non-survivors with aGvHD. Abundances of *Lactobacillaceae* increased predominantly after aGvHD onset in patients who died during the follow-up period. Patient P24 developed chronic GvHD on day +187, relapsed on day +548 and died on day +784 due to graft rejection after re-transplantation. Patient P26 died on day +602 after a relapse on day +442 followed by re-transplantation on day +518. Patient P30 died on day +192 after relapse on day +77. Patient P26 and P30 had no reported bacterial infections during the depicted period. Abbreviations: InvSimpson, inverse Simpson diversity index; hBD2, human beta-defensin 2; aGvHD, acute graft-versus-host-disease; inf, bacterial infection.

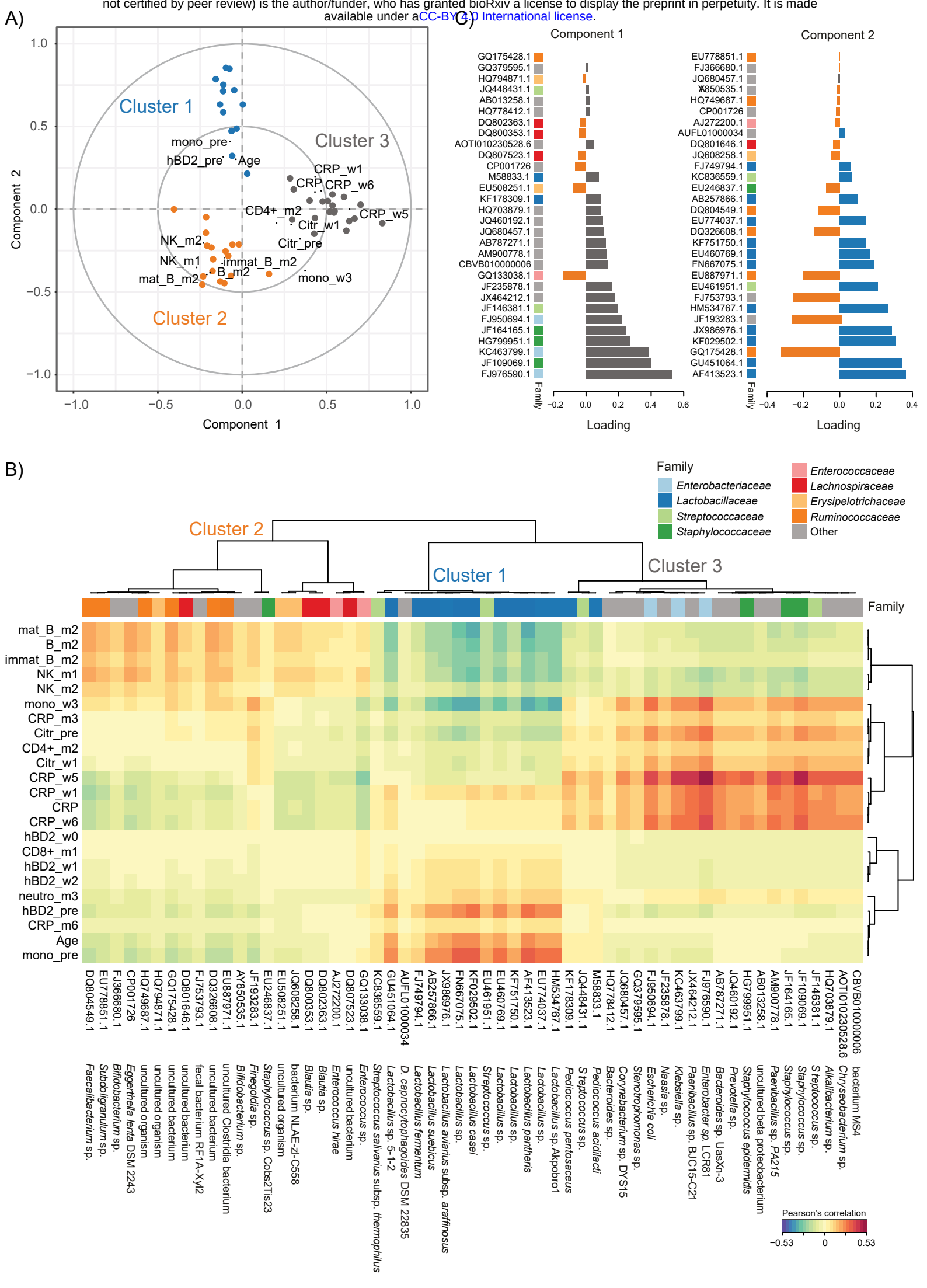


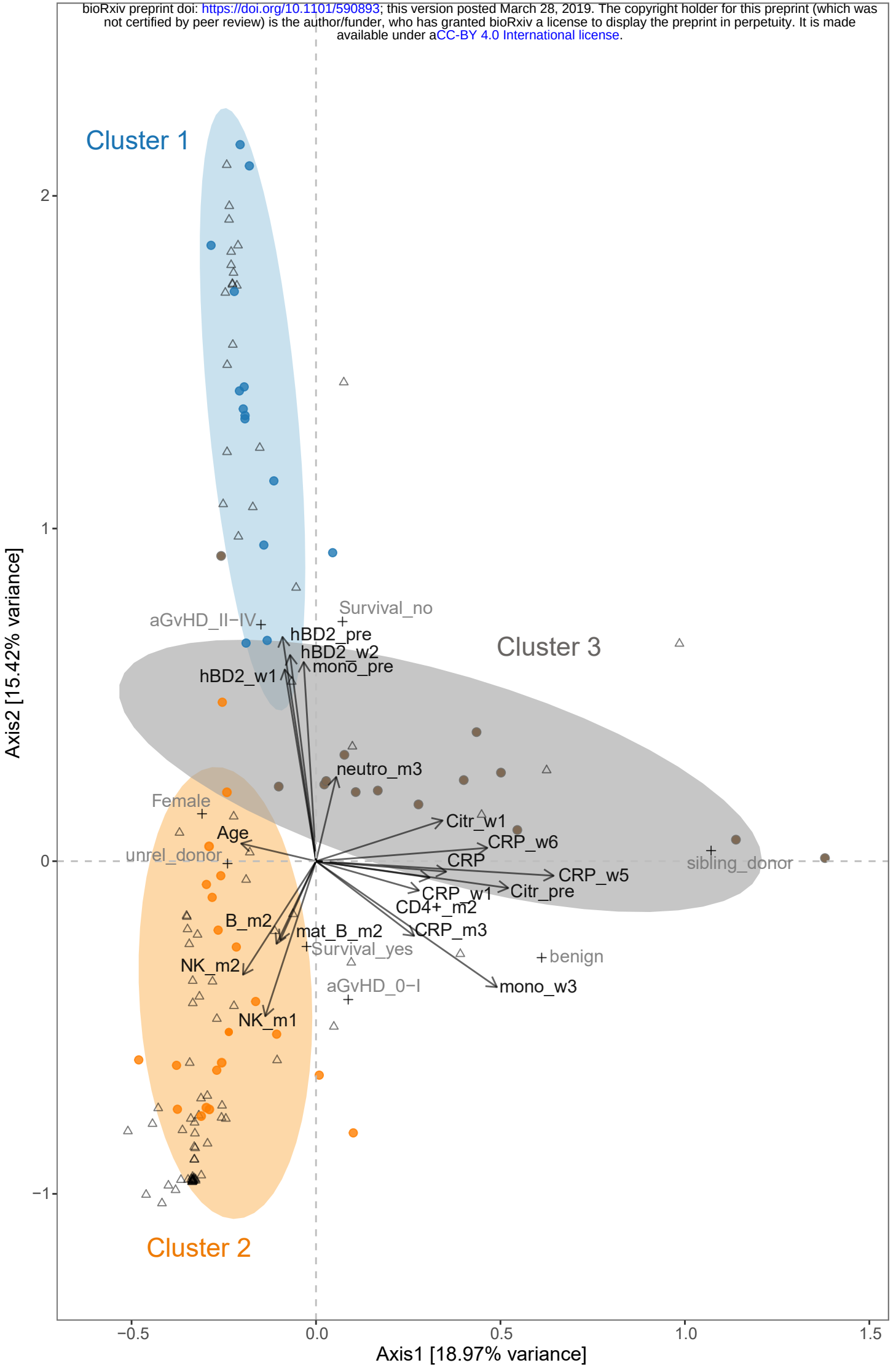


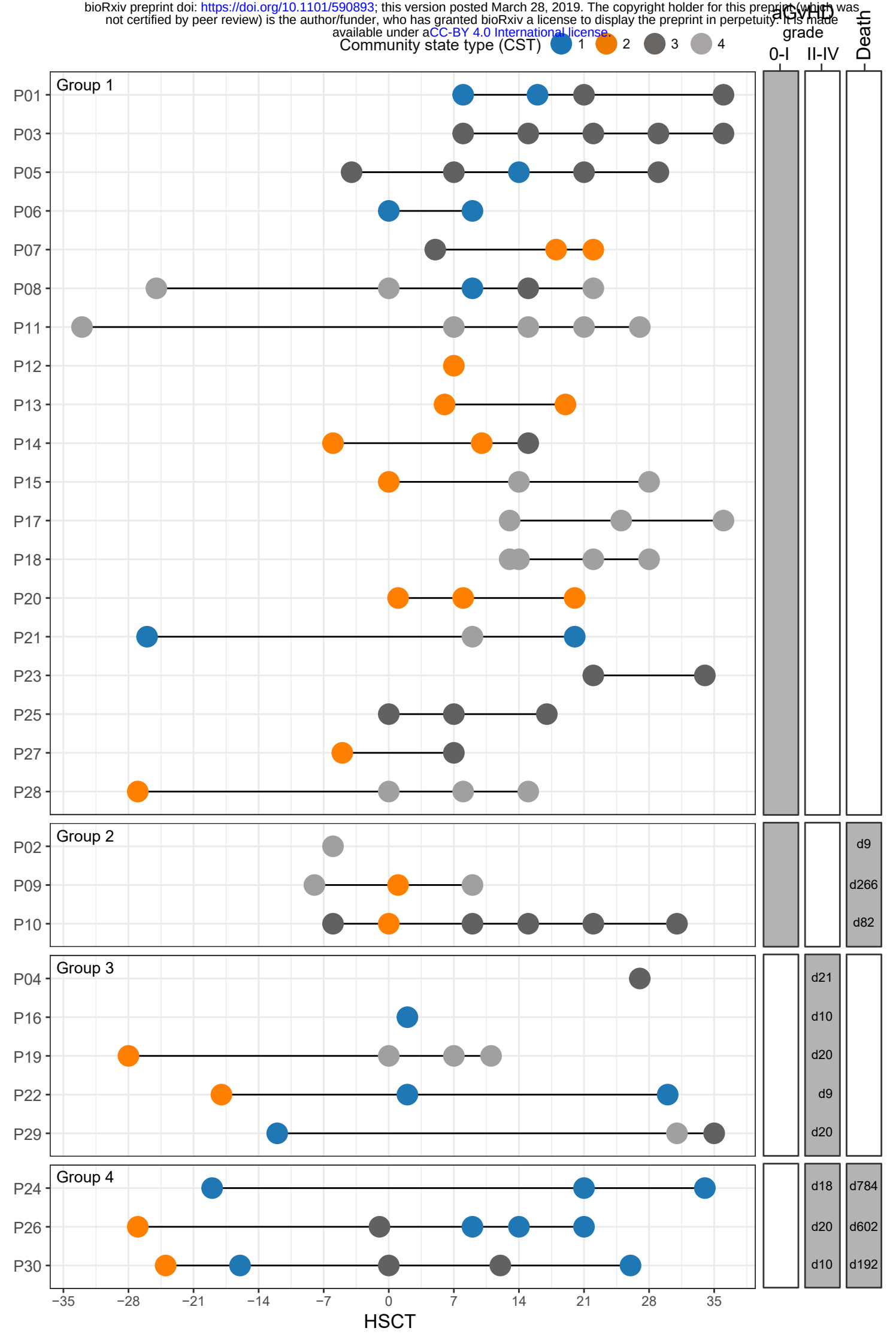




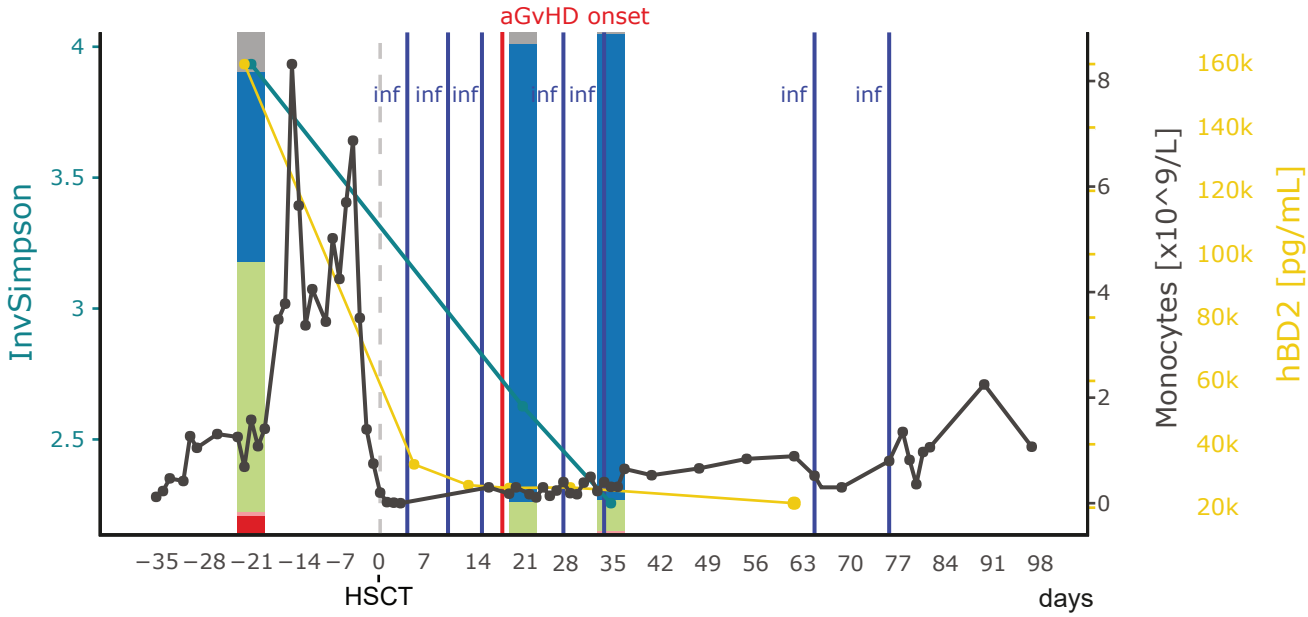

A)
hBD2_pre
hBD2_w0
hBD2_w1
hBD2_w2
mono_pre
neutro_m3
CD8+_m1
Age
NK_m1
NK_m2
B_m2
mat_B_m2
immat_B_m2
CD4+_m2
Citr_pre
Citr_w1
mono_w3
CRP_w1
CRP_w5
CRP_w6
CRP_m3
CRP_m6
CRP
Spearman correlation
−1
−0.5
0
0.5
1
B)
NK cells
*
NK cell count [x10^9/L] +2 months
1.6
1.2
0.8
0.6
0.4
0.3
0.2
0.1
B cells
***
B cell count [x10^9/L] +2 months
3.5
2.0
1.0
0.5
0.2
0.1
0.05
0.015
0.005
0.001
0 - I
II - IV
acute GVHD grade



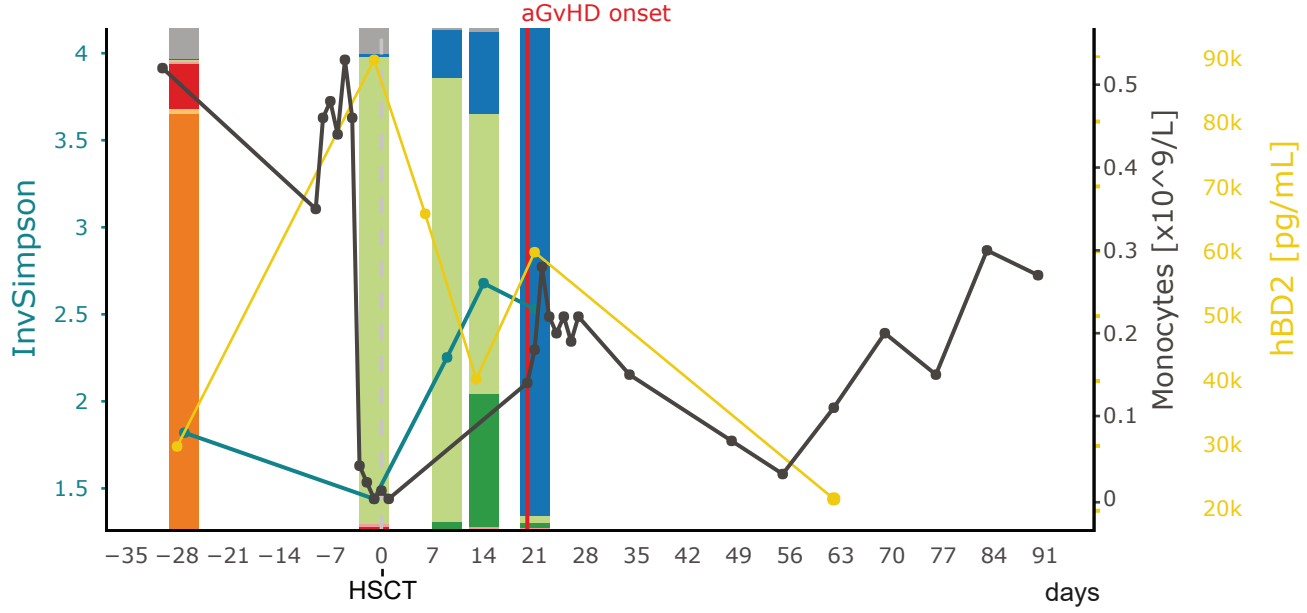




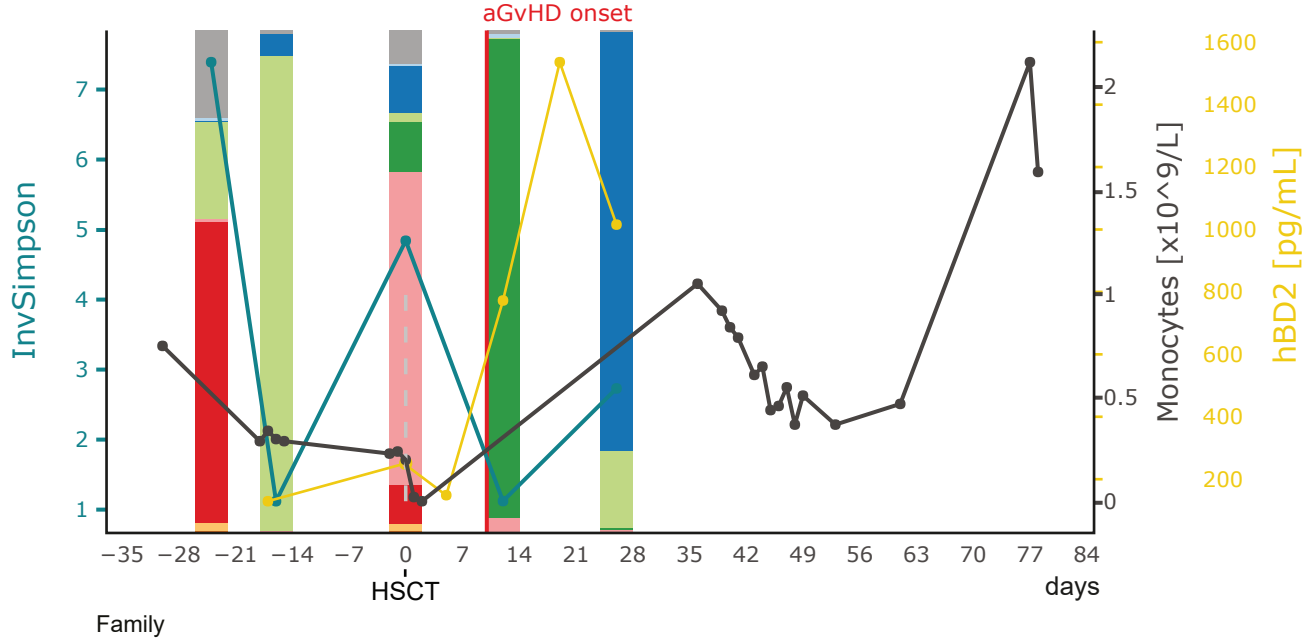
Patient 24



Patient 26



Patient 30

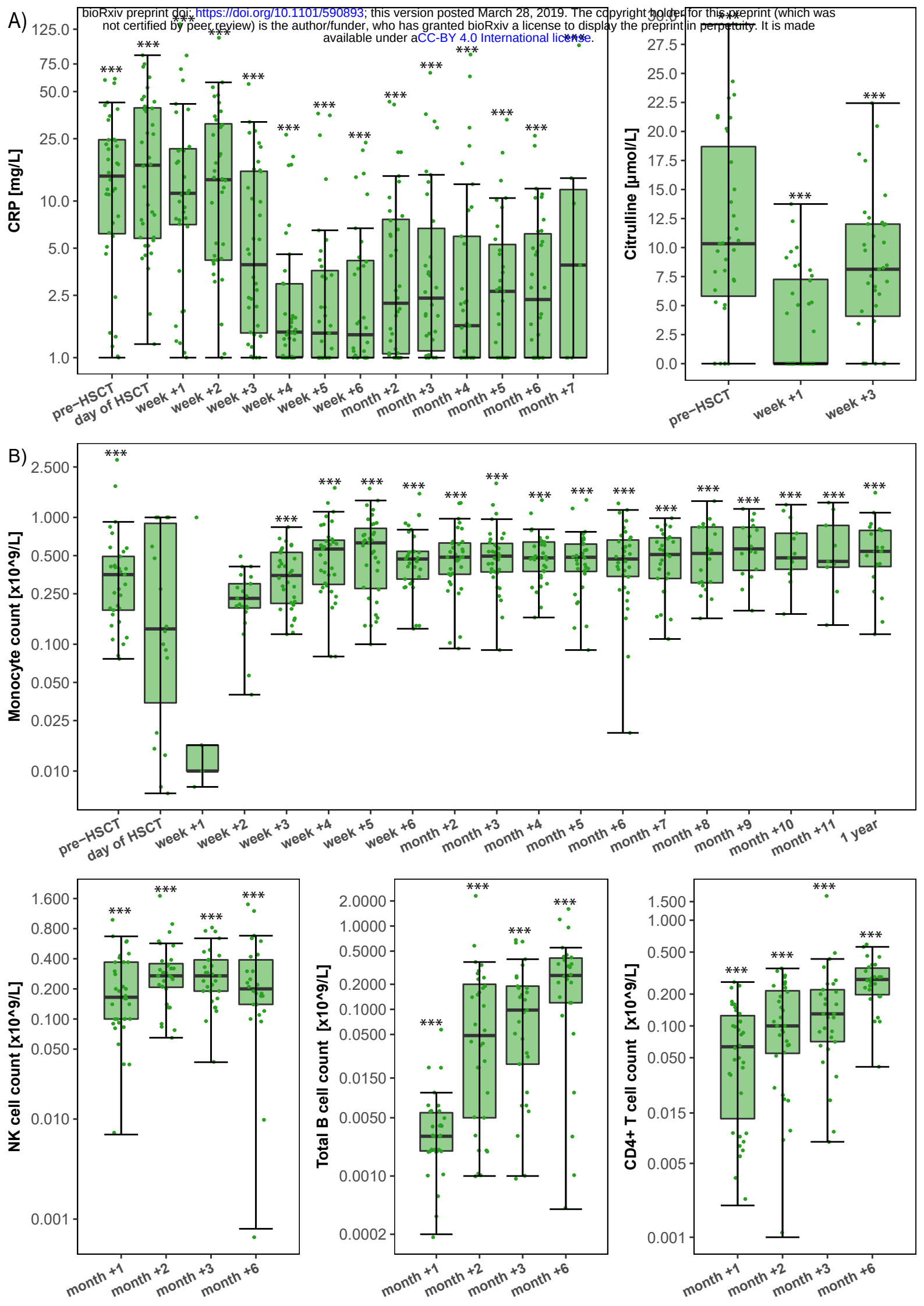


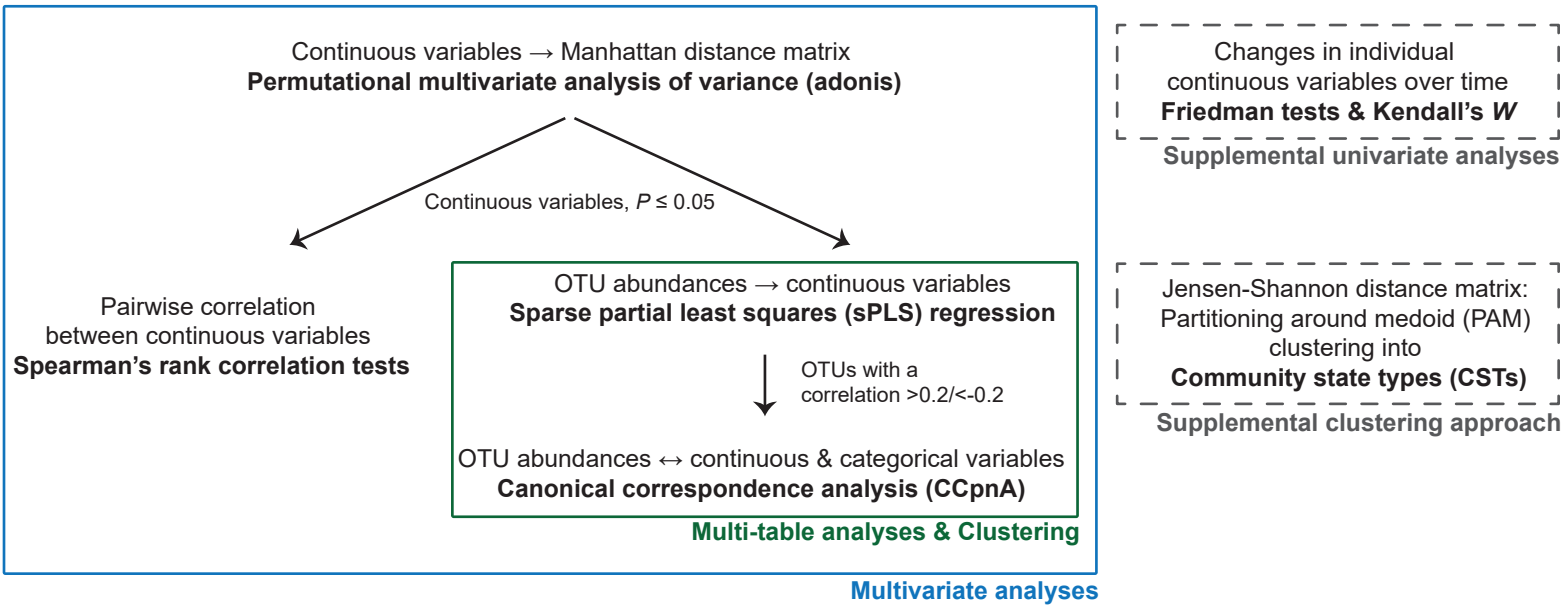
Characteristics		Number of patients (all)	Percentage of all patients (%)	Number of patients (Sub-cohort for which the microbiota was characterized)	Percentage of sub-cohort (%)
Transplant recipients		37	100	30*	81.1
Average recipient age in years		8.2 (Range: 1.1-18.0)	NA	7.8 (Range: 1.1-16.5)	NA
Intestinal microbiome characterized	**At 1-2 timepoints**	NA	NA	8	26.7
	At 3-4 timepoints	NA	NA	15	50
	At 5-6 timepoints	NA	NA	8	26.7
Patient sex	**Female**	15	40.5	12	40
	Male	22	59.5	18	60
Disease at transplantation	**malignant hematologic diseases**	23	62.2	17	56.7
	Severe aplastic anemia	5	13.5	5	16.7
	Other benign disorders including immunodeficiencies	9	24.3	8	26.7
Donor type	**HLA-matched sibling**	7	18.9	6	20
	HLA-matched unrelated donor (9/10 or 10/10 match)	27	73	22	73.3
	HLA-mismatched umbilical cord blood donor (8/10 match)	3	8.1	2	6.7
Stem cell source	**Bone marrow**	30	81.1	25	83.3
	Umbilical cord blood	4	10.8	3	10
	Peripheral blood	2	5.4	1	3.3
	Bone marrow and umbilical cord blood	1	2.7	1	3.3
Conditioning regimen	**TBI + CY / TBI + VP16**	10	27.1	10	33.3
	Combinations of BU, CY, VP16 and MEL	18	48.7	12	40
	Fludarabine-based conditioning	9	24.3	8	26.7

Anti-thymocyte globulin treatment		28	75.7	23	76.7
Antibiotics pre- and post-HSCT		37	100	30	100
Sex mismatch (female donor to male recipient)		4	10	3	10
Acute GvHD	**Grade 0-I**	26	70.3	22	73.3
	Grade II-IV	11	29.2	8	26.7
Chronic GvHD within 24 months		8	18.9	5	16.7
Bacterial infections	**At least one registered bacterial infection**	25	67.6	19	63.3
	No registered bacterial infection	12	32.4	11	36.7
Overall Survival**	**Alive**	30	81.1	24	80
	Dead	7	18.9	6	20
Relapse of primary malignancy		6	16.2	7	23.3
Non-relapse mortality		2	5.4	1	3.3
Re-transplantation		3	8.1	4	13.3

*29/37 +1, one patient's microbiome was characterized but for which hBD2 was not measured

** Mean follow-up time after HSCT: 62.1 months (range: 48.9 – 75.8 months)





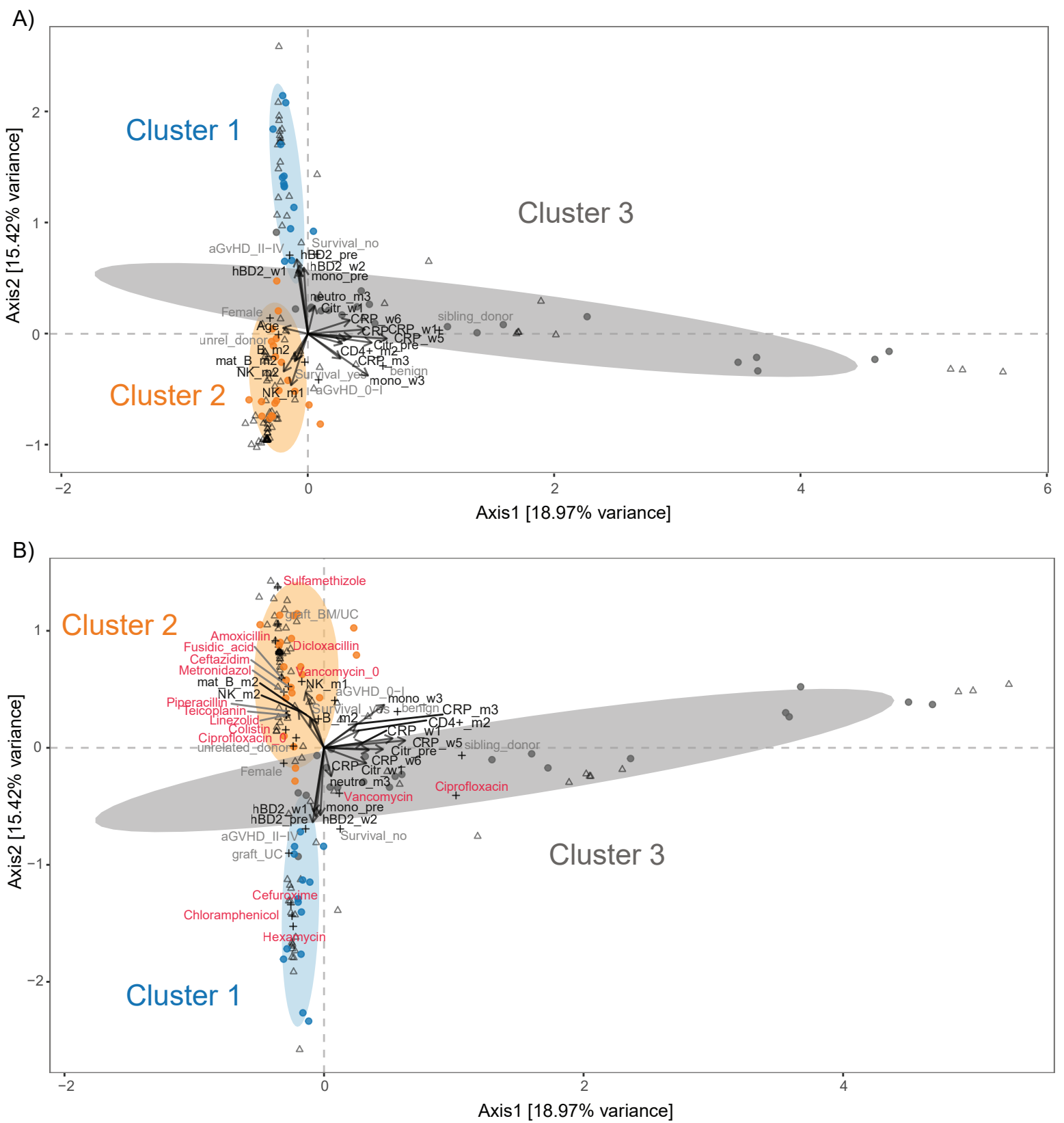
Variable	SumOfSqs	F	*P*-value
hBD2_sim	9763	0.4635	0.915
hbd2_pre	89958	4.3373	**0.001**
hbd2_w0	39432	1.9012	**0.037**
hbd2_w1	62425	3.0098	**0.007**
hbd2_w2	70010	3.3755	**0.004**
hbd2_w3	32106	1.5480	0.088
hbd2_m2	33806	1.6300	0.104
CRP_sim	28707	1.3629	**0.015**
CRP_pre	15714	0.8102	0.631
CRP_w0	22573	1.1639	0.280
CRP_w1	51055	2.6324	**0.011**
CRP_w2	34122	1.7593	0.098
CRP_w3	41269	2.1278	0.055
CRP_m1	27106	1.3976	0.169
CRP_w5	41965	2.1637	**0.026**
CRP_w6	38860	2.0036	**0.029**
CRP_m2	31527	1.6255	0.095
CRP_m3	39972	2.0610	**0.034**
CRP_m4	23961	1.2354	0.252
CRP_m6	39945	2.0596	**0.024**
pIL6_w1	23180	1.1431	0.311
Citr_pre	78105	3.8519	**0.002**
Citr_w1	42029	2.0727	**0.021**
Citr_w3	33872	1.6705	0.069
CD3+_m1	28654	1.4500	0.123
CD3+_m2	17891	0.9054	0.501
CD4+_m1	30266	1.5316	0.090
CD4+_m2	38132	1.9296	**0.037**
CD8+_m1	38615	1.9541	**0.029**
CD8+_m2	21275	1.0766	0.337
B_m1	26588	1.3454	0.166
B_m2	41803	2.1154	**0.025**
mat_B_m1	26622	1.3472	0.167
mat_B_m2	41697	2.1100	**0.027**
immat_B_m1	28078	1.4208	0.133
immat_B _m2	40142	2.0313	**0.029**
NK_m1	63891	3.2331	**0.002**
NK_m2	44566	2.2552	**0.020**
Lymphocyte_count_sim	21541	1.0227	0.610
Lymphocyte_m2	22196	1.1232	0.324

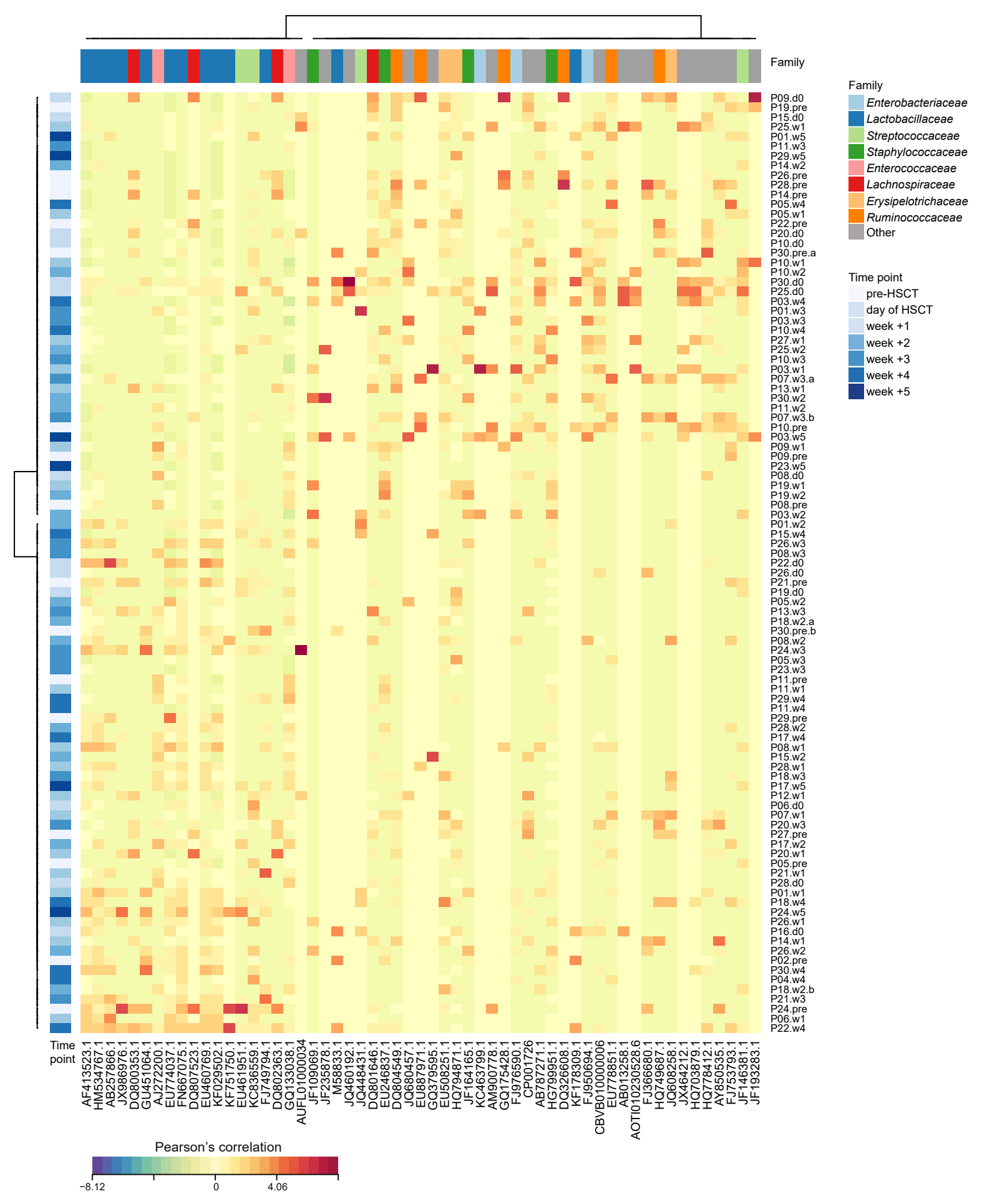
Lymphocyte_m3	24450	1.2373	0.224
Lymphocyte_m6	26278	1.3298	0.177
Lymphocyte_1y	29658	1.5008	0.114
mean_mono_pre	47349	2.3109	**0.013**
mean_mono_w3	64242	3.1353	**0.002**
mean_mono_m1	17924	0.8748	0.561
mean_mono_m2	25930	1.2655	0.187
mean_mono_m3	18269	0.8916	0.516
mean_mono_m6	1579	0.7708	0.654
mean_neutro_pre	25198	1.2298	0.225
mean_neutro_w3	31161	1.5208	0.149
mean_neutro_m1	29036	1.4171	0.137
mean_neutro_m2	16656	0.8129	0.611
mean_neutro_m3	48466	2.3653	**0.010**
mean_neutro_m6	15960	0.7789	0.653

OTU number	Phylum	Order	Family	Genus/Species/Description	Cluster	
					sPLS	CST
HM534767.1	*F*	*Lactobacillales*	*Lactobacillaceae*	*Lactobacillus* sp. Akpobro1	1	1
AF413523.1	*F*	*Lactobacillales*	*Lactobacillaceae*	*Lactobacillus pantheris*	1	1
KF751750.1	*F*	*Lactobacillales*	*Lactobacillaceae*	*Lactobacillus* sp.	1	1
AB257866.1	*F*	*Lactobacillales*	*Lactobacillaceae*	*Lactobacillus suebicus*	1	1
JX986976.1	*F*	*Lactobacillales*	*Lactobacillaceae*	*Lactobacillus aviarius* subsp. *araffinosus*	1	1
GU451064.1	*F*	*Lactobacillales*	*Lactobacillaceae*	*Lactobacillus* sp. 5-1-2	1	1
EU774037.1	*F*	*Lactobacillales*	*Lactobacillaceae*	*Lactobacillus* sp.	1	1
FN667075.1	*F*	*Lactobacillales*	*Lactobacillaceae*	*Lactobacillus* sp.	1	1
EU461951.1	*F*	*Lactobacillales*	*Streptococcaceae*	*Streptococcus* sp.	1	1
EU460769.1	*F*	*Lactobacillales*	*Lactobacillaceae*	*Lactobacillus* sp.	1	1
KC836559.1	*F*	*Lactobacillales*	*Streptococcaceae*	*Streptococcus salivarius* subsp. *thermophilus*	1	1
FJ749794.1	*F*	*Lactobacillales*	*Lactobacillaceae*	*Lactobacillus fermentum*	1	1
KF029502.1	*F*	*Lactobacillales*	*Lactobacillaceae*	*Lactobacillus casei*	1	1
AUFL01000034	*B*	*Bacteroidales*	*Porphyromonadaceae*	*Dysgonomonas capnocytophagoides* DSM 22835	1	1
DQ801646.1	*F*	*Clostridiales*	*Lachnospiraceae*	uncultured bacterium	2	2
EU246837.1	*F*	*Bacillales*	*Staphylococcaceae*	*Staphylococcus* sp. *Cobs2Tis23*	2	4
DQ804549.1	*F*	*Clostridiales*	*Ruminococcaceae*	*Faecalibacterium* sp.	2	2
FJ366680.1	*A*	*Bifidobacteriales*	*Bifidobacteriaceae*	*Bifidobacterium* sp.	2	2

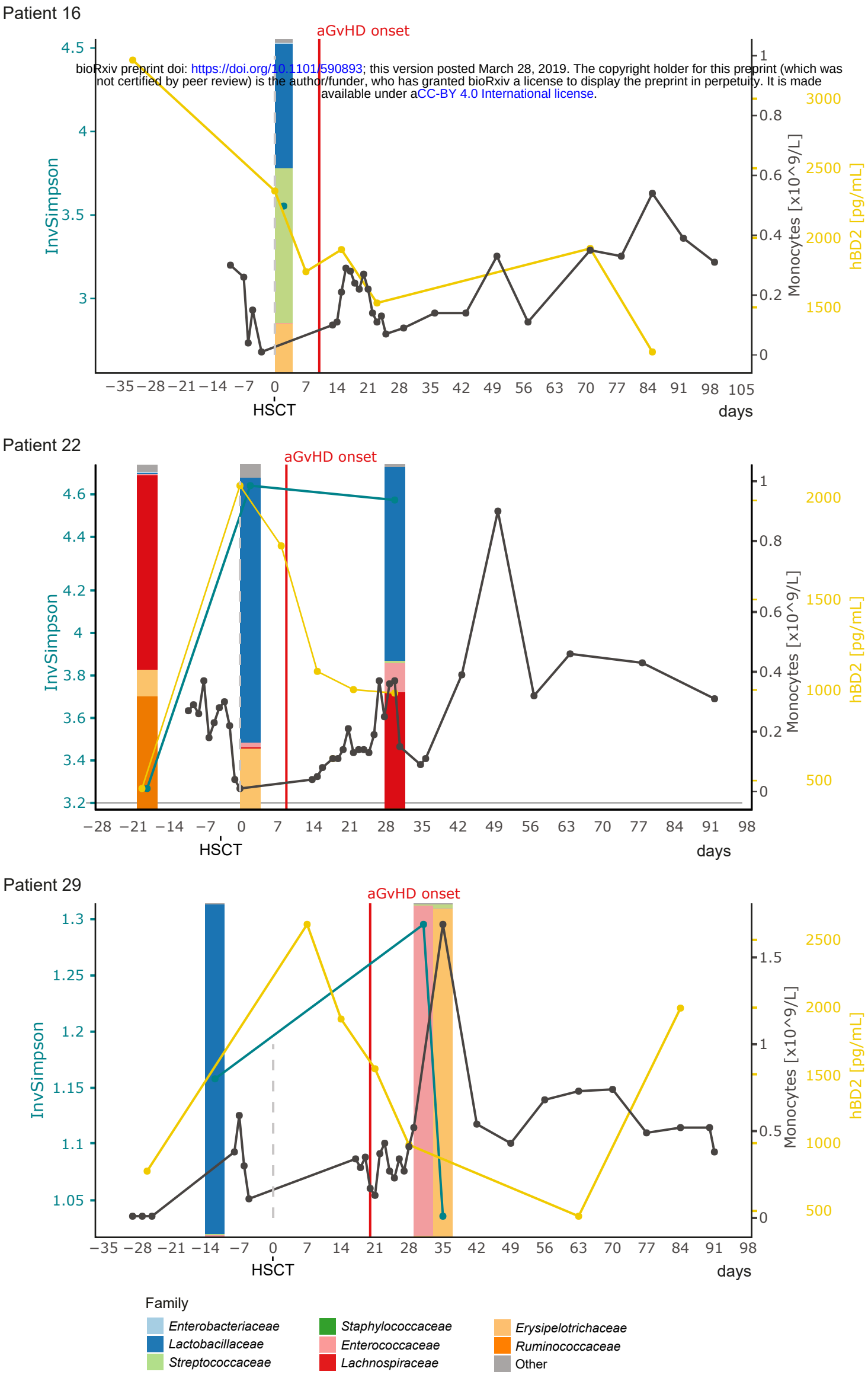
HQ749687.1	*F*	*Clostridiales*	*Ruminococcaceae*	uncultured bacterium	2	2
FJ753793.1	*F*	*Clostridiales*	*Peptostreptococcaceae*	swine fecal bacterium RF1A-Xyl2	2	2
DQ800353	*F*	*Clostridiales*	*Lachnospiraceae*	*Blautia* sp.	2	2
EU887971.1	*F*	*Clostridiales*	*Ruminococcaceae*	uncultured Clostridia bacterium	2	2
JQ608258.1	*F*	*Erysipelotrichales*	*Erysipelotrichaceae*	bacterium NLAE-zl-C558	2	2
AY850535.1		*Bifidobacteriales*	*Bifidobacteriaceae*	*Bifidobacterium* sp.	2	2
DQ326608.1	*F*	*Clostridiales*	*Ruminococcaceae*	uncultured bacterium	2	2
AJ272200.1	*F*	*Lactobacillales*	*Enterococcaceae*	*Enterococcus hirae*	2	4
EU508251.1	*F*	*Erysipelotrichales*	*Erysipelotrichaceae*	uncultured bacterium	2	2
DQ807523.1	*F*	*Clostridiales*	*Lachnospiraceae*	uncultured bacterium	2	2
CP001726	*A*	*Coriobacteriales*	*Coriobacteriaceae*	*Eggerthella lenta* DSM 2243	2	2
HQ794871.1	*F*	*Erysipelotrichales*	*Erysipelotrichaceae*	*uncultured organism*	2	2
JF193283.1	*F*	*Clostridiales*	*Family XI*	*Finegoldia* sp.	2	2
DQ802363.1	*F*	*Clostridiales*	*Lachnospiraceae*	*Blautia* sp.	2	2
GQ175428.1	*F*	*Clostridiales*	*Ruminococcaceae*	uncultured bacterium	2	2
EU778851.1	*F*	*Clostridiales*	*Ruminococcaceae*	*Subdoligranulum* sp.	2	2
GQ133038.1	*F*	*Lactobacillales*	*Enterococcaceae*	*Enterococcus* sp.	2	4
JQ448431.1	*F*	*Lactobacillales*	*Streptococcaceae*	*Streptococcus* sp.	3	1
JQ460192.1	*B*	*Bacteroidales*	*Prevotellaceae*	*Prevotella* sp.	3	3
M58833.1	*F*	*Lactobacillales*	*Lactobacillaceae*	*Pediococcus acidilactici*	3	1

HG799951.1	*F*	*Bacillales*	*Staphylococcaceae*	*Staphylococcus epidermidis*	3	3
JF235878.1	*A*	*Micrococcales*	*Microbacteriaceae*	*Naasia* sp.	3	3
JQ680457.1	*A*	*Corynebacterial es*	*Corynebacteriaceae*	*Corynebacterium* sp. DYS15	3	3
GQ379595.1	*P*	*Xanthomonadal es*	*Xanthomonadaceae*	*Stenotrophomonas* sp.	3	4
AB787271.1	*B*	*Bacteroidales*	*Bacteroidaceae*	*Bacteroides* sp. UasXn-3	3	3
AOTI010230528. 6	*B*	*Flavobacteriales*	*Flavobacteriaceae*	*Chryseobacterium Triticum urartu*	3	3
JF109069.1	*F*	*Bacillales*	*Staphylococcaceae*	*Staphylococcus* sp.	3	3
JX464212.1	*F*	*Bacillales*	*Paenibacillaceae*	*Paenibacillus* sp. *BJC15-C21*	3	3
HQ778412.1	*B*	*Bacteroidales*	*Bacteroidaceae*	*Bacteroides* sp.	3	2
FJ976590.1	*P*	*Enterobacteriale s*	*Enterobacteriaceae*	*Enterobacter* sp. *LCR81*	3	3
KF178309.1	*F*	*Lactobacillales*	*Lactobacillaceae*	*Pediococcus pentosaceus*	3	3
HQ703879.1	*F*	*Lactobacillales*	*Carnobacteriaceae*	*Alkalibacterium* sp.	3	3
FJ950694.1	*P*	*Enterobacteriale s*	*Enterobacteriaceae*	*Escherichia coli*	3	3
JF164165.1	*F*	*Bacillales*	*Staphylococcaceae*	*Staphylococcus* sp.	3	3
KC463799.1	*P*	*Enterobacteriale s*	*Enterobacteriaceae*	*Klebsiella* sp.	3	3
JF146381.1	*F*	*Lactobacillales*	*Streptococcaceae*	*Streptococcus* sp.	3	3
AM900778.1	*F*	*Bacillales*	*Paenibacillaceae*	*Paenibacillus* sp. PA215	3	3
CBVB010000006	*B*	*Bacteroidales*	*Bacteroidaceae*	bacterium MS4	3	3
AB013258.1	*P*	*Neisseriales*	*Neisseriaceae*	uncultured beta proteobacterium	3	3


A)
Cluster 1
Cluster 2
Cluster 3
aGvHD_II−IV
Survival_no
hBD2_pre
hBD2_w2
hBD2_w1
mono_pre
neutro_m3
Citr_w1
Female
CRP_w6
sibling_donor
Age
CRP_w1
unrel_donor
CRP_w5
Citr_pre
B_m2
CD4+_m2
mat_B_m2
CRP_m3
NK_m2
benign
Survival_yes
mono_w3
NK_m1
aGvHD_0−I
Axis2 [15.42% variance]
Axis1 [18.97% variance]
B)
Cluster 2
Cluster 3
Cluster 1
Sulfamethizole
graft_BM/UC
Amoxicillin
Fusidic_acid
Dicloxacillin
Ceftazidim
Metronidazol
Vancomycin_0
mat_B_m2
NK_m1
NK_m2
aGVHD_0−I
Piperacillin
mono_w3
Teicoplanin
Survival_yes
benign
CRP_m3
Linezolid
B_m2
CD4+_m2
Colistin
CRP_w1
Ciprofloxacin_0
CRP_w5
sibling_donor
unrelated_donor
Citr_pre
CRP_w6
Female
CRP
Citr_w1
Ciprofloxacin
neutro_m3
Vancomycin
hBD2_w1
mono_pre
hBD2_pre
hBD2_w2
aGVHD_II−IV
Survival_no
graft_UC
Cefuroxime
Chloramphenicol
Hexamycin
Axis2 [15.42% variance]
Axis1 [18.97% variance]


Family
Family
Enterobacteriaceae
Lactobacillaceae
Streptococcaceae
Staphylococcaceae
Enterococcaceae
Lachnospiraceae
Erysipelotrichaceae
Ruminococcaceae
Other
Time point
pre-HSCT
day of HSCT
week +1
week +2
week +3
week +4
week +5
P09.d0
P19.pre
P15.d0
P25.w1
P01.w5
P11.w3
P29.w5
P14.w2
P26.pre
P28.pre
P14.pre
P05.w4
P05.w1
P22.pre
P20.d0
P10.d0
P30.pre.a
P10.w1
P10.w2
P30.d0
P25.d0
P03.w4
P01.w3
P03.w3
P10.w4
P27.w1
P25.w2
P10.w3
P03.w1
P07.w3.a
P13.w1
P30.w2
P11.w2
P07.w3.b
P10.pre
P03.w5
P09.w1
P09.pre
P23.w5
P08.d0
P19.w1
P19.w2
P08.pre
P03.w2
P01.w2
P15.w4
P26.w3
P08.w3
P22.d0
P26.d0
P21.pre
P19.d0
P05.w2
P13.w3
P18.w2.a
P30.pre.b
P08.w2
P24.w3
P05.w3
P23.w3
P11.pre
P11.w1
P29.w4
P11.w4
P29.pre
P28.w2
P17.w4
P08.w1
P15.w2
P28.w1
P18.w3
P17.w5
P12.w1
P06.d0
P07.w1
P20.w3
P27.pre
P17.w2
P20.w1
P05.pre
P21.w1
P28.d0
P01.w1
P18.w4
P24.w5
P26.w1
P16.d0
P14.w1
P26.w2
P02.pre
P30.w4
P04.w4
P18.w2.b
P21.w3
P24.pre
P06.w1
P22.w4
Time point
AF413523.1
HM534767.1
AB257866.1
JX986976.1
DQ800353.1
GU451064.1
AJ272200.1
EU774037.1
FN667075.1
DQ807523.1
EU460769.1
KF029502.1
KF751750.1
EU461951.1
KC836559.1
FJ749794.1
DQ802363.1
GQ133038.1
AUFL01000034
JF109069.1
JF235878.1
M58833.1
JQ460192.1
JQ448431.1
DQ801646.1
EU246837.1
DQ804549.1
JQ680457.1
EU887971.1
GQ379595.1
EU508251.1
HQ794871.1
JF164165.1
KC463799.1
AM900778.1
GQ175428.1
FJ976590.1
CP001726
AB787271.1
HG799951.1
DQ326608.1
KF178309.1
FJ950694.1
CBVB01000006
EU778851.1
AB013258.1
AOTI01023528.6
FJ366680.1
HQ749687.1
JQ608258.1
JX464212.1
HQ703879.1
HQ778412.1
AY850535.1
FJ753793.1
JF146381.1
JF193283.1
Pearson's correlation
−8.12
0
4.06


acute GvHD
5-year survival
Community state type
1
2
3
4
grade 0-I
grade II-IV
alive
dead
GQ133038.1
GQ156251.1
AF413523.1
JF496526.1
EU774037.1
KF029502.1
HQ794871.1
JF071738.1
HG799951.1
EU467701.1
GU166170.1
EU508251.1
DQ801646.1
FN667075.1
HM534767.1
AM278246.1
AY977101.1
EU459226.1
JF803601.1
JF189758.1
AX491683
FJ370624.1
HG798420.1
DQ804549.1
HQ795338.1
EU774890.1
KC836559.1
EU460769.1
CBVB010000006
FJ976590.1
JQ608086.1
HQ778412.1
DQ799171.1
EU772409.1
CP001726
AJ272200.1
JQ607879.1
DQ796873.1
HQ816624.1
EU774788.1
AF349930.1
EU778851.1
JF241304.1
EU776581.1
AB696384.1
AY850535.1
FJ749794.1
HQ767494.1
HQ742544.1
JF223020.1
Enterococcus sp.
Streptococcus sp.
Lactobacillus pantheris
Enterococcus faecium
Lactobacillus sp.
Lactobacillus casei
uncultured bacterium
Streptococcus sp.
Staphylococcus epidermidis
uncultured bacterium
Enterococcus sp.
uncultured bacterium
uncultured bacterium
Lactobacillus sp.
Lactobacillus sp. Akpobro1
Lactococcus sp.
uncultured bacterium
Clostridium sensu stricto 1
Actinomyces odontolyticus
Streptococcus sp.
Bifidobacterium longum subsp. longum
uncultured bacterium
Enterococcus faecium
Faecalibacterium sp.
Subdoligranulum sp.
Weissella sp.
Streptococcus salivarius subsp. thermophilus
Lactobacillus sp.
Bacteroidaceae bacterium MS4
Enterobacter sp. LCR81
bacterium NLAE−zl−C377
Bacteroides sp.
uncultured bacterium
uncultured bacterium
Eggerthella lenta DSM 2243
Enterococcus hirae
Blautia sp.
bacterium NLAE−zl−C140
Blautia sp.
uncultured bacterium
Enterococcus sp. CR−303S
Subdoligranulum sp.
Granulicatella sp.
uncultured bacterium
uncultured bacterium
Bifidobacterium sp.
Lactobacillus fermentum
uncultured organism
Blautia sp.
uncultured bacterium
P01.w1 P06.w1 P24.w5 P26.w1 P06.d0 P01.w2 P16.d0 P22.w4 P05.w2 P24.w3 P24.pre P21.pre P26.w2 P22.d0 P30.w4 P08.w1 P26.w3 P21.w3 P30.pre.b P29.pre P30.pre.a P15.d0 P20.w3 P10.d0 P19.pre P19.d0 P27.pre P20.d0 P07.w3.a P22.pre P20.w1 P07.w3.b P14.pre P13.w1 P12.w1 P28.pre P26.pre P13.w3 P14.w1 P10.w2 P10.w1 P07.w1 P30.d0 P25.w1 P05.pre P05.w1 P25.d0 P05.w4 P03.w4 P10.pre P10.w3 P01.w3 P03.w3 P01.w5 P08.w2 P10.w4 P03.w5 P30.w2 P27.w1 P25.w2 P10.w3 P03.w1 P14.w2 P26.d0 P04.w4 P23.w3 P29.w5 P05.w3 P23.w4 P17.w2 P09.w1 P21.w1 P28.d0 P08.d0 P18.w4 P19.w1 P17.w5 P18.w3 P28.w1 P15.w2 P18.w2.a P02.pre P11.w2 P19.d0 P17.w4 P18.w2.b P28.w2 P19.w2 P08.pre P11.w4 P29.w4 P11.w1 P11.pre P08.w3 P09.pre P11.w3
Sample
Family
Enterobacteriaceae
Lactobacillaceae
Streptococcaceae
Staphylococcaceae
Enterococcaceae
Lachnospiraceae
Erysipelotrichaceae
Ruminococcaceae
Other
Variance stabilized counts
5
10
15



Our results suggest that a high *Lactobacillaceae* abundance prior to the onset of aGvHD may point to a preventive effect, as these patients survived. A human clinical trial of *Lactobacillus rhamnosus* GG prebiotic gavage to HSCT patients at time of engraftment demonstrated no protection against GvHD [1]. This could mean that *Lactobacillaceae* may not play a key role in aGvHD development, at least not the particular *Lactobacillus rhamnosus* strain under the conditions used in this group of patients. However, the administered probiotic did not alter the abundance of *Lactobacillus* spp. in the patients' guts [1], suggesting that the strain was not able to establish and proliferate in the host environment in this situation. An intrinsic increase of *Lactobacillaceae* prior to aGvHD onset, as observed here, therefore might still play a role in reducing aGvHD. A recent study related to the use of a probiotic given to infants to prevent sepsis suggested that the time point of application of a specific *Lactobacillus* sp. strain as a synbiotic played a critical role in positive clinical outcomes [2]. Furthermore, a study on gut microbial immunomodulation emphasized the importance of characterizing bacteria at the strain-level, because individual strains can have different modulatory effects on the immune system [3]. Therefore, it would be of great interest to determine the identity and predicted function of the specific *Lactobacillus* spp. strains in our patients, and in particular, in those who exhibited an early high abundance of *Lactobacillus* spp., as compared with those who experienced an expansion of *Lactobacillus* spp. after aGvHD and who later died.

Interestingly, *Enterococcus* was not among the most relevant taxa identified by our multivariate analyses. Intestinal domination of *Enterococcus* spp. was not clearly associated with adverse outcomes in our subgroup of 30 patients, in contrast to previous findings [4–6]. It should be noted that these previous observations were made in adult allo-HSCT patients and were dependent on

the type and amount of antimicrobial treatment. In addition, to elucidate this discrepancy further, we are currently characterizing *Enterococcus* isolates from fecal samples of our patient group, to gain insight into bacterial strain-level differences.

Supplementary references

1. Gorshein E, Wei C, Ambrosy S, Budney S, Vivas J, Shenkerman A, et al. *Lactobacillus rhamnosus GG* probiotic enteric regimen does not appreciably alter the gut microbiome or provide protection against GVHD after allogeneic hematopoietic stem cell transplantation. Clin Transplant. 2017;31:e12947. doi:10.1111/ctr.12947.

2. Panigrahi P, Parida S, Nanda NC, Satpathy R, Pradhan L, Chandel DiS, et al. A randomized synbiotic trial to prevent sepsis among infants in rural India. Nature. 2017;548:407–12. doi:10.1038/nature23480.

3. Geva-Zatorsky N, Sefik E, Kua L, Pasman L, Tan TG, Ortiz-Lopez A, et al. Mining the Human Gut Microbiota for Immunomodulatory Organisms. Cell. 2017;168:928–943.e11. doi:10.1016/j.cell.2017.01.022.

4. Shono Y, van den Brink MRM. Gut microbiota injury in allogeneic haematopoietic stem cell transplantation. Nat Rev Cancer. 2018. doi:10.1038/nrc.2018.10.

5. Taur Y, Xavier JB, Lipuma L, Ubeda C, Goldberg J, Gobourne a., et al. Intestinal Domination and the Risk of Bacteremia in Patients Undergoing Allogeneic Hematopoietic Stem Cell Transplantation. Clin Infect Dis. 2012;55:905–14. doi:10.1093/cid/cis580.

6. Holler E, Butzhammer P, Schmid K, Hundsrucker C, Koestler J, Peter K, et al. Metagenomic Analysis of the Stool Microbiome in Patients Receiving Allogeneic Stem Cell Transplantation: Loss of Diversity Is Associated with Use of Systemic Antibiotics and More Pronounced in Gastrointestinal Graft-versus-Host Disease. Biol Blood Marrow Transplant. 2014;20:640–5. doi:10.1016/j.bbmt.2014.01.030.