Downloaded from orbit.dtu.dk on: Apr 20, 2024

DTU Library

Interactions between the Microbiota and the Host Immune System in Hematopoietic Stem Cell Transplantation

Masche, Anna Cäcilia

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Masche, A. C. (2018). *Interactions between the Microbiota and the Host Immune System in Hematopoietic Stem Cell Transplantation*. Technical University of Denmark.

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Interactions between the Microbiota and the Host Immune System in Hematopoietic Stem Cell Transplantation

PhD Thesis

Anna Cäcilia Masche

December 2018

Research Group for Genomic Epidemiology
National Food Institute
Technical University of Denmark

Academic supervisors

Principal Supervisor:

Associate Professor Sünje Johanna Pamp, PhD. Research Group for Genomic Epidemiology, National Food Institute, Technical University of Denmark, Kongens Lyngby, Denmark.

Co-supervisors:

Professor Klaus Gottlob Müller, MD, PhD. Institute for Inflammation Research, Department of Rheumatology and Spine Disease; and Department of Pediatrics and Adolescent Medicine, Copenhagen University Hospital Rigshospitalet, Copenhagen, Denmark.

Professor Frank M. Aarestrup, PhD. Research Group for Genomic Epidemiology, National Food Institute, Technical University of Denmark, Kongens Lyngby, Denmark.

Professor Ole Lund, PhD. Research Group for Genomic Epidemiology, National Food Institute, Technical University of Denmark, Kongens Lyngby, Denmark.

Assessment committee

Professor Tine Rask Licht, PhD. Research Group for Gut, Microbes and Health, National Food Institute, Technical University of Denmark, Kongens Lyngby, Denmark.

Marcel R. M. van den Brink, MD, PhD. Memorial Sloan Kettering Cancer Center, New York, USA.

Affiliate Professor Paal Skytt Andersen, PhD. Statens Serum Institute; and Department of Veterinary and Animal Sciences, Copenhagen University, Copenhagen, Denmark.

Table of contents

Preface i

Acknowledgements iii

Summary iv

Dansk resumé vi

Research articles included in this thesis viii

Abbreviations ix

1. Introduction 1

1.1 Motivation for this PhD project 1

1.2 Objectives of this PhD project 1

2. Background 2

2.1 The human microbiota 2

2.1.1 The gut microbiota 2

2.1.2 The oral microbiota 3

2.1.3 The nasal microbiota 3

2.2 Interactions of the microbiota and the immune system in the healthy host 4

2.2.1 The role of the microbiota in immune system maturation 4

2.2.2 Interactions of the microbiota and the innate immune system 5

2.2.3 Interactions of the microbiota and the adaptive immune system 7

2.2.4 Specific interactions of the oral and nasal microbiota with the immune system 8

2.3 The microbiota in acute and chronic disease 9

2.4 Hematopoietic stem cell transplantation 10

2.4.1 Principles of allogeneic HSCT 10

2.4.1.1 Acute graft-versus-host disease after allogeneic HSCT 11

2.4.2 Principles of autologous HSCT 13

2.4.3 Infectious complications and antimicrobial treatment in HSCT 13

2.4.4 Immune reconstitution after HSCT 15

2.5 Current knowledge about the role of the microbiota in HSCT 17

3. Methodological considerations 20

3.1 Study populations 20

3.2 Longitudinal study design and sample collection 20

3.3 DNA extraction 21

3.4 16S rRNA gene sequencing 21

3.5 Variant inference from 16S rRNA sequences 22

3.6 Integration, transformation, and multivariate multi-table analyses of data 23

3.7 Considerations for time series data analyses and predictive modeling 24

4. Manuscripts 25

4.1 Manuscript I 25

4.2 Manuscript II 60

4.3 Manuscript III 100

5. Discussion 131
5.1 Microbiota reconstitution: Temporal dynamics of bacterial diversity and bacterial abundances in HSCT 131
5.2 Host-gut microbial associations in HSCT 132
5.2.1 The gut microbiota is associated with immune reconstitution after allogeneic HSCT 132
5.2.2 Host-gut microbial associations in autologous HSCT – The relevance of obligate anaerobes across transplantation modes 133
5.3 Beyond the gut – The role of the oral and nasal microbiota in allogeneic and autologous HSCT 134
5.4 Prediction of aGvHD from pre-transplant *Lactobacillus* sp. and other gut, oral, and nasal microbial taxa 136
5.5 Strengths and limitations 137
6. Conclusions 138
7. Future perspectives 139
References 140
Appendix 156
Supplementary Material – Manuscript I 156
Supplementary Material – Manuscript II 172
Supplementary Material – Manuscript III 186

Preface

This thesis presents the work I conducted between October 2015 and December 2018 in the Research Group for Genomic Epidemiology, National Food Institute, Technical University of Denmark. The project was carried out in collaboration with the Copenhagen University Hospital Rigshospitalet, Copenhagen, Denmark. The 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.

Chapter 1 introduces the motivation and objectives of this PhD project. In chapter 2, a general overview is provided of the human microbiota and its interactions with the host immune system in health and disease. Subsequently, an introduction is given to hematopoietic stem cell transplantation (HSCT), and the role of the microbiota in HSCT. The latter presents knowledge from studies other than this thesis. The contributions of the work within this thesis is related to recent literature in chapter 5 (see below). In Chapter 3, I discuss the methodological challenges faced during this project, and the decisions that were made to overcome them. In the fourth chapter, the three manuscripts that are part of this thesis are presented. Manuscript I has been submitted and is currently undergoing the second round of review. Manuscripts II and III are in preparation.

Chapter 5 constitutes a review of host-microbiota interactions in HSCT, discussing current knowledge including the contributions of this PhD thesis. This chapter is also intended to serve as a concise summary of the findings presented in the three manuscripts. However, to avoid redundancy, the discussion focusses on findings across studies, and does not repeat every individual finding. This serves the purpose of answering the key questions of this PhD project with respect to different study populations (allogeneic versus autologous HSCT patients; children versus adults), and microbiota at different body sites (gut, and oral and nasal cavities). Chapter 6 comprises concise, general conclusions corresponding to the key questions of the project stated in chapter 1. More specific conclusions serve as the base for future perspectives in chapter 7.

Ethics approval and consent to participate

For all studies included in this thesis, written informed consent for participation was obtained from patients and/or their legal guardians. Study protocols were approved by the local ethics committee and the Danish Data Protection Agency. Approval numbers are provided within the respective manuscripts.

Acknowledgements

I would like to thank a number of people who have been part of my PhD journey and without whom it would not have been possible to carry it out. I have had great supervisors in Sünje Johanna Pamp, Klaus Gottlob Müller, Frank Møller Aarestrup, and Ole Lund.

First of all, I am thanking my main supervisor Sünje Johanna Pamp for her incredible support and guidance. You inspired and motivated me throughout my time as a PhD student, and created great opportunities for me. You taught me how to stand up for my decisions and to step out of my comfort zone, and reassured me that I am capable whenever I was in doubt. Thank you for being a great supervisor, and for taking the time to step back and reflect on the process together with me whenever needed.

I am thanking my co-supervisor Frank Møller Aarestrup for giving me the opportunity to perform my PhD project in his group, and for once in a while reminding me to step back and see the bigger picture in my project. Thank you for making time for invaluable support during the final phase of the project. Furthermore, I would like to thank my co-supervisor Klaus Gottlob Müller for providing the clinical perspective for this project, and for his great support and insights. I would also like to thank Katrine Kielsen for tirelessly supporting my research with her indispensable clinical and immunological knowledge.

Many thanks go to the patients and their parents for agreeing to participate in a research project during such hard times. Furthermore, I would like to thank all nurses and doctors at Rigshospitalet who fitted the collection of samples for this PhD project into their busy clinical routines.

I would like to thank Professor Susan Holmes for giving me the unique opportunity to spend my external stay in her lab at Stanford University. I have learned a lot about statistics for microbiome research from her and Pratheepa Jeganathan.

All past and present members of the Research Group for Genomic Epidemiology have at all times made me feel welcome and created a great working environment. Thanks to Vibeke Dybdahl Hammer and Jeffrey Edward Skiby for their invaluable administrative support. Many thanks go to Marie Stengaard Jensen for helping with sample transport and especially for keeping me company in the cold hospital basement. I am grateful for the help of Hanne Mordhorst in sharing the load of extracting DNA from over a thousand samples. Thanks also to the team of lab technicians for welcoming me in the lab, for helping with one or the other task, and for letting me take advantage of them for practicing my Danish.

I could not have had better office mates. In the small office next to the coffee machine, shared with Mathis and Casper, we never ran short on laughs and adventurous stories. Casper kept being my desk neighbor, and ggplot-partner in crime for 3 years. Nanna, Marie, and Ann-Sofie have always been there when I needed to be cheered up over a therapeutic piece of canteen cake. I had great discussions about statistics with Patrick N, mixed with exciting kindergarden stories. Jolene has recently become a great part of the team and provided helpful language support. Also, thank you for always lending an ear for questions related to bioinformatics, R, or microbiome analysis to Oksana, Rolf, Shinny, Patrick M, Philipp K, and Saria.

I would like to thank my mother Susanne for believing in me and encouraging me to follow my dreams, even though that involved moving 400 km away. And last but not least, I thank my fiancé Dennis who was by my side through all the highs and lows of PhD-life with incredible patience and optimism, and who kept reminding me of the importance to celebrate the small daily successes.

Summary

The commensal human microbiota and the human immune system form an inseparable unit. Reciprocal effects ensure their homeostasis. The microbiota is crucial for immune system maturation and contributes directly and indirectly to innate and adaptive immune defense mechanisms. In turn, the immune system tolerates and selectively spares commensal microbial colonizers, and shapes the community structure. Immune-microbial interactions are crucial in the healthy host. Upon acute or chronic disease, especially when involving infections and inflammation, changes in bacterial diversity and in the residential microbial community structure have become evident. In a number of conditions including cancer, the state of the microbiota is associated with clinical outcomes and the patients' recovery from disease might to a certain extent depend on it. This attracts attention to the microbiota as a promising resource for future clinical predictive and prognostic markers, and even as a target for therapeutic modulation. However, the underlying mechanisms that play a role in homeostatic and disturbed immune-microbial interactions in the human host are not completely understood. A concept that is of interest for studying the role of the microbiota constitutes hematopoietic stem cell transplantation (HSCT). It involves the treatment of patients with e.g. hematologic malignancies, such as leukemia, through the infusion of a donor's (allogeneic) or their own (autologous) stem cells. Preceding HSCT, patients undergo preparative immunosuppressive and chemotherapeutic regimens, and antimicrobial prophylaxis. These treatments introduce immuno-compromisation and changes in the commensal microbiota.

The aim of this thesis was to characterize interactions of the microbiota and the human immune system in the context of HSCT. For cohorts of allogeneic and autologous HSCT patients, we implemented longitudinal study designs spanning the time from before, at the time of, and after transplantation. By using 16S rRNA gene profiling, we assessed temporal dynamics of the gut, oral, and nasal microbiota. We determined changes in bacterial alpha diversity, proportional changes on taxonomic family level, and temporal abundance trajectories of specific phylogenetically closely related sequence variants. In multivariate multi-table analyses, we examined whether changes in the microbial community structure are associated with immune cell counts, immune markers and clinical outcomes after transplantation. We accounted for patient characteristics, such as age and sex, underlying disease, transplantation modalities, such as stem cell source, and treatment regimens. Moreover, we assessed in a machine learning approach whether microbial abundances prior to transplantation can be used to predict the development of graft-versus-host disease (GvHD) after allogeneic HSCT.

In the first study, we investigated host-microbial associations in children undergoing allogeneic HSCT. We focused on dynamics in the gut microbiota around the time of transplantation and during the first month post-HSCT. We found that high abundances of the *Lactobacillaceae* family were associated with the occurrence of moderate to severe acute GvHD and high mortality. These concerned patients also had high plasma levels of the antimicrobial peptide human beta-defensin 2 (hBD2), which we assessed for the first time in the HSCT context. High abundances of the families *Ruminococcaceae* and *Lachnospiraceae* were associated with rapid reconstitution of NK and B cells. Moreover, treatment with specific antibiotics were associated with high *Enterobacteriaceae* abundances.

In the second study, we were interested in studying long-term microbial abundance dynamics in relation to pediatric allogeneic HSCT and therefore monitored the microbiota over a period of one year. We also examined the microbiota at additional mucosal sites, namely the oral and nasal cavities. With the aid of

phylogenetic trees, we identified taxa that discriminated the microbial community structure between time points. These taxa included for instance *Enterococcus* spp. and *Blautia* spp. in the gut. While the temporal trajectory of *Enterococcus* spp. showed an expansion after transplantation, Blautia spp. decreased in abundance. Similarly, abundances of *Actinomycetaceae* in the oral cavity and *Corynebacteriaceae* in the nasal cavity decreased post-transplant. Common to all was the regain of pre-transplant like abundance levels after approximately three months. Furthermore, pre-transplant abundance of specific taxa at all three body sites, e.g. intestinal *Parabacteroides distasonis*, in this cohort showed potential to predict subsequent acute GvHD development. We contextualized these findings by integrating microbial abundances with the above named host parameters in multivariate analyses. This revealed similarities in host-microbial associations at the gut, oral, and nasal sites.

The third study comprised investigations within a cohort of adult autologous HSCT recipients. Previous research including our first two studies focused primarily on allogeneic HSCT. Here, we revealed associations of gut, oral, and nasal microbial abundances with host immune parameters in autologous HSCT patients. For instance, the abundances of oral *Prevotella* spp. correlated with counts of T lymphocyte subpopulations. Since the findings in this study are based on a rather small group of patients, further research is needed to validate our observations.

This thesis extends the knowledge about interactions of the microbiota and the host immune system in hematopoietic stem cell transplantation. We extend the knowledge about commensal microbial shifts after HSCT by characterizing not only patterns of disturbance early after transplantation, but also the reconstitution of the microbiota at late follow-up time points. The 'three months' time point might provide a crucial temporal threshold for evaluating whether microbial homeostasis has been regained. Microbial associations with specific immune markers, such as hBD2, were assessed, which are novel within the HSCT context. We suggest further targeted research to evaluate the clinical relevance of these markers. The correlation between abundances of specific commensal taxa and adaptive and innate immune cell reconstitution after HSCT constitutes an interesting novel finding that is in line with the role of the microbiota in immune system maturation in the health host. This emphasizes the importance of microbial homeostasis in HSCT patients and warrant careful considerations with regards to antimicrobial treatment in these patients. We revealed host-microbial associations shared across body sites, e.g. certain immune cell counts that correlate with microbial abundances of gut, oral, and nasal taxa. Since the feasibility of collecting oral and nasal swab samples is higher compared with fecal sample collection, these findings might open opportunities for future clinical routine screening of the microbiota, and thereby to take advantage of its possible value in guiding personalized treatment strategies. We showed that acute GvHD post-transplantation might be predicted from microbial abundances pre-HSCT in our cohort, and propose that the predictive potential of specific taxa is worth validating in further cohorts. Specific microbial markers might in the future aid with identifying patients at risk of adverse outcomes and allow for early preventive interventions. Our initial findings in a small cohort of autologous HSCT patients highlight the relevance of investigating the role of the microbiota in this patient group, which is also of importance as a reference for allo-HSCT cohorts concerning donor-related effects.

Dansk resumé

Den kommensale humane mikrobiota og det humane immunsystem danner en uadskillelig enhed og sammenspillet mellem dem sikrer deres homøostase. Mikrobiotaen er afgørende for modning af immunsystemet og bidrager direkte og indirekte til medfødte og adaptive immunforsvarsmekanismer. Immunsystemet tåler og skåner selektivt kommensale bakterier og former den mikrobielle samfundsstruktur. Immun-mikrobielle interaktioner er afgørende for den sunde humane vært. Ved akut eller kronisk sygdom, især når det drejer sig om infektioner og betændelse, er ændringer i bakteriel diversitet og i den mikrobielle samfundsstruktur blevet vist. I en række sygdomme, herunder kræft, spiller mikrobiotaen en rolle, og patienternes helbredsmæssige bedring kan i et vist omfang afhænge af den. Netop dette gør mikrobiotaen til en potentiel kilde for fremtidige kliniske prædiktive og prognostiske markører, og også til en komponent som muligvis kan moduleres som en del af behandlingen. Imidlertid er de underliggende mekanismer, der spiller en rolle i homøostatiske og forstyrrede immun-mikrobielle interaktioner i den menneskelige vært, ikke fuldstændig forstået. Et koncept, som er af interesse for at studere mikrobiotaens rolle, udgør hæmatopoietisk stamcelletransplantation (HSCT). Dette involverer behandling af patienter med f.eks. hæmatologiske maligne sygdomme, såsom leukæmi, ved infusion af en donors (allogen) eller deres egne (autolog) stamceller. Forudgående HSCT undergår patienter præparative immunosuppressive og kemoterapeutiske regimer og antimikrobielle profylakse. Disse behandlinger gør at patienterne bliver immunsvækkede, og at der efterfølgende sker ændringer i deres kommensale mikrobiota.

Formålet med denne afhandling var at karakterisere interaktioner mellem mikrobiotaen og det humane immunsystem i forbindelse med HSCT. For kohorter af allogen og autolog HSCT-patienter gennemførte vi longitudinelle studier, der dækker perioden før, under, og efter transplantationen. Ved at anvende 16S rRNA gen sekventering, vurderede vi de tidsmæssige forandringer i tarm samt forandringer af oral og nasal mikrobiota. Vi viste ændringer i bakteriel alfa-diversitetet, proportionale ændringer på taksonomisk familieniveau, og tidsmæssige ændringer af specifikke fylogenetisk nært beslægtede 'sekvensvarianter'. I multivariate multi-tabel analyser undersøgte vi, om ændringer i den mikrobielle samfundsstruktur korrelerede med antallet af immunceller, immunmarkører og kliniske 'outcomes' efter transplantation. Vi tog højde for patienternes alder, køn, underliggende sygdomme, transplantations modaliteter, såsom stamcellekilde og behandlingsregimer. Desuden vurderede vi, ved brug af 'machine learning', om antallet af visse bakterier før transplantationen kan anvendes til at forudsige udviklingen af graft-versus-host-sygdom (GvHD) efter allogen HSCT.

I det første studie undersøgte vi værtsmikrobielle korrelative sammenhænge hos pædiatrisk allogen HSCT patienter. Vi fokuserede på forandringer i tarmens mikrobiota omkring transplantationstidspunktet og i den første måned efter HSCT. Vi fandt, at et højt antal af bakterier tilhørende *Lactobacillaceae*-familien opstod sammen med forekomsten af moderat eller alvorlig akut GvHD og høj dødelighed. Disse patienter havde også høje plasma niveauer af det antimikrobielle peptid human beta defensin 2 (hBD2), som vi vurderede for første gang i HSCT-konteksten. Et højt antal af bakterier tilhørende familierne *Ruminococcaceae* og *Lachnospiraceae* blev observeret sammen med hurtig rekonstitution af NK og B-celler. Desuden var behandling med specifikke antibiotika forbundet med et højt antal af *Enterobacteriaceae*.

I vores anden studie var vi interesserede i at undersøge langsigtede mikrobielle forandringer i forhold til pædiatrisk allogen HSCT og overvågede derfor mikrobiotaen over en periode på et år. Vi undersøgte også mikrobiotaen på andre slimhinder, nemlig mund- og næsehule. Ved hjælp af fylogenetiske træer identificerede vi taxa, der adskilte den mikrobielle samfundsstruktur på forskellige tidspunkter. Disse taxa omfattede f.eks. *Enterococcus* spp. og *Blautia* spp. i tarmen. Vi fandt at antallet af *Enterococcus* spp. steg efter transplantationen mens antallet af *Blautia* spp. faldt. Tilsvarende faldt antallet af *Actinomycetaceae* i mundhulen og *Corynebacteriaceae* i næsehulen efter transplantationen. Fælles for dem alle var, at antallet af bakterier tilsvarende til før transplantationen blev opnået igen efter ca. tre måneder. Endvidere var det muligt i denne kohorte, at forudsige udvikling af akut GvHD fra antallet af specifikke taxa i tarmen, og mund- og næsehulen, f.eks. *Parabacteroides distasonis* i tarmen, før HSCT. Vi kontekstualiserede disse resultater ved at integrere antaller af bakterier med ovennævnte værtsparametre i multivariate analyser. Dette afslørede overenstemmelser i værtsmikrobielle korrelative sammenhænge vedrørende bakterier i tarmen, og i mund- og næsehulen.

Det tredje studie omfattede undersøgelser inden for en kohorte af voksne autologe HSCT-patienter. Tidligere undersøgelser, inklusiv vores første to studier, fokuserede primært på allogen HSCT. Her viste vi korrelationer mellem antaller af bakterier i tarmen, og i mund- og næsehulen og værtsparametre hos autologe HSCT-patienter. For eksempel, korrelerer antallet af oral *Prevotella* spp. med antallet af T-lymfocyt-subpopulationer. Fordi resultaterne i dette studie er baseret på en ret lille gruppe patienter, er der behov for yderligere forskning for at validere vores observationer.

Denne afhandling bidrager til øget viden om interaktioner mellem mikrobiotaen og værtsimmunsystemet i forbindelse med hæmatopoietisk stamcelletransplantation. Vi udvider kendskabet til kommensale mikrobielle forandringer efter HSCT ved at karakterisere ikke kun forstyrrelsen af den mikrobielle samfundsstruktur tidligt efter transplantationen, men også rekonstituering af mikrobiotaen sener efter HSCT. 'Tre måneders tidpunktet' kunne give en afgørende tidsgrænse for at vurdere, om mikrobiel homøostase er blevet genvundet. Mikrobielle korrelationer med visse immunmarkører, såsom hBD2, blev for første gang undersøgt i HSCT-konteksten. Vi foreslår yderligere målrettet forskning for at vurdere den kliniske relevans af disse markører. Den potentielle sammenhæng mellem antallet af specifikke kommensale taxa, og rekonstituering af adaptiv og medfødt immunsystemet efter HSCT udgør et interessant og nyt resultat, som er i overensstemmelse med mikrobiotaens rolle i immunsystemets modning i den sunde vært. Dette understreger vigtigheden af mikrobiel homøostase hos HSCT patienter og fremtidelige overvejelser med hensyn til antimikrobiel behandling hos disse patienter anbefales derfor. Vi afslørede værtsmikrobielle korrelationer, f.eks. antal af visse immunceller, der korrelerer med antal af specifikke bakterier i tarmen, og mund- og næsehulen. Da der er bedre mulighed for at indsamle oral og nasal svabeprøver sammenlignet med afføringsprøver, kan disse resultater påpege muligheder for fremtidig undersøgelser af mikrobiotaen på hospitalet og derved muligvis være til gavn af styring af personlige behandlingsstrategier. Vi viste, at akut GvHD kunne forudsiges baseret på antal af visse bakterier forud for HSCT i vores kohort, og vi foreslår, at det prediktive potentiale af specifikke taxa skal valideres i yderligere større kohorter. Specifikke mikrobielle markører kan i fremtiden muligvis hjælpe med at identificere patienter med risiko for negative 'outcomes' og give mulighed for tidlige forebyggende indgreb. Vores forløbige resultater af en lille gruppe autologe HSCT-patienter fremhæver relevansen af at undersøge mikrobiotaens rolle i denne patientgruppe, hvilket også er vigtigt som reference for allo-HSCT kohorter vedrørende donorrelaterede effekter.

Research articles included in this thesis

I. **Masche A C**, Kielsen K, Cilieborg M S, Lund O, Holmes S, Aarestrup F M, Müller K G, Pamp S J. Specific gut microbiome members are associated with distinct immune markers in allogeneic hematopoietic stem cell transplantation. *Under review.*

II. **Masche A C**, Kielsen K, Mordhorst H, Aarestrup F M, Müller K G, Pamp S J. The gut, oral, and nasal microbiota in pediatric allogenic HSCT and microbial marker candidates for predicting acute GvHD severity. *Manuscript in preparation.*

III. **Masche A C**, Kielsen K, Mordhorst H, Aarestrup F M, Müller K G, Pamp S J. Bacterial alpha diversity trajectories and immune-microbial associations of the gut, oral, and nasal microbiota in adult autologous HSCT patients. *Manuscript in preparation.*

Abbreviations

3-IS: 3-indoxyl sulfate
ALL: Acute lymphoblastic leukemia
AML: Acute myeloid leukemia
AMP: Antimicrobial peptide
APC: Antigen-presenting cells
ATG: Antithymocyte globulin
ASV: Amplicon sequence variant
BSI: Blood stream infections
CCR2: CC-chemokine receptor 2
CD: Crohn's disease
CLL: Chronic lymphocytic leukemia
CMV: Cytomegalovirus
CRP: C-reactive protein
DC: Dendritic cell
FMT: Fecal microbiota transplantation
GI: Gastrointestinal
GF: Germ-free
GPR43: G protein-coupled receptor 43
(a)GvHD: (acute) Graft-versus-host disease
GvL effect: Graft-versus-leukemia effect
hBD2/3: Human beta-defensin 2/3
HD5/6: Human defensin 5/6
HDACi: Histone deacetylase inhibitor
HIV: Human immunodeficiency virus
HL: Hodgkin lymphoma
HLA: Human leukocyte antigen
HSC: Hematopoietic stem cell
HSCT: Hematopoietic stem cell transplantation
IBD: Inflammatory bowel disease
IEC: Intestinal epithelial cell
(S)Ig: (secretory) Immunoglobulin
LDA: Linear discriminant analysis
LPS: Lipopolysaccharides
MCP1: Monocyte chemotactic protein-1
MDS: Myelodysplastic syndrome
MHC: Major histocompatibility complex
ML: Machine learning
MM: Multiple Myeloma
NALT: Nasal associated lymphoid tissue
NGS: Next-generation sequencing

NHL: Non-Hodgkin lymphoma
NK cell: Natural killer cell
NKT cell: Natural killer T cell
Nod1: Nucleotide-binding oligomerization domain–containing protein-1
OTU: Operational taxonomic unit
PAMP: Pathogen-associated molecular pattern
PRR: Pattern recognition receptors
RegIIIγ: Regenerating islet-derived protein III-gamma
SCFA: Short-chain fatty acid
SFB: Segmented filamentous bacteria
SLC5A8: Sodium-coupled monocarboxylate transporter 1
sPLS regression: Sparse partial least squares regression
TBI: Total body irradiation
T_H17 cell: T helper 17 cell
TLR: Toll-like receptor
T_{reg} cell: T regulatory cell
TRM: Treatment-related mortality

1. Introduction

1.1 Motivation for this PhD project

The development of the human immune system, both throughout the evolution of our species and in each newborn child, is highly dependent on the microorganisms that the human body comes to harbor. From this ongoing co-evolution emerge a homeostatic symbiosis and beneficial host-microbial interactions. However, the underlying mechanisms and mutual effects of the interplay between the human immune system and the commensal microbiota in health and disease are not completely understood. In this PhD project, these interactions were studied in a unique setting of particular relevance where components of the immune system are replaced, namely in patients undergoing hematopoietic stem cell transplantation.

1.2 Objectives of this PhD project

The key role of a homeostatic interplay between the residential microbiota and the host immune system in protection from inflammation and infection has been proposed by previous research. This PhD project aimed to investigate this interplay by making use of a scenario of 'host-transplantation', i.e. hematopoietic stem cell transplantation.

The objectives of this PhD thesis were to answer the following key questions:

1. How does a stem cell graft shape the recipients' microbiota after transplantation and what does that teach us about the influence of the host's adaptive immune system on the formation of a homeostatic microbiota in general?

2. How does a stem cell graft influence the recipients' innate immune components such as antimicrobial peptides and innate immune cells, and their capability to maintain microbial homeostasis during the transplant process?

3. Does the presence and abundance (and/or changes thereof) of certain bacterial taxa before and after transplantation correlate with and potentially predict the status of distinct immune markers and HSCT-associated side effects, such as infections and acute graft-versus-host disease?

4. Does the presence and abundance (and/or changes thereof) of certain bacterial taxa before and after transplantation correlate with and potentially predict clinical outcomes, such as relapse, and survival?

2. Background

2.1 The human microbiota

The human body is colonized by distinct bacterial communities at various body sites. Their numbers might give a glimpse of their importance: For every human cell, 1.3 bacterial cells are estimated to be present in the human body, constituting the bacterial part of the human microbiota [1]. The collection of their genes, the bacterial part of the microbiome, outnumbers us to an even larger extent. In the gut alone, the microbiome is suggested to comprise as many as 2,000,000 genes, 100 times more than human genes (estimated 20,000) [2, 3]. Distinct niches select for highly adapted sets of commensal colonizers at different body sites. In the following paragraphs, the microbial communities of the human gut, oral cavity, and nasal cavity, which were subject to research in this thesis, will be introduced in more detail.

2.1.1 The gut microbiota

The gut is the body's largest surface and is suggested to harbor up to 1,000 different bacterial species providing various crucial functionalities to their hosts [2, 3]. These include, among others, digestion, nutrient metabolism, as well as immune system maturation and modulation [4, 5]. The latter is facilitated by close, mutual interactions of the gut microbial community and the host immune system at the intestinal mucosal surface [6]. The current consensus points to the absence of a microbiota *in utero,* even though a few studies have reported bacteria present in the placenta or the amniotic fluid [7]. Either way, the well-studied succession of the human gut microbiota starts latest at birth, when the newborn child is exposed to maternal vaginal and skin commensals [8]. During the first year of life, the gut is predominantly inhabited by members of the phyla *Proteobacteria*, *Actinobacteria*, *Firmicutes*, *Bacteroidetes*, and *Verrucomicrobia* [9]. Among the first gut colonizers right after birth are aerobic and facultative anaerobic bacteria, e.g. *Enterobacteriaceae* (*Proteobacteria*) [8]. A few days later, depletion of initially high oxygen levels in the infant gut allows the growth of obligate anaerobes, such as *Bifidobacterium* spp. and *Bacteroides* spp. [8]. These human milk oligosaccharide utilizing bacteria typically dominate the infant's gut microbiota during the period of breast-feeding [8]. With the introduction of solid food, predominance shifts to taxa affiliated with the class *Clostridia*, such as *Blautia* spp. and *Ruminococcus* spp., as well as *Bacteroides* spp., which can process plant-derived polysaccharides [6, 10]. Even though the infant's diet is closely related to gut microbiota maturation, mode of delivery, gestational age at birth, maternal diet, host genetics, and environmental factors also play important roles, and are subject to intense research [6]. It is suggested that the gut microbiota acquires an adult-like composition within the first three years of life [7, 10]. However, the gut microbiota undergoes further essential changes from childhood to adulthood. In both, older children and young adolescents (7-14 years), and adults, the phyla *Firmicutes*, *Bacteroidetes*, *Proteobacteria*, *Verrucomicrobia*, and *Actinobacteria* represent the majority of gut commensals [11, 12]. Among older children and young adolescents, *Bacteroides* and *Alistipes* (both *Bacteroidetes*), the *Firmicutes* affiliated genera *Faecalibacterium*, *Dialister* and *Roseburia,* as well as genera within the *Ruminococcaceae* family are the most abundant [11, 12]. In adults, the abundance of *Bacteroides* increases compared to children and adolescents, while the abundance of *Bifidobacterium* spp. decreases [12].

2.1.2 The oral microbiota

After the gut, the oral cavity is the environment within the human body harboring the highest abundance of microbes. The oral microbiota differs by niche, e.g. saliva, dental plaque and buccal mucosa. Within the oral cavity, the environment of the buccal mucosa is unique as it is shedding and due to its direct interaction with the innate and the adaptive immune system, comparable to the intestinal mucosa. Furthermore, the composition of the oral microbiota varies with age as children undergo different dentition stages, from absence of dentition in neonates to primary, mixed, and eventually permanent dentition, constantly changing the oral environment [13].

In healthy children and adults, the oral microbiota is dominated by members of the phyla *Firmicutes, Proteobacteria, Bacteroidetes, Actinobacteria, Spirochaetes, Fusobacteria,* and *Saccharibacteria* (formely named TM7) [13, 14]. Major genera comprised in the oral 'core' microbiome of adults are *Streptococcus* and *Veillonella* (*Firmicutes*), *Campylobacter*, *Cardiobacterium*, *Haemophilus*, and *Neisseria* (*Proteobacteria*), *Actinomyces, Bergeyella, Capnocytophaga,* and *Prevotella (Bacteroidetes),* as well as *Atopobium, Corynebacterium,* and *Rothia (Actinobacteria)* [13, 15].

With increasing age, dominance of the major phyla persists, with an increased relative abundance of *Firmicutes* and *Proteobacteria,* and a decreased relative abundance of *Bacteroidetes* in adults compared to 4-6 year old children [14]. However, a more notable change of the oral microbial composition with age becomes evident with higher phylogenetic resolution and within individual niches of the oral cavity. The genus *Streptococcus*, a highly abundant member of the oral microbiota, was the only genus identified as part of a 'core' microbiome of the buccal mucosa independent of age [14]. Even though they are known members of the overall oral microbiota in adults, a number of genera that accounted for the 'core' microbiome of the buccal mucosa in neonates, children and young adults (up to 17 years) did not persist in the buccal mucosa 'core' microbiome of adults. Among those genera were e.g. *Granulicatella (Firmicutes)*, *Veillonella (Firmicutes), Haemophilus (Proteobacteria),* and *Prevotella (Baceroidetes)* [14].

2.1.3 The nasal microbiota

The various microenvironments within the nose play key roles in both, the respiratory and the immune system. The nasal cavity represents the part that is most accessible for studying the nasal microbiota. Among the different niches within the nasal cavity, the anterior naris is the most exposed to external influences. The non-keratinized, stratified squamous epithelium lining the anterior naris resembles that of the skin and is lined with sebaceous glands and vibrissae (coarse hairs) [16, 17].

Bacterial colonizers belonging to three phyla, namely *Actinobacteria*, *Firmicutes*, and *Proteobacteria*, account for the major part the of nasal cavity microbiota [16]. An interesting observation within this habitat is the colonization with *Staphylococcus aureus* in an estimated 30% of the human population [18]. *Staphylococcus aureus* carrier status has been related to various host and environmental factors, as well as to niche competition with commensals [16, 18–20].

Nasal cavity colonization undergoes a succession from infancy to adulthood [17]. During the first year of life, members of the *Moraxellaceae* family increase in abundance in healthy infants, while *Corynebacteriaceae* and *Staphylococcaceae* decrease [21, 22]. This may partially reflect a development away from an initial skin microbiota-like composition induced by close contact between mother and child

[21]. In prepubertal children (2-9 years), colonizers of the nares predominantly include taxa affiliated with the phyla *Firmicutes (*e.g. *Streptococcus, Dolosigranulum, Gemella,* and *Granulicatella)* and *Proteobacteria (e.g. Moraxella, Haemophilus,* and *Neisseria)* [23]. Dominance shifts to *Actinobacteria* in adults, with particularly high abundances of the genera *Corynebacterium*, and *Propionibacterium*, followed by *Staphylococcus* or *Dolosigranulum* (both *Firmicutes*, dependent on *Staphylococcus aureus* carrier status), and *Moraxella* (*Proteobacteria*) [16, 23]. The abundance of *Propionibacterium* members, which are sebum-utilizers, is particularly high in the anterior naris [16].

2.2 Interactions of the microbiota and the immune system in the healthy host

The human immune system does not only co-exist with and tolerate the commensal microbiota. The two components rather depend on each other and form an inseparable physiological unit that maintains and promotes homeostasis. That is, ensuring protection against pathogens while restricting inflammation to a minimum. The following paragraphs will elucidate reciprocal shaping effects of the gut, oral, and nasal microbiota and the immune system that are relevant in the context of this thesis. Many of the insights described below are based on studies performed in mouse models and therefore the corresponding knowledge about host-microbial interactions in humans is merely translational [24]. Of note, in this thesis we chose a strategy to investigate host-microbial interactions feasible in human samples (namely during HSCT), to contribute to closing this translational gap.

2.2.1 The role of the microbiota in immune system maturation

The microbiota plays a crucial role in the development of the host immune system. Various animal model studies have demonstrated severely impaired immune maturation, both structurally and functionally, in germ-free (GF) mice [25]. The presence and structure of the commensal microbiota is thought to have such a critical impact on the immune system, that the period in early life where microbes facilitate immunological development of potentially life-spanning importance, has been termed the 'window of opportunity' [8, 25, 26].

GF mice exhibit immunological defects on several levels, affecting both the innate and the adaptive immune system. For instance, innate immune cell counts, e.g. neutrophils and monocytes, are reduced in GF and antibiotic-treated mice [27, 28]. Of particular importance to the adaptive immune system, the small intestine and mesenteric lymph nodes of GF mice are smaller and have reduced cell numbers compared to conventionally colonized animals [29]. Within these organs, certain lymphocyte populations are reduced in number or function [25, 29]. For instance, a lower frequency of certain intestinal CD4+ and CD8+ T lymphocyte subsets are characteristic for GF mice, as well as reduced production of immunoglobulin A (IgA) by plasma cells [25, 29]. Colonization of GF mice with mouse-specific gut microbiota could to a large extend rescue T cell numbers in the small intestine lamina propria, as well as intestinal epithelial lymphocyte numbers [5]. However, colonization of GF mice with rat-specific or human-specific gut microbiota failed to restore these lymphocyte deficiencies, indicating that not only the presence of any microbiota, but of a host-specific microbiota is key to immune system development [5]. Learning tolerance through antigen exposure is a crucial step in T cell activation, especially in infants [25]. Therefore, the environment in which the T cells develop is pivotal. The microbiota is an important

characteristic of the gut environment and it was suggested that host-specific gut colonizers might stimulate antigen uptake by antigen-presenting cells (APCs) which then activate T cells [5].

2.2.2 Interactions of the microbiota and the innate immune system

The innate immune system constitutes the first line of defense against pathogens invading the host's body. GF mice studies indicate that the microbiota is crucial for innate immune cell differentiation, not only for embryo-derived cells in early life stages, but also for hematopoiesis throughout host life [27, 30]. Hence, the microbiota contributes to innate immunological homeostasis, even in steady-state [31].

Several innate immune cell types have been suggested to interact with the commensal microbiota. Neutrophils are considered the first responders to infection and are therefore key players in the innate immune response. Neutrophils are activated via stimulation of the nucleotide-binding oligomerization domain–containing protein-1 (Nod1) receptor with bacterial-derived peptidoglycan during infection [32]. Under steady-state conditions, neutrophil activity underlies regulation through gut commensal-derived peptidoglycan [33]. Neutrophils are also characterized by a particularly rapid turnover rate [34]. In this context, residential microbes in mice have been shown to promote ageing of neutrophils via Toll-like receptor (TLR) signaling, thereby elevating pro-inflammatory neutrophil activity [34]. Together, these mechanisms of neutrophil regulation by the microbiota are important for maintaining homeostasis of pathogen protection and inflammation-induced tissue damage [33, 34].

Monocytes, another important innate immune cell type, are being recruited from the bone marrow via stimulation of CC-chemokine receptor 2 (CCR2) by monocyte chemotactic protein-1 (MCP1) [35]. Bacteria are suggested to drive monocyte recruitment from the bone marrow via tonic sensing of bacterial cell wall components, such as lipopolysaccharides (LPS), by mesenchymal stem cells in the bone marrow. Upon LPS-stimulation, these cells produce MCP1, thereby enhancing CCR2-mediated monocyte recruitment [31, 35]. Even very low LPS-concentrations can facilitate monocyte migration, indicating not only a response to pathogenic stimulation, but a response also to minimal fluctuations in the commensal microbial community structure [35]. Subsequent to their release into the bloodstream, monocytes migrate to target tissues where they differentiate into macrophages or dendritic cells, both in steady state and during infection and inflammation [36]. This, too, underlies residential microbial influence: The adult mouse microbiota promotes the recruitment of a sufficient number of monocytes from the bloodstream to constantly replenish the intestinal mucosa with macrophages, ensuring an adequately equipped first line of defense in the event of infection [30].

Another myeloid innate immune cell type regulated by the microbiota are basophils. Commensal microbes are involved in preventing allergic inflammation by restricting the steady-state population of this cell type [37]. Furthermore, the microbiota modulates natural killer (NK) cells, a lymphocyte subpopulation critically involved in the rapid innate immune response, particularly the anti-viral response [38]. Antiviral immunity through NK cells depends on commensal stimulation of interferon gene expression in macrophages and dendritic cells, which in turn activate NK cells - a mechanism that is impaired in GF mice [39, 40].

An important non-cellular part of the innate immune system is represented by antimicrobial peptides (AMPs) [41, 42]. These small cationic and amphiphilic proteins contribute to the innate immune defense by immunomodulation and antimicrobial activity against a broad spectrum of bacteria, viruses, and fungi

[41–43]. In AMP-mediated killing, electrostatic attraction and binding of cationic AMPs by anionic phospholipids in the bacterial and other target cell membranes is followed by membrane penetration and microbial cell lysis [44]. Human AMPs comprise the groups cathelicidins, defensins, and histatins [45]. The defensins are divided into two types, human alpha- and beta-defensins. The alpha-defensins human defensin 5 and 6 (HD5 and HD6) are being produced by Paneth cells in the crypts of the small intestine (Figure 1) [45, 46]. Beta-defensins are more ubiquitously expressed by various epithelial cells, as well as by antigen presenting cells, such a macrophages and dendritic cells [47–49]. Certain AMPs, including HD5 and HD6, are constitutively expressed while the production of others, e.g. human beta-defensins 2 and 3 (hBD2 and hBD3), requires microbial stimulation of AMP-producing cells through pattern recognition receptors (PRRs) [47, 49, 50]. In the intestine, enterocytes are predominant producers of AMPs and secrete them to form a chemical barrier to invading pathogens, but also to compartmentalize and shape the commensal microbiota, facilitating host-microbial homeostasis [45, 47, 50]. However, the interaction of AMPs and the microbiota exceeds the regulation of commensal homeostasis. The microbiota also contributes to AMP-mediated innate immune function: For instance, the expression of regenerating islet-derived protein III-gamma (RegIIIγ), an AMP produced by Paneth cells and intestinal epithelial cells (IECs), is dependent on commensal TLR-mediated stimulation [51, 52].

Another crucial mode of influence that the microbiota exerts on the host immune system involves the modification or de novo synthesis of metabolites [53]. For instance, gut commensals utilize dietary compounds to produce short chain fatty acids (SCFAs), including butyrate produced by *Firmicutes* such as *Faecalibacterium prausnitzii*, as well as propionate and acetate produced by *Bacteroidetes* such as *Bacteroides thetaiotaomicron* [53]. SCFAs modulate innate immune cells and are for instance involved in neutrophil chemotaxis, and reduction of pro-inflammatory cytokine production in LPS-stimulated neutrophils and macrophages [53–56]. Furthermore, SCFAs serve as an energy source for enterocytes that oxidize e.g. butyrate to carbon dioxide [53, 57, 58].

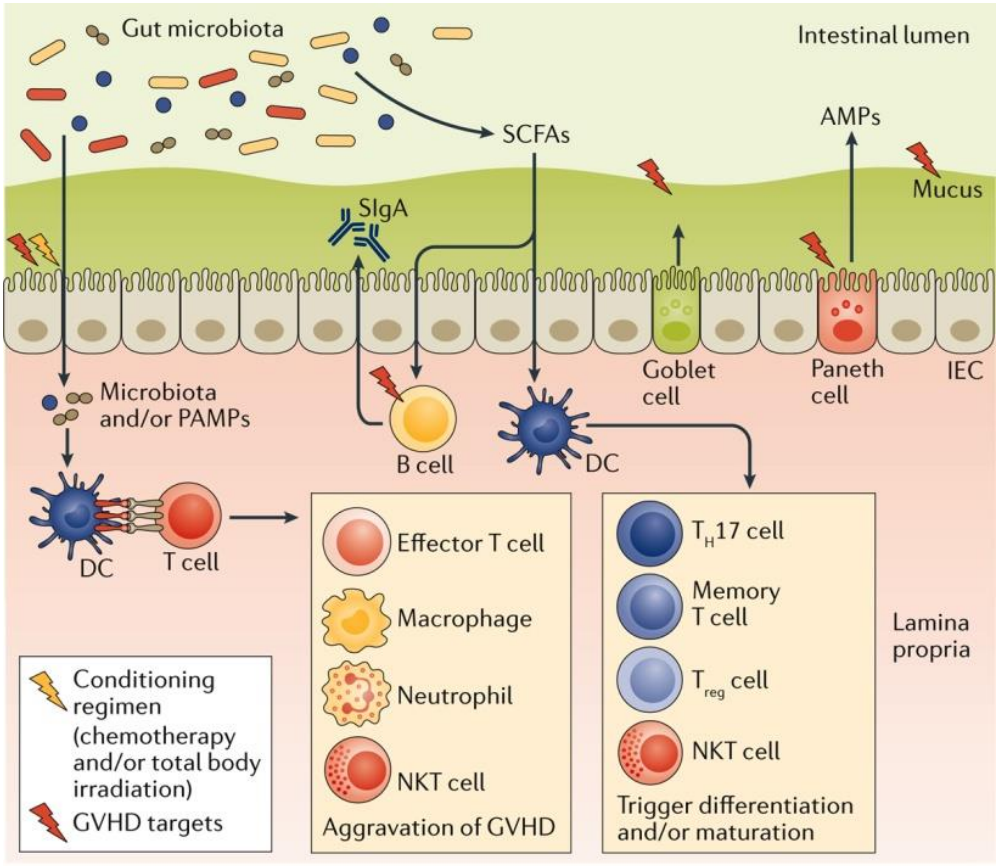


Figure 1. Interactions of the microbiota and the immune system in the healthy host, and in HSCT patients. The commensal gut microbiota retains a homeostatic relationship with the healthy human immune system, which is disturbed under the influence of conditioning regimens and GvHD in HSCT patients. Abbreviations: secretory immunoglobulin A (SIgA); pathogen-associated molecular patterns (PAMPs); NKT, natural killer T cell; DC, dendritic cell. Figure from Shono *et al.*, 2018 [59].

2.2.3 Interactions of the microbiota and the adaptive immune system

The adaptive immune system is a key player in maintaining host-microbial homeostasis, characterized by a diverse microbiota, integrity of mucosal barrier function, and minimal inflammation [26, 60, 61]. By producing IgA, plasma cells contribute to regulating the microbiota, e.g. in the gut. Firstly, IgA coats and agglutinates bacteria, both pathogens and commensals, and thereby promotes targeted effector cell reactions rather than direct contact with the mucosa, e.g. in the gut or nose [26, 60]. Like pathogens, commensals also induce IgA production, but to different degrees [61, 62]. Especially microbes residing in close proximity to the intestinal epithelium promote IgA responses via SCFA-mediated differentiation of naïve B cells to plasma cells, and enhancement of IgA secretion [26, 63]. Members of the microbiota underlie a constant selective pressure to evade the host immune system. In that context, it has been suggested that commensals gain a competitive advantage by reducing the induction of IgA production

[62]. This is achieved by strain diversification, which also benefits the host as it broadens the range of functionality that the microbiota can put at the host's disposal. However, the interplay of IgA with the residential microbiota is yet more complex. For instance, IgA has been shown to influence gene expression in *Bacteroides thetaiotaomicron* in mono-colonized gnotobiotic mice, as a means of circumventing pro-inflammatory signaling [26, 62]. The homeostasis regulation facilitated by IgA is enhanced by antibody diversification and affinity maturation which allow for highly specific targeting of commensals [64]. In summary, these findings give insights into crucial reciprocal effects between the microbiota and the humoral adaptive immune response.

Of no less importance are interrelations of the microbiota and the cell-mediated part of the adaptive immune system. In mice, specific commensals have been reported to activate T cells, especially the CD4+ T cell subpopulations of T helper 17 (T_H17) cells and T regulatory (T_{reg}) cells (Figure 1) [26, 65]. Murine host-specific segmented filamentous bacteria (SFB) induced the differentiation of naïve CD4+ cells into T_H17 cells by binding to the gut epithelium [66–68]. An equivalent effect in the human host remains to be investigated. However, human gut microbial strains from an ulcerative colitis patient were shown to induce T_H17 activation in mice, indicating that a similar effect in humans is likely [66]. By activating T_H17 cells, the microbiota promotes homeostasis as T_H17 cells contribute to mucosal integrity, e.g. in the gut and the oral cavity, by protecting against extracellular bacterial and fungal pathogens [69]. However, T_H17 cells can become pathogenic upon high cytokine stimulation, leading to pro-inflammatory signaling and autoimmune responses [26, 70]. To maintain homeostasis, anti-inflammatory counter-action by immunosuppressive T_{reg} cells is crucial [26, 70, 71]. Expansion of T_{reg} cells underlies commensal microbial influence, as well. Several *Bacteroides* and *Clostridia* cluster IV, XIVa and XVIII strains originating from human feces were shown to induce T_{reg} cells in GF mice, e.g. through SCFA production [53, 72, 73]. These interactions are suggested to be highly strain and cell type specific [61]. For instance, specific human fecal strains of *Bacteroides thetaiotamicron* and *Clostridium ramosum* were shown to specifically induce the subpopulation of colonic RORg+Helios- T_{reg} cells in monocolonized mice [61]. Even though the tested strains originated from the human intestine, the question remains to which extent findings from mouse models are translatable to humans. Furthermore, monocolonization experiments clearly omit microbe-microbe interactions, which play an important role in homeostasis given the vast diversity of commensal colonizers. However, these findings provide important insights and guide the direction of investigating human host immune-microbial interactions.

2.2.4 Specific interactions of the oral and nasal microbiota with the immune system

Even though the majority of studies on host immune system-microbial interactions focus on the gut microbiota, other niches should not be neglected. Since we know that immunomodulation by the gut microbiota can be systemic, microbiotas residing at other body sites might be involved to a similar extent. Moreover, the microbial community structure at one site is not entirely independent of that at another site, e.g. demonstrated by the finding that gut and oral community state types could predict each other [74]. The oral microbiota is suggested to enhance host defense by promoting inflammation in the oral cavity [17]. For instance, neutrophil chemoattractants and pro-inflammatory interleukin 1 beta in the gingiva are upregulated to a larger extent in conventional compared with GF mice, suggesting important immunomodulatory function of the oral microbiota [75, 76]. The nasal microbiota has been proposed to

structurally and immuno-functionally shape the nasal cavity [77]. GF mice were found to have both a thinner epithelium and a thinner mucosa in the paranasal sinuses, as well as less nasal associated lymphoid tissue (NALT) compared with conventional mice [77]. Furthermore, a commensal *Lactobacilllus* strain in the murine nose could stimulate CD4+ T cells in NALT to a lesser degree compared with a pathogenic *Streptococcus* strain, and without causing disease [78]. This suggests that nasal residents can facilitate immune responses without being invasive [17, 78].

2.3 The microbiota in acute and chronic disease

The residential microbiota is involved in the immune response against pathogenic infection, both directly and through its immunomodulatory properties. By outcompeting pathogens for the colonization of the mucosal niche, commensals exert a direct defense mechanism called colonization resistance [79, 80]. Indirectly, the microbiota contributes to maintaining the balance between defense against the invading pathogen and inflammation-mediated collateral tissue damage during infection. For instance, during gastrointestinal (GI) infection commensals stimulate inflammatory monocytes to produce the lipid mediator prostaglandin E2, which suppresses neutrophil activation and thereby limits tissue-damage [81]. Bearing in mind the numerous complex interactions between the microbiota and the host immune system under steady-state conditions and during acute disease summarized above, commensal involvement in complex chronic diseases understandably has come more into the focus of research. The microbiota is extensively researched in the context of for instance inflammatory bowel disease (IBD), cancer, human immunodeficiency virus (HIV) infection, type 1 diabetes, and obesity [82, 83]. In addition to disease-specific microbial associations, meta-studies have recently identified microbial patterns shared across diseases, attributing comprehensive importance to microbial influence on host disease [82, 83]. For instance, reduced abundances of the taxonomic families *Ruminococcaceae* and *Lachnospiraceae* (order *Clostridiales*) are commonly found across various diseases, including IBD, HIV infection, and obesity [82, 84, 85]. In contrast, abundances of *Enterobacteriaceae* and *Lactobacillaceae* are increased in association with several diseases, such as IBD, diarrhea, and HIV infection [82–84, 86].

With the loss of health-associated *Clostridia,* the host is being deprived of major SCFA-producers and therefore more prone to inflammation [56, 87]. Moreover, the lack of SCFAs has been suggested to contribute to increased oxygen levels in the intestinal lumen [88]. When butyrate is unavailable, enterocytes switch to fermentation of glucose, promoting oxygenation of the epithelium and oxygen diffusion to the lumen [58, 88]. As a consequence, microbial growth conditions shift in favor of facultative anaerobic bacteria, promoting overgrowth of disease-associated taxa, such as *Enterobacteriaceae* [57, 89].

2.4 Hematopoietic stem cell transplantation

In the physiological process of hematopoiesis, hematopoietic stem cells (HSCs) give rise to blood cell progenitors that differentiate further into mature cells. Due to the self-renewal capabilities of HSCs, hematopoietic stem cell transplantation (HSCT) was introduced over 60 years ago as the first regenerative clinical treatment [90]. It has since been applied as a curative treatment in various hematologic diseases, including myeloid and lymphoid malignancies, as well as non-malignant conditions, such as immune deficiency syndromes [91]. HSCs for transplantation can be of allogeneic (a donor's cells) or autologous (the recipient's own cells) origin (Figure 2).

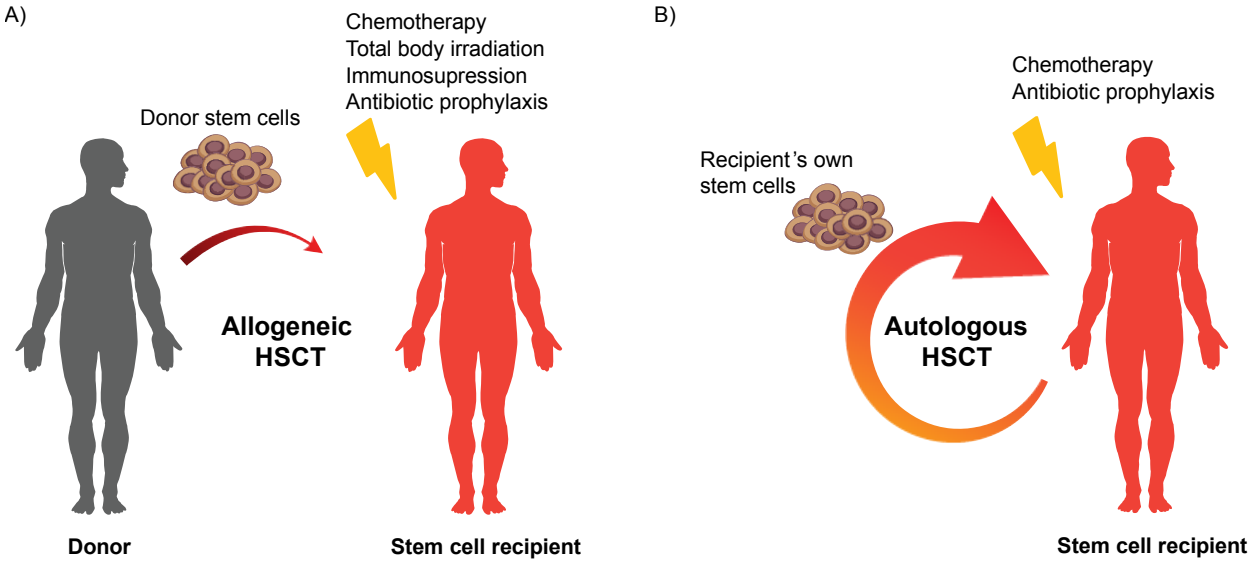


Figure 2. General concepts of allogeneic and autologous hematopoietic stem cell transplantation. A) Allogeneic HSCT. B) Autologous HSCT. Created with vector graphics from: colourbox.com

2.4.1 Principles of allogeneic HSCT

In allogeneic HSCT, the recipient is infused with a stem cell graft derived from a foreign donor. The immunotherapeutic aims of allo-HSCT are (1) to restore the recipient's depleted hematopoietic system and (2) to achieve a graft-versus-leukemia (GvL) effect of donor-derived cells against malignant cells in the recipient [90]. An important prerequisite to allo-HSCT is the choice of a suitable donor. To prevent cytotoxic donor-derived T lymphocytes from recognizing healthy recipient cells as 'foreign targets' (alloreactivity), matching of human leukocyte antigen (HLA) genes between donor and recipient is the key procedure of donor selection [92]. HLA-identical siblings are the most desirable type of donor, but are only available for a fraction of patients [92]. Alternatives include HLA-matched or HLA-mismatched (1 or 2 mismatches) unrelated donors, and haploidentical related donors [92, 93]. Prior to the transplantation, recipients undergo a preparative myeloablative conditioning regimen that can comprise stand-alone treatment with cytotoxic chemotherapeutic agents or a combination of chemotherapy and total body

irradiation (TBI) (Figure 2A) [90]. The aims of preparative conditioning are eradication of malignant cells for disease control, and immunosuppression in the recipient to minimize the graft rejection risk [94]. Among the most widely used conditioning regimens are stand-alone chemotherapy with combinations of busulfan and other cytotoxic agents, and TBI in combination with e.g. cyclophosphamide [94]. More recently, reduced-intensity fludarabine-based regimens have increasingly been used and were suggested to improve overall survival as compared to cyclophosphamide/TBI treatment [94]. In addition, patients may be treated with antithymocyte globulin (ATG), an immunosuppressant which depletes the recipient's T cells, with the intention to lower graft rejection and GvHD risks [95]. Another important consideration is the choice of stem cell source. Options comprise the donor's bone marrow, peripheral blood, or umbilical cord blood [96]. Peripheral blood derived grafts contain the highest number of residual mature, differentiated donor-derived cells (accessory cells), followed by bone marrow, and umbilical cord blood derived grafts [96]. These accessory cells, which for instance include T lymphocytes and Natural killer (NK) cells, can contribute to faster engraftment (i.e. proliferation of donor-derived HSCs in the recipient) and immune reconstitution, but also to high alloreactivity in the recipient [96, 97]. Furthermore, a high number of CD34+ progenitor cells in the graft (which is lowest in umbilical cord blood and highest in peripheral blood) is suggested to prevent graft rejection and reduce overall mortality [96]. Other factors, such as malignancy and stage of disease, weigh into the decision for a suitable stem cell source as well [96]. The most common indications for allo-HSCT are myeloid malignancies, such as acute myeloid leukemia (AML) and myelodysplastic syndrome (MDS), as well as lymphoid malignancies, including acute lymphoblastic and chronic lymphocytic leukemia (ALL and CLL) [91]. The advantage of allogeneic compared to autologous HSCT lies in a strong GvL effect since donor-derived immune cells can recognize residual malignant cells in the recipient as 'non-self' and attack them [90]. Therefore, usually only allogeneic HSCT is suitable for the treatment of leukemia [98]. Furthermore, an allo-graft, in contrast to an auto-graft, is free from malignant cell contamination and therefore facilitates replenishment of the recipient's hematopoietic tissue with functional, healthy HSCs [90]. However, being of foreign origin also comes with challenges concerning transplantation of allo-grafts. To start with, the risk of graft rejection is higher than in auto-HSCT patients which is being counter-acted by finding the donor with the highest HLA-match [90]. Secondly, the GvL effect is often accompanied by activity of donor-derived cells against the recipient's healthy cells, a condition known as graft-versus-host disease (GvHD) [90]. Finally, slow engraftment extends the period of immunodeficiency in recipients after allogeneic transplantation [90].

2.4.1.1 Acute graft-versus-host disease after allogeneic HSCT

Acute GvHD is one of the major complications and the main cause of non-relapse related death following HSCT [99, 100]. Leading up to aGvHD is the presence of immunocompetent donor-derived T cells in the stem cell graft, which cannot be rejected by the immunocompromised recipient [99]. Immunological disparity of recipient- and donor-derived cells, i.e. the presence of different major and minor human leukocyte antigens (HLA), causes activation of the donor-derived cells and eventually attack of healthy recipient tissue [99]. The pathogenesis of acute GvHD may be described in a three-phased model mainly based on animal studies: (1) Treatment-induced tissue damage in the recipient, activating APCs; (2) Activation of donor T cells by recognition of APCs; and (3) Target cell lysis through cytotoxic activity and inflammation [99]. In general, high age of both donor and recipient, and transplantation from a female

donor to a male recipient pose an increased aGvHD risk [100]. Specific risk factors come into effect in each phase of aGvHD pathology. In the first phase, the degree of tissue damage depends on preceding infections, the underlying primary disease, and the conditioning regimen [94, 99]. Release of pro-inflammatory cytokines from the damaged tissue leads to activation of recipient APCs. Gut epithelial injury caused by high-toxicity conditioning is suggested to enhance APC activation through bacteria-derived pathogen associated molecular patterns released by residential microbes that translocate to the bloodstream [99]. However, this concept is controversially discussed in terms of its universal validity for all gut bacteria and the role of the microbiota in aGvHD that will be discussed in more detail in paragraph 2.5 is suggested to be more complex [101, 102]. In the second phase, donor-derived T cells recognize recipient APCs as 'foreign' based on the HLA-encoded major histocompatibility complex (MHC) molecules they present on their surfaces [99]. Related to this phase, patients with unrelated, HLA-mismatched donors are at increased risk for aGvHD [103, 104]. However, disparity in the minor HLA can still occur and facilitate aGvHD, even in patients with donors fully matched in the major HLA [99, 105]. Prophylactic ATG treatment addresses this issue by depleting the recipient's T cells prior to graft infusion, whereas cyclosporine treatment post-transplant, which might be substituted with e.g. methotrexate or corticosteroids, aims at suppressing donor T cells [106]. During the third phase, donor-derived T cells exert cytotoxic activity on healthy recipient tissue and induce inflammation. In this context, risk factors also comprise non-HLA genotypes, such as polymorphisms in cytokine-encoding genes, as well as interleukin secretion levels [107–110].

Table 1. Grading of acute graft-versus-host disease. GI: gastrointestinal. Table from Nassereddine *et al.*, 2017 [100].

Overall clinical grade	lower GI	Upper GI	Liver	Skin
I	0	0	0	1-2
II	1	1	1	3
III	2-3		2-4	-
IV	4		-	4

Acute GvHD affects the skin, liver, and gut where epithelial cells within the epidermis, the biliary duct epithelium, and the intestinal epithelium are targeted, respectively [99, 100]. Cutaneous aGvHD manifests as erythematous maculopapular rash, covering different proportions of the body surface depending on severity, and can be accompanied by keratinocyte necrosis leading to erythroderma with desquamation in severe cases [111]. Hepatic aGvHD involves increased serum bilirubin levels, indicating impaired liver function [99]. In the GI tract, the base of the crypts is most severely affected by aGvHD, where apoptosis is induced in intestinal stem cells and Paneth cells [112, 113]. Paneth cells provide crucial innate immune functionality by producing AMPs, such as alpha-defensins and RegIIIγ whose expression decreases in the small intestine during aGvHD, impairing the antimicrobial immune defense repertoire (Figure 1) [114, 115]. However, increased levels of RegIIIγ were found in the large intestine and systemically during aGvHD, which is suggested to serve as a survival signal for intestinal stem cells and

Paneth cells in the crypts [114, 116]. Acute GvHD in the GI tract manifests as diarrhea, anorexia, nausea, abdominal pain, and vomiting [99]. The severity of aGvHD can be described by an overall grading, the Glucksberg score, which takes into account the organ involvement (of skin, liver, and gut) and clinical performance of the patient (Table 1) [117]. Grade I denotes mild aGvHD involving only the skin, whereas grade II-IV involve skin, liver, and gut with increasing severity [100].

2.4.2 Principles of autologous HSCT

Autologous HSCT refers to the transplantation grafts consisting of the recipients' own HSCs, which are harvested and cryopreserved before the onset of chemotherapy, and expected to be free from significant amounts of malignant cells (Figure 2B) [118]. The main stem cell source utilized in auto-HSCT is peripheral blood which yields higher numbers of progenitor cells than bone marrow and promotes faster hematopoietic reconstitution [96]. The conditioning regimen preceding auto-HSCT predominantly consists of high-dose chemotherapy alone, without TBI (Figure 2B) [119]. TBI is associated with adverse clinical outcomes, such as neutropenic fever, relapse and mortality, following autologous transplantation [119, 120]. Auto-HSCT is the standard treatment for lymphoid malignancies, such as multiple myeloma (MM), as well as Hodgkin and non-Hodgkin lymphoma (HL and NHL) [91]. MM patients usually receive high-dose melphalan as a standard conditioning regimen, while the standard treatment for lymphoma patients consists of a combination of carmustine, etoposide, cytarabine, and melphalan ('BEAM') [121, 122]. The clear advantage of autologous over allogeneic HSCT lies in self-tolerance of the graft as the HSCs originate from the recipient [90]. This promotes engraftment and lowers the risk of adverse events associated to allo-transplantation, such as graft rejection and GvHD [90]. Challenges of auto-HSCT include the sufficient acquisition of stem cells to transplant and the risk of graft contamination by residual malignant cells in the peripheral blood or bone marrow of the recipient [90]. Residual malignant cells that are infused together with the auto-graft state the main cause of increased relapse risk in autologous compared to allogeneic transplantation recipients [90]. Moreover, auto-transplantation is mostly ineffective for the treatment of leukemia due to the lack of a GvL-effect [98].

2.4.3 Infectious complications and antimicrobial treatment in HSCT

Infectious complications contribute decisively to morbidity and mortality after allogeneic HSCT [123, 124]. Inherent characteristics of allo-HSCT, comprising myeloablation, immunosuppression, and subsequent delayed immune reconstitution are major risk factors for infections [123]. Moreover, patients are more susceptible, particularly to bacterial blood stream infections (BSI), when they are adults (>18 years), have a malignant primary disease, receive a graft from a mismatched unrelated donor, and undergo steroid treatment and myeloablative (versus reduced-intensity) conditioning [123, 124]. However, in children with immunodeficiency disorders, reduced-intensity conditioning with melphalan posed a higher risk for BSI compared with myeloablative regimens [125]. The most common Gram-positive bacterial pathogens infecting allo-HSCT patients comprise *Staphylococcus* spp., *Streptococcus* spp., and *Enterococcus* spp. [126, 127]. Frequent Gram-negative pathogens include *Klebsiella* spp., *Enterobacter* spp., *Pseudomonas* spp., *Stenotrophomonas* spp., as well as *Escherichia coli* and *Acinetobacter baumanii* [126, 127]. During conditioning prior to HSCT, patients usually receive prophylactic antimicrobial treatment, e.g. with

fluoroquinolones, such as ciprofloxacin, or with trimethoprim-sulfamethoxazole, to limit infections due to immunosuppression [128–130]. Immediately post HSCT, patients usually receive broad spectrum beta-lactam antimicrobial monotherapy, e.g. with ceftazidime or meropenem, especially during neutropenia [123, 131]. At all times, standard prophylactic antimicrobial regimens are supplemented by or substituted with antimicrobial agents active against specific pathogens in case of fever or when indicated by blood culture results [131]. These antimicrobial agents comprise for instance vancomycin, metronidazole, and linezolid [130, 131]. During the first month after HSCT, patients that develop neutropenia or mucositis are generally at higher risk of infectious complications [123, 127]. Opportunistic pathogens (mostly Gram-negatives) originating from the commensal microbial community, and infectious agents introduced by vascular catheters (mostly Gram-positives) are dominating during this early phase [123]. Two to three months post-HSCT, infections are most common in patients affected by aGvHD as the associated increased gut epithelial permeability, Paneth cell and alpha defensin reduction, and immunosuppressive treatment of aGvHD with e.g. cyclosporine facilitate gut microbial translocation to the bloodstream, combined with an impaired immune defense [102, 123, 124]. From four months after transplantation and thereafter, it is particularly important how rapid the adaptive immune system is reconstituted and regains the ability to avert persisting and newly introduced infectious agents [123, 132].

Viral infections are a huge burden and a major cause of morbidity and mortality following allo-HSCT [123]. Herpes simplex virus-1 and -2 infections predominate immediately after HSCT, followed by cytomegalovirus (CMV) and human herpes virus-6 infections during the first two months post HSCT, and Epstein-Barr virus and varicella-zoster virus infections from three months onwards [123]. Of those, CMV infection compromises transplantation success to the largest degree by causing tissue-invasive disease, promoting secondary bacterial and fungal infections, and increasing the risk of GvHD and treatment-related mortality (TRM) [123, 133]. Most often, CMV infection results from reactivation of a latent CMV infection in immunocompromised allo-HSCT recipients. Therefore, patients are being screened for anti-CMV IgG serology prior to transplantation to estimate CMV infection risk, with the aim to match seronegative recipients with seronegative donors to prevent CMV contamination [123, 133]. Post engraftment, i.e. 2 - 3 months after HSCT, persisting immunodeficiency, potentially enhanced by GvHD treatment, also increases susceptibility to fungal infections, predominantly caused by *Aspergillus* spp. and *Candida* spp. [123].

The distinct procedures of allogeneic and autologous HSCT entail differences in the infectious complications of the respective patients [127]. Since autologous patients do not receive foreign, but rather their own stem cells, the incidence of GvHD after auto-HSCT is much lower compared with allo-HSCT. Whereas moderate to severe GvHD constitutes a major risk factor for infections in allo-HSCT, it is consequently less pivotal for infections in auto-HSCT [127, 134]. Major risk factors in auto-HSCT comprise not receiving prophylactic antimicrobial therapy, having high serum levels of the inflammatory marker C-reactive protein (CRP) prior to transplantation, and prolonged duration of engraftment [127, 134]. In autologous patients, BSI following transplantation occur less often, but earlier compared with allogeneic patients [125, 127, 134].

2.4.4 Immune reconstitution after HSCT

An essential element of successful allo-HSCT is the timely reconstitution of a functional immune system built up from sufficient numbers of functionally mature donor-derived immune cells [132, 135]. In the post-engraftment phase (four months post HSCT and thereafter), the ability to avert recurring infections from the period of immunosuppression, as well as to prevent newly introduced infectious agents, is crucial for the patients' recovery and long-term survival [123, 132]. Innate immune cells are the first to be reconstituted within 20-30 days post-transplantation [132, 136]. Before engraftment of donor cells, i.e. during the first 2-4 weeks after HSCT, patients suffer from neutropenia [132]. Subsequent replenishment of neutrophils serves as an initial indicator for transplantation success and depends on the stem cell source [132, 136]. Neutrophil counts on average increase after two, three, or four weeks when derived from a peripheral blood graft, bone marrow graft, or umbilical cord blood graft, respectively, and increase further during the following months [132, 137]. Simultaneously, monocyte counts recover within the first month post-transplant. Chemotherapy-resistant recipient-derived macrophages might still circulate during this time, while donor-derived monocytes differentiate into macrophages and gradually substitute the recipients' cell population [137]. NK cells are the next innate immune cell type to be reconstituted between the second and third month after HSCT [136, 137]. NK cell numbers during the first month post-transplantation are rarely examined, even though they are major contributors to the GvL effect, prevent viral infections, and their rapid recovery and functional maturation has been suggested to be predictive for survival [136–139].

The reconstitution of the adaptive immune system is a prolonged process of one to two years after transplantation, taking an even longer time until full restoration of cellular and humoral immune functionality and normal cell numbers [135]. T cells play a central role in the GvL effect and defense against infectious agents after allo-HSCT [132, 135]. During the first six months post-transplant, T cells derived either from the donor graft or from remaining recipient cells that have not been destroyed during conditioning, increase in numbers via homeostatic peripheral expansion [135, 136]. Recovery is slower in case of ATG treatment or transplantation of a T cell depleted graft [140]. Since the resulting T cell population in this early phase predominantly comprises CD8+ memory T cells, the MHC-mediated antigen recognition in response to newly introduced infectious agents is limited [135]. Moreover, these T cells have not undergone the maturation steps selecting for self-tolerance in the thymus and therefore pose a risk for the recipient to develop aGvHD [135]. Thus, sustained restoration of T cell immune function requires the production of naïve CD4+ and CD8+ T cells from donor-derived lymphocyte progenitors and hematopoietic stem cells in the thymus [132, 135, 140]. Successful thymopoiesis resulting in rapid and stable reconstitution of a functional T cell repertoire is crucial for survival, and to prevent GvHD, relapse, and infectious complications [135, 136].

Varying reconstitution times have been reported for B cells that are responsible for the humoral part of adaptive immunity. Depending on stem cell source, the recovery of the B cell compartment takes between two months (following umbilical cord blood transplantation) and one year (after bone marrow transplantation), but complete reconstitution is often achieved late after 2-5 years [132, 136]. GvHD and its accompanying steroid treatment have been suggested to slow down B cell recovery [136, 141]. Similar to T cell reconstitution, donor graft-derived B cells dominate the B cell population after allo-HSCT at first, and are subsequently replaced by stem-cell derived naïve B cells originating from the thymus [132, 137].

After allo-HSCT, the recipients' immunoglobulin repertoire is confined and recovers in the same order (IgM – IgG – IgA) that is observed during establishment of adaptive immunity in young children [132, 137]. However, compared with the ontogeny in children, serum level elevation of antibody isotypes after allo-HSCT is slower because isotype switching depends on signaling from T helper cells, which are also depleted after transplantation [132, 137].

Compared with allogeneic HSCT, immune reconstitution after autologous HSCT is faster [140, 142]. The main reason for this may be that GvHD and the accompanying immunosuppressive treatment do not interfere with immune reconstitution after auto-HSCT as the prevalence of GvHD is negligible compared with allo-HSCT [143]. For instance, the T cell compartment is replenished more rapidly within the first two months after auto-HSCT, compared with allo-HSCT, and comprises a higher proportion of naïve T cells compared with memory T cells, indicating effective thymopoiesis [142]. Despite varying reported times for B cell reconstitution, recovery from as early as one month and up to 1.5 - 2 years after auto-HSCT have been observed, compared to 2 to >24 months after allo-HSCT [142, 143].

2.5 Current knowledge about the role of the microbiota in HSCT

The procedure of hematopoietic stem cell transplantation and its accompanying treatments interfere in various ways with the complex host-microbial interactions presented in chapter 2.2 (and summarized below in Figure 3). Within this process, disturbance of homeostasis is introduced in both directions: On the one hand, immune system components, altered by conditioning treatment and stem cell engraftment, may induce changes in the microbial community structure. On the other hand, the microbiota, altered by extensive antibiotic prophylaxis, may influence immune parameters after HSCT. The hitherto acquired knowledge (other than this thesis) about these bi-directional relations with a focus on the neutropenic phase will be the topic of this chapter.

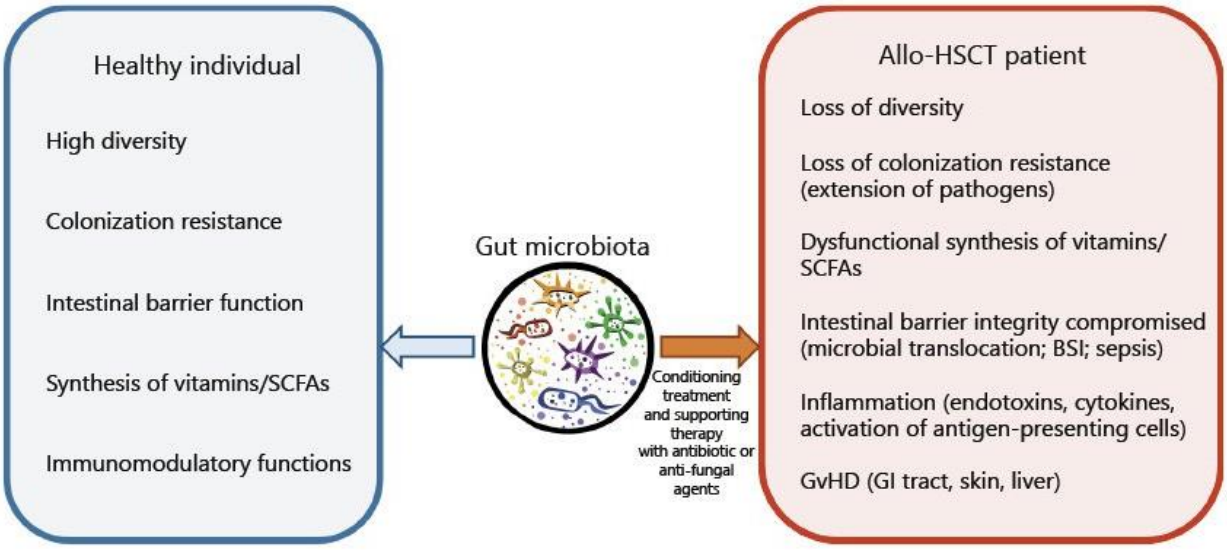


Figure 3. Summary of major host-gut microbial interactions in healthy individuals versus allogeneic HSCT patients. Conditioning and antibiotic treatment (orange arrow) of HSCT patients interfere with host – microbial homeostasis. Abbreviations: SCFAs, short-chain fatty acids; BSI, blood stream infections; GvHD; graft-versus-host-disease; GI tract; gastrointestinal tract. Figure from Noor *et al.,* 2018 [144].

Before advanced next-generation sequencing (NGS) techniques were available, the initial approach of accounting for gut microbial effects in HSCT involved 'gut decontamination' [59, 145]. Under the assumption that an immunocompromised host would give way to exclusively adverse effects of presumably uniformly pathogenic anaerobic bacteria, high-dose antibiotic regimens were used, aiming at the depletion of any bacteria in the gut [145]. Eventually, controlled clinical studies were not able to document any benefit of 'gut decontaminantion' [146]. While prophylactic antibiotic regimens are still standard of care for HSCT patients today, recent research has raised awareness that the microbiota needs to be assessed differentially [147–149]. A number of specific observations have been re-occurring in several studies. Generally, a loss of bacterial alpha diversity in the gut occurs during the neutropenic phase

of HSCT (Figure 3) [150–152]. Loss of diversity is more pronounced in patients with GvHD and is associated with increased mortality [153–156].

Moreover, the community structure changes after HSCT and specific taxa typically expand, while others are depleted. Microbiota members of the phylum *Proteobacteria*, in particular the family *Enterobacteriaceae*, and the order *Lactobacillales,* in particular *Enterococcus* spp. and *Lactobacillus* spp., have been found to increase in abundance after transplantation [150, 153, 154, 157, 158]. In contrast, a post-transplant decrease of taxa affiliated with the order *Clostridiales*, in particular the families *Lachnospiraceae* and *Ruminococcaceae*, has been shown [151, 159, 160]. The expansion of *Enterococcus* spp. and loss of *Clostridiales* early after HSCT is especially pronounced in patients with GvHD [154, 161–164]. Furthermore, this pattern has been associated with high TRM and BSI [150, 165, 166]. It has been suggested that the antibiotic treatment accompanying HSCT is promoting this shift in the gut microbial community structure [148, 158, 160]. Antimicrobial agents frequently used in prophylactic regimens, such as vancomycin, ciprofloxacin, and metronidazole, are targeting commensal obligate anaerobe bacteria, including *Clostridiales* members [148, 160]. A central role has been attributed to SCFAs (especially butyrate) produced by *Clostridiales* members. Antibiotic-induced depletion of SCFA-producers entails a switch to glucose fermentation by enterocytes and increased oxygen availability in the intestinal lumen (as discussed in paragraph 2.3) (Figure 3) [58]. This facilitates the expansion of facultative anaerobes such as *Enterobacteriaceae* (phylum *Proteobacteria*), and *Enterococcus* spp. [57]. Within the order of *Clostridiales*, maintenance of high abundances of the butyrate-producers *Blautia* spp. and *Faecalibacterium prausnitzii* has been demonstrated to be of particular importance for positive clinical outcomes in HSCT [167, 168].

Besides the involvement of SCFAs in shaping intestinal microbial growth conditions, direct immunomodulatory effects have been described. Butyrate is a histone deacetylase inhibitor (HDACi) and, when taken up by IECs, promotes acetylation of histone H4 [169, 170]. Histone acetylation generally facilitates chromatin relaxation and thereby transcription [56, 170]. Reduced H4 acetylation in IECs after allogeneic HSCT has been related to decreased butyrate levels within these cells, but not in the intestinal lumen in mice [169]. It was suggested that decreased butyrate production due to depleted butyrate-producing bacteria after HSCT would not significantly affect intestinal butyrate levels because of a concurrent decrease in expression of butyrate transporter sodium-coupled monocarboxylate transporter 1 (SLC5A8) and G protein-coupled receptor 43 (GPR43), leading to reduced uptake [169]. Conversely, a human study found reduced butyrate and propionate levels in feces after allo-HSCT in pediatric patients [171]. No change in SLC5A8 and GPR43 expression was found in this study [171]. These contradictive findings warrant caution when deriving conclusions for the human HSCT setting from mouse models. Importantly, these and other studies indicate a crucial role of commensal-derived endogenous butyrate in protection from GvHD [169, 171, 172]. Several mechanisms have been suggested. Intestinal butyrate levels two weeks post-transplant were significantly lower in children with subsequent GVHD [171]. In mice, butyrate-induced histone acetylation in IECs has been shown to upregulate expression of genes involved in junctional function and downregulate pro-apoptotic genes [169]. This suggest a direct protection against alloreactive T cells [169]. Another mechanism might involve the induction of T_{reg} cells mediated by butyrate and propionate, facilitating an anti-inflammatory milieu that protects from GvHD [172, 173].

In addition to butyrate, another microbial metabolite, namely the intercellular signal molecule indole and its derivate 3-indoxyl sulfate (3-IS), has been attributed a role in HSCT and GvHD [154, 165, 174, 175]. Low 3-IS levels in urine early after HSCT have been associated with low abundances of intestinal *Lachnospiraceae* and *Ruminococcacea* and high mortality in adult allo-HSCT patients [165]. Indole and/or its derivates can promote tight junction integrity, hamper the growth of Gram-negative enterobacteria, and induce anti-inflammatory cytokines [175–177]. Therefore, a protective effect against GvHD has been proposed for 3-IS [165]. Consistently, low 3-IS levels are particularly pronounced in allo-HSCT patients with GvHD [154, 174].

Another proposed mechanism contributing to microbial community changes after HSCT involves the reduced Paneth cell numbers and accompanying decrease in human alpha defensin production during GvHD [112, 115]. As discussed in paragraphs 2.2.2 and 2.4.1.1, Paneth-cell derived AMPs contribute to gut microbial homeostasis and the lack thereof during GvHD facilitates overgrowth and domination of opportunistic pathogens, such as *E. coli* in mice [112, 115]. Consistently, a human study showed that adult allo HSCT-patients with severe gastrointestinal GvHD exhibited low Paneth cells numbers, low Paneth- cell derived AMP expression, and low urinary 3-IS, the latter pointing to low bacterial diversity [114]. The role of more ubiquitously expressed AMPs, such as beta defensins, in HSCT and GvHD has to our knowledge not been described prior to this thesis. However, involvement of beta defensins in chronic inflammatory diseases, e.g. Crohn's disease, has been shown and a cell type dependent HDACi-mediated regulation was recently revealed for hBD2 [178–180].

The host-microbial interactions described above mainly apply to the gut microbiota in allogeneic HSCT. A few studies have characterized the oral microbiota in relation to HSCT prior to this thesis [181–183]. After HSCT, a substitution of streptococci with staphylococci was found in the oral cavity, and high *Enterococcus* spp. were associated with oral mucositis [181, 183]. With regards to the mode of transplantation (allogeneic versus autologous), few studies examined the microbiota in autologous HSCT patients prior to this thesis. A recent study has described a post-transplant loss of gut bacterial diversity in auto-HSCT patients, comparable to the pattern observed in allo-HSCT [184]. Taken together, the microbiota, especially in the gut of allo-HSCT patients, has been well characterized during the neutropenic phase. A number of studies have related the pre-transplant microbial community structure to clinical outcomes after HSCT [156, 163, 185, 186]. Moreover, the focus of research is recently shifting to the post-engraftment phase to investigate how the microbiota and the host immune system mutually undergo reconstitution. These aspects will be discussed in more detail in chapter 5.

3. Methodological considerations

3.1 Study populations

Our first two studies comprise patients undergoing allogeneic HSCT, whereas in the third study we describe an autologous HSCT cohort. In an ideal scenario, it would be of great interest to treat the two cohorts that were established during the course of this thesis [**Masche II and III**] as a single study population, differing only by the mode of transplantation (allogeneic versus autologous). This way, the autologous group could serve as a control group for a key element of this thesis: the 'host-transplantation'. Thereby, one could disentangle which changes that occur in the microbiota post-transplantation can be attributed to the fact that foreign cells (allogeneic HSCT) have been introduced to the recipient, and which effects are owed to the procedure of transplantation itself and its accompanying treatments. However, the autologous cohort comprised adult patients while the allogeneic patients were children. On this account, and due to further substantial differences in e.g. primary diseases, conditioning regimens, and immunosuppressive treatment (only in allo-HSCT), it would have been inadequate to analyze the allogeneic and autologous cohorts jointly. This led us to the informed decision of analyzing the two cohorts in separate studies [**Masche II and III**]. However, results of the two studies will be jointly discussed in paragraph 5.2.2, taking into consideration the named differences between the two cohorts. Of note, the autologous group in itself is of interest, since this mode of transplantation hitherto has been underrepresented in studies investigating the microbiota in HSCT patients.

3.2 Longitudinal study design and sample collection

The aim of this thesis was to assess interactions of the microbiota with the human immune system. This was done in the context of hematopoietic stem cell transplantation. To be able to compare the microbiota and the immune system before and after the transplantation, we chose a longitudinal study design for all studies included in this thesis. In the first study, fecal samples were collected once within the month preceding HSCT, once on the day of transplantation (+/- 2 days) and once weekly during the first month following transplantation. This allowed us to assess host-microbial interactions during the period immediately after allogeneic HSCT [**Masche I**]. In the second and third study, we collected buccal and anterior naris swabs in addition to fecal samples, and added three follow-up sampling time points after three, six, and twelve months [**Masche II and III**]. This allowed us to describe long-term microbial abundance patterns, and to make predictions of events at late time points from observations at early time points [**Masche II and III**]. For all studies that are part of this thesis, samples were collected by nurses at the hospital during hospitalization of the patients, immediately stored at 4°C, and transferred to long-term storage at -80°C on the same or consecutive day. For the collection of samples at late time points after hospitalization in **Masche II and III**, a strategy was developed in cooperation with Rigshospitalet to teach patients or their guardians how to collect samples at home. Following the instructions, they stored the samples in their refrigerator and transported them to the hospital in a cooling bag on the following day, where they were placed into the freezer (-80°C). The optimal feasible storage conditions were chosen based on a pilot study in which we compared different storage durations at several temperatures. Samples were thawed at room temperature prior to DNA extraction on the respective days.

3.3 DNA extraction

The choice of DNA extraction technique is a crucial step when researching the microbiota. Purified DNA obtained via different DNA extraction kits can vary in quantity (DNA yield), and in quality (proportions of represented bacteria) [187]. The main steps common to the majority of kits on the market are bacterial membrane disruption, protein denaturation, and nucleic acid purification [187]. Automated DNA isolation was performed from the fecal samples of the first study on a Maxwell 16 instrument (Promega Corporation, USA) [**Masche I**]. For the isolation of DNA from fecal samples in **Masche II and III**, we used the QIAamp Fast DNA Stool Mini kit (Qiagen, Venlo, Netherlands) with an additional bead-beating step to enhance the extraction of DNA from Gram-positive bacteria, as suggested in a study that aimed to optimize this procedure [188]. An important consideration when choosing the DNA extraction method is the sample type, and in particular its inherent bacterial biomass [187]. Especially for samples from low-biomass sources, such as the oral and nasal cavities examined in **Masche II and III**, it is pivotal to account for a low signal-to-noise ratio in order to approximate the true signal. Beside environmental and cross-contamination, extraction reagents can harbor bacteria, and studies suggest that there even are microbial profiles specific to manufacturers, kits, and production batches [189, 190]. To minimize the risk of kit-derived contamination especially in the low-biomass samples, we used the QIAamp UCP Pathogen Mini kit (Qiagen, Venlo, Netherlands), which comes with ultraclean production reagents intended for e.g. nasal swabs. In addition, the DNA extraction from oral and nasal swabs was performed in a pre-PCR laboratory spatially separated from high-biomass (e.g. fecal) sample handling. Furthermore, for all sample types, we included one blank control in each extraction round, as well as one mock community sample (Biodefense and Emerging Infectious Research (BEI) Resources of the American Type Culture Collection (ATCC) (Manassas, VA, USA), Catalog No. HM-276D) per sequencing run, and two fecal/swab collection tube controls, and unused swabs, each. How the sequencing results of these controls were utilized to handle environmental and cross-contaminations computationally is described in paragraph 3.6.

3.4 16S rRNA gene sequencing

For a number of reasons, 16S rRNA gene profiling was chosen over metagenomic sequencing in this thesis. The aim of this thesis was to generate phylogenetic surveys of microbial communities at a number of body sites. With the same financial funds available, 16S rRNA gene sequencing would be more suitable to capture different taxa present in a diverse community compared with shallow metagenomic sequencing that might not allow for the assembly of whole genomes. However, deep metagenomic sequencing that allows for whole genome assembly would provide higher phylogenetic resolution.

The conserved regions of the 16S rRNA genes serve as a universal markers of the bacterial and archaeal kingdoms, while the hypervariable regions of the gene can provide up to species-level information. However, a taxonomic resolution on species-level is by far not always reached in 16S rRNA gene sequencing. Taxonomic characterization of a microbial community using 16S rRNA gene sequencing benefits from the large size of data bases available for the 16S rRNA gene, compared to a much smaller number of bacteria for which the entire genome has been described [191, 192]. Of importance, a significantly cheaper per-sample price for 16S rRNA gene profiling compared to metagenomics facilitates an investigation of a higher sample size within the scope of a single project, and thereby a broader insight

into the characterized community and an increased statistical power in the subsequent data analysis [191, 192]. A number of sequencing techniques are available for 16S rRNA profiling, including e.g. Ilumina sequencing technologies and Ion Torrent [193]. In **Masche I**, 454 pyrosequencing has been used, which has by now been outperformed by newer technologies with regards to its error rate, and support for this platform has been discontinued [187]. We employed Illumina paired-end sequencing on a MiSeq instrument in **Masche II and III**. It requires caution to compare findings from studies based on different sequencing technologies. However, the observation of a number of results in **Masche II** that were consistent with **Masche I** (as discussed in paragraph 5.2.1) suggests that we were able to obtain reproducible results that are to a certain extent independent of the sequencing platform. Of course, other methodological differences play a role, too. For instance, we used different sets of primers for the amplification of hypervariable regions in **Masche I** vs in **Masche II and III.** We amplified regions V4 and V5 in **Masche I**, while regions V3 and V4 were targeted in **Masche II and III.** The V4 region, covered in all studies comprised in this thesis, is generally considered an adequate choice for 16S rRNA gene profiling [187, 194].

3.5 Variant inference from 16S rRNA sequences

After sequencing, a number of filtering steps need to be performed, including primer truncation and chimera removal. The detailed procedures of pre-processing are described in the individual manuscripts [**Masche I, II, and III**]. One of the most crucial steps constitutes the inference of biological variation from 16S rRNA gene sequences without interpreting sequencing errors as biological signal. Until recently, the most frequently used method was the construction of operational taxonomic units (OTUs) and subsequent taxonomy assignment of a representative sequence [195]. This method was originally developed for sequencing data generated by 454 pyrosequencing and was used in our first study within the Quantitative Insights Into Microbial Ecology (QIIME) bioinformatics pipeline [see details in **Masche I**] [196]. An OTU typically comprises sequences with <3% dissimilarity [195]. This way, dimensionality is reduced to thousands of OTUs, and noise due to sequencing errors is diminished [195]. However, this compromises the detection of biological variation on a finer scale [195, 197]. Within the first study, we implemented oligotyping, an approach to further resolve OTUs by identifying highly variable nucleotide positions with potential indications for true signal, from sequencing errors at random positions [198]. However, these results are not presented within this thesis. More recently, clustering-free approaches, such as the divisive amplicon denoising algorithm 2 (DADA2), have been developed that allow the identification of exact amplicon sequence variants (ASVs) [197]. In brief, an error model is implemented that is proposed to be capable of disentangling true biological variation from Illumina-specific error profiles in single nucleotide resolution. In addition to the advantages of 16S rRNA gene profiling discussed in paragraph 3.4, a higher resolution is of great relevance for describing complex microbial communities on a fine scale. Moreover, this approach facilitates comparability between studies since it provides consistent labels for ASVs anchored in an adapted version of the SILVA reference data base [199, 200]. Accordingly, we decided to use DADA2 in our second and third study (Figure 4) [**Masche II and III**].

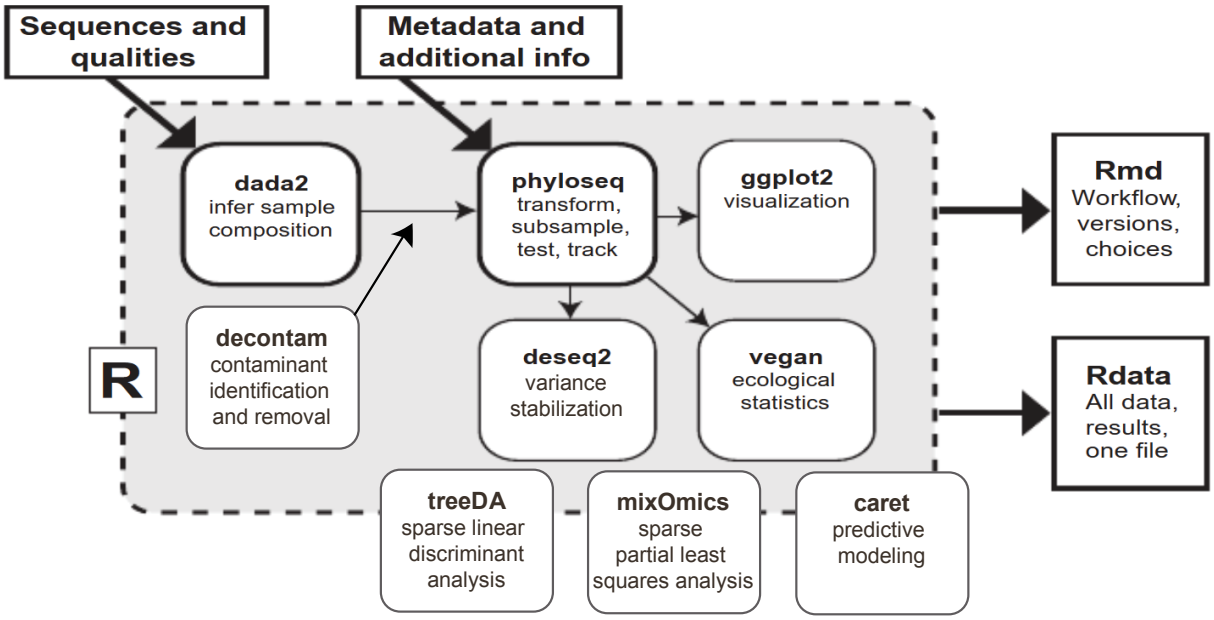


Figure 4. Workflow for data integration, management, and analysis in R. The R packages used for major analysis steps include: DADA2, decontam, phyloseq, ggplot2, DESeq2, vegan, mixOmics, treeDA, and caret [197, 201–207]. Figure modified from Callahan *et al.,* 2016 [208].

3.6 Integration, transformation, and multivariate multi-table analyses of data

In recent years, efforts have been made in the microbiome field to promote reproducible research in a way that published studies should provide access to data and analysis workflows [202, 208–211]. Accordingly, data, R scripts and reports created within the knitr framework [212] have been made available online for the first study already [see details in **Masche I**], and will be made available for the second and third study upon publication. When working with the program R, the package phyloseq is a useful tool to integrate and manage data [202, 213]. The studies within this thesis comprised different data types, including OTU/ASV count tables, categorical data of patient characteristics and clinical outcomes, continuous data of immune cell counts and immune marker levels, and in addition phylogenetic trees in the second study. The phyloseq package facilitated the integration of these distinct data types into a single 'object', collectively linking all data points to the corresponding sample identifiers (Figure 4). Before proceeding to statistical analyses, several crucial steps need to be incorporated. As indicated in paragraph 3.3, contaminant filtering needs to be approached computationally, which we did with the package decontam and/or careful manual curation (see details in individual manuscripts **Masche I, II, and III**) [201]. Decontam utilizes the concept that contaminants occur at higher frequencies in low-concentration samples and blank extraction controls to statistically segregate them from sample-derived variants [201]. Moreover, it is pivotal to transform OTU/ASV count data to account for differences in both, library size and biological variance, often referred to as 'normalization' [214]. Different down-stream analyses require different preceding transformations, and the respective strategies we used are described

in detail in the individual manuscripts [**Masche I, II, III**] [214]. However, whenever applicable we implemented variance-stabilizing transformation within the R package DESeq2, which is based on a Negative Binomial model and, in contrast to rarefying, does not discard valid data (Figure 4) [215]. For subsequent multivariate analysis, we utilized the well-established community ecology package vegan with visualization by ggplot2, as well as the more recently released package mixOmics [203–205]. The latter provides a number of multi-table approaches suitable for our studies, i.e. the joint analysis of for instance OTU/ASV count data and other data types (often collective referred to as 'metadata').

3.7 Considerations for time series data analyses and predictive modeling

Time series studies are gaining increasing interest in microbiome research and offer great potential for detecting temporal dynamics, but also pose challenges concerning implementation and statistical analysis [216–218]. In all studies comprised in this thesis, we used a longitudinal study design, including time points spanning from before HSCT to one [Masche I], or twelve months [Masche II and III] post-transplant. Analyzing these longitudinal data came with a number of challenges that we addressed in multiple ways. The time points where immune cell count and immune marker data was available were to a certain extent subject to clinical routine procedures. We therefore transformed variables with systematically missing time points to combinations of the respective variable and each available time point (e.g. cell count in week +1, cell count in week +2, etc.). In the first study, these variable-time point combinations have been used as input to multivariate analyses, allowing us to gain insight into temporal dependencies of host-microbial associations [**Masche I**]. In the second and third study, a higher number of variables were available at all time points. Therefore, we used the time point itself as a categorical variable in multivariate analyses in **Masche II and III**, and could show how certain host-microbial associations depended on the time point. This approach might require further optimization, either computationally or by better synchronizing sampling schedules of different data types. However, this way, we could extent the current knowledge of host-microbial associations by taking into account temporal dependencies. Our approach might complicate the interpretation of individual findings, but serves the purpose of drawing a holistic picture of time-dependent host-microbial associations in HSCT. In the second study, we additionally made use of our time series data to predict clinical outcomes from pre-transplant microbial abundances [**Masche II**]. For this purpose, we implemented a machine learning (ML) approach within the R package caret [207]. Within marker gene studies, ML is an approach with the potential to increase robustness and clinical applicability of findings that relate microbial community structure profiles with e.g. clinical outcomes [209, 219].

4. Manuscripts

4.1 Manuscript I

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.

Under review

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).

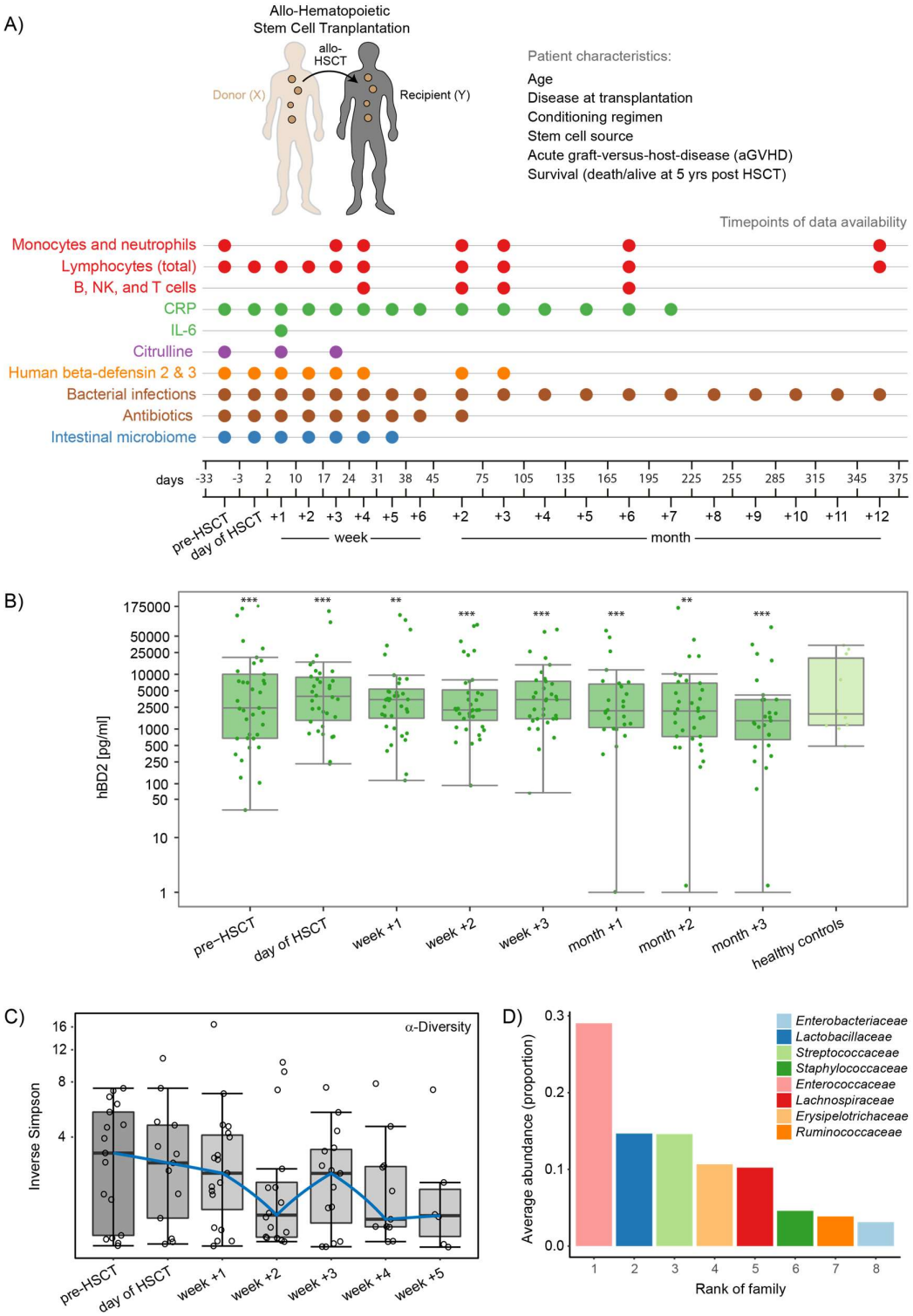


Figure 1. Monitoring of the host immune system and intestinal microbiome in hematopoietic stem cell transplantation (HSCT). (A) Study outline. A cohort of 37 pediatric recipients of allogeneic HSCT was monitored prior to HSCT, at the time point of HSCT, and post HSCT (median follow-up time 5.2 years). 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).

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). 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 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 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 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 as well as CD4+ T cell counts 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).

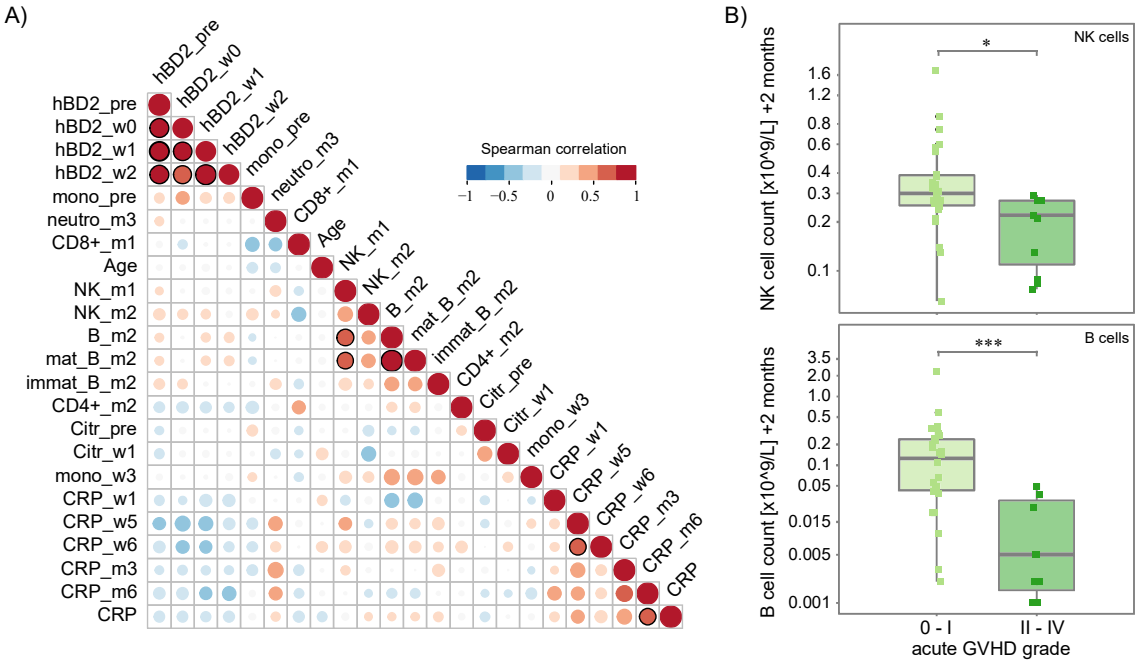


Figure 2. Correlations between immune markers, immune cell counts, and outcomes in patients undergoing HSCT. A) Pairwise Spearman's correlation between immune markers and immune cell counts in HSCT patients (*n=37*) that were determined to be significant in a permutational multivariate analysis of variance 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 *** $P < 0.001$.

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: 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 4: 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 overall model was significant ($P = 0.001$) as well as the first two components (Table 1). The CCpnA confirmed the separation of the data into three clusters as observed in the sPLS regression (Figures 3A, 4, and

Additional file 5: Figure S2A), 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.

Table 1: ANOVA of the Canonical Correspondence Analysis (CCpnA) model, and significant and trending variables.

		F	***X²***	***P*-value**
CCpnA	Overall model	2.4131	4.1983	0.001
	Component 1	12.8964	0.8013	0.001
	Component 2	10.4837	0.6514	0.001
Continuous constrains	CD8+ T cells in month +1	2.8107	0.1746	0.021
	NK cells in month +2	2.2697	0.1410	0. 074
	monocytes pre-HSCT	2.0925	0.1300	0.079

Immune cell counts contributed most to the fit of the ordination. NK cells: natural killer cells.

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. The latter showed a trending contribution to the fit of the ordination (P = 0.079, Table 1). 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 6: Figure S3). 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).

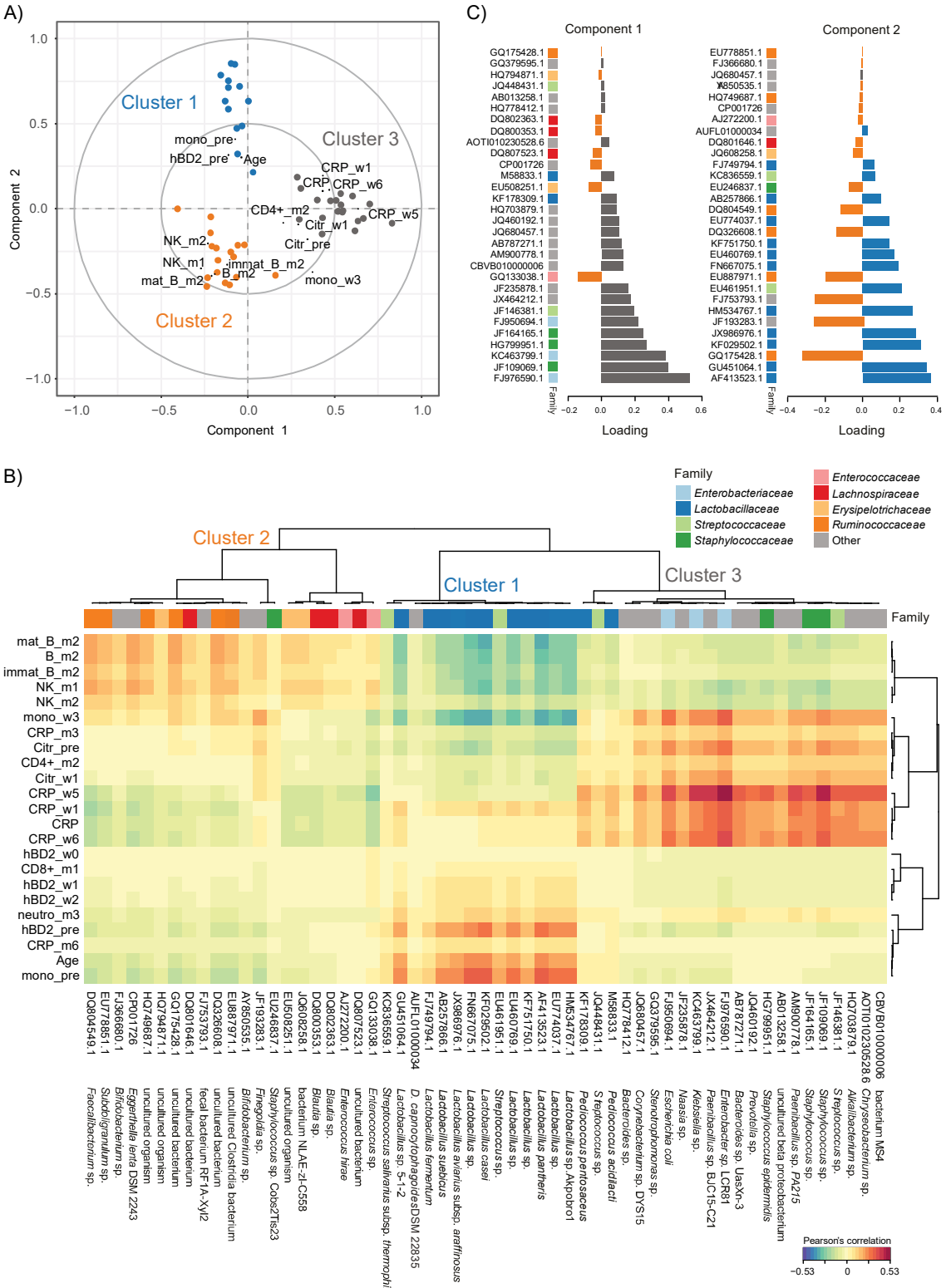


Figure 3. Sparse partial least squares (sPLS) regression of immune parameters and intestinal bacterial taxa during HSCT. (A) Correlation circle plot for the first two sPLS dimensions with correlations displayed for >0.2/<-0.2. The two grey circles indicate correlation coefficient radii at 0.5 and 1.0. 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.

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 4: Table S3 and Additional file 7: Figure S4). 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 6: Figure S3).

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 8: Figure S5). 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 4: 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 4: 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 4: Table S3).

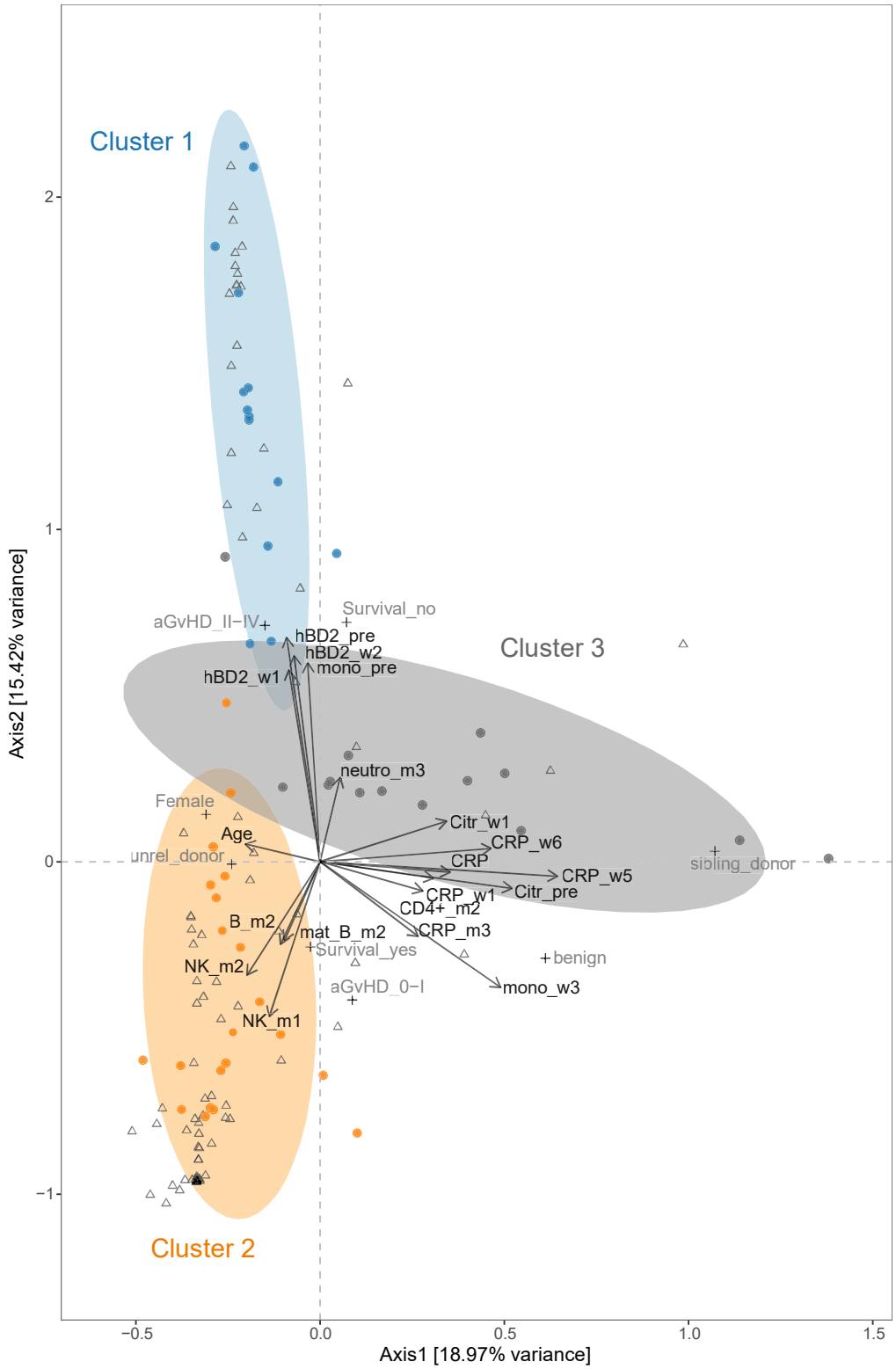


Figure 4. Canonical correspondence analysis (CCpnA) of immune parameters and intestinal bacterial taxa in patients undergoing HSCT. Triplot showing dimension 1 and 2 of the CCpnA that included continuous clinical variables (arrows), categorical variables (+), and OTUs (circles). Samples are depicted 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.

In support of the sPLS regression, the CCpnA also revealed high NK and B cell counts in cluster 2 in months +1 and +2. NK cell counts in month +2 showed a trending contribution to the fit of the ordination (P = 0.074, Table 1). 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 5: Figure S2B). The same pattern was observed for treatment with ciprofloxacin (Additional file 5: Figure S2B).

Community state typing also revealed a state type dominated by *Ruminococcaceae* and *Lachnospiraceae* family members (CST 2) (Additional file 7: Figure S4). 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.

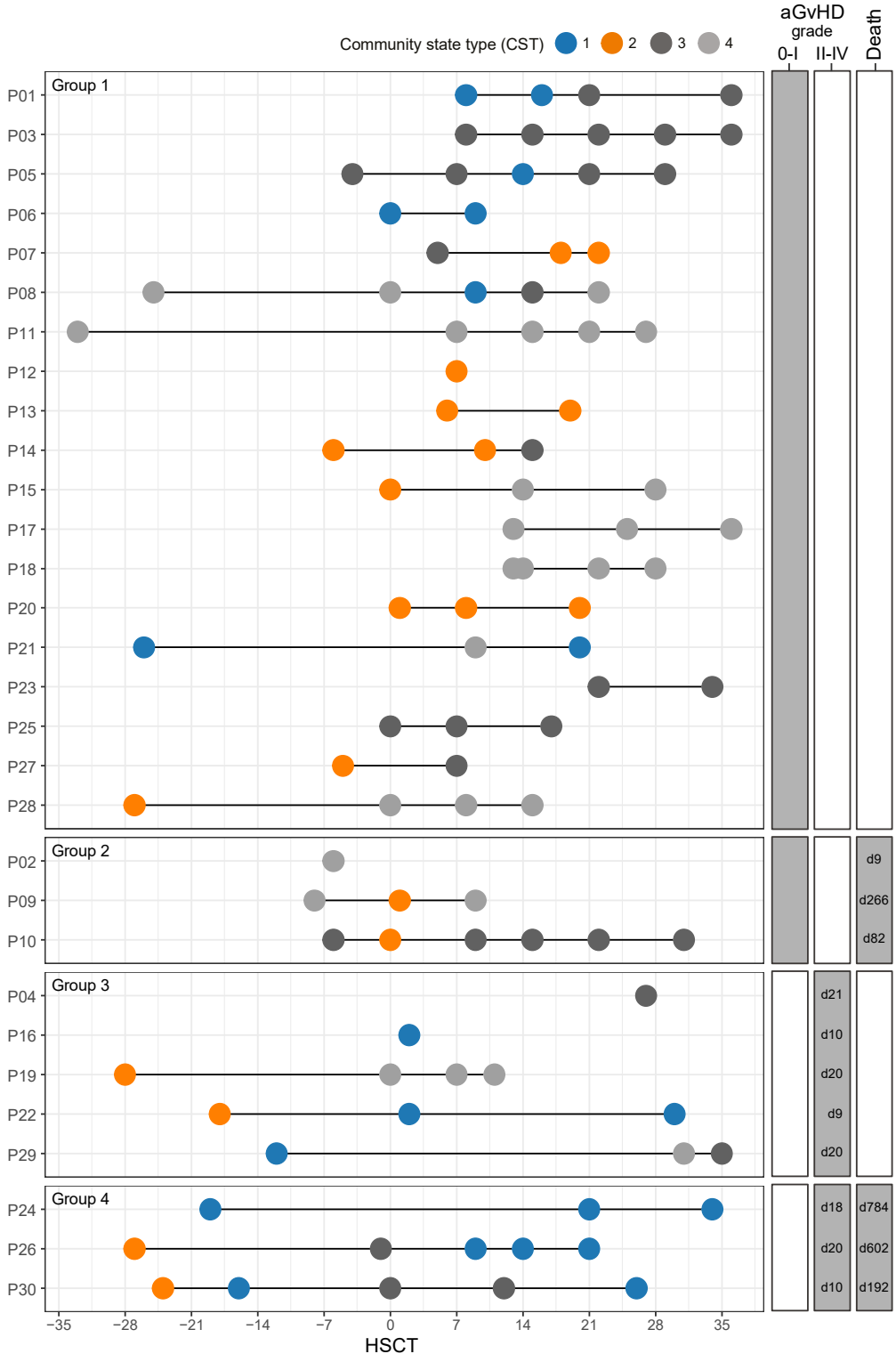


Figure 5. Bacterial community state types over time in patients undergoing allogeneic HSCT. Patients are grouped into four outcome groups: 1. Patients who developed no or mild aGvHD (grade 0-I) and survived vs. 2. Patients who did not survive; 3. Patients who developed moderate to severe aGvHD (grade 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.

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 7: Figure S4). 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 4: 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 4: 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 5: Figure S2A). These samples were characterized by high CRP levels (especially in week +1, +5, and +6). 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 5: Figure S2A). Of note, patients represented in cluster 3 were younger compared to those in the other two clusters (Figure 4).

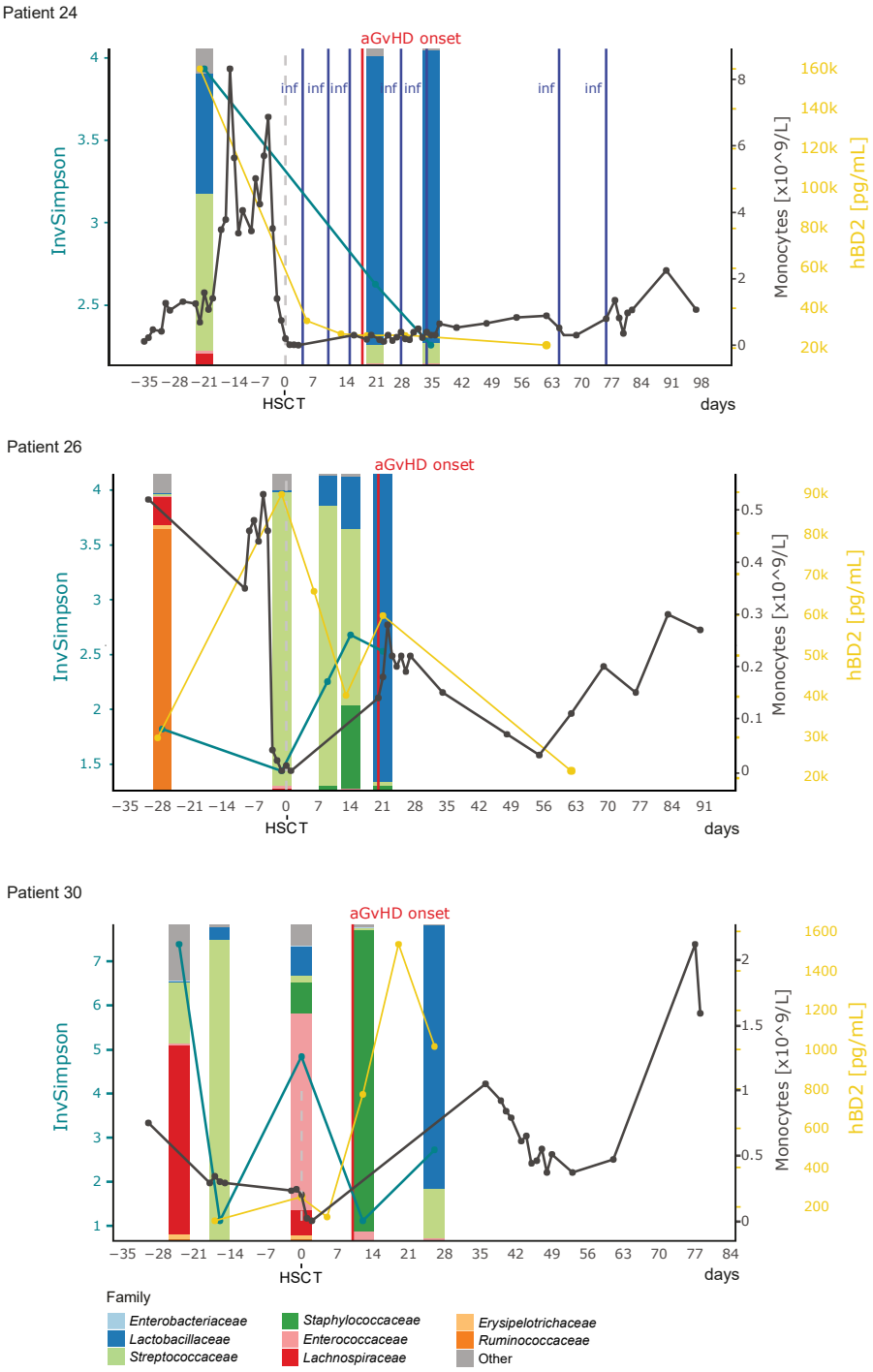


Figure 6. Longitudinal profiles of microbial community composition and immune markers in non-survivors with aGvHD. Abundances of *Lactobacillaceae* increased predominantly after aGvHD onset in patients who died during the follow-up period. Patient P24 developed chronic GvHD on day +187, relapsed on day +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.

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, 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 9: 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 9: 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 in 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*). 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 rheumatoid arthritis patients was previously shown to be associated with an expansion of *Proteobacteria* [42]. The use of a prophylactic ciprofloxacin treatment to prevent chemotherapy- and transplantation-related bloodstream infections is an established method [43], and it was proposed that fluoroquinolones (the antibiotic class comprising ciprofloxacin) can prevent intestinal domination of *Proteobacteria* in HSCT patients [15]. However, ciprofloxacin treatment in healthy subjects has been associated with decreased microbial diversity and decreased *Ruminococcaceae* and *Lachnospiraceae* abundances [44], indicating microbial community disruption. Therefore, our findings further challenge the

choice of antibiotics, such as vancomycin and ciprofloxacin, in patients undergoing chemotherapy and HSCT. However, we did not take the mode of application of the antibiotics into account here, which might limit the strength of our conclusion. It remains to be determined whether certain antibiotics modulate the patients' microbiome and how this might lead to either positive or adverse clinical outcomes.

Conclusions

Our findings support the increasing evidence of microbial involvement in the context of HSCT in cancer patients. We provide evidence for the association between specific microbial taxa and host immune markers. In particular, we examined the prognostic potential of immune markers and gut microbial community dynamics for immune reconstitution and outcomes after HSCT by revealing multivariate associations. We observed increased human beta-defensin 2 in patients with moderate to severe aGvHD and high mortality. In those patients, NK and B cell reconstitution was slow compared to patients with low mortality. These associations only applied when distinct gut microbial abundance patterns were observed, namely low abundances of *Ruminococcaceae* and high abundances of *Lactobacillaceae*. Therefore, hBD2, in connection with longitudinal microbial community pattern surveillance, could be further evaluated as a potential novel candidate marker to identify patients at risk of adverse outcomes (e.g. aGvHD) and slow immune cell reconstitution after HSCT, contributing to improved clinical outcomes. Of note, our cohort comprised a relatively small number of patients with different primary diseases and conditioning regimens, and our findings would therefore benefit from being assessed in larger, more homogenous patient groups. Importantly, microbial abundances also depended on antibiotic treatment. Our findings suggest that certain antimicrobial agents might contribute to a shift from obligate to facultative anaerobes. This highlights the need to assess the usage of specific antibiotics in more detail and to take antibiotic treatment into consideration when describing microbial communities in HSCT recipients.

Methods

Patient recruitment and sample collection

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

Infections and antibiotics

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

Analysis of T, B and NK cells in peripheral blood

T, B and NK cell counts were determined in months +1, +2, +3, and +6 post transplantation. Lymphocyte subsets in peripheral blood were quantified using Trucount Tubes (Becton Dickinson, Albertslund, Denmark) together with the following 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 [45].

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, 46]. 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, 46]. Citrulline levels have previously been described for patients of this cohort in a different context [17, 46].

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 [48]. 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 [49]. 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 database, *Silva_119_rep_set97)*. OTU tables in BIOM format were created with *make_otu_table.py* (and subsequently converted to JSON BIOM format to be compatible with analysis in R [50] with the package *phyloseq* [51]). The OTU table and the taxonomy table are available from figshare (https://dx.doi.org/10.6084/m9.figshare.6508187).

Statistical analyses

Statistical analyses and creation of graphs were performed with the program R (Version 3.4.0, R Foundation for Statistical Computing, Vienna, Austria) [50]. All R scripts documenting our statistical analyses are available from figshare (https://dx.doi.org/10.6084/m9.figshare.6508238). Sequencing data and all related experimental and clinical data (data sets available from figshare, https://dx.doi.org/10.6084/m9.figshare.6508232) were integrated for analysis with the R package

phyloseq and its dependencies [51] (Additional file 10). The resulting phyloseq objects are provided through figshare (https://dx.doi.org/10.6084/m9.figshare.6508235). Plots were generated with the packages *ggplot2* [52], *plotly* [53], and *mixOmics* [54]. Dose-response analysis of the ELISA data was performed with four-parameter log-logistic models in the R package *drc* [55].

Alpha diversity (measured by inverse Simpson index), levels of human beta-defensin 2 (hBD2) concentration, citrulline and C-reactive protein (CRP), as well as monocyte counts, NK cell counts, total B cell counts, and CD4+ T cell counts at different time points were compared using Friedman tests with Benjamini-Hochberg correction for multiple testing (Additional files 11 and 12). As an exception, hBD2 in healthy controls was compared with hBD2 in patients at individual time points by using Wilcoxon rank-sum tests, because hBD2 was only measured once in the healthy control individuals and can therefore not be analyzed in a Friedman test designed for repeated measurements. Monocyte counts in Additional file 2: Figure S1B are depicted at more time points than indicated in Figure 1A because not all time points were included into further analyses. The day of HSCT, week +1 and week +2 were excluded as data were missing for ≥40% of the patients. Months +2, +3, and +6, as well as 1 year were chosen as representative follow-up time points for further analyses as indicated in Figure 1A.

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* [56]. 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* [57] (Additional file 10). The function implements a Gamma-Poisson mixture model [58] 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 13).

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* [51]. Subsequently, permutational multivariate analysis of variance using distance matrices (adonis) for model selection was performed by applying the *adonis2()* function in the package *vegan* [59] (Additional file 14). 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 14).

Correlations among the selected clinical variables were assessed in correlation matrices based on Spearman's rank correlation tests (Additional file 13). Matrices were calculated using the *rcorr()* function of the R package *Hmisc* [60] and displayed with the package *corrplot* [61]. *P*-values were calculated with the *rcorr.adjust()* function with correction for multiple testing (method "Holm").

Sparse PLS regression was performed by applying the *spls()* function in the package *mixOmics* [54] (Additional file 14). The sPLS regression allows the integration of the microbial community data matrix and the clinical variable matrix for multiple regression. It is robust enough to handle collinearity and noise in the data and is suitable to model multiple response variables [62]. The number of clinical variables to be kept in the model for each component (*keepY*) was set to 23, corresponding to the number of variables pre-selected with adonis. We ran the sPLS regression with a range of numbers (20-40) of OTUs to be kept for each component (*keepX*). As the results were robust to this choice, *keepX* was set to 30. The number of components to choose was estimated with the *perf*() function and set to *ncomp* = 2. The sPLS model was run in regression mode. Thereafter hierarchical clustering was performed within the mixOmics *cim()* function based on the sPLS regression model with the clustering method "complete linkage" and the distance method "Pearson's correlation". Coefficients of pairwise correlations between OTU abundances and clinical variables were thereby obtained. Furthermore, loading plots were generated with the function *plotLoadings()* (method = "mean") to visualize loading vectors of specific OTUs that contribute most to the separation of variables in components 1 and 2.

Canonical (i.e., bidirectional) correspondence analysis (CCpnA), a multivariate constrained ordination method, was performed by using the *cca()* function in the package *vegan* [59] (Additional file 14). In this method, the microbial community data matrix is Chi-square transformed and weighted linear regression on pre-selected constraining variables is performed. The resulting fitted values are used for correspondence analysis by singular value decomposition. CCpnA is a constrained method in the sense that it does not aim at depicting all variation in the data, but only the variation directly explained by the constraints (i.e., the provided set of pre-selected variables). The resulting triplot is not displayed as a square representation, but rather corresponds to the percentage of variance explained by axis 1 and 2, respectively, as previously suggested [63]. 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. The significance of the overall CCpnA model, the significance of the components and the significance of the constraints, were assessed in ANOVA using the *anova.cca()* function in package *vegan* [59].

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* [64]) based on a Jensen-Shannon distance of the variance stabilized microbial count data (R code modified after [65]) (Additional file 15). 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 3: 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 16).

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.

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

15. 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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

35. Weber D, Oefner PJ, Hiergeist A, Koestler J, Gessner A, Weber M, et al. Low 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.

1112 38. Rivera-Chávez F, Lopez CA, Bäumler AJ. Oxygen as a driver of gut dysbiosis. Free Radic Biol Med.
1113 2017;105:93–101. doi:10.1016/J.FREERADBIOMED.2016.09.022.

1114 39. Fouhy F, Guinane CM, Hussey S, Wall R, Ryan CA, Dempsey EM, et al. High-throughput
1115 sequencing reveals the incomplete, short-term recovery of infant gut microbiota following
1116 parenteral antibiotic treatment with ampicillin and gentamicin. Antimicrob Agents Chemother.
1117 2012;56:5811–20. doi:10.1128/AAC.00789-12.

1118 40. Weber D, Jenq RR, Peled JU, Taur Y, Hiergeist A, Koestler J, et al. Microbiota Disruption Induced
1119 by Early Use of Broad Spectrum Antibiotics is an Independent Risk Factor of Outcome after
1120 Allogeneic Stem Cell Transplantation. Elsevier Inc.; 2017. doi:10.1016/j.bbmt.2017.02.006.

1121 41. Simms-Waldrip TR, Sunkersett G, Coughlin LA, Savani MR, Arana C, Kim J, et al. Antibiotic-
1122 Induced Depletion of Anti-inflammatory Clostridia Is Associated with the Development of Graft-
1123 versus-Host Disease in Pediatric Stem Cell Transplantation Patients. Biol Blood Marrow Transplant.
1124 2017;23:820–9. doi:10.1016/J.BBMT.2017.02.004.

1125 42. Isaac S, Scher JU, Djukovic A, Jiménez N, Littman DR, Abramson SB, et al. Short- and long-term
1126 effects of oral vancomycin on the human intestinal microbiota. J Antimicrob Chemother.
1127 2017;72:128–36. doi:10.1093/jac/dkw383.

1128 43. Yeh T-C, Liu H-C, Hou J-Y, Chen K-H, Huang T-H, Chang C-Y, et al. Severe infections in children
1129 with acute leukemia undergoing intensive chemotherapy can successfully be prevented by
1130 ciprofloxacin, voriconazole, or micafungin prophylaxis. Cancer. 2014;120:1255–62.
1131 doi:10.1002/cncr.28524.

1132 44. Dethlefsen L, Relman DA. Incomplete recovery and individualized responses of the human distal
1133 gut microbiota to repeated antibiotic perturbation. Proc Natl Acad Sci U S A. 2011;108 Suppl
1134 Supplement 1:4554–61. doi:10.1073/pnas.1000087107.

1135 45. Kielsen K, Jordan KK, Uhlving HH, Pontoppidan PL, Shamim Z, Ifversen M, et al. T cell
1136 Reconstitution in Allogeneic Haematopoietic Stem Cell Transplantation: Prognostic Significance of
1137 Plasma Interleukin-7. Scand J Immunol. 2015;81:72–80. doi:10.1111/sji.12244.

1138 46. Jordan K, Pontoppidan P, Uhlving HH, Kielsen K, Burrin DG, Weischendorff S, et al.
1139 Gastrointestinal Toxicity, Systemic Inflammation, and Liver Biochemistry in Allogeneic Hematopoietic
1140 Stem Cell Transplantation. Biol Blood Marrow Transplant. 2017;23:1170–6.
1141 doi:10.1016/j.bbmt.2017.03.021.

1142 47. Kielsen K, Ryder LP, Lennox-Hvenekilde D, Gad M, Nielsen CH, Heilmann C, et al. Reconstitution
1143 of Th17, Tc17 and Treg cells after paediatric haematopoietic stem cell transplantation: Impact of
1144 interleukin-7. Immunobiology. 2018;223:220–6. doi:10.1016/j.imbio.2017.10.023.

1145 48. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur:
1146 open-source, platform-independent, community-supported software for describing and comparing
1147 microbial communities. Appl Environ Microbiol. 2009;75:7537–41. doi:10.1128/AEM.01541-09.

49. 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.

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

51. 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.

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

53. 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.

54. 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.

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

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

57. 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.

58. 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.

59. 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.

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

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

62. 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.

63. 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.

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

65. 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.

4.2 Manuscript II

The gut, oral, and nasal microbiota in pediatric allogenic HSCT and microbial marker candidates for predicting acute GvHD severity

Anna Cäcilia Masche[1], Katrine Kielsen[2, 3], Hanne Mordhorst[1], 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.

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

Manuscript in preparation

Abstract

Background

After allogeneic hematopoietic stem cell transplantation, the community structure of the stem cell recipient's microbiota undergoes major shifts. Important knowledge has been acquired concerning changes early after transplantation. However, stronger focus on the subsequent reconstitution of the microbiota is needed. Evaluating the resilience of the commensal community is of importance since it provides a crucial functional repertoire to the human host. This applies not only to the microbiota residing in the gut, but also to other body sites, such as the oral and nasal cavities. For instance, immunomodulatory properties of the microbiotas at these sites might be of relevance locally and systemically in HSCT, but have hitherto not been studied as extensively as the gut in HSCT patients. Here, we have tracked gut, oral, and nasal microbial changes over a period of one year and assessed host-microbe associations at all three body sites. Allogeneic HSCT is often accompanied by side effects, such as bloodstream infections and graft-versus-host disease. Specific gut bacterial taxa are depleted, or expand after HSCT, which has been linked to aGvHD. To advance our knowledge about host immune-microbial associations in HSCT and GVHD towards clinical prognostic applications, we used pre-HSCT gut, oral, and nasal bacterial abundances in machine learning models for the prediction of acute GvHD severity post-transplant.

Results

For each body site, we employed sparse linear discriminant analysis informed by phylogenetic trees to determine which bacterial taxa are discriminating the community structure between time points. For amplicon sequence variants affiliated with intestinal *Blautia* spp., oral *Actinomycetacea* and *Streptococcacea*, and nasal *Corynebacteriaceae*, we revealed 'smile'-shaped temporal bacterial abundance trajectories. That is, their abundances decreased after HSCT and returned to pre-transplant levels after three months. In contrast, we identified 'frown'-shaped trajectories for intestinal *Enterococcus* spp. and *Lactobacillus* spp., i.e. a post-transplant expansion followed by a decrease in abundance after three months. Moreover, by implementing support vector machines with linear kernel, we predicted high acute GvHD severity from high pre-transplant abundances of intestinal *Parabacteroides distasonis*, and oral and nasal *Actinomyces* spp.. In multivariate multi-table analyses, we showed host immune-microbial associations shared between body sites. For instance, CD4+ T cell and T_H17 cells counts were positively correlated with abundances of distinct bacterial taxa in the gut, and the oral and nasal cavities.

Conclusions

Our analysis indicates that the period between one and three months after allo-HSCT might be a crucial time for the reconstitution of the microbiota in the gut, and the oral and nasal cavities. Denser sampling during this period might allow determining an accurate temporal threshold that can be used as a reference in the clinical setting. We show that pre-transplant bacterial abundances predict acute GvHD severity post-transplant, which might serve as a starting point for employing early microbial predictive markers for preventive management of adverse clinical outcomes. Fecal sample collection during clinical routine depends on the patient's compliance and bowel movements, which is often impaired in allo-HSCT recipients. Therefore, our findings regarding the microbiota at body sites that are easier accessible for

sampling, i.e. the oral and nasal cavities, are of value for a potential future implementation of regular microbial screening in the clinical setting. We point out that patients' clinical baseline parameters, such as primary disease, need to be taken into consideration when comparing host-microbial associations at different body sites.

Background

In allogeneic hematopoietic stem cell transplantation (HSCT), the infusion of donor derived stem cells is employed as a curative approach for various types of hematologic disorders [1]. In allo-HSCT patients, the community structure of the human microbiota changes subsequent to transplantation [2]. This is to a large extend attributed to preceding antimicrobial treatment and chemotherapeutic conditioning regimens [2, 3]. The microbiota in the gut, and at other mucosal body sites, such as the oral and nasal cavities, is in general affected by treatment with antibiotics to different degrees [4]. Especially the depletion of butyrate-producing bacteria affiliated with the *Clostridiales* class are depleted in the gut early after HSCT, while *Proteobacteria*, and *Lactobacillales* such as *Enterococcus* spp. expand, possibly due to increased oxygen levels in the intestinal lumen in the absence of butyrate [2, 3, 5]. Conditioning-induced intestinal epithelial permeability promotes bacterial translocation to the bloodstream where opportunistic pathogens can cause bacteremia [6]. This might be aggravated in patients with graft-versus-host disease (GvHD), which is a common side effect of allo-HSCT where alloreactive donor T cells exhibit cytotoxic activity against healthy tissue in the host, including the gut epithelium [7]. Inflammation both facilitates the development of GvHD, and is enhaced during GvHD [7]. Especially the T helper 17 (T_H17) cell/T regulatory (T_{reg}) cell ratio has been attributed a role in GvHD, which is however controversially discussed [8–10]. While T_H17 cells in steady state foster mucosal integrity, they can switch to be pathogenic and promote pro-inflammtory signaling upon intense cytokine stimulation [11]. In GvHD, a high T_H17/Treg ratio might promote inflammation, contributing to GvHD development [9]. However, a low T_H17/Treg has also been observed in relation to GvHD, suggesting a compensatory increase of T regulatory (T_{reg}) cells upon inflammation [8, 10]. The microbiota exerts immunomodulatory function on the host's adaptive immune system, e.g. on T cells [11]. For instance, in germ-free mice it has been shown that human commensal gut strains affiliated with *Bacteroides* and *Clostridia* can induce T_{reg} cells [12]. The microbiota at body sites other than the gut, e.g. the oral and nasal cavities, have also been suggested to be involved in immunomodulation [13]. We have previously proposed that the gut microbiota is associated with immune cell reconstitution after allo-HSCT [14]. It is of interest to investigate whether microbiotas at other mucosal sites also play a role in immune reconstitution, and whether they have an influence on the T_H17/Treg ratio and thereby GvHD. In this context, the changes in microbial abundances early after HSCT need to be taken into consideration, but also the recovery of the microbiota at late follow-up time points. This requires long-term monitoring of the microbiota after HSCT, in addition to the focus on early post-transplant time points. Moreover, pre-transplant time points are of great importance to make predictions from microbial abundances on subsequent adverse clinical outcomes, such as GvHD, which in the future might aid early preventive management in the clinical setting.

Here, we tracked changes in the community structure of the gut, oral, and nasal microbiota in pediatric allogeneic HSCT patients over a period of one year. We inferred amplicon sequence variants (ASVs) from 16S rRNA gene sequences and generated phylogenetic trees. By implementing tree-based sparse linear

discriminant analyses (LDA), we reveal sets of phylogenetically closely related ASVs that characterize differential community structures between time points. At the three examined body sites, we identify 'smile'- and 'frown'-shaped temporal microbial abundance trajectories. For instance, distinct intestinal *Lachnospiraceae* and *Ruminococcacea* ASVs, as well as oral *Actinomycetacea* and *Streptococcaceae* ASVs exhibited 'smile'-shaped trajectories. That is, their abundances decreased early after HSCT and recovered from month+3 onwards. I contrast, we show 'frown'-shaped trajectories of intestinal *Enterococcaceae* and *Lactobacillaceae* ASVs that expanded after HSCT and decreased in abundance from month +3 onwards. Furthermore, by using a machine learning approach, we predict aGvHD post-transplant from pre-transplant ASV abundances in the gut, and the oral and nasal cavities. By relating ASV abundances to immune cell counts, immune markers, clinical outcomes, and patients' baseline parameters, we uncover similarities in host-microbial associations at different body sites. For instance, we demonstrate associations with T_H17 cell counts and thereby propose new microbiota-related considerations for the controversially discussed role of those T cell subtypes in GvHD. We suggest that associations of the microbiota at different body sites with clinical outcomes after allo-HSCT might depend on baseline parameters, such as the underlying primary disease.

Results

We characterized long-term microbiota dynamics in the context of pediatric, allogeneic hematopoietic stem cell transplantation (HSCT) at three body sites: the gut, and the oral and nasal cavities (Figure 1). Fecal samples, buccal swabs, and anterior naris swabs were collected from 29 pediatric allo-HSCT recipients at 10 time points over a one-year period: 2 times prior to HSCT, on the day of HSCT, weekly during the first month after HSCT, and at three follow-up time points up to twelve months (Figure 1). Microbial community dynamics in these samples were determined by 16S rRNA gene profiling. We utilized these comprehensive data to predict the development of acute GvHD, one of the most common and severe side effects following HSCT, from pre-transplant fecal, oral, and nasal microbial abundances. Furthermore, we assessed multivariate associations between the microbiota and immune reconstitution, immune markers, and outcomes in relation to HSCT. As a means of gaining insight into immune reconstitution, we determined counts of T, B, and NK cells, as well as leukocyte subpopulations in peripheral blood (Figure 1). We assessed a number of kidney and liver function markers (e.g. plasma cystatin C and triglyceride levels), as well as plasma immunoglobulin (Ig) levels (IgA, IgG, and IgM) as markers of infection (Figure 1, see Methods).

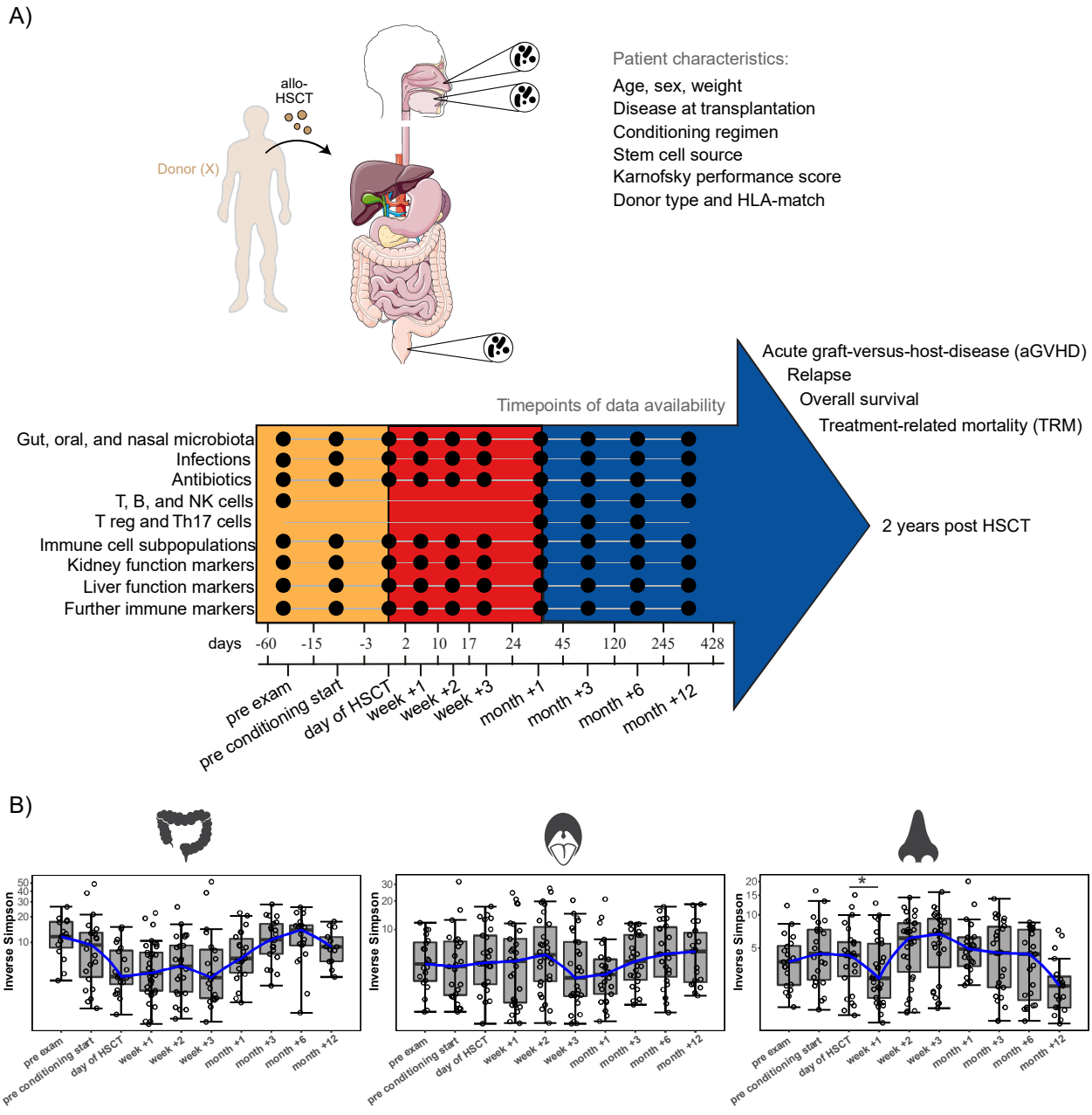


Figure 1. Monitoring gut, oral, and nasal microbiota and the host immune system in allogeneic hematopoietic stem cell transplantation (HSCT). A) Twenty-nine children were monitored before, at the time of, and immediately post allogeneic HSCT, as well as at late follow-up time points. Patients' baseline characteristics, clinical outcomes, as well as immune cell counts, and immune markers over time were monitored. Patient characteristics are described in detail in Table S1. Host immune system parameters were related to longitudinal dynamics of the gut, oral, and nasal microbiota that was assessed at the denoted time points. B) Bacterial alpha diversity before, at the time of, and after HSCT at each body site, displayed on a log10 transformed y-axis for visualization purposes. Asterisks indicate significant differences in median inverse Simpson index between time points * $P < 0.05$, ** $P < 0.01$ and *** $P < 0.001$. ('Human body'-vector graphic in A) from: smart.servier.com. 'Body site'-icons in B) and the following figures from: colourbox.com.)

Patient cohort and outcomes

The cohort comprised 29 children, aged 9.5 years on average (range: 2.5-16.4) at the time of HSCT. At least one bacterial infection throughout the monitored period was reported for 25 patients (86.2%). Nine patients (31%) had no or mild aGvHD (grade 0 or I) and 20 patients (69%) developed moderate to severe aGvHD (grade II-IV) following HSCT at median 14 days (range: day +9 to day +61) (Table S1). During the follow-up period of 21.4 months on average (range: 10.1 – 32.7 months), two patients (7%) relapsed and one patient underwent retransplantation. Three patients (10%) died (one relapse-related death at day +91 and two treatment-related death at days +111 and +241, respectively). All patients were treated with antibiotics prophylactically, and during febrile neutropenia or upon infections indicated by positive culture. Trimethoprim-sulfamethoxazole was given to all patients in advance of transplantation. In case of fever, the prophylactic antimicrobial treatment was substituted with meropenem, vancomycin, ciprofloxacin, phenoxymethylpenicillin, or other antibiotics corresponding to culture-based results.

Bacterial alpha diversity changes in the gut, and the oral and nasal cavities

In order to gain insight into the gut, oral, and nasal microbial community structures in relation to allogeneic HSCT, 212 fecal samples, 248 oral swabs, and 249 nasal swabs from 10 time points were characterized by 16S rRNA gene sequence analysis. After all filtering steps (see Methods), we retained 2465, 377, and 197 ASVs in the fecal, oral, and nasal core microbiota sets, respectively. We assessed microbial alpha-diversity by means of the inverse Simpson index. Alpha diversity in the gut was generally the highest, followed by the oral cavity, and the nose (Figure 1B). We observed the lowest alpha diversity at a time point within the first month post HSCT at all three body sites (gut and oral cavity: week +3, nasal cavity: week +1). However, a significant decrease in diversity was only found in the nasal cavity (Figure 1B). Median alpha diversity in the nasal cavity decreased significantly from 4.43 (range: 1.38 – 48.66) at conditioning start to 2.65 (range: 1.05 – 22.2) in week +1 (P = 0.02) (Figure 1B).

To elucidate changes of the gut, oral, and nasal microbial communities over time, before, at the time of and after HSCT, we initially assessed relative abundances of the 12 most abundant families at each body site (Figures S1, 2A, 3A, and S3A). In the gut, we observed a reduction of *Lachnospiraceae* immediately after HSCT from 13% relative abundance at pre-examination to 4.7% in week +1, followed by a recovery to 27.5% in month +3 (Figure 2A). Concurrently, an expansion in proportion of *Enterococcaceae* (pre-examination: 6.1%, week +1: 22.8%) and *Lactobacillaceae* (pre-examination: 2%, week +1: 7%), followed by a reduction from month +3 onwards to 0.2% and 0.6%, respectively (Figure 2A).

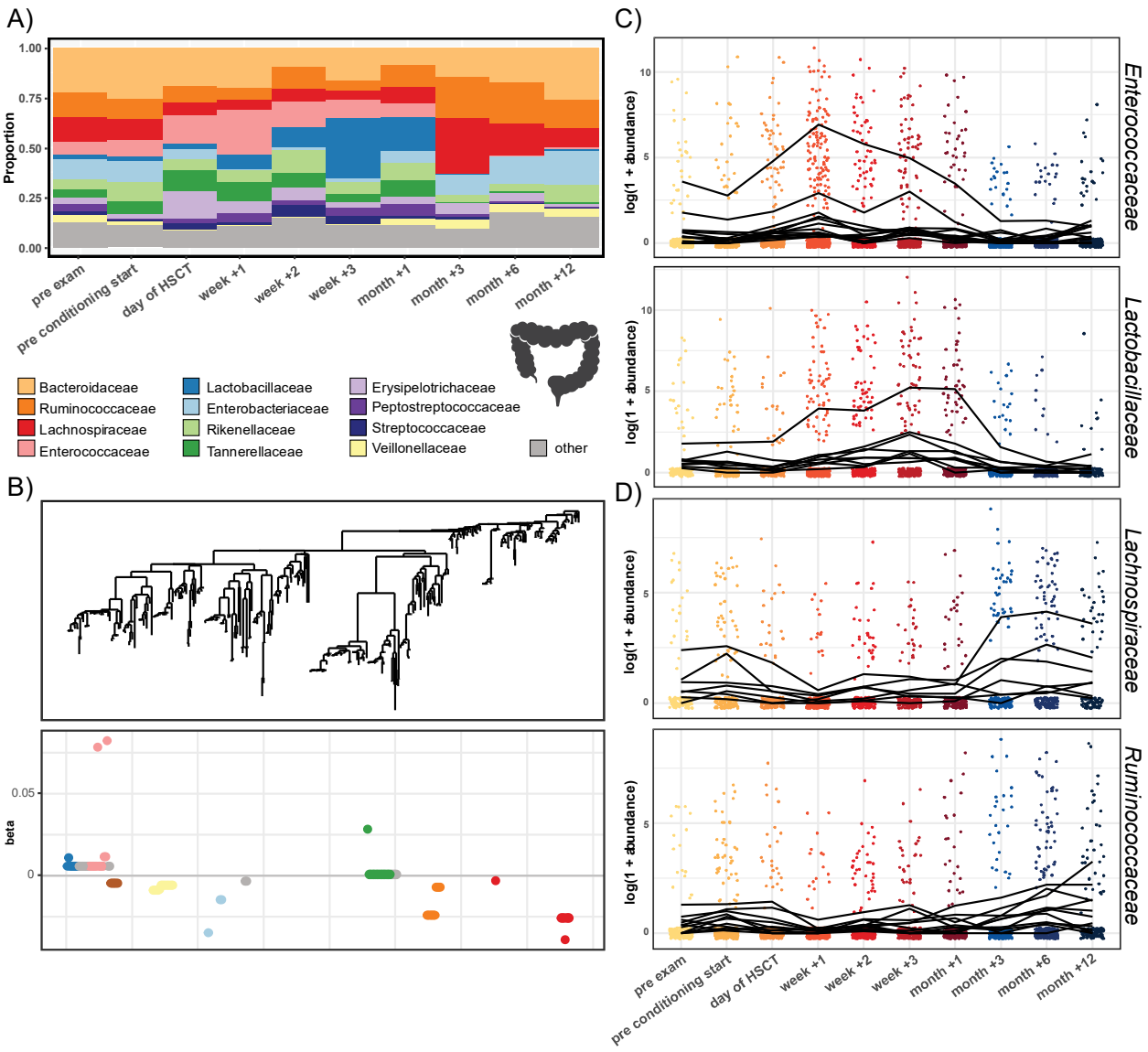


Figure 2. Tree-based sparse linear discriminant analysis revealing gut ASVs that distinguish time points in relation to HSCT. A) Relative abundances over time of the 12 most abundant families in the gut. B) Coefficients of discriminating clades of ASVs on the first LDA axis, colored by taxonomic family, and plotted along the phylogenetic tree. C) Trajectories of ASVs in two discriminating groups, affiliated with the families *Enterococcaceae* and *Lactobacillaceae,* with increasing abundances after HSCT. D) Trajectories of ASVs in two discriminating groups, affiliated with the families *Lachnospiraceae* and *Ruminococcaceae*, with decreasing abundances after HSCT and recovery at late follow-up time points. Detailed taxonomic information and LDA-coefficients of the displayed ASVs are listed in Table S2.

Blautia spp., _Enterococcus_ spp. and _Lactobacillus_ spp. discriminate the gut microbial community structure between time points

For a more detailed assessment of clades of ASVs in the gut, oral cavity, and nasal cavity that best characterized the separation of samples from different time points, we performed tree-based sparse linear discriminant analyses (LDA) for each body site. This supervised method implements prior information about phylogenetic relationships between ASVs and induces sparsity constraints to increase interpretability. When plotting the sample scores of the first and second LDA-axis against each other, we observed at all three body sites that samples at pre-examination and from month +3 onwards clustered together, and were separated from samples from the day of HSCT and the first month thereafter (Figure S2). The first LDA-axis best represented the maximized separation by time points of fecal, oral, and nasal samples, respectively (Figure S2). Therefore, we focused on microbial predictors that were identified on the first axes.

The LDA revealed 19 discriminating clades of fecal ASVs on the first axis that best separated samples by time point (Figure 2B). The two most discriminating groups of phylogenetically closely related ASVs with positive LDA-coefficients comprised a group of ASVs of the family *Enterococcaceae,* and a group of *Lactobacillaceae* (Figure 2B). In support of the proportions of these families (Figure 2A), the abundances of ASVs within these two discriminating groups increased from conditioning start (*Enterococcaceae*) and day of HSCT (*Lactobacillaceae*), respectively (Figure 2C). From month +3 onwards, their abundances decreased again to levels comparable to the time of pre-examinantion (Figure 2C). All members of the *Enterococcaceae* group were *Enterococcus* spp. (Table S2). Specific individual ASVs (ASV 1 (*Enterococcus* sp.), detected at least once in all patients, and ASV 3 (*Lactobacillus* sp.), detected at least once in 26 patients) exhibited higher abundances and higher coefficients compared with the other group members and were particularly characteristic of the observed pattern of expansion after transplantation (Figure 2B, Table S2). The two most discriminative groups of ASVs with negative LDA-coefficients included two individual ASVs and one clade of the *Lachnospiraceae* family*,* and two *Ruminococcaceae* clades (Figure 2B). The abundances of these ASVs decreased in week +1 and recovered from month +1 onwards, returning to abundances comparable with before HSCT or higher (Figure 2D), in agreement with the initial patterns observed for those families (Figure 2A). All ASVs within the *Lachnospiraceae* group belonged to the genus *Blautia*.

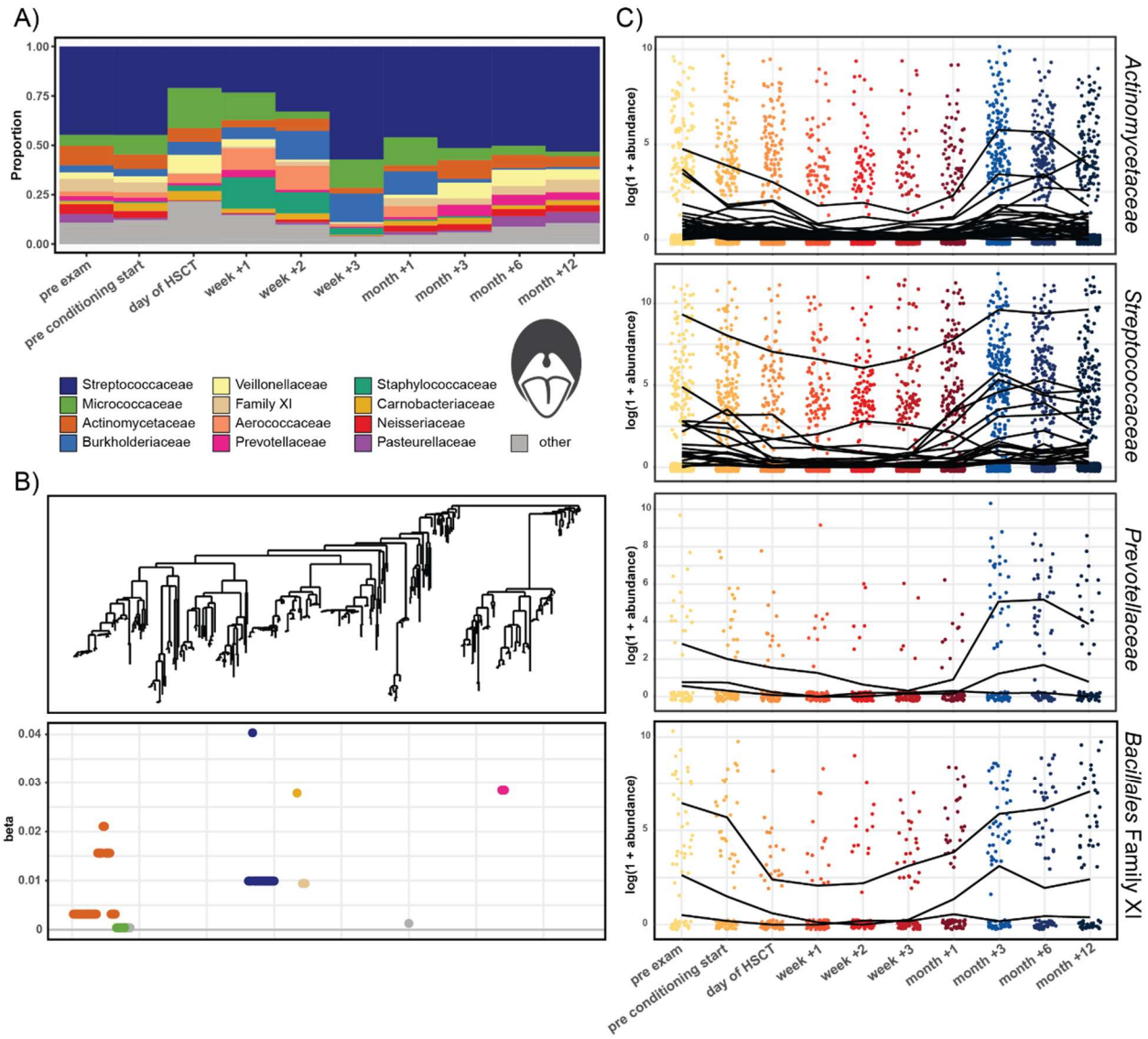


Figure 3. Tree-based sparse linear discriminant analysis revealing oral ASVs that distinguish time points in relation to HSCT. A) Relative abundances over time of the 12 most abundant families in the oral cavity. B) Coefficients of discriminating clades of ASVs on the first LDA axis, colored by taxonomic family, and plotted along the phylogenetic tree. C) Trajectories of ASVs in four discriminating groups, affiliated with the families *Actinomycetaceae, Streptococcaceae, Prevotellaceae, and* Family XI (Class *Bacillales),* with decreasing abundances after HSCT and recovery at late follow-up time points. Detailed taxonomic information and LDA-coefficients of the displayed ASVs are listed in Table S2.

Actinomyces spp. and *Streptococcus* spp. discriminate the oral bacterial community structure between time points

In the oral cavity, we observed reduced relative abundance of *Actinomycetaceae*, especially in weeks +1 and +3 and in month +1 (Figure 3A). Furthermore, *Streptococcaceae* abundances were lower from the day of HSCT until week +2 compared with before HSCT and late follow-up time points (Figure 3A). Tree-based

sparse LDA identified 10 clades of ASVs at the oral site that best separated samples by time points along the first axis. The most discriminating clades on the first LDA axis had exclusively positive coefficients (Table S2). The two largest discriminating groups of ASVs were affiliated with two families, namely *Actinomycetaceae*, and *Streptococcaceae* (Figure 3B). Additional discriminating ASV's were affiliated with *Prevotellaceae*, and Family XI (*Gemella* spp.), respectively. In agreement with relative abundance dynamics observed for the respective families (Figure 3A), all clades shared a pattern of depletion from the day of HSCT or week+1 onwards, until their abundances recovered from month +1 onwards, similar to the temporal pattern observed for *Ruminococcaceae* and *Lachnospiraceae* in the gut.

Corynebacteriaceae and _Streptococcaceae_ abundances in the nasal cavity decrease after HSCT

In the nasal cavity, the proportion of *Corynebacteriaceae* was decreased during the first month after HSCT and recovered from month +3 onwards (Figure S3A). The LDA revealed 30 discriminating nasal clades on axis 1 many of which consisted of individual ASVs (Figure S3B). High abundances of closely related ASVs affiliated with the same family were not always associated with the same time points, i.e. some had positive coefficients, while others had negative coefficients. A group of *Corynebacteriaceae* ASVs and a group of *Streptococcaceae* ASVs all had positive coefficients (Figure S3B). Corynebacteriaceae abundances decreased after HSCT and increased again from month+1 onwards (Figure S3C).

High pre-HSCT abundances of distinct _Parabacteroides_ spp. in the gut predict moderate to severe aGvHD

To reveal potential associations between the bacterial community structure and the severity of acute GvHD, we examined relative abundances of the 12 most abundant families at each body site in patients with no or mild (grade 0-I) and moderate to severe (grade II-IV) aGvHD. In the gut, *Tannerellaceae* were more abundant at time points before HSCT in patients with grade II-IV compared to grade 0-I, especially at pre-examination and at conditioning start (Figure 4A). In order to predict the development of aGvHD (grade 0-I versus grade II-IV) following HSCT from gut, oral, and nasal microbial abundances at time points up until the time of stem cell infusion, we implemented machine-learning models. We identified support vector machines with linear kernel (svmLinear) as the best performing classifier out of four (see Methods). In the fecal data set, the final cross-validated svmLinear model predicted aGvHD grade (0-I versus II-IV) from the abundances of distinct gut ASVs pre-HSCT with 86% accuracy (95% CI: 65% to 97%) (Figure 4B). Of the 50 most important gut ASVs identified by the svmLinear model, 23 were confirmed by Boruta feature selection (Table S3). Five of these ASVs (22%) were *Parabacteroides* spp. (family *Tannerellaceae*), and 3 ASVs were affiliated with *Lachnospiraceae* (13%). A conditional inference tree (CTREE) was then fit for those 23 predictive ASVs to assess their effect on the outcome (aGvHD grade) via nonparametric regression. In each step of the CTREE analysis, it is evaluated whether any of the ASVs' abundances explain the aGvHD grade. If so, a node is created that is split by the ASV that best explains the aGvHD grade. In the next recursive step, the procedure is repeated with the remaining set of ASVs. The recursion is stopped as soon as no significant association between any of the remaining ASVs and aGvHD is detected anymore. The CTREE analysis revealed 3 significant predictive ASVs in the gut: ASV 128 (*Parabacteroides distasonis,* $P < 0.01$), ASV 268 (*Lachnospiraceae* NK4A136 group sp., $P = 0.01$) and ASV 3 (*Lactobacillus* sp., $P < 0.01$) (Figure 4C). This means, high abundances of these ASVs before HSCT were associated with the subsequent development aGvHD grade II-IV (Figure 4C). For instance, all samples with a variance stabilized abundance

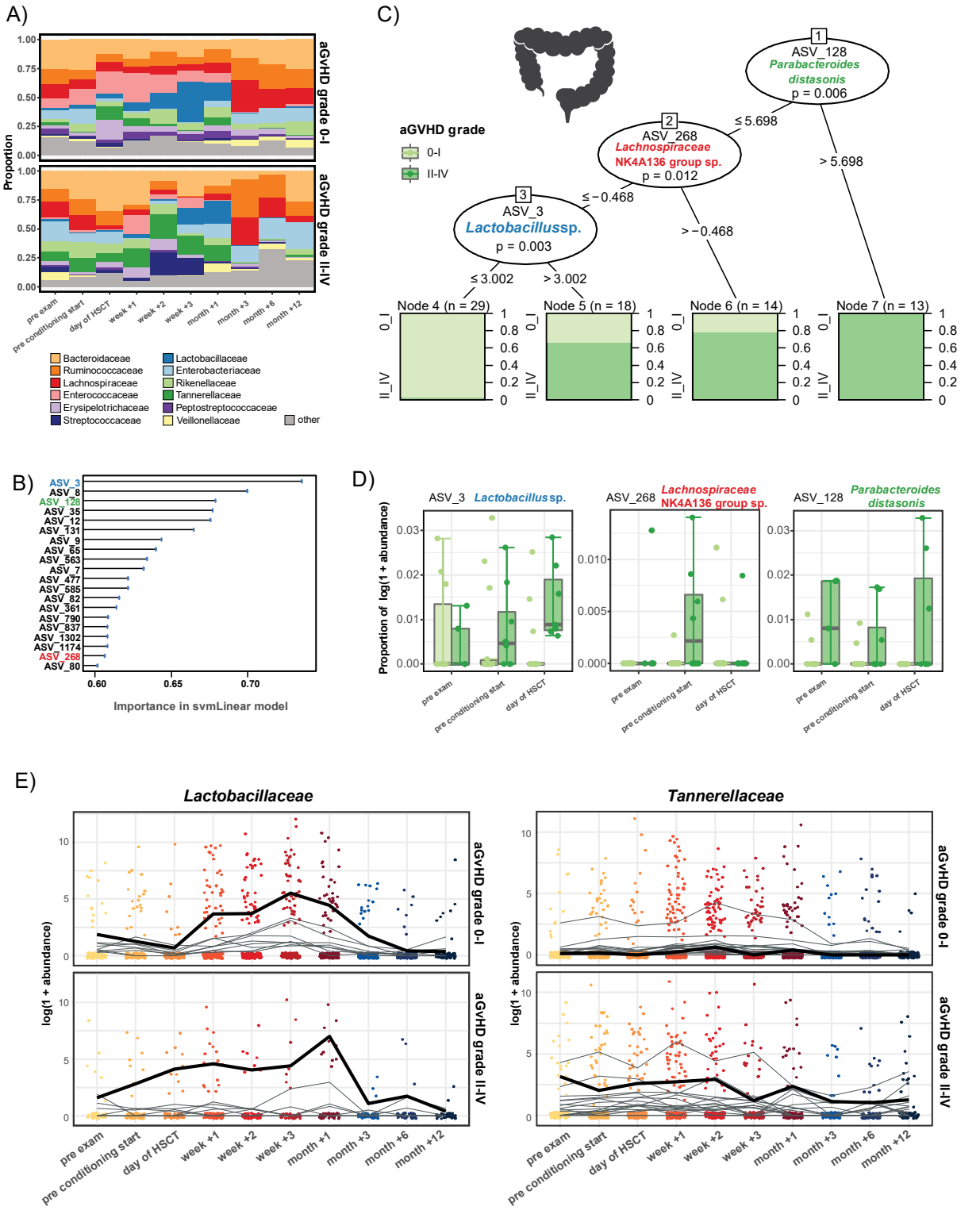


Figure 4. Machine learning-based prediction of aGvHD severity from gut microbial abundances pre-HSCT. A) Relative abundances over time of the 12 most abundant families in the gut in patients with aGvHD grade 0-I versus II-IV. B) Importance plot of top 20 predictive gut ASVs identified by the svmLinear model with importance scores indicating the mean decrease in prediction accuracy in case the respective ASV would be excluded from the model.

C) Conditional inference tree (CTREE) displaying ASVs identified as significant split nodes by nonparametric regression for prediction of aGvHD. Numbers along the branches indicate split values of variance stabilized bacterial abundances. The terminal nodes show the proportion of samples originating from patients with aGvHD grade 0-I vs II-IV. D) Boxplots depict the log transformed relative abundances of the predictive ASVs at time points up to the transplantation in aGvHD grade 0-I compared with grade II-IV patients. E) Trajectories of *Lactobacillaceae* and *Tannerellaceae* ASVs that were identified by tree-based sparse LDA, including those predictive for aGvHD (bold lines), in patients with aGvHD grade 0-I vs II-IV.

>5.7 of ASV 128 (*Parabacteroides distasonis*) and 67% of samples with a variance stabilized abundance >3 of ASV 3 (*Lactobacillus* sp.) originated from patients who later on developed aGvHD grade II-IV (Figure 4C). In agreement, log transformed relative abundances of these ASVs were mostly higher at pre-examination, conditioning start, and the day of HSCT, in patients who later on developed aGvHD grade II-IV compared with those exhibiting grade 0-I (Figure 4D). For instance, the average abundance of ASV 128 *(Parabacteroides distasonis*) was 5.5 times higher at pre-examinantion in grade II-IV versus in grade 0-I patients (Figure 4D). We observed ASV 3 (*Lactobacillus* sp.) as a member of the group of *Lactobacillaceae* that discriminated samples from different time points (Figure 2B). Within this group, ASV 3 was the single most discriminating ASV with the highest LDA-coefficient (0.01 compared to 0.006 for the remaining ASVs, Table S2). When examining temporal trajectories of this group of *Lactobacillaceae* in relation to aGvHD, we again observed that the abundance of ASV 3 was higher at time points up to the transplantation in patients with grade II-IV versus those with grade 0-I (Figure 4E). Within this group of *Lactobacillaceae*, this pattern seemed to be restricted to this particular *Lactobacillus* sp. ASV (Figure 4E). Moreover, all five *Parabacteroides* spp. ASVs were part of the discriminating group of *Tannerellaceae* identified in the LDA (Figure 4E). Their trajectories facetted by aGvHD severity confirmed the observation of increased pre-HSCT abundances of these ASVs in patients with subsequent development of aGvHD grade II-IV (Figure 4E).

High pre-HSCT abundances of distinct oral ASVs predict moderate to severe aGvHD

In the oral cavity, the bacterial community before HSCT in patients with grade II-IV was characterized by e.g. a lower relative abundance of *Neisseriaceae*, and higher relative abundances of *Aerococcaceae* and *Prevotellaceae*, compared with grade 0-I patients, especially at pre-examination and conditioning start (Figure 5A). Here, we generated a cross-validated svmLinear model, which predicted aGvHD grade (0-I versus II-IV) from the abundances of selected oral ASVs pre-HSCT with 92% accuracy (95% CI: 73% to 99%) (Figure 5B). Boruta feature selection confirmed 26 of the 50 most important oral ASVs identified by the svmLinear model (Table S3). The CTREE fitted on those 26 ASVs found 3 significant oral ASVs that predicted aGvHD: ASV 568 (*Actinomyces* sp., $P < 0.001$), ASV 226 (*Prevotella melaninogenica*, $P < 0.001$) and ASV 500 (*Pseudopropionibacterium propionicum*, $P < 0.001$) (Figure 5C). High abundances of these ASVs before transplantation predicted the development of aGvHD grade II-IV after HSCT (Figure 5C). For instance, 91% of samples with a variance stabilized abundance >0.4 of ASV 568 (*Actinomyces* sp.) and 92% of samples with a variance stabilized abundance >6.1 of ASV 226 (*Prevotella melaninogenica*) originated from patients with subsequent development of aGvHD grade II-IV (Figure 5C). Likewise, pre-HSCT log transformed relative abundances of these ASVs were higher in those patients, e.g. the average abundance

of ASV 500 *(Pseudopropionibacterium propionicum*) on the day of HSCT was 10 times higher in grade II-IV versus in grade 0-I patients (Figure 5D). ASV 226 and ASV 568 also contributed to the separation of oral samples by time points (Figure 3C). Temporal trajectories of the associated oral *Prevotellaceae* and *Actinomycetaceae* groups in relation to aGvHD, again showed that the abundances of ASV 226 and ASV 568 were higher at time points up to the transplantation in patients with grade II-IV versus those with grade 0-I (Figure 5E).

High pre-HSCT abundances of distinct nasal ASVs predict moderate to severe aGvHD

The proportion of nasal *Neisseriaceae* was lower pre-HSCT in patients with aGvHD grade II-IV vs 0-I (Figure S4A). In contrast, *Actinomycetaceae* at conditioning start and *Corynebacteriaceae* at the day of HSCT were higher abundant in aGvHD grade II-IV patients compared to grade 0-I (Figure S4A). The final cross-validated svmLinear model predicted aGvHD grade with 76% accuracy (95% CI: 56% to 90%) from variance stabilized abundances of selected nasal ASVs (Figure S4B). Twelve out of the 50 most important ASVs were also selected by Boruta feature selection. Three out of these 12 ASVs were *Rothia* spp. (family *Micrococcaceae*, Table S3). By fitting a CTREE, we found two ASVs significantly affecting aGvHD grade with opposite effects: High pre-HSCT abundance of ASV 66 (*Actinomyces* sp., $P = 0.03$) predicted development of aGvHD grade II-IV and 94% of samples with a variance stabilized abundance >6.4 originated from patients with subsequent development of aGvHD grade II-IV (Figure S4C). Congruently, pre-HSCT log transformed relative abundances of ASV 66 were higher in those patients, e.g. the average abundance at pre-examination was 2.3 times higher in grade II-IV versus in grade 0-I patients (Figure S4D). In contrast, high pre-HSCT abundance of ASV 47 (*Rothia* sp., $P = 0.03$) strongly predicted that patients would be spared from aGvHD. All nasal samples with a variance stabilized pre-HSCT abundance >-3.05 of ASV 47 originated from patients who subsequently developed no or mild aGvHD (grade 0-I) (Figure S4C).

Reconstitution of CD4+ T cells and the T_H17 subpopulation is associated with fecal, oral, and nasal microbiota

In order to characterize associations between the microbiota and immune cell counts, immune markers, and clinical outcomes in HSCT, we implemented two multivariate multi-table approaches. Describing host-microbial associations in a multivariate setting gave us insight into which immunological and clinical parameters might impose conditions on the predictions we have made from the microbiota about aGvHD. Furthermore, it allowed us to reveal associations that are either shared by or exclusive to the three body sites we examined - gut, oral cavity, and nasal cavity. For each body site, we first performed a sparse partial least squares regression (sPLS) that explained variation in bacterial alpha-diversity, immune cell counts, and immune markers by bacterial abundances, followed by hierarchical clustering. Subsequently, we implemented canonical correspondence analyses (CCpnA), identifying bidirectional (i.e. canonical) relationships between abundances of ASVs selected in the sPLS analysis and immune cell counts, immune markers, as well as clinical outcomes. At each body site, we identified three clusters of ASVs through hierarchical clustering based on the first three latent dimensions of each sPLS analysis (Figures 6A, S5A and S6A). The CCpnA analyses reinforced the cluster separations and additionally provided insight into associations with categorical variables, including patient baseline parameters, the occurrence of infections, antibiotics treatment, and clinical outcomes (Figures 6B, S5B and S6B). Below, we focus on reporting the associations that we observed at more than one body site.

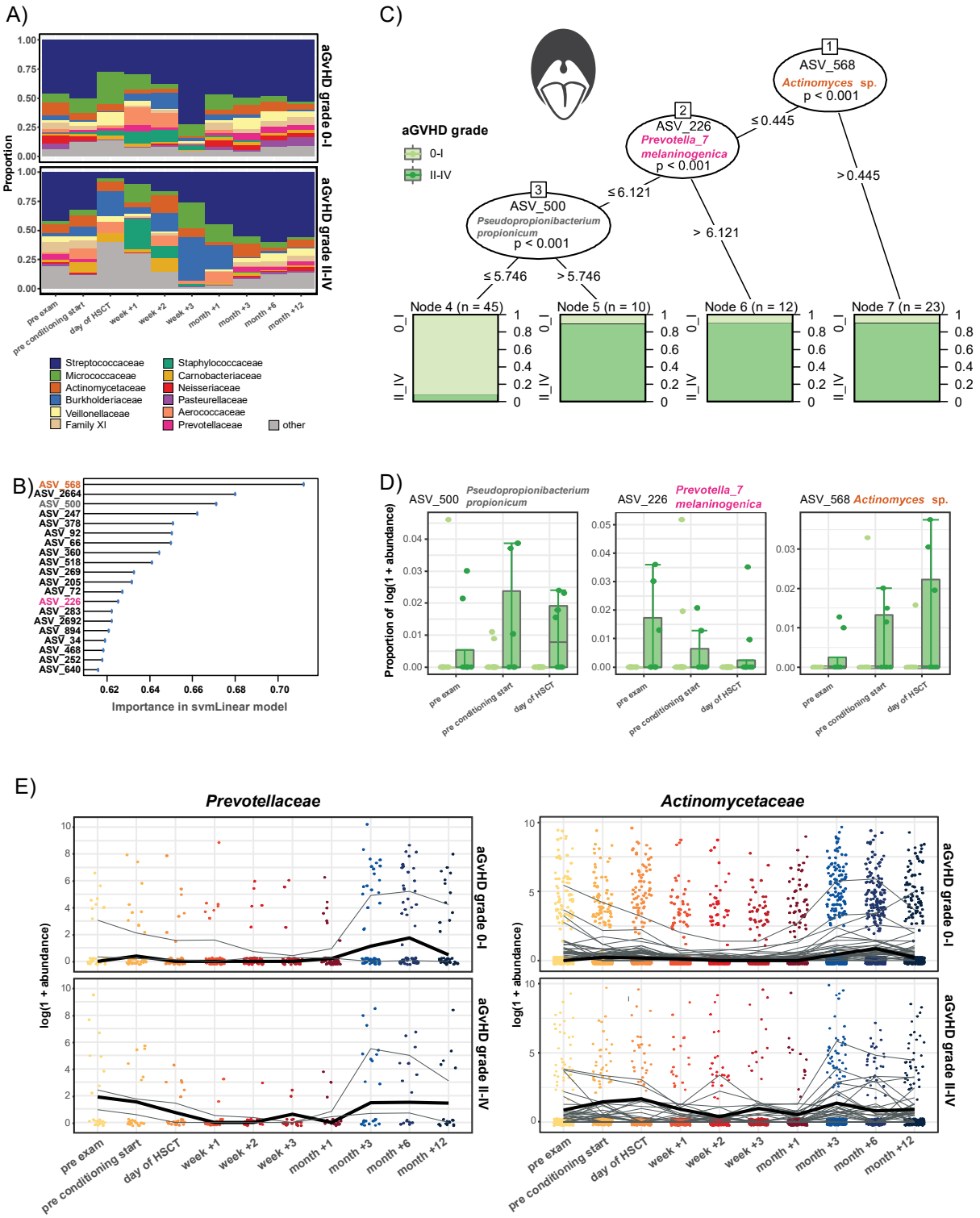


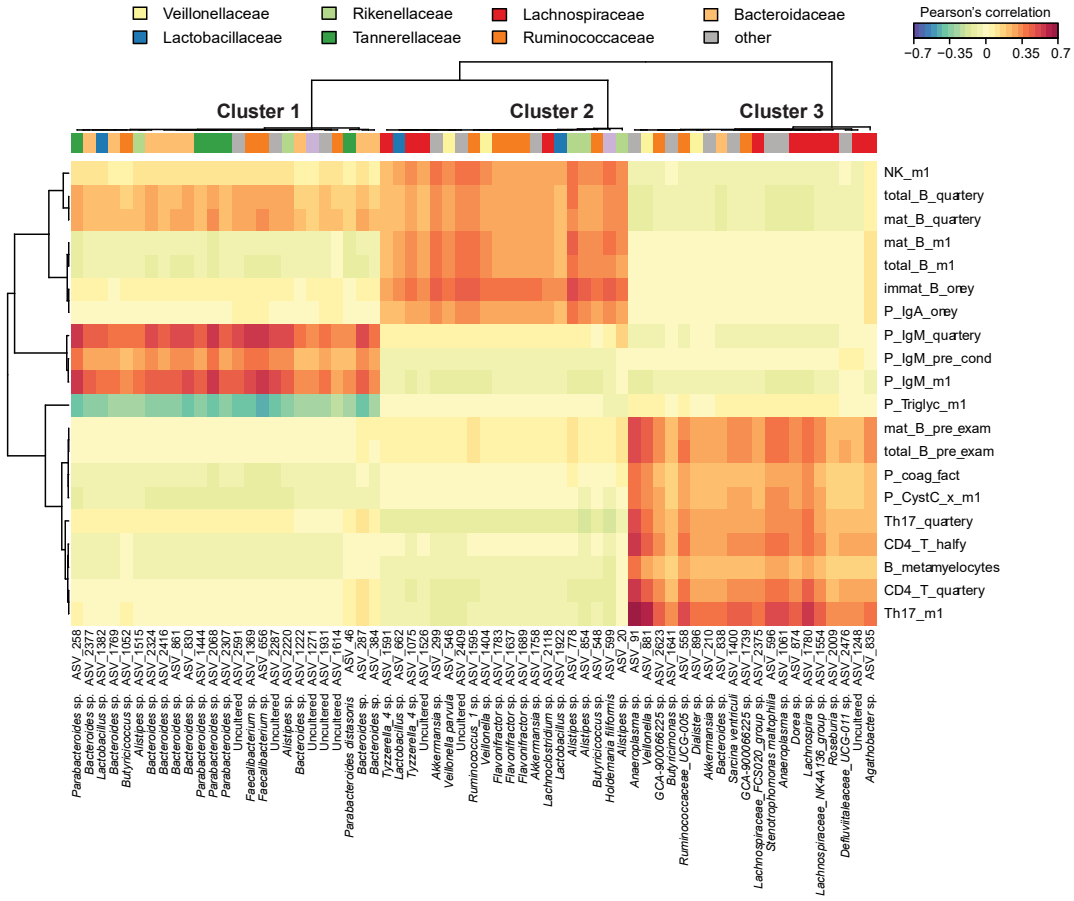
Figure 5. Machine learning-based prediction of aGvHD severity from oral microbial abundances pre-HSCT. A) Relative abundance over time of the 12 most abundant families in the oral cavity in patients with aGvHD grade 0-I versus II-IV. B) Importance plot of top 20 predictive oral ASVs identified by the svmLinear model with importance scores indicating the mean decrease in prediction accuracy in case the respective ASV would be excluded from the

model. C) Conditional inference tree (CTREE) displaying ASVs identified as significant split nodes by nonparametric regression for prediction of aGvHD. Numbers along the branches indicate split values of variance stabilized bacterial abundances. The terminal nodes show the proportion of samples originating from patients with aGvHD grade 0-I vs II-IV. D) Boxplots depict the log transformed relative abundances of the predictive ASVs at time points up to the transplantation in aGvHD grade 0-I compared with grade II-IV patients. E) Trajectories of *Prevotellaceae* and *Actinomycetaceae* ASVs that were identified by tree-based sparse LDA, including those predictive for aGvHD (bold lines), in patients with aGvHD grade 0-I vs II-IV.

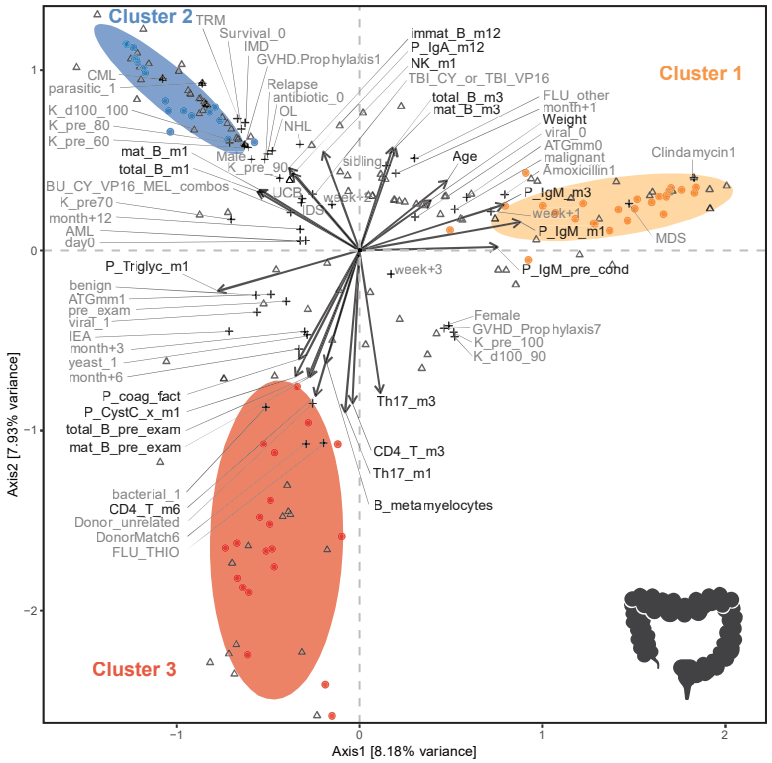
Several cell populations of the adaptive immune response were associated with one cluster each at all 3 body sites according to the sPLS analysis. These included total and mature B cell counts at the pre-examination time point, as well as T cell counts at late follow-up time points, particularly CD4+ T cells in months +3 and +6, and the subpopulation of T_H17 cells in months +1 and +3. In the gut, high numbers of these adaptive immune cell populations were associated with high abundances of *Lachnospiraceae*, *Ruminococcaceae*, and few *Veillonellaceae* ASVs (gut cluster 3, Figure 6A). In the oral cavity, the same cell types were positively correlated with specific *Flavobacteriaceae*, *Prevotellaceae*, *Veillonellaceae*, and *Neisseriaceae* ASVs (oral cluster 1, Figure S5A). Moreover, the same ASVs were associated with younger age and lower C-reactive protein (CRP) levels in the oral cavity. The nasal cluster 2 that was affiliated with high T cells counts and young age comprised predominantly *Veillonellaceae* (Figure S6A).

When taking into account patient baseline parameters, the occurrence of infections, antibiotics treatment, and clinical outcomes in the CCpnA, we observed that samples in gut cluster 3 belonged to patients with benign primary diseases, particularly inherited abnormalities of erythrocyte differentiation or function (IEA), who received fludarabine-based conditioning including thiothepa substituted with anti-thymocyte globulin (ATG), and who had unrelated donors (Figure 6B). Moreover, these patients had a high number of bacterial and viral infections. The concerned samples were primarily collected at pre-examination and in months +3 and +6 (Figure 6B). In the oral cavity, samples associated with CD4+ T cell reconstitution stemmed from similar time points (mainly pre-examination and month +12). Like in the CCpnA of the gut data, the patients in oral cluster 1 mostly had underlying benign diseases such as IEA and immunodeficiency syndromes (IDS), but leukemia also occurred (Figure S5B). Patients in oral cluster 1 received umbilical cord blood (UCB) grafts, were generally treated with few antibiotics, and were associated with relapse. However, relapse was only observed in 2 patients overall during the monitored period. In the CCpnA of the nasal data set, CD4+ T cells and T_H17 cells fell below the displayed threshold of 0.3, suggesting only weak associations with the *Veillonellaceae* – dominated nasal cluster 2 (Figure S6B). Most samples in this cluster were collected in week +2 and were affiliated primarily with non-Hodgkin lymphoma (NHL) - patients.

A)


Veillonellaceae
Lactobacillaceae
Rikenellaceae
Tannerellaceae
Lachnospiraceae
Ruminococcaceae
Bacteroidaceae
other
Pearson's correlation
−0.7 −0.35 0 0.35 0.7
Cluster 1
Cluster 2
Cluster 3
NK_m1
total_B_quartery
mat_B_quartery
mat_B_m1
total_B_m1
immat_B_orey
P_IgA_orey
P_IgM_quartery
P_IgM_pre_cond
P_IgM_m1
P_Triglyc_m1
mat_B_pre_exam
total_B_pre_exam
P_coag_fact
P_CystC_x_m1
Th17_quartery
CD4_T_halfy
B_metamyelocytes
CD4_T_quartery
Th17_m1

B)


Cluster 2
Cluster 1
Cluster 3
TRM
Survival_0
IMD
GVHD.Prophylaxis1
Relapse
antibiotic_0
OL
NHL
CML
parasitic_1
K_d100_100
K_pre_80
K_pre_60
mat_B_m1
total_B_m1
BU_CY_VP16_MEL_combos
K_pre70
month+12
AML
day0
Male
K_pre_90
sibling
JCR
week+2
IDS
immat_B_m12
P_IgA_m12
NK_m1
TBI_CY_or_TBI_VP16
total_B_m3
mat_B_m3
Age
FLU_other
month+1
Weight
viral_0
ATGmm0
malignant
Amoxicillin1
Clindamycin1
P_IgM_m3
week+1
P_IgM_m1
MDS
week+3
P_IgM_pre_cond
P_Triglyc_m1
benign
ATGmm1
pre_exam
viral_1
IEA
month+3
yeast_1
month+6
Female
GVHD_Prophylaxis7
K_pre_100
K_d100_90
Th17_m3
CD4_T_m3
Th17_m1
B_metamyelocytes
P_coag_fact
P_CystC_x_m1
total_B_pre_exam
mat_B_pre_exam
bacterial_1
CD4_T_m6
Donor_unrelated
DonorMatch6
FLU_THIO
Axis2 [7.93% variance]
Axis1 [8.18% variance]

467

Figure 6. Multivariate associations of the gut microbiota with immune and clinical parameters in HSCT. A) Clustered image map (CIM) based on sparse partial least squares (sPLS) regression analysis dimensions 1, 2, and 3, displaying pairwise correlations >0.3/<-0.3 between ASVs (right), and continuous immune and clinical parameters (bottom). Red indicated positive correlation, and blue indicates negative correlation, respectively. Based on the sPLS regression model, hierarchical clustering (clustering method: complete linkage, distance method: Pearson's correlation) was performed resulting in the three depicted clusters. B) Canonical correspondence analysis (CCpnA) relating gut microbial abundances (circles) to continuous (arrows) and categorical (+) immune and clinical parameters. ASVs and variables with at least one correlation >0.3/<-0.3 in the sPLS analysis were included in the CCpnA. The triplot shows variables and ASVs with a score >0.3/<-0.3 on at least one the first three CCpnA axes, displayed on axis 1 versus 2 with samples depicted as triangles. The colored ellipses (depicted with 80% confidence interval) correspond to the clusters of ASVs identified by the sPLS-based hierarchical clustering. Abbreviations not mentioned in text: ATGmm, anti-thymocyte globulin; B_, blood;BU, busulfan; CY, Cyclophosphamide; DonorMatch6, matched unrelated donor; FLU_other, fludarabine combinations without thiotepa; GvHD.Prophylaxis1, treatment with cyclosporine; GvHD.Prophylaxis7, treatment with cyclosporine and methotrexate; immat_B, immature B cells; K_d100, Karnofsky score on day +100; K_pre, Karnofsky score before HSCT; m1, month+1; m3, month+3; m6, month+6; m12, month+12; mat_B, mature B cells; MEL, melphalan; total_B, total B cells; P_, plasma; parasitic, parasitic infection; pre_cond, before conditioning start; pre_exam, pre-examination; THIO, thiotepa; Triglyc, triglyceride; viral, viral infection; VP16, Etoposide; yeast, yeast infection.

Reconstitution of innate immune cell populations is reflected in oral and nasal microbial patterns

In the sPLS analyses, specific bacterial clusters in both the oral and nasal cavity exhibited associations with overall lymphocyte and leukocyte counts, as well as with a set of innate immune cells types: monocytes, neutrophils, basophils, eosinophils, and thrombocytes. The same clusters comprised associations with blood hemoglobin and plasma albumin levels. In the nasal cavity, counts of these cell types were positively correlated with the abundances of particular *Corynebacterium* spp. (*Corynebacteriaceae,* ASVs 14 and 360), *Streptococcaceae*, and *Micrococcaceae* ASVs, including ASV 47 (*Rothia* sp.) that was predictive for aGvHD grade 0-I. The sPLS analysis of the oral set suggested that these immune cells and blood markers were associated with two oral clusters comprising *Flavobacteriaceae*, *Prevotellaceae*, *Veillonellaceae*, and *Neisseriaceae* ASVs in oral cluster 1 and ASVs affiliated with *Streptococcaceae* and *Gemella* spp. (*Bacillales* Family XI) in oral cluster 2. The CCpnA supported the positive correlations of ASVs in oral cluster 1, which was also associated with CD4+ T cell reconstitution, and to a lesser degree with oral cluster 2. Both the oral and nasal CCpnA suggested that patients whose samples were associated with high innate immune cell counts received generally few antibiotics and that the associations primarily occurred at late follow-up time points (Figures S5B and S6B).

Infection-markers after HSCT are associated with the gut and nasal microbiota in distinct patient groups

In both the gut, and the nasal cavity, plasma IgM and triglyceride levels in month +1 were among the variables with the strongest associations with microbial abundances. IgM levels were positively correlated with high abundances of *Tannerellaceae* and *Ruminococcaceae* in the gut (cluster 1), and ASVs within cluster 3 in the nasal cavity, particularly ASV 481 (*Streptococcus peroris*) and ASV 288 (*Neisseria* sp.) (Figures 6A and S6A). In contrast, triglyceride levels were negatively correlated with the same ASVs at both body sites, respectively (Figures 6A and S6A). The CCpnA indicated that samples in gut cluster 1 were

taken predominantly in week +1, and samples in nasal cluster 3 were mainly collected pre-HSCT or on the day of HSCT (Figures 6B and S6B). In the CCpnA, we observed that microbial associations with IgM and triglyceride levels in month +1 in the gut and the nasal cavity did not necessarily occur in samples from the same patients or time points. For instance, the CCpnA on the fecal data set revealed that samples in this cluster originated predominantly from comparatively older patients diagnosed with malignant diseases, such as myelodysplastic or myeloproliferative disorders (MDS), and were treated with Amoxicillin more often than the overall population (Figure 6B). In contrast, the nasal CCpnA indicated that these associations occurred in younger patients with chronic myeloid leukemia (CML), severe aplastic anemia (SAA), or acute myeloid leukemia (AML) who received UCB grafts (Figure S6B). These patients had a comparatively high number of parasitic infections. Moreover, samples from the few patients that underwent a relapse (n=2), and from those who died due to treatment-related (n=2) or relapse-related causes (n=1) were grouped into nasal cluster 3 (Figure S6B).

Body site-specific immune-microbial associations

In addition to immune-microbial associations shared between two or three of the examined body sites, we observed a number of patterns that were exclusive to individual sites. In cluster 2 in the gut, we observed ASVs affiliated with *Ruminococcaceae*, *Lachnospiraceae*, and *Rikenellaceae*, as well as few *Veillonellaceae* and *Lactobacillaceae, and* their abundances showed positive correlations with mature, immature and total B cell counts at several late follow up time points, as well as with NK cell counts in month +1. Furthermore, high IgA levels in month +12 characterized this cluster. According to the CCpnA, samples grouped into gut cluster 2 originated mainly from male patients diagnosed with inherited disorders of metabolism (IMD), NHL, CML, or other leukemias (OL) who underwent total body irradiation (TBI)-based conditioning or non-fludarabine based combinations without TBI and primarily received UCB - derived grafts (Figure 6B). These patients had a comparatively high number of parasitic infections, were treated with comparatively few antibiotics, and received cyclosporine for aGvHD prophylaxis (Figure 6B). Furthermore, this cluster was associated with relapse, and mortality.

In the oral cavity, the sPLS analysis revealed a cluster (oral cluster 3) dominated by *Bacteroidaceae* and *Ruminococcaceae* ASVs that were positively correlated with high bacterial alpha diversity denoted by the inverse Simpson index, and to a lesser degree with CRP levels (Figure S5A). The CCpnA indicated that these associations occurred predominantly from the day of HSCT onwards and throughout the first three weeks post HSCT (Figure S5B). Patients whose samples were grouped into oral cluster 3 were treated with Vancomycin more often than other patients were and had a high number of fungal infections (Figure S5B).

Discussion

When a patient undergoes allogeneic HSCT, both the microbiota and the immune system are subject to major changes. Failure to re-stablish host-microbial homeostasis might have adverse consequences for the patients. Long-term surveillance of microbial dynamics is required to understand not only the shifts in the microbial community structure induced by HSCT and its accompanying treatments, but also when and under which conditions re-establishment of immunological and microbial homeostasis occurs. Such knowledge may be of prognostic value and may assist in guiding personalized treatment strategies. Here, we present a comprehensive assessment of temporal microbial abundance trajectories from before, at the time of, and after HSCT, to late follow-up time points up to one year. We identified AVSs with the potential to predict post-transplant aGvHD, which might open opportunities to improved preventive clinical management. We related our findings of microbial dynamics to patient baseline parameters, immune cell counts, immune markers, and clinical outcomes of HSCT, and did so for the gut microbiota and the so far comparatively less described oral and nasal microbiota. Like the gut microbiota, the oral and nasal commensal residents might be of systemic relevance, and a more holistic picture of microbial influence might be drawn by examining various niches with bacterial communities potentially interacting across body sites.

Here, we have identified a group of *Ruminococcaceae*, and a clade of *Blautia* spp *(Lachnospiraceae)*, temporally discriminating microbial community structure in the gut in relation to HSCT. We show a clear pattern of depletion of fecal *Blautia* spp. immediately post HSCT, as well as their recovery from month +3 post HSCT onwards. One could describe the trajectories of these potentially beneficial taxa as a “smile”-shape. Previous studies have associated the taxonomic families of *Ruminococcaceae* and *Lachnospiraceae* (both class *Clostridia*), and especially the genus *Blautia* (family *Lachnospiraceae*), with lower mortality, lower GvHD, and higher bacterial diversity in adult allo-HSCT recipients [14–18]. In turn, a loss of those taxa after HSCT was associated with subsequent adverse outcomes. Our findings extend the potential of *Blautia* spp. abundances as an indicator of favorable clinical outcomes, as we characterize abundance dynamics in children and provide important insight into the time point for the expected return to abundances comparable to pre-HSCT time points.

Adverse effects, like bacteremia and GvHD, have been found to accompany an expansion of the genus *Enterococcus* post transplantation [2, 3, 6, 19]. We have found a characteristic expansion of this genus, as well as of certain *Lactobacillaceae* after HSCT, in agreement with previous studies [6]. In addition, here we were able to show a decrease of *Enterococcus* spp. and *Lactobacillaceae* from month +3 to abundances comparable to pre-HSCT levels. The abundance of these taxa over the course of one year one could describe as a “frown”-shaped trajectory. As for the “smile”-trajectory of potentially beneficial taxa, the “frown”-trajectories of these taxa could be the first step towards a novel basis to evaluate the reestablishment of patients’ microbial homeostasis and associated convalescence. This could be based on whether e.g. *Enterococcus* spp. abundances have returned to pre-HSCT levels within the expected time. Of note, despite the expansion of *Enterococcus* being in line with previous studies, our multivariate analyses did not reveal direct detrimental host-microbial associations of *Enterococcus* in the present cohort. The particular *Lactobacillaceae* ASVs did not exhibit any correlations with immune cell counts, immune markers, or clinical outcomes either. However, it has been suggested that the expansion of

Lactobacillus, a genus commonly associated with probiotic properties promoting immune homeostasis, might exert a protective effect to limit *Enterococcus* expansion or GvHD risk [14, 19, 20].

We found associations of high *Lachnospiraceae* and *Ruminococcacea* in the gut with rapid B and NK cell reconstitution and high IgA levels after one year in the present study, which is in support of our previous study [14]. These two Clostridia families play an important role in providing the host with short-chain fatty acids (SCFAs), such as butyrate [5, 21]. A study demonstrated that SCFAs can facilitate the differentiation of human naïve B cells to plasma cells in culture and additionally enhance immunoglobulin production, which provides a potential explanation for our finding [22]. Moreover, we found that the patients for which we observed these associations were treated with comparatively few antibiotics at time points early post-HSCT. This is in line with our previous study that revealed high *Lachnospiraceae* and *Ruminococcacea* abundances and rapid B and NK cell reconstitution in the absence of vancomycin and ciprofloxacin treatment [14]. Here, our multivariate analysis did not point to an association with specific antimicrobial agents within the concerned cluster. However, vancomycin and ciprofloxacin were among the most frequently applied agents (at least once during the monitored period in 28 and in 22 patients, respectively).

Furthermore, we have to our knowledge for the first time determined long-term dynamics of the oral and nasal microbiota in allogeneic HSCT patients. Interestingly, we identified abundance trajectories of phylogenetically closely related groups of *Actinomycetaceae*, *Streptococcaceae*, *Prevotellaceae, and* Family XI (*Gemella* spp., Class *Bacillales)* in the oral cavity, resembling the "smile"-shaped trajectories observed in gut. These taxa are part of the normal oral microbiota. Our findings are in agreement with previous studies reporting the detection of fewer *Prevotella* spp. and *Streptococcus* spp. in the oral cavity during the first month post HSCT [23, 24]. In addition, our current study provides insight into the time of recovery of these taxa in month +3 after HSCT. In relation to host-microbial interactions, two *Actinomycetaceae* ASVs, six *Streptococcaceae* ASVs, and one *Prevotellaceae* ASV characterizing these temporal patterns were positively correlated with high innate immune cell counts and with plasma albumin levels throughout the monitored period in our multivariate analyses. Albumin levels give insight into liver function and nutritional status of HSCT patients [24].

For the oral cavity, a post-transplant expansion of *Enterococcus* spp. and *Staphylococcus* spp. has been reported previously [24, 25]. Consistently, we observed an increased relative abundance of *Staphylococcaceae* during the first month post HSCT, but we did not identify *Enterococcus* spp. or *Staphylococcus* spp. as significant drivers of temporal dynamics in the oral cavity. Previously, increased *Enterocoocus* abundances post HSCT were found predominantly in patients who developed oral mucositis, which was not directly assessed in our study [24, 26]. Oral mucositis has been found to prolong the requirement of parenteral nutrition which is frequently accompanied by a reduced nutritional status indicated by low albumin levels [24]. Therefore, our findings suggest that it could be relevant in relation to oral mucositis to further investigate taxa that exhibit "smile"-like abundance trajectories. Especially low abundances of *Actinomycetaceae*, *Streptococcaceae*, and *Prevotellaceae,* might be candidates for microbial predictors of oral mucositis, and furthermore might be employed to facilitate preventive management.

In the nasal cavity, the microbiota did not exhibit temporal patterns as distinct as the "smile"- and "frown"- shaped trajectories in the gut and the oral cavity. However, certain host-microbial associations observed in the gut were reflected in the nasal cavity. For instance, reconstitution of CD4+ T cells and the T_H17 subset were associated with distinct groups of ASVs at all three body sites (Figures 6A and S5A and B). Counts of these adaptive immune cell types were positively correlated with a set of ASVs in the nasal cavity affiliated with *Veillonellaceae*, a family commonly colonizing the oral cavity, but also occurring in the nasal cavity [27]. Host–nasal microbiota relations in HSCT are so far poorly described. However, a high intestinal abundance of *Veillonellaceae* has been identified as a microbial marker of Crohn's disease, a condition characterized by a state of chronic inflammation [28]. Further studies will have to elucidate whether high nasal *Veillonellaceae* abundances can similarly indicate a state of high inflammation in HSCT patients, which the associated high T_H17 cell counts might indicate.

Together, these findings suggest that the oral and the nasal cavity might constitute accessible microbial niches suitable for investigating host-microbial associations in the context of HSCT, similar to the research-based assessments currently employed for the gut. While all three body niches are characterized by mucous membranes that are in close association with distinct microbial communities, it is more feasible to collect buccal and anterior naris swabs during clinical routine as compared to collecting fecal samples. Fecal sample collection is dependent on bowel movements that are often impaired in this patient group. Therefore, our study provides valuable knowledge for possible future applications that could include the determination of distinct oral microbial dynamics in clinical routine, which might be easier to implement than routine fecal sampling.

Associations between acute GvHD severity and the microbiota have to date merely been based on logistic regression and correlation analyses [29–32]. In addition, microbial abundances at the time of neutrophil recovery or engraftment were assessed, i.e. at time points shortly before, concurrent to, or potentially after aGvHD onset [9, 30, 32]. Here, we have implemented machine learning techniques to take the assessment of microbiota-aGvHD relations from correlative to predictive modeling: We presented evidence that aGvHD severity may be predicted from pre-HSCT microbial abundances. This could open up opportunities for the future where microbial markers guide early interventions to prevent aGvHD. Furthermore, we have to our knowledge for the first time revealed microbial taxa in the oral and nasal cavity that may predict aGvHD.

Most prominent among the gut bacteria was a group of five distinct *Parabacteroides* spp. ASVs (family *Tannerellaceae*) that appeared predictive for moderate to severe aGvHD in the machine learning analysis. Our multivariate analyses indicated that a set of other *Parabacteroides* spp. ASVs was associated with high plasma IgM levels. Previous studies indicated that aGvHD is associated with low immunoglobulin levels [33–35]. However, this was suggested to be owed to ATG and steroid treatment and a correspondingly impaired B cell reconstitution and higher rate of opportunistic bacterial infections, which in our study was not associated with high IgM levels [33–36]. Similarly, the association we observed between *Parabacteroides* spp. abundances and low triglyceride levels contradicts previous descriptions of hypertriglyceridemia in aGVHD patients [37, 38]. It should be noted however, that the *Parabacteroides* spp. ASVs in the multivariate analyses were not associated directly with aGvHD severity and comprised a

set of ASVs different from those revealed by the aGvHD prediction model. Interestingly, *Parabacteroides* spp. can have both beneficial and adverse impact on the host: *Parabacteroides* members are providing propionate, thereby stimulating T_{reg} cells, which in turn prevent T_H17-induced inflammation [39, 40]. On the one hand, this might contribute to an environment that favors the re-establishment of microbial and immune system homeostasis after HSCT [31]. On the other hand, it might limit the recovery of the T_H17 subset, crucial for a sufficient graft-versus-leukemia (GvL) effect [39, 41]. The role of the T_H17/Treg ratio in GvHD has been controversially discussed, suggesting for instance a higher ratio in moderate to severe aGvHD, potentially explained by T_H17 cells promoting inflammation and thereby aGvHD [9]. In contrast, several studies suggested that a low T_H17/Treg ratio is associated with more severe aGvHD, explaining increased Treg cell counts as a compensatory reaction towards GvHD-related inflammation [8, 10]. The latter is in line with our finding and provides a potential explanation for our observation of a distinct set of *Parabacteroides* spp. predicting moderate to severe aGvHD and generally increased *Tannerellaceae* abundances in those patients. The likely ambiguous role of *Parabacteroides* spp. in HSCT and aGvHD underlines the importance of assessing this group with high taxonomic resolution, as performed here by choosing the use of distinct amplicon sequence variants (ASVs) over operational taxonomic units (OTUs). However, further characterization of *Parabacteroides* spp. at the strain level might be required to draw even clearer conclusions.

A similar ambiguity might have to be attributed to the family of *Lachnospiraceae*. We have found one *Lachnospiraceae* ASV in the gut that when highly abundant predicts subsequent moderate to severe aGVHD. This family has previously been reported in connection with reduced GvHD [9, 30]. However, here we examined pre-HSCT abundances to make predictions, i.e. we did not aim at elucidating *Lachnospiraceae* concurrent to aGvHD. This bacterial family is a crucial producer of the SCFA butyrate, which has been suggested to prevent aGvHD [21]. The mechanism by which butyrate counteracts GvHD might be direct by improving epithelial integrity in GvHD-target tissue, and/or indirect by inducing T reg cells that reduce inflammation [21, 40]. However, as mentioned above, the role of T reg cells might also be compensatory during inflammation and therefore associated with more severe aGvHD [8]. We actually observed high CD4+ T cell and T_H17 cell counts at late follow-up time points associated with the abundances of a group of *Lachnospiraceae* at pre-examination and after month +3 in our multivariate analyses, but no direct association of these ASVs with Treg cells or aGvHD. Moreover, these ASVs did not comprise the ASV predictive for moderate to severe aGvHD. Therefore it could be of interest for future studies to reveal the exact mechanisms by which certain *Lachnospiraceae*, and specifically the predictive ASV we identified, might promote aGvHD, or whether their increase might be compensatory.

In addition to ASVs in the gut, we also identified oral and nasal cavity ASVs with pre-transplant abundances predicting subsequent aGvHD. For instance, in the oral cavity, an ASV affiliated with *Prevotella melaninogenica* strongly predicted aGvHD. This species is known to cause oral mucosal infections, and to be elevated in abundance in oral cancer [26, 42]. One prominent oral complication in allogeneic HSCT is oral mucositis, but has to our knowledge not been investigated whether *P. melaninogenica* is involved in oral mucositis. However, *Prevotella intermedia* has previously been detected in low numbers in oral mucosal lesions [43]. Oral mucositis has been linked to elevated aGvHD risk, with the potential explanation that chemotherapy-induced tissue injury of the oral cavity might reflect a damaged gut epithelium, more prone to aGvHD [44, 45]. This is in line with our findings that high pre-HSCT abundances of *P.*

718 *melaninogenica* in the oral cavity predicted subsequent moderate to severe aGvHD. However, the cluster
719 of oral ASVs comprising a number of *Prevotellaceae*, including another *P. melaninogenica* ASV, did not
720 exhibit any strong associations with the conditioning regimen, again highlighting the need for direct
721 evaluation of oral mucositis in this context to strengthen our conclusions.
722
723 Other microbial predictors of aGvHD in the oral and nasal cavity included members of the order
724 *Actinomycetales* (*Actinomyces* spp., *Pseudopropionibacterium propionicum*, *Rothia* sp.). The order
725 *Actinomycetales* comprises numerous commensals colonizing the oral and nasal cavity in healthy
726 individuals, but can cause opportunistic infections e.g. in allo-HSCT patients [46]. It has been
727 demonstrated that aGvHD increased the susceptibility to infections in allogeneic HSCT patients, but a
728 predisposition to aGvHD due to a preceeding infection has also been suggested for *Clostridium difficile*
729 [47]. Whether actinomycosis might play a role in facilitating aGvHD development remains to be
730 investigated. Interestingly, an increase of oral Actinobacteria in the gut at the time of neutrophil recovery
731 has been found to be correlated with subsequent severe aGvHD [30]. One potential explanation involves
732 the ability of *Actinomyces* spp. to drive biofilm formations [48]. This is commonly observed in
733 periodontitis, an inflammatory condition of the tissue around the teeth, and has been suggested to also
734 occur when oral biofilm colonizers shift to the gut, e.g. during colorectal cancer [48]. *Actinomycelates*
735 have also been found to play a role in epithelial DNA damage during the initiation of colorectal cancer
736 [49]. High *Actinomyces* abundances in the oral and nasal cavities might induce similar epithelial damage
737 at those sites, potentially promoting e.g. oral manifestations of GvHD. However, we cannot draw
738 conclusions concerning this within the scope of the present study.
739
740 Interestingly, among all ASVs predicting aGvHD at any body site, we only found one nasal *Rothia* sp. ASV
741 with high pre-HSCT abundances predicting that patients would be spared from subsequent aGvHD. *Rothia*
742 *mucilaginosa* was among the genera that when present in the gut post-transplant correlated with
743 moderate to severe aGvHD in a previous study [30] However, it has to date not been assessed how
744 different *Rothia* spp. in the nasal cavity might be associated with aGvHD. This discrepancy again
745 emphasizes the importance of high taxonomic resolution in this context, and the examination of the
746 microbiota at different body sites.
747
748 In summary, with the present study we bring forward a comprehensive framework of host-microbial
749 associations in allogeneic HSCT. We focused on long-term microbial dynamics, demonstrating distinct
750 microbial abundance patterns of disturbance and recovery, as well as making predictions about aGvHD
751 from the pre-transplant microbiota. We integrated microbial abundance patterns with a number of
752 patient baseline parameters, immune cell counts, immune markers, and clinical outcomes in multiple
753 multivariate analyses. Of note, we have included relapse and mortality as clinical outcomes into our
754 multivariate analyses in addition to aGvHD. However, we currently cannot draw meaningful conclusions
755 from associations with these clinical outcomes as only a few patients relapsed or died. These observed
756 associations should be reevaluated in the same patient cohort after a longer follow-up period or in
757 another larger cohort with a more balanced design. We have discovered host-microbial associations
758 shared between two or more of the examined body sites, opening opportunities for implementing more
759 feasible oral and nasal swab sampling into research and clinical diagnostic activities. It should be noted

that shared associations were to a certain extent dependent on patient characteristics, such as primary disease. Therefore, these parameters should always be accounted for when examining HSCT-related host-microbial interactions.

Conclusions

We investigated the role of the microbiota in allogeneic HSCT and confirmed the depletion of *Blautia* spp. and expansion of *Enterococcus* spp. in the gut after HSCT. We expand this knowledge by precisely defining which phylogenetically closely related sequence variants of these genera are characteristic for those patterns, and when they return to pre-HSCT levels. We identified similar patterns for members of the oral and nasal microbiota and propose month+3 post-transplant as a possible universally crucial time point for microbiota reconstitution in HSCT. We demonstrate that high abundances of e.g. an intestinal *P. distasonis* ASV, an oral *P. melaninogenica* ASV predict the development of moderate to severe aGvHD post-transplant. When relating microbial abundances with immune cell counts, we found rapid B and NK cell reconstitution and high IgA levels to be associated with high abundances of *Lachnospiraceae* and *Ruminococcacea*, especially in patients treated with few antibiotics. Distinct ASVs at all three body sites were associated with T_H17 cell counts, suggesting future research on a potential immunomodulatory involvement of the microbiota in inflammation regulation, which might play a role for GvHD development.

Materials and Methods

Patient recruitment and sample collection

We recruited 29 children (age range: 2.5 - 16.4 years) who underwent their first myeloablative allogeneic hematopoietic stem cell transplantation at Copenhagen University Hospital Rigshospitalet (Denmark) between November 2015 and October 2017. We provide detailed information about the patients' clinical characteristics in Table S1. Every patient underwent a myeloablative conditioning regimen starting on day -10 for patients receiving a graft from a haploidentical donor, and on day -7 for patients with sibling or matched unrelated donors (Table S1). One patient was retransplanted on day +223 after the first transplantation. Data from day -7 prior to retransplantation for this patient was excluded from our analysis. The following sampling time points were defined: pre-examination (collected between day -57 and day -15), around the starting date of conditioning (collected between day -14 and day -3 and latest 2 days after conditioning start), at the time of HSCT, preferably before graft infusion (collected between day -2 and day +2), and weekly during the first 3 weeks after transplantation (week +1: day +3 to day +10, week +2: day +11 to day +17, week +3: day +18 to day +24) (Figure 1A). Broader intervals applied to follow-up time points: Month +1 (between days +25 and +45), month +3 (between days +46 and +120), month +6 (between days +121 and +245), and month +12 (between day +246 and +428).

Infections and antibiotics

Records of bacterial, fungal, viral, and parasitic infections and antibiotic treatment from before HSCT (from day -30 or at the collection time of the first microbiota sample in case this was earlier) until month +12 (day +428) were taken into consideration (or as long as data was available for the most recent patients; data accessed in July 2018). This corresponds to the length of the sampling period of fecal and swab samples.

Analysis of immune cell subpopulations

Cell counts of leukocytes, thrombocytes, and erythrocytes were recorded daily during hospitalization starting prior to HSCT, and later weekly in the outpatient clinic by flow cytometry (Sysmex XN) or microscopy (CellaVision DM96 microscope) in case of very low counts. Monitored subpopulations included: Monocytes, neutrophils, basophils, eosinophils, blast cells, plasma cells, erythroblasts, erythrocytes, reticulocytes, myelocytes, metamyelocytes, and promyelocytes.

Analysis of T, B and NK cells in peripheral blood

Total lymphocyte counts, and T, B, and NK cell counts in $x10^9$/L were determined at pre-examination, and in month +1, +3, +6, and +12. Trucount Tubes (Becton Dickinson, Albertslund, Denmark) were used to quantify these cell types in peripheral blood on a FC500 flow cytometer (Beckman Coulter, Copenhagen, Denmark). For immunofluorescence staining, the following conjugated monoclonal antibodies were used for CD3+ T cells, CD3+CD4+ T cells and CD3+CD8+ T cell quantification: CD3-PerCP, CD3-FITC, CD4-FITC, CD8-PE (Becton Dickinson). CD45-PerCP, CD16/56-PE antibodies were used to determine NK cells based on their CD45+CD16+CD56+ phenotype. For B cells, total B cells (CD45+CD19+), mature B cells (CD45+CD19+CD20+) and immature B cells (CD45+CD19+CD20-) were differentiated by using CD20-FITC and CD19-PE antibodies.

Subtyping of T cells

We collected peripheral blood samples in month +1, +3 and +6. Subsequently, we isolated peripheral blood mononuclear cells (PBMCs) by gradient centrifugation of heparinized blood with Lymphoprep™ (Axis-Shield, Oslo, Norway) that was washed in PBS (Life Technologies, Invitrogen, Paisley, U.K.) three times and then resuspended in RPMI 1640 buffer containing HEPES (Biological Industries Israel Beit-Haemek Ltd, Kibbutz Beit-Haemek, Israel), L-glutamine (GIBCO, Invitrogen, Carlsbad, CA) and Gentamycin (GIBCO), 30% fetal bovine serum (Biological Industries) and 10% Dimethyl Sulfoxide (VWR, Herlev, Denmark). For cryo-preservation, PBMCs were then aliquoted and stored in liquid nitrogen. T cell subsets, i.e. T_H17 cells and T_{reg} cells, were quantified from frozen PBMCs by flow cytometry on a FACS Fortessa III flow cytometer (Becton Dickinson, Albertslund, Denmark). PBMCs were thawed and washed before incubation with Fixable viability stain 620 (Becton Dickinson) and a set of conjugated monoclonal antibodies for 30 minutes on ice: CD3-APC-A750 (Beckmann Coulter), CD4-PE-Cy7 (Beckmann Coulter), CD8-A700 (Becton Dickinson), CD25-PE (Becton Dickinson), CD39-PerCP-Cy5.5 (Beckmann Coulter), CD196-BV510 (Biolegend, San Diego, USA), CD127-BV711 (Biolegend), CD161-BV650 (Becton Dickinson) and CD45RA-BV786 (Becton Dickinson). Next, PBMCs were washed and incubated with transcription factor buffer set (BD) for 45 minutes on ice. Afterwards, PBMCs were washed twice and intracellular monoclonal antibodies were added and incubated for 45 minutes on ice: RORγT-A488 (Becton Dickinson), FOXp3-A647 (Becton Dickinson) and Helios-PB (Beckmann Coulter). TH17 cells were determined by the CD4+ RORγT+ phenotype, and T reg cells by the $CD4+CD25^{high}FOXp3+$ phenotype. Absolute cell counts in x109/L were obtained by multiplying the frequency of TH17 and T reg cells with the CD4+ T cell count from the same time point.

Quantification of immune markers

Immune markers were measured at the Department of Clinical Biochemistry, Copenhagen University Hospital Rigshospitalet, Denmark. Plasma levels of immunoglobulins A, G, and M in g/L were measured by immunoturbidimetry. As a marker of infection, plasma procalcitonin in µg/L was determined by sandwich electrochemiluminescence immunoassays (ECLIA). As a marker of inflammation, we determined CRP in mg/L by latex immunoturbidimetric assays (LIA).

Routine clinical assessment tests

The following kidney function markers were assessed in plasma: Creatinine in µmol/L was measured using an amperometric biosensor; Cystatin C in mg/L was determined by immunoturbidimetry. The following liver function markers were assessed in plasma: alanine transaminase (ALT), aspartate transaminase (ASAT), and basic phosphatase in U/L were measured by absorption spectroscopy. Bilirubin levels in µmol/L were determined by absorption spectroscopy. Furthermore, we measured the following markers in plasma (units and detection methods denoted in brackets): calcium (mmol/L, absorption spectroscopy), ion-free calcium (mmol/L, absorption spectroscopy), cortisole (nmol/L, ECLIA), hemoglobin (µmol/L, photometry), potassium (mmol/L, ion selective electrodes measurements), lactate dehydrogenase (U/L, absorbtion photometry), magnesium (mmol/L, absorption photometry), sodium (mmol/L, ion selective electrodes measurements), cholesterol (mmol/L, absorption photometry), glucose (mmol/L, absorption photometry), ferritin (µg/L, sandwich ECLIA), phosphate (mmol/L, absorption spectroscopy), carbamide (mmol/L, absorption spectroscopy), albumin (g/L, absorption spectroscopy), triglyceride (mmol/L, absorption spectroscopy), haptoglobin (g/L, immunoturbidimetry), and the fraction of coagulation factors II, VII, and X.

DNA isolation from fecal, oral, and nasal samples and 16S rRNA gene sequencing

212 fecal samples for analysis of the intestinal microbiome were collected from 29 patients at the 10 time points described above. The gut microbiota was characterized at ≤6 time points in 9 patients (31%), at 7-8 time points in 13 patients (45%) and at 9-10 time points in 7 patients (24%) (Table S1). DNA from fecal samples, one blank control per extraction round (thereof sequenced: 14), one mock community sample (Biodefense and Emerging Infectious Research (BEI) Resources of the American Type Culture Collection (ATCC) (Manassas, VA, USA), Catalog No. HM-276D) per sequencing run and two collection tube controls was isolated using the QIAamp Fast DNA Stool Mini kit (Qiagen, Venlo, Netherlands), following the manufacturer's instructions with modifications according to [50].

We collected 248 buccal swabs (3x at ≤6 time points (10%), 11x at 7-8 time points (38%), 15x at 9-10 time points (52%)) and 249 anterior naris swabs (3x at ≤6 time points (10%), 9x at 7-8 time points (31%), 17x at 9-10 time points (59%)). DNA from swab samples, one blank control per extraction round (therof sequenced: 28), one mock community sample per run, two collection tube controls, and two sampling swab controls was isolated using the QIAamp UCP Pathogen Mini kit (Qiagen, Venlo, Netherlands), with the 'Protocol: Pretreatment of Microbial DNA from Eye, Nasal, Pharyngeal, or other Swabs (Protocol without Pre-lysis)' and subsequently the 'Protocol: Sample Prep (Spin Protocol)', following the manufacturer's instructions with the following modifications: 550µl instead of 500µl Buffer ATL was used during pretreatment; DNA was eluted twice with 20µl Buffer AVE into 1.5 ml DNA LoBind tubes (Eppendorf, Hamburg, Germany) instead of the tubes provided with the kits.

Library construction and sequencing on an Illumina MiSeq instrument (Illumina Inc., San Diego, CA, USA) was performed at the Multi Assay Core facility (DMAC), Technical University of Denmark. DNA concentration of each sample was measured using a NanoDrop spectrophotometer (Thermo Scientific, Waltham, MA, USA). Library construction was performed according to the *16S Metagenomic Sequencing Library Preparation* protocol by Illumuna [51]: The V3-V4 region of the 16S ribosomal RNA gene were amplified in a PCR in each sample and in the controls, using the following previously evaluated primers [52]: 341F (5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG-3') and 805R (5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC-3'). Amplicons were then analyzed for quantity and quality in an Agilent 2100 Bioanalyzer with the use of an Agilent RNA 1000 Nano Kit (Agilent Technology, Santa Clara, CA, USA). Subsequently, the amplicons were purified on AMPure XP Beads (Beckman Culter, Copenhagen, Denmark) according to the manufacturer's instructions. Illumina adapters and dual-index barcodes were then added to the amplicon target in a PCR according to Illumina [51] using the 96 sample Nextera XT Index Kit (Illumina, FC-131–1002). A final clean-up of the libraries was performed in another PCR step, using AMPure XP Beads (Beckman Culter, Copenhagen, Denmark) according to the manufacturer's instructions, followed by a confirmation of the target size in an Agilent 2100 Bioanalyzer (Agilent Technologies). Before sequencing, DNA concentration was determined with a Qubit (Life Technologies, Carlsbad, CA, USA) and libraries were pooled. In preparation for sequencing, the pooled libraries were denatured with NaOH, diluted with hybridization buffer, and heat denatured. 5% PhiX was included as an internal control for low-diversity libraries. Paired-end sequencing with 2 × 300bp reads was performed with a MiSeq v3 reagent kit on an Illumina MiSeq instrument (Illumina Inc., San Diego, CA, USA).

16S rRNA gene sequence pre-processing

Raw sequence reads were demultiplexed based on sample-specific barcodes and 'read 1' and 'read 2' FASTQ files for each sample were generated on the Illumina MiSeq instrument by the MiSeq reporter software. Primers were removed by using cutadapt (version 1.16) [53] at a tolerated maximum error rate of 15% for matching the primer sequence anchored in the beginning of each read. In the case that at least one read of a pair did not contain the primer, the pair was discarded. Only pairs in which the forward read contained the forward primer (341F) and the reverse read contained the reverse primer (805R) were retained.

The resulting reads were further processed using the R package DADA2 (version 1.8) to infer high-resolution amplicon sequence variants (ASV) [54]. Forward and reverse reads were truncated at 280 bp and 200 bp respectively. This way, the majority of reads retained a quality score >25 according to MultiQC analysis [55]. These truncation thresholds also ensured an overlap of 480 bp (expected amplicon length of 460 bp + 20 bp), allowing to merge forward and reverse reads. Samples were pooled for the sample inference step (*dada*() function) to increase the power for detecting rare variants. Default values were used for all other quality filtration parameters in DADA2. DNA from samples with a read count <10,000 after preliminary chimera and contaminant removal were re-sequenced. DNA from feces samples with a read count <5,000 were re-extracted. Eventually, chimeras were identified by sample and removed from the whole data set (over all sequencing runs) based on a consensus decision (*removeBimeraDenovo*() function, method "consensus"). Taxonomic assignment on ASVs was done by using the Silva reference data base (version 132), formatted for DADA2 [56]. Additional species assignment by exact reference

strain matching was performed using the Silva species-assignment training data base, formatted for DADA2 [56].

The resulting ASV and taxonomy tables were integrated with the R package phyloseq and its dependencies (version 1.24.0) [57]. The data was split into two data sets, one containing feces sample data and one containing nasal and oral swab data. Subsequently, contaminant removal was performed with the R package decontam [58]. Potential technical batch effects by sequencing run, 96-well plate, extraction kit, extraction round, experimenter, and extraction date were assessed by ordination (Principal Coordinates Analysis (PCoA)).

For both, the fecal sample data set and the swab data set, contaminants were identified by sequencing run as a batch effect and a subsequent calculation of a consensus probability. For the feces sample data set, contaminants were identified by both, increased prevalence in 14 blank extraction controls and by relating ASV frequency to post-PCR sample DNA concentration, assuming inverse correlation (method "both", frequency threshold: 0.2, prevalence threshold: 0.075) [58]. After manual evaluation of edge cases, 89 ASVs were removed from the fecal sample data set as contaminants. In an additional step, we identified 7 contaminants from 2 sampling tube controls (method and thresholds as stated above). In total, 96 ASVs were removed as contaminants from the fecal sample data set.

For the swab sample data set, contaminants were identified by both, increased prevalence in 28 blank extraction controls and by relating ASV frequency to post-PCR sample DNA concentration (method "both", frequency threshold: 0.1, prevalence threshold: 0.6) [58]. A more stringent threshold for prevalence compared to frequency was chosen here, given the low biomass of the swab samples, accompanied by post-PCR DNA concentrations similar to those in blank controls. After manual evaluation of edge cases, 1137 ASVs were removed from the swab sample data set as contaminants. In an additional step, we identified 16 contaminants from 2 sampling tube controls and 2 swab controls (method "both", frequency threshold: 0.075, prevalence threshold: 0.5). In total, 1153 ASVs were removed as contaminants from the oral and nasal swab sample data set.

For each subset, we created a phylogenetic tree by de novo alignment of the inferred ASVs, following a previously described workflow [59]. First, we performed multiple alignment with the package DECIPHER [60]. Subsequently, we built a neighbor-joining tree using the package phangorn [61], based on which we fitted a GTR+G+I (Generalized time-reversible with Gamma rate variation) maximum likelihood tree. The phylogenetic tree for each data set (fecal, oral, and nasal) was then integrated with the respective phyloseq object.

Next, we took core subsets of the ASVs remaining after contaminant removal using the function *kOverA*() from R package genefilter [62]. In the fecal set, 2465 ASVs with ≥5 reads in ≥2 samples were retained. With ≥5 reads in ≥10 samples, 509 ASVs were retained from the oral sample set, and 602 ASVs from the nasal sample set. Additional manual contaminant filtering was applied to the oral and nasal core sets. ASVs affiliated with taxonomic families commonly found in both the oral or nasal cavity and the gut were only retained in the oral sample set in case they had ≥10 reads in ≥10 samples. ASVs of families only expected in the gut were removed from the oral and nasal sample sets after manually assessing their abundances. Subsequently, we retained 377 ASVs in the oral sample set, and 197 ASVs in the nasal sample set.

Statistical analysis

Statistical analyses and generation of graphs was performed in R (version 3.5.1, R Foundation for Statistical Computing, Vienna, Austria) [63]. Sequencing data and all related experimental and clinical data were integrated for analysis with the R package phyloseq [57]. Plots were generated with the packages ggplot2 [64], mixOmics [65], treeDA [66], caret [67], and partykit [68, 69].

From the core sets of ASV counts for each body site, bacterial alpha diversity (denoted by the inverse Simpson index) was calculated and compared between time points by using a Friedman test with Benjamini-Hochberg correction for multiple testing, and a post-hoc Conover test. To gain insight into changes of microbial abundances over time in relation to HSCT, we agglomerated ASV counts on family levels with the function *tax_glom()* in phyloseq [57]. Thereafter, we displayed the relative abundances of the 12 most abundant families at each body site for each time point. We also depicted relative abundances over time on family level in patients with aGvHD grade 0-I versus grade II-IV.

In order to determine which particular ASVs are relevant in temporal microbial abundance dynamics at each body site, we implemented tree-based sparse linear discriminant analysis (LDA) with the package treeDA [66]. This method utilizes phylogenetic information to perform supervised discrimination of classes, here time points [70]. Leaves and nodes of the phylogenetic tree, representing log+1 transformed ASV abundances and the sums thereof respectively, were used as predictive features. The core oral and nasal sets were used as input as described above, while the fecal set was further reduced to 389 ASVs with >5 reads in >10 samples for this analysis. Leave-one-out cross validation (LOOCV) was performed to choose the optimal minimum number of predictive features ensuring sparse, interpretable models. The resulting LDA models had 9 components. By default, this number corresponds to the number of predicted classes (here 10 time points) less one. To identify relevant components, we plotted sample scores colored by time points along each component and plotted the components pairwise against each other (Figure S2). Thereby, we revealed that the first LDA-component for each body site showed the highest sample scores and best separated the samples by time point. Therefore, we proceeded with displaying temporal trajectories of clades of predictive features (ASVs) on the first component. For selected groups of predictive ASVs we displayed trajectories for patients with aGvHD grade 0-I versus with grade II-IV.

Next, we implemented machine learning models to predict aGvHD grade post-transplant from preceeding ASV abundances. The strategy and R code for the machine learning approach was partially adapted from a previous approach [71, 72]. As a preparative step for this analysis, we variance-stabilized the ASV count data. To do so, we first performed size factor estimation for zero-inflated data on the core data sets for each bbody site with the package GMPR [73]. Subsequently, we transformed the data by using the function *varianceStabilizingTransformation*() in the package DESeq2 [74]. The function implements a Gamma-Poisson mixture model to account for both library size differences and biological variability [75].

For the prediction of aGvHD grade, we compared the performances of four different classifiers (random forest (rf), boosted logistic regression (LogitBoost), support vector machines with linear kernel (svmLinear), and support vector machines with radial basis function kernel (svmRadial)) using the package caret [67]. We took subsets of the phyloseq objects comprising only the time points preceding aGvHD onset: pre-examination, conditioning start, and at the time of HSCT. Prior to fitting the models, we excluded ASVs with near zero variance, i.e. those that were not differentially abundant between any samples, by using the function *nearZeroVar*() in package caret [67]. Thereby we obtained sets of 238, 186, and 100 ASVs for the fecal, oral, and nasal data set, respectively, which were then assessed as potential

predictors of subsequent aGvHD. All classifiers were trained on a randomly chosen subset of 70% of the data to build a predictive model evaluated on a test set (30% of the data). Splitting was performed in a way that samples from the same patient at different time points were kept together in either the testing or training set to ensure that the outcome of a patient can only appear in either the testing set or the training set, but not both. Thirty iterations of 10-fold cross-validation were performed for each classifier, both with and without up-sampling. Up-sampling refers to the process of replacement-based sampling of the class with fewer samples (here aGvHD grade II-IV) to the same size as the class with more samples (here aGvHD grade 0-I) to achieve a balanced design. SvmLinear on up-sampled data was chosen as the best performing predictive model for all three data sets (gut, oral, and nasal). Subsequently, we performed Boruta feature selection using the package Boruta [76]. The Boruta algorithm is a Random Forest classification based wrapper that compares the importance of real features to that of so called 'shadow attributes' with randomly shuffled values. Features that are less important than the 'shadow attributes' are iteratively removed. Here, we retained those ASVs in each data set that were both, among the 50 most important predictors in the svmLinear model and confirmed by the Boruta algorithm. Subsequently, we fitted a CTREE on each set of selected predictors (23 fecal, 26 oral, and 12 nasal ASVs) by using the package partykit [68, 69]. In the CTREE analysis the effect of the predictive ASVs on aGvHD grade is evaluated in a nonparametric regression framework. CTREE iteratively tests if the abundance of any ASV has a significant effect on aGvHD grade. In the case that a significant relation is found, the ASVs with the largest effect is picked as a node for the tree. The procedure is then recursively repeated until no further significant effect of any ASV on aGvHD is found. We plotted the result as a tree featuring the significant split nodes, represented by the ASVs and the Bonferroni-corrected p-values indication significant influence of their abundance on aGvHD grade. The terminal nodes of the tree show the proportion of samples stemming from patients with aGvHD grade 0-I versus II-IV, under the condition of the abundance split criterion described on each branch. Since we used-variance stabilized bacterial abundances as input for the machin learning analyses, abundances can be presented as negative values in some cases and are therefore not easy to interpret intuitively. Therefore, we additionally displayed the log-transformed relative abundances of all ASVs significantly predicting aGvHD in boxplots at the three investigated time points (pre-examination, conditioning start, and at the time of HSCT).

Subsequently, we were interested in associations between the fecal, oral, and nasal microbiota and immune cell counts, and clinical outcomes in HSCT. Records of immune markers, and immune cell counts contained left- and right-censored measurements, i.e. observations below or above the detection (or recording) limit, respectively. In order to use these data in analyses that do not tolerate censored records, we needed to impute the censored data. Therefore, we first fitted the non-parametric maximum likelihood estimator (NPMLE, also called Turnbull estimator) for univariate interval censored data on each variable that contained censored records, using the function *ic_np*() in the R package icenReg [77]. Subsequently, censored records were imputed, informed by the model that was fitted on the entity of observed and censored data of each variable, using the *imputeCens()* function [77]. Next, we took the median of measurements for the time points defined above for those immune markers, and immune cell counts that have been measured more frequently than that. This way, we obtained comparable data sets. Continuous immune marker and cell count data that was systematically missing for certain sampling time points was split by time points and unavailable time points were excluded. Missing values in continuous immune marker and cell count data were imputed for variables with ≤ 50% missingness. Simultaneous

multivariate non-parametric imputation was performed using the R package missForest [78]. Variables with more than 50% missing values were excluded from the analysis.

To explore associations between patients' baseline parameters, immune cell counts, immune markers, and clinical outcomes with the overall microbial community structure, we implemented permutational multivariate analysis of variance using distance matrices (adonis) in the package vegan [79] based on Manhattan distance matrices of the variance-stabilized bacterial community data for each body site. Permutation design was set up with respect to repeated measurements within the same patients and the intact chronological order of time points. Thereby, we estimated the number of relevant variables to be included in subsequent multivariate analyses.

Next, we implemented two multivariate multi-table approaches to gain a more detailed understanding of how the fecal, oral, and nasal microbiota might be associated with immune cell counts, immune markers, and clinical outcomes in HSCT. Evaluated clinical outcomes comprised acute GvHD (grade 0-I versus II-IV), relapse, overall survival, and treatment-related mortality. Furthermore, we included bacterial alpha diversity (inversed Simpson index), antibiotic treatment, infections, Karnofsky scores before conditioning and at day +100, and patients' baseline parameters (age, weight, sex, primary disease, malignant versus benign primary disease, conditioning regimen, chemotherapeutic agents' dosages, TBI treatment and dosage, stem cell source, cell dose per kg, engraftment, GvHD prophylactic regimen, recipients' and donors' cytomegalovirus antibody status, donor type (sibling/matched unrelated/haploidentical), donor HLA-match, and donor sex).

For each body site, we performed sparse partial least squares (sPLS) regression by using the function *spls*() in the package mixOmics [65]. In sPLS regression, two matrices are being integrated and both their structures are being modelled. Here, we used variance stabilized ASV abundances as explanatory variables and all continuous clinical and immune parameters as response variables. The method allows multiple response variables. Collinear, and noisy data can be handled by this method as well [80]. We specified the number of response variables to be kept for each component (keepY) according to the number of trending and significant variables revealed by the adonis analysis. The number of explanatory variables (ASVs) to be kept on each component (keepX) was set to 25 after running the sPLS regression models for each body site with a range of values between 20 and 40 for keepX, showing results robust to keepX. The *perf()* function was used to inform the choice of 3 relevant components. Based on the sPLS regression models for each body site, we then performed hierarchical clustering with the *cim()* function, using the clustering method "complete linkage" and the distance method "Pearson's correlation". Thereby, we generated matrices of coefficients indicating correlations between ASV abundances and continuous clinical and immune parameters.

Subsequently, we carried out canonical (i.e. bidirectional) correspondence analysis (CCpnA), which is a multivariate constrained ordination method. This method allow us to assess associations of both categorical and continuous clinical and immune parameters to ASV abundances. We included ASVs and variables with a correlation of >0.2/<-0.2 (oral and nasal data set) or >0.3/<-0.3 (fecal data set) in the sPLS analysis into the CCpnA, and additionally included categorical variables that could not be included in the sPLS. The method was implemented with the *cca()* function in package vegan [79]. It implements a Chi-square transformation of the log+1 transformed ASV count matrix and subsequent weighted linear regression, followed by singular value decomposition. We depicted the CCpnA results as a triplot with plot dimensions corresponding in length to the percentage of variance explained by each axis. In an ANOVA,

we tested for the overall significance of the CCpnA models at each body site, as well as for the significance of the components, and the significance of the variables. For this purpose we used the function *anova.cca()* in vegan [79].

List of abbreviations

ALT: Alanine transaminase
AML: Acute myeloid leukemia
ASAT: Aspartate transaminase
ASV: Amplicon sequence variant
ATG: Anti-thymocyte globulin
CCpnA: Canonical correspondence analysis
CML: Chronic myeloid leukemia
CRP: C-reactive protein
CTREE: Conditional inference tree
ECLIA: Electrochemiluminescence immunoassays
GvL effect: Graft-versus-leukemia effect
(a)GvHD: (acute) graft-versus-host disease
HSCT: Hematopoietic stem cell transplantation
IDS: Immunodeficiency syndromes
IEA: Inherited abnormalities of erythrocyte differentiation or function
Ig: Immunoglobulin
IMD: Inherited disorders of metabolism
LIA: Latex immunoturbidimetric assay
LDA: Linear discriminant analysis
LogitBoost: Boosted logistic regression
LOOCV: Leave-one-out cross validation
MDS: Myelodysplastic or myeloproliferative disorders
MM: Multiple myeloma
NHL: Non-Hodgkin lymphomas
NPMLE: Non-parametric maximum likelihood estimator
OL: Other leukemia
OTU: Operational taxonomic unit
PBMC: Peripheral blood mononuclear cell
PCoA: Principal Coordinates Analysis
Rf: Random forest
SAA: Severe aplastic anemia
SCFA: Short-chain fatty acid
sPLS: Sparse partial least squares analysis
svmLinear: Support vector machines with linear kernel
svmRadial: Support vector machines with radial basis function kernel
TBI: Total body irradiation

T_H17 cell: T helper 17 cell

T_{reg} cell: T regulatory cell

UCB: Umbilical cord blood

Declarations

Acknowledgements

We thank the patients and their families for their participation in this study. We are grateful for the help of Marianne Ifversen (Department of Pediatrics and Adolescent Medicine, Copenhagen University Hospital Rigshospitalet) with clinical data registration and logistics of sample collection. We also thank Marlene Danner Dalgaard and Neslihan Bicen (Multi Assay Core facility (DMAC), Technical University of Denmark) for library construction and sequencing. Furthermore, we would like to thank Patrick Murigu Kamau Njage (Technical University of Denmark) for helpful discussions related to machine learning models.

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 will be made available through the European Nucleotide Archive (ENA) at the European Bioinformatics Institute (EBI) at publication. The datasets generated and/or analysed during the current study as well as the R code used to analyze the data will be available from the figshare repository at publication.

Author's contributions

A.C.M., K.K., K.G.M., and S.J.P. designed the research; A.C.M., K.K., H.M, and S.J.P. performed the research; A.C.M., and S.J.P. contributed analytic tools; A.C.M., and S.J.P. analyzed the data; A.C.M. and S.J.P. wrote the manuscript; and K.K., F.M.A., and K.G.M. edited the manuscript.

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-7-2014-016) and the Danish Data Protection Agency.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Chabannon C, Kuball J, Bondanza A, Dazzi F, Pedrazzoli P, Toubert A, et al. Hematopoietic stem cell transplantation in its 60s: A platform for cellular therapies. Sci Transl Med. 2018;10:eaap9630. doi:10.1126/scitranslmed.aap9630.

2. 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.

3. 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.

4. Bik EM. The Hoops, Hopes, and Hypes of Human Microbiome Research. Yale J Biol Med. 2016;89:363–73. http://www.ncbi.nlm.nih.gov/pubmed/27698620. Accessed 15 Aug 2018.

5. 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.

6. 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.

7. Ghimire S, Weber D, Mavin E, Wang X nong, Dickinson AM, Holler E. Pathophysiology of GvHD and Other HSCT-Related Major Complications. Front Immunol. 2017;8:79. doi:10.3389/fimmu.2017.00079.

8. 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.

9. Han L, Jin H, Zhou L, Zhang X, Fan Z, Dai M, et al. Intestinal Microbiota at Engraftment Influence Acute Graft-Versus-Host Disease via the Treg/Th17 Balance in Allo-HSCT Recipients. Front Immunol. 2018;9:669. doi:10.3389/fimmu.2018.00669.

10. Ratajczak P, Janin A, Peffault de Latour R, Leboeuf C, Desveaux A, Keyvanfar K, et al. Th17/Treg ratio in human graft-versus-host disease. Blood. 2010;116:1165–71. doi:10.1182/blood-2009-12-255810.

11. Honda K, Littman DR. The microbiota in adaptive immune homeostasis and disease. Nature. 2016;535:75–84. doi:10.1038/nature18848.

12. Atarashi K, Tanoue T, Oshima K, Suda W, Nagano Y, Nishikawa H, et al. Treg induction by a rationally selected mixture of Clostridia strains from the human microbiota. Nature. 2013;500:232–6. doi:10.1038/nature12331.

13. Larsen JM. The immune response to *Prevotella* bacteria in chronic inflammatory disease. Immunology. 2017;151:363–74. doi:10.1111/imm.12760.

14. Masche AC, Kielsen K, Cilieborg MS, Lund O, Holmes S, Aarestrup FM, et al. Specific gut microbiome

members are associated with distinct immune markers in allogeneic hematopoietic stem cell transplantation. Under Rev. 2018.

15. 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.

16. Weber D, Oefner PJ, Hiergeist A, Koestler J, Gessner A, Weber M, et al. Low 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.

17. 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.

18. Andermann TM, Peled JU, Ho C, Reddy P, Riches M, Storb R, et al. The Microbiome and Hematopoietic Cell Transplantation: Past, Present, and Future. Biol Blood Marrow Transplant. 2018. doi:10.1016/J.BBMT.2018.02.009.

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

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

21. 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.

22. Kim M, Qie Y, Park J, Kim CH. Gut Microbial Metabolites Fuel Host Antibody Responses. Cell Host Microbe. 2016;20:202–14. doi:10.1016/j.chom.2016.07.001.

23. Verma D, Garg PK, Dubey AK. Insights into the human oral microbiome. Arch Microbiol. 2018;200:525–40. doi:10.1007/s00203-018-1505-3.

24. Osakabe L, Utsumi A, Saito B, Okamatsu Y, Kinouchi H, Nakamaki T, et al. Influence of Oral Anaerobic Bacteria on Hematopoietic Stem Cell Transplantation Patients: Oral Mucositis and General Condition. Transplant Proc. 2017;49:2176–82. doi:10.1016/j.transproceed.2017.09.012.

25. Soga Y, Maeda Y, Ishimaru F, Tanimoto M, Maeda H, Nishimura F, et al. Bacterial substitution of coagulase-negative staphylococci for streptococci on the oral mucosa after hematopoietic cell transplantation. Support Care Cancer. 2011;19:995–1000.

26. Olczak-Kowalczyk D, Daszkiewicz M, Krasuska-Slawińska, Dembowska-Bagińska B, Gozdowski D, Daszkiewicz P, et al. Bacteria and Candida yeasts in inflammations of the oral mucosa in children with secondary immunodeficiency. J Oral Pathol Med. 2012;41:568–76.

27. Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R. Bacterial community variation in

human body habitats across space and time. Science. 2009;326:1694–7. doi:10.1126/science.1177486.

28. Gevers D, Kugathasan S, Denson LA, Vázquez-Baeza Y, Van Treuren W, Ren B, et al. The treatment-naive microbiome in new-onset Crohn's disease. Cell Host Microbe. 2014;15:382–92. doi:10.1016/j.chom.2014.02.005.

29. 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.

30. Golob JL, Pergam SA, Srinivasan S, Fiedler TL, Liu C, Garcia K, et al. Stool Microbiota at Neutrophil Recovery Is Predictive for Severe Acute Graft vs Host Disease After Hematopoietic Cell Transplantation. Clin Infect Dis. 2017;65:1984–91. doi:10.1093/cid/cix699.

31. Biagi E, Zama D, Nastasi C, Consolandi C, Fiori J, Rampelli S, et al. Gut microbiota trajectory in pediatric patients undergoing hematopoietic SCT. Bone Marrow Transplant. 2015;50:992–8. doi:10.1038/bmt.2015.16.

32. Mancini N, Greco R, Pasciuta R, Barbanti MC, Pini G, Morrow OB, et al. Enteric Microbiome Markers as Early Predictors of Clinical Outcome in Allogeneic Hematopoietic Stem Cell Transplant: Results of a Prospective Study in Adult Patients. Open Forum Infect Dis. 2017;4. doi:10.1093/ofid/ofx215.

33. Sahin U, Toprak SK, Atilla PA, Atilla E, Demirer T. An overview of infectious complications after allogeneic hematopoietic stem cell transplantation. J Infect Chemother. 2016;22:505–14. doi:10.1016/J.JIAC.2016.05.006.

34. Forcina A, Rancoita PMV, Marcatti M, Greco R, Lupo-Stanghellini MT, Carrabba M, et al. A New Clinicobiological Scoring System for the Prediction of Infection-Related Mortality and Survival after Allogeneic Hematopoietic Stem Cell Transplantation. Biol Blood Marrow Transplant. 2017;23:2151–8. doi:10.1016/J.BBMT.2017.08.010.

35. Ogonek J, Kralj Juric M, Ghimire S, Varanasi PR, Holler E, Greinix H, et al. Immune Reconstitution after Allogeneic Hematopoietic Stem Cell Transplantation. Front Immunol. 2016;7:507. doi:10.3389/fimmu.2016.00507.

36. de Koning C, Plantinga M, Besseling P, Boelens JJ, Nierkens S. Immune Reconstitution after Allogeneic Hematopoietic Cell Transplantation in Children. Biol Blood Marrow Transplant. 2016;22:195–206. doi:10.1016/j.bbmt.2015.08.028.

37. Blaser BW, Kim HT, Alyea EP, Ho VT, Cutler C, Armand P, et al. Hyperlipidemia and Statin Use after Allogeneic Hematopoietic Stem Cell Transplantation. Biol Blood Marrow Transplant. 2012;18:575–83. doi:10.1016/J.BBMT.2011.08.003.

38. Marini BL, Choi SW, Byersdorfer CA, Cronin S, Frame DG. Treatment of dyslipidemia in allogeneic hematopoietic stem cell transplant patients. Biol Blood Marrow Transplant. 2015;21:809–20. doi:10.1016/j.bbmt.2014.10.027.

39. Kverka M, Zakostelska Z, Klimesova K, Sokol D, Hudcovic T, Hrncir T, et al. Oral administration of Parabacteroides distasonis antigens attenuates experimental murine colitis through modulation of

immunity and microbiota composition. Clin Exp Immunol. 2011;163:250–9. doi:10.1111/j.1365-2249.2010.04286.x.

40. Arpaia N, Campbell C, Fan X, Dikiy S, Van Der Veeken J, Deroos P, et al. Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation. Nature. 2013;504:451–5. doi:10.1038/nature12726.

41. Pitt JM, Vétizou M, Waldschmitt N, Kroemer G, Chamaillard M, Boneca IG, et al. Fine-Tuning Cancer Immunotherapy: Optimizing the Gut Microbiome. Cancer Res. 2016;76:4602–7. doi:10.1158/0008-5472.CAN-16-0448.

42. Mager DL, Haffajee AD, Devlin PM, Norris CM, Posner MR, Goodson JM. The salivary microbiota as a diagnostic indicator of oral cancer: a descriptive, non-randomized study of cancer-free and oral squamous cell carcinoma subjects. J Transl Med. 2005;3:27. doi:10.1186/1479-5876-3-27.

43. Laheij AMGA, de Soet JJ, von dem Borne PA, Kuijper EJ, Kraneveld EA, van Loveren C, et al. Oral bacteria and yeasts in relationship to oral ulcerations in hematopoietic stem cell transplant recipients. Support Care Cancer. 2012;20:3231–40. doi:10.1007/s00520-012-1463-2.

44. Correa ME, Schubert MM, Storer B, Martin PJ, Flowers MED. Correlation Between Severity of Oral Mucositis and Both Overall and Gut Acute GvHD After High Intensity Conditioning Hematopoietic Stem Cell Transplantation. Blood. 2012;120. http://www.bloodjournal.org/content/120/21/3057?sso-checked=true. Accessed 25 Nov 2018.

45. Chaudhry HM, Bruce AJ, Wolf RC, Litzow MR, Hogan WJ, Patnaik MS, et al. The Incidence and Severity of Oral Mucositis among Allogeneic Hematopoietic Stem Cell Transplantation Patients: A Systematic Review. Biol Blood Marrow Transplant. 2016;22:605–16. doi:10.1016/J.BBMT.2015.09.014.

46. Barraco F, Labussière-Wallet H, Valour F, Ducastelle-Leprêtre S, Nicolini FE, Thomas X, et al. Actinomycosis after allogeneic hematopoietic stem cell transplantation despite penicillin prophylaxis. Transpl Infect Dis. 2016;18:595–600.

47. Chakrabarti S, Lees A, Jones S, Milligan D. Clostridium difficile infection in allogeneic stem cell transplant recipients is associated with severe graft-versus-host disease and non-relapse mortality. Bone Marrow Transplant. 2000;26:871–6. doi:10.1038/sj.bmt.1702627.

48. Flynn KJ, Baxter NT, Schloss PD. Metabolic and Community Synergy of Oral Bacteria in Colorectal Cancer. mSphere. 2016;1. doi:10.1128/mSphere.00102-16.

49. Tjalsma H, Boleij A, Marchesi JR, Dutilh BE. A bacterial driver-passenger model for colorectal cancer: Beyond the usual suspects. Nat Rev Microbiol. 2012;10:575–82. doi:10.1038/nrmicro2819.

50. Knudsen BE, Bergmark L, Munk P, Lukjancenko O, Prieme A, Aarestrup FM, et al. Impact of Sample Type and DNA Isolation Procedure on Genomic Inference of Microbiome Composition. bioRxiv. 2016;1:64394. doi:10.1101/064394.

51. 16S Metagenomic Sequencing Library Preparation. https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-

guide-15044223-b.pdf. Accessed 17 Apr 2018.

52. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res. 2013;41:e1–e1. doi:10.1093/nar/gks808.

53. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 2011;17:10. doi:10.14806/ej.17.1.200.

54. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13:581–3. doi:10.1038/nmeth.3869.

55. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32:3047–8. doi:10.1093/bioinformatics/btw354.

56. Callahan B. Silva taxonomic training data formatted for DADA2 (Silva version 132). 2018. doi:10.5281/ZENODO.1172783.

57. 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.

58. Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome. 2018;6:226. doi:10.1186/s40168-018-0605-2.

59. 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.

60. Wright ES. Using DECIPHER v2.0 to Analyze Big Biological Sequence Data in R. R J. 2016;8:352–9.

61. Schliep KP. phangorn: phylogenetic analysis in R. Bioinformatics. 2011;27:592–3. doi:10.1093/bioinformatics/btq706.

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

63. R Core Team. R: A Language and Environment for Statistical Computing. 2018. https://www.r-project.org/.

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

65. Rohart F, Gautier B, Singh A, Lê Cao K-A. mixOmics: An R package for 'omics feature selection and multiple data integration. PLOS Comput Biol. 2017;13:e1005752. doi:10.1371/journal.pcbi.1005752.

66. Fukuyama J. treeDA: Tree-Based Discriminant Analysis. 2017. https://github.com/jfukuyama/treeda.

67. Kuhn M, Wing J, Weston S, Williams A, Keefer C, Engelhardt A, et al. caret: Classification and Regression Training. R package version 6.0-80. 2018. https://cran.r-project.org/package=caret.

68. Hothorn T, Zeileis A, Cheng E:, Ong S. partykit: A modular toolkit for recursive partytioning in R. J Mach Learn Res. 2015;16:3905–9.

69. Hothorn T, Hornik K, Zeileis A. Unbiased Recursive Partitioning: A Conditional Inference Framework. J Comput Graph Stat. 2006;15:651–74. doi:10.1198/106186006X133933.

70. Fukuyama J, Rumker L, Sankaran K, Jeganathan P, Dethlefsen L, Relman DA, et al. Multidomain analyses of a longitudinal human microbiome intestinal cleanout perturbation experiment. PLoS Comput Biol. 2017;13:e1005706. doi:10.1371/journal.pcbi.1005706.

71. Njage PMK, Henri C, Leekitcharoenphon P, Mistou M, Hendriksen RS, Hald T. Machine Learning Methods as a Tool for Predicting Risk of Illness Applying Next-Generation Sequencing Data. Risk Anal. 2018;:risa.13239. doi:10.1111/risa.13239.

72. Njage PMK, Leekitcharoenphon P, Hald T. Improving hazard characterization in microbial risk assessment using next generation sequencing data and machine learning: Predicting clinical outcomes in shigatoxigenic Escherichia coli. Int J Food Microbiol. 2019;292:72–82. doi:10.1016/J.IJFOODMICRO.2018.11.016.

73. Chen J, Zhang L. GMPR: Geometric Mean of Pairwise Ratios. R package version 0.1.3. 2017.

74. 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.

75. 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.

76. Kursa MB, Rudnicki WR. Feature Selection with the Boruta Package. Journal of Statistical Software. 2010;36:1–13.

77. Anderson-Bergman C. **icenReg** : Regression Models for Interval Censored Data in *R*. J Stat Softw. 2017;81. doi:10.18637/jss.v081.i12.

78. Stekhoven DJ, Buhlmann P. MissForest--non-parametric missing value imputation for mixed-type data. Bioinformatics. 2012;28:112–8. doi:10.1093/bioinformatics/btr597.

79. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan: Community Ecology Package. R package version 2.5-2. 2018. https://cran.r-project.org/package=vegan.

80. 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.

4.3 Manuscript III

Bacterial alpha diversity trajectories and immune-microbial associations of the gut, oral, and nasal microbiota in adult autologous HSCT patients

Anna Cäcilia Masche[1], Katrine Kielsen[2, 3], Hanne Mordhorst[1], 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.

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

Manuscript in preparation

Abstract

Background

The importance of the human microbiota in health and disease, and its role in immune homeostasis is increasingly recognized. A number of studies have provided insight into characteristic shifts in the gut microbial community of patients undergoing hematopoietic stem cell transplantation, and linked microbial abundance patterns to clinical outcomes. However, so far the focus has been on allogeneic transplantation, i.e. the infusion of stem cells derived from a foreign donor. Autologous HSCT, involving the transplantation of the patient's own cells, extracted and preserved before the onset of chemotherapy, has been studied to a lesser degree with regards to the microbiota. Recipients of auto-HSCT receive prophylactic antimicrobial treatment prior to transplantation, indicating that the microbiota might be affected and potentially plays a role in this patient group.

Results

Here, we have monitored long-term gut, oral, and nasal microbial abundances in a cohort of adult autologous HSCT recipients over the period of one year. We show decreased bacterial alpha diversity early after transplantation and rapid recovery within one to three weeks at all three body sites. Multivariate multi-table approaches revealed a number of host-microbial associations that were partly shared between body sites. High abundances of intestinal *Bacteroides* spp. in the gut were associated with a high number of bacterial infections and frequent antibiotic treatment. The abundances of oral *Prevotella* spp. were positively correlated with T_H17 and T_{reg} cell counts.

Conclusions

With the exploration of host-microbial associations in this relatively small group of adult autologous HSCT patients, we demonstrate that the microbiota undergoes changes and exhibits correlations with immune cell counts and immune markers. For instance, correlations of microbial abundances with T_H17 and T_{reg} cell counts points to a potential microbial involvement in mucosal or systemic inflammatory regulation in the context of auto-HSCT. We emphasize the need to put focus on autologous HSCT patients in addition to the allogeneic setting, both as a reference group for donor-related effects and as a relevant patient group in itself. Further research might enable the identification of microbial markers based on the starting points we provide here, which might in the future be of clinical use.

Background

Autologous hematopoietic stem cell transplantation (HSCT) is a form of therapy used in lymphoid malignancies, such as multiple myeloma (MM), as well as Hodgkin and non-Hodgkin lymphoma (HL and NHL) [1]. It involves the extraction of the recipient's own stem cells prior to high-dose chemotherapy, and subsequent infusion of these cells to re-establish the compromised immune system [2]. The role of the microbiota in autologous HSCT is vastly understudied compared with allogeneic HSCT. One could argue that more substantial changes are expected in the commensal microbial community structure in relation to allogeneic HSCT due to the encounter with foreign immune cells. In this light, autologous HSCT recipients still constitute a valuable control population to relativize the extent of donor-related effects. However, the impact on the microbiota in allo-HSCT can be attributed to antibiotic treatment to a high degree [3]. Auto-HSCT patients likewise undergo antimicrobial treatment, both prophylactically and upon the occurrence of infections, even though the choice of antimicrobial agents might differ [4]. Therefore, the microbiota might be similarly affected and involved in auto-HSCT as was shown in allo-HSCT, and on its own constitutes a patient group of interest. One important aspect for the success of HSCT is the timely reconstitution of the immune system. Microbial involvement in the recovery of B and NK cells has been shown in pediatric allo-HSCT patients [5]. In addition, the T cell compartment, and especially the balance between T helper 17 (T_H17) and T regulatory (T_{reg}) cells, are of interest. T_H17 cells are important for the defense against e.g. extracellular bacterial pathogens [6]. However, pro-inflammatory signaling of T_H17 cells induces autoimmunity if not counter-acted by T_{reg} cell activity [6]. Members of the host microbiota are able to induce both, T_H17 cells by epithelial adhesion, and T_{reg} cells via the production of short-chain fatty acids [7–9]. These mechanisms and their stimulation by members of the commensal microbiota at mucosal sites might be of relevance to understand inflammatory processes in HSCT. Inflammatory complications in HSCT occur for instance in relation to chemotherapy-induced gut toxicity and oral mucositis [10, 11]. With regards to microbial involvement, *Prevotella* spp. have been proposed to play a role in both T_H17-mediated systemic and local mucosal inflammation, as well as in oral inflammatory disorders, such as periodontitis, suggesting a host-microbial link that might be worth investigating in the context of HSCT [12].

Here, we provide comprehensive insight into changes of gut, oral, and nasal microbial abundances, which we tracked over a period of one year in adult recipients of autologous HSCT. We show a decrease in bacterial alpha diversity within the first two weeks post-transplant at all three body sites. Diversity levels recovered rapidly, especially in the nasal cavity. In the gut, we reveal that amplicon sequence variants (ASVs) affiliated with the family *Enterococcaceae* increase in abundance early after auto-HSCT. Furthermore, *Streptococcaceae* expanded concurrent to the post-transplant loss of diversity at all three body sites. We integrated ASV count data with immune cell counts, immune markers, and clinical outcomes. For instance, we included plasma glucose and phosphate levels. Both markers might be of relevance, especially during the neutropenic phase: Hypophosphatemia has been described as a marker of rapid neutrophil recovery after HSCT, and hyperglycemia has been related to an increased risk of infections post-transplantation [13, 14]. Here we demonstrate that a high number of bacterial infections and frequent antibiotic treatment through the one year period, as well as high plasma glucose levels were associated with high abundances of distinct sets of ASVs, including intestinal *Bacteroides* spp., at all three body sites. In addition, high plasma phosphate levels and high reticulocyte counts were

correlated with intestinal *Ruminococcaceae* and *Lachnospiraceae* in NHL patients, whereas the same markers were associated with oral and nasal microbial abundances in MM patients. Moreover, abundances of oral *Prevotellaceae* ASVs, and nasal *Corynebacteriaceae* ASVs were correlated with T_{reg} and T_H17 cell counts. With regards to the relatively small size of our cohort, we restrain from drawing definite conclusions from the revealed associations. This study might be considered as an initial exploration of the role of the microbiota in autologous HSCT, and demonstrates its relevance. We provide a number of starting points for further investigations, which are required to obtain more conclusive evidence of host-microbial interactions in auto-HSCT.

Results

The gut, oral, and nasal microbiota of adult recipients of autologous hematopoietic stem cell transplantation (auto-HSCT) was monitored for one year. At nine time points, we collected fecal samples, and buccal and anterior naris swabs: once before HSCT, on the day of HSCT, weekly during the first month post transplantation, and at three later time points. We used 16S rRNA gene profiling to describe how the microbial community changes over time in auto-HSCT patients. Associations between microbial dynamics and immune reconstitution, immune markers, and clinical outcomes in relation to auto-HSCT were investigated in two multivariate multi-table approaches. To elucidate immune reconstitution, we quantified a range of immune cell types, including reticulocytes and neutrophils, in peripheral blood (see Methods). Moreover, we examined biochemical blood markers, for instance plasma phosphate and bilirubin levels, among other immune markers (see Methods).

Patient cohort and outcomes

The cohort consisted of 13 adult patients (average age: 54.2, range: 28.7 - 70.4) whose microbiota was monitored in the oral and nasal cavities, and the microbiota of 11 of these was in addition monitored in the gut. From before auto-HSCT and during the first year post transplantation, eight patients (62%) had at least one bacterial infection and two patients (15%) had at least one viral infection (Table S1). Within an average follow-up time of eight months, six patients (46%) went into complete remission, while six (46%) never reached complete remission (Table S1). Two patients (15%) died due to relapse. All autologous stem cell recipients were treated with amoxicillin from day +4. In case of fever, the patients were administered intra-venous piperacillin-tazobactam. Upon lack of effect, this was substituted with meropenem.

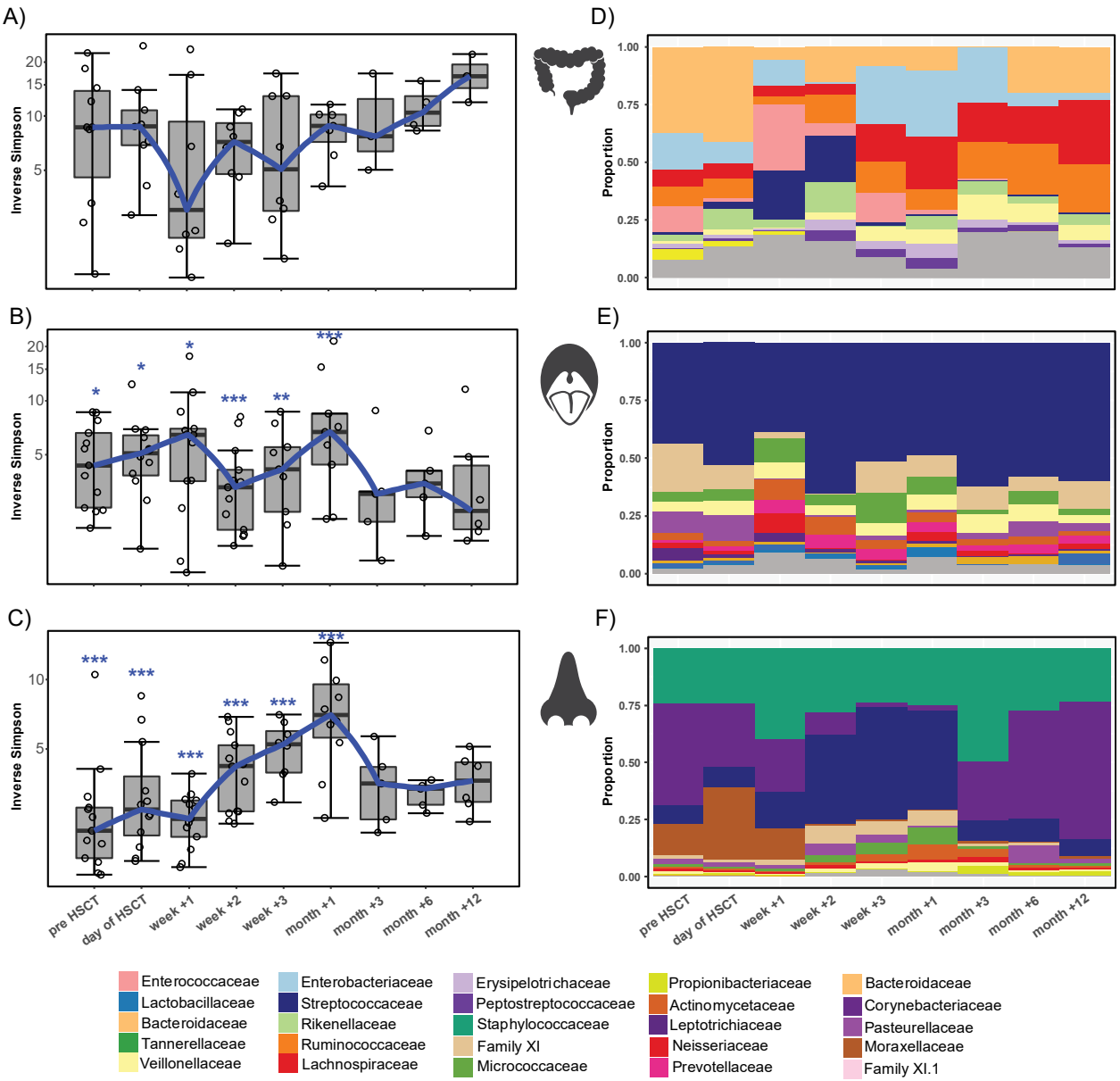


Figure 1. Bacterial alpha diversity and major taxonomic families in the gut, oral, and nasal microbiota of autologous HSCT patients. A) Bacterial alpha diversity before, at the time of, and after autologous HSCT in the gut, E) the oral cavity, and F) the nasal cavity. Alpha diversity is displayed on a log10-transformed y-axis for visualization purposes. Asterisks indicate significant differences in median inverse Simpson index between time points (showing the maximum significance level). * $P < 0.05$, ** $P < 0.01$ and *** $P < 0.001$. D) Relative abundances of the 12 most abundant families at each body site are shown over time for the gut, E) the oral cavity, and F) the nasal cavity. ('Body site'-icons in this figure and the following figures from: colourbox.com.)

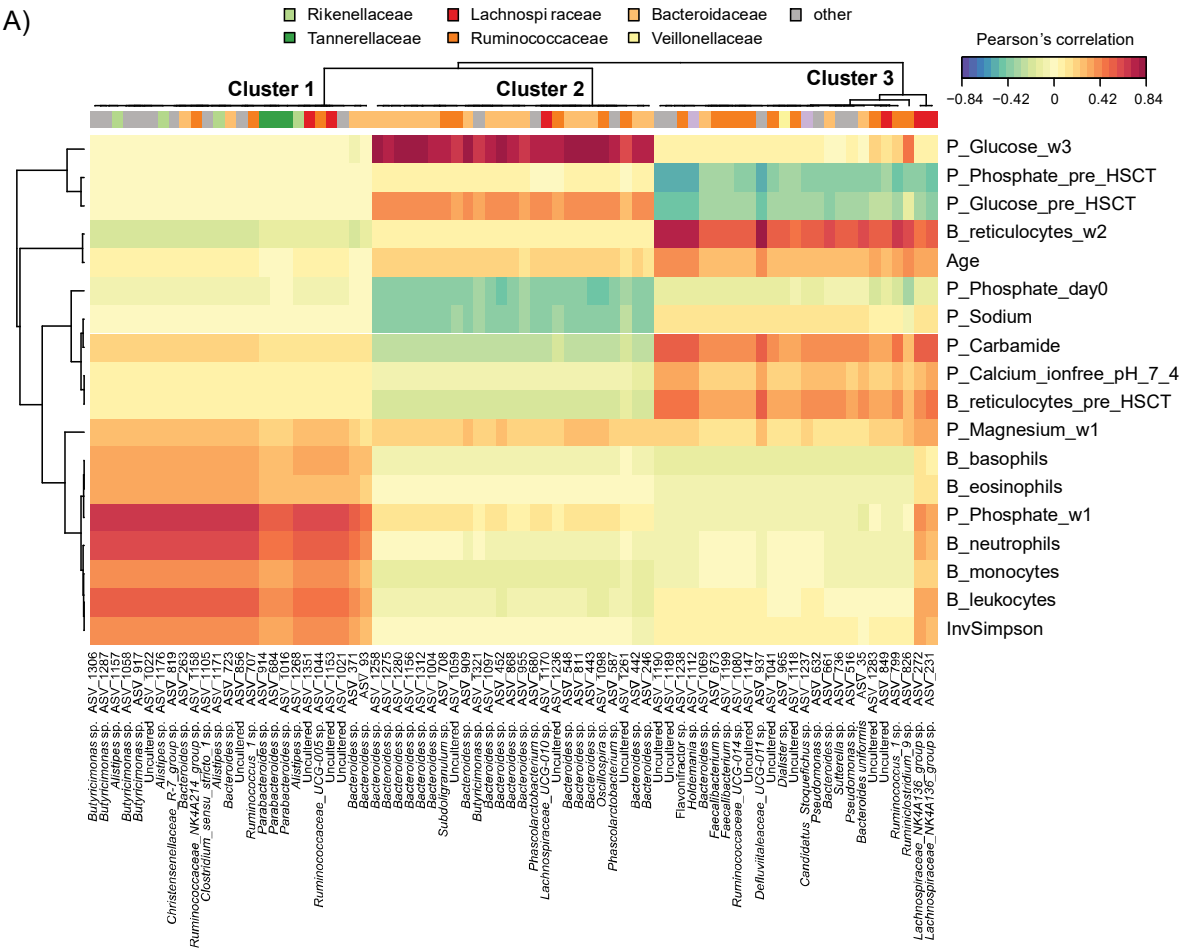
Bacterial alpha diversity changes upon HSCT in the gut, and the oral and nasal cavities

We obtained 60 fecal, 85 buccal swab, and 85 anterior naris swab samples that were examined by 16S rRNA gene profiling in order to understand dynamics of microbial abundance patterns over time before, at the time of, and after auto-HSCT. We obtained 1,030 ASVs in the fecal data set, 372 ASVs in the oral data set, and 219 ASVs in the nasal data set (see Methods). In the gut, alpha diversity (inverse Simpson) decreased from a median inverse Simpson index of 8.74 (range: 2.81 -24.66) on the day of HSCT to a minimum of 3.0 in week+1 (range: 1.26 – 23.58), and increased in month +1 to a level comparable to pre HSCT (8.84 (range: 4.05 – 10.22)) (Figure 1A). We did not detect significant differences in diversity between time points for the gut microbiota. In contrast, in the oral cavity, a significant decrease in median alpha diversity occurred from 6.47 in week +1 (range: 1.1 – 17.72) to 3.29 in week +2 (range 1.56 – 8.15, $P = 0.02$, Figure 1B). Diversity successively increased again to 6.71 (range 2.19 – 21.53) in month +1 ($P < 0.001$, Figure 1B). In the nasal cavity, we observed a significant decline in median alpha diversity from 2.73 (range: 1.63 – 8.51) on the day of HSCT to 2.48 (range: 1.53 – 3.9) in week +1 ($P = 0.03$, Figure 1C). Nasal diversity levels recovered rapidly to 4.21 (range: 2.37 – 6.9) in week +2 ($P < 0.01$, Figure 1C), exceeding median alpha diversity levels before and at the time of transplantation.

Microbial community dynamics in the gut, and the oral and nasal cavities

As an initial indication of microbial compositional changes over time, we transformed ASV counts by taxonomic family to relative abundance. Overall, the gut was dominated by a number of families, comprising e.g. *Bacteroidaceae*, *Enterobacteriaceae,* and *Lachnospiraceae* (Figure S1A). Concurrent to the drop in gut bacterial alpha diversity in week +1, we observed an expansion in the proportion of *Streptococcaceae* from 3.3% to 21.2%, and of *Enterococcaceae* from 1.4% to 28.4%. In contrast, relative abundance of *Bacteroidaceae* dropped from 41.2% on the day of HSCT to 5.4% in week +1 (Figure 1D). In the oral cavity, members of the family *Streptococcaceae* were dominant at all time points (Figure S1B). An expansion in relative abundance of this family from 38.4% to 65% accompanied the decrease in diversity at this site in week +2 (Figure 1E). The nasal cavity generally exhibited high relative abundances of *Streptococcaceae*, *Staphylococcaceae*, and *Corynebacteriaceae* (Figure S1C). At the time of decreased alpha diversity in week +1, the proportion of *Streptococcaceae* increased from 8.7% to 16.1%, while we observed a reduction in *Moraxellaceae* (from 31.5% to 13.8%) and to a lesser degree in *Corynebacteriaceae* (from 27.8% to 22.8%). Between week +2 and month +1 where we observed nasal bacterial diversity exceeding pre-transplant levels, *Streptococcaceae* expanded further to 43.5%. Moreover, other families including *Pasteurellaceae* and *Micrococcaceae* increased in relative abundance during this period, while *Moraxellaceae* and *Corynebactericeae* were further reduced (Figure 1F).

A)


Rikenellaceae
Tannerellaceae
Lachnospiraceae
Ruminococcaceae
Bacteroidaceae
Veillonellaceae
other
Pearson's correlation
−0.84 −0.42 0 0.42 0.84
Cluster 1
Cluster 2
Cluster 3
P_Glucose_w3
P_Phosphate_pre_HSCT
P_Glucose_pre_HSCT
B_reticulocytes_w2
Age
P_Phosphate_day0
P_Sodium
P_Carbamide
P_Calcium_ionfree_pH_7_4
B_reticulocytes_pre_HSCT
P_Magnesium_w1
B_basophils
B_eosinophils
P_Phosphate_w1
B_neutrophils
B_monocytes
B_leukocytes
InvSimpson

B)

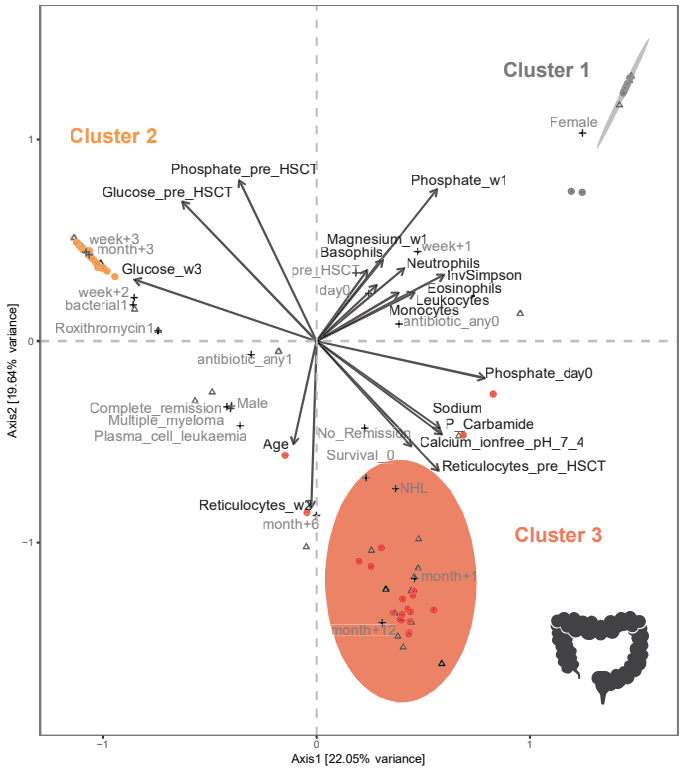

Cluster 1
Cluster 2
Cluster 3
Female
Phosphate_pre_HSCT
Glucose_pre_HSCT
Phosphate_w1
week+3
month+3
Glucose_w3
Magnesium_w1
Basophils
week+1
Neutrophils
InvSimpson
Eosinophils
Leukocytes
Monocytes
antibiotic_any0
pre_HSCT
day0
week+2
bacterial1
Roxithromycin1
antibiotic_any1
Phosphate_day0
Complete_remission
Multiple_myeloma
Plasma_cell_leukaemia
Male
Age
No_Remission
Survival_0
Sodium
P_Carbamide
Calcium_ionfree_pH_7_4
Reticulocytes_pre_HSCT
NHL
Reticulocytes_w2
month+6
month+12
Axis1 [22.05% variance]
Axis2 [19.64% variance]

Figure 2. Multivariate host – gut microbial associations in autologous HSCT patients. A) Pairwise correlations (>0.3/<-0.3) of gut ASVs (bottom) and continuous immune and clinical parameters (right) displayed in a hierarchically clustered image map (CIM). Complete linkage clustering was based on the first three dimensions of the partial least squares (sPLS) regression model. Red and blue coloring indicates positive and negative Pearson's correlation, respectively. B) Triplot of the canonical correspondence analysis (CCpnA), showing associations between gut ASV abundances (circles) and continuous (arrows) and categorical (+) immune and clinical parameters. ASVs and variables were included in the CCpnA if they had at least one correlation >0.3/<-0.3 in the sPLS regression analysis. Variables and ASVs with a score >0.3/<-0.3 on at least one the first three CCpnA axes are displayed in dimension 1 versus 2. Samples are shown as triangles. Clusters of ASVs identified by the sPLS-based hierarchical clustering are shown as colored ellipses (with 80% confidence interval). Abbreviations not mentioned in text: bacterial, bacterial infection; P_, plasma.

Glycemic and infectious status are associated with the microbiota across body sites

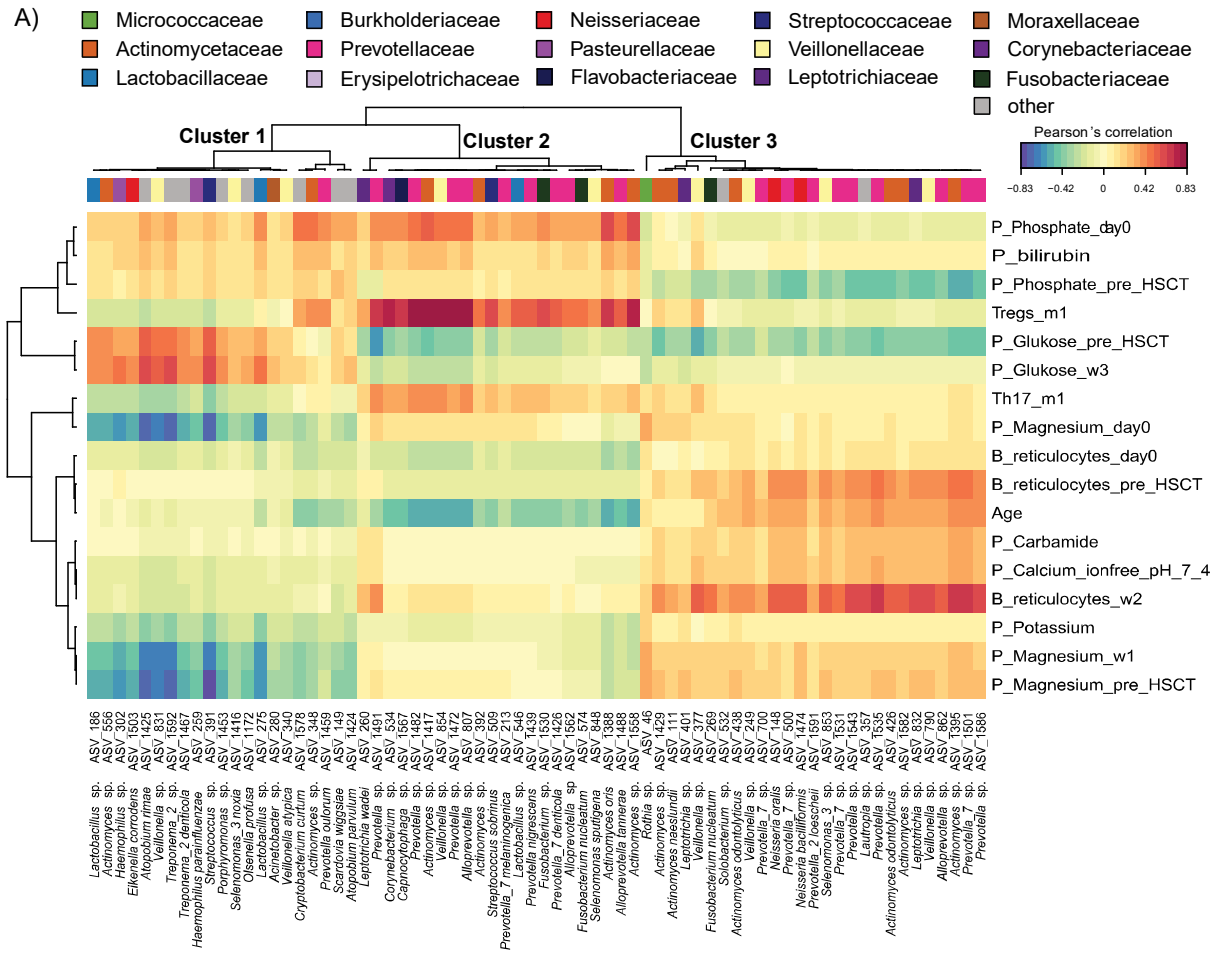
We were interested in how the microbiota in the gut, and the oral and nasal cavities might be correlated with immune reconstitution, immune markers, and clinical outcomes in the context of autologous HSCT. Two multivariate multi-table approaches allowed us to discover associations and compare them across the three investigated body sites. In a first step, we used sparse partial least squares (sPLS) analysis separately for each body site, with subsequent hierarchical clustering. The method was run in regression mode to explain variation in multiple continuous outcomes (immune cell counts, immune marker levels, and bacterial alpha diversity) by variance stabilized ASV abundances. In addition, we performed canonical correspondence analysis (CCpnA) on the subset of ASVs and immune parameters with the highest correlations in the sPLS analyses (see Methods). CCpnA facilitates the inclusion of categorical variables, such as clinical outcomes, and assesses relationships between the microbiota, and immune and clinical parameters in a bidirectional manner. Three clusters of AVSs were revealed at each body sites by sPLS-based hierarchical clustering, which were confirmed by the respective CCpnAs (Figures 2, 3, and S2). The major associations shared between two or three of the investigated body sites will be described in detail in the following paragraphs.

The sPLS analysis suggested a strong positive correlation between plasma glucose levels prior to HSCT and in week +3, and distinct sets of ASVs in the gut (cluster 2), the oral cavity (cluster 1), and the nasal cavity (cluster 1) (Figures 2A, 3A, S2A). Gut cluster 2 primarily comprised members of the families *Bacteroidaceae* (exclusively *Bacteroides* spp.) and *Ruminococcaceae* (Figure 2A). In the oral cavity, ASVs in cluster 1 were affiliated with a number of families including *Lactobacillaceae*, *Streptococcaceae*, and *Veillonellaceae* (Figure 3A). Nasal cluster 1 mainly consisted of *Corynebacteriaceae,* and *Moraxellaceae* ASVs (Figure S2A). No ASVs in these clusters were shared between body sites. In the oral, and nasal cavities, but not the gut, the same ASVs additionally exhibited negative correlations with plasma levels of magnesium prior to HSCT, at the time of HSCT, and in week +1. The CCpnA confirmed the associations observed in the sPLS analysis. In addition, our analysis revealed that samples affiliated with the respective cluster at each body site (gut cluster 2, oral cluster 1, nasal cluster 1) originated from patients that suffered many bacterial infections and were treated with antibiotics more often than the rest of the cohort. In the gut and oral cavity, roxithromycin-treatment was associated with the respective clusters

(Figures 2B and 3B), whereas frequent metronidazole treatment was observed when analyzing the nasal cavity data set (Figure S2B). In total, four patients were treated with at least one of those antimicrobial agents and had at least two bacterial infections during the monitored period (Figure 4). In the gut and the nasal cavity, samples in these clusters originated mainly from week +2 or week +3 according to the CCpnAs. Additionally samples from months +3 and +6 occurred here at all three body sites.

Reticulocyte counts and phosphate levels are associated with site-specific microbiota of distinct patient groups

A second set of shared associations across body sites was revealed by the sPLS analysis, showing positive correlations between counts of reticulocytes prior to HSCT and in week +2 in older patients compared to the rest of the cohort, and cluster 3 at each body site, respectively (Figures 2A, 3A, and S2A). In contrast, the same clusters of ASVs were negatively correlated with pre-transplant plasma levels of phosphate and glucose. In the gut, cluster 3 was dominated by *Ruminococcaceae*, and additionally comprised a few *Lachnospiraceae* and *Bacteroidaceae* ASVs(Figure 2A). A large group of ASVs constituted oral cluster 3, representing e.g. *Prevotellaceae*, *Actinomyceataceae*, *Veillonellaceae*, and *Neisseriaceae* (Figure 3A). A number of different families was observed in nasal cluster 3, including *Corynebacteriaceae*, *Moraxellaceae*, *Actinomycetaceae*, and *Neisseriaceae* (Figure S2A). ASV 516 (*Pseudomonas* sp.) occurred in both the gut, and the nasal microbial cluster 3 (Figures 2A and S2A). The CCpnA indicated that patients whose samples were affiliated with gut cluster 3 were mainly Non-Hodgkin lymphoma patients who never reached complete remission and had high mortality (Figure 2B). In contrast, in the oral and nasal cavities, we observed that multiple myeloma patients treated with high-dose melphalan were grouped into cluster 3 who did achieve complete remission (Figures 3B and S2B). At all three body sites, the associated ASVs stemmed from late time points (month +1 and thereafter).


A)
Micrococcaceae
Actinomycetaceae
Lactobacillaceae
Burkholderiaceae
Prevotellaceae
Erysipelotrichaceae
Neisseriaceae
Pasteurellaceae
Flavobacteriaceae
Streptococcaceae
Veillonellaceae
Leptotrichiaceae
Moraxellaceae
Corynebacteriaceae
Fusobacteriaceae
other
Cluster 1
Cluster 2
Cluster 3
Pearson's correlation
−0.83 −0.42 0 0.42 0.83
P_Phosphate_day0
P_bilirubin
P_Phosphate_pre_HSCT
Tregs_m1
P_Glukose_pre_HSCT
P_Glukose_w3
Th17_m1
P_Magnesium_day0
B_reticulocytes_day0
B_reticulocytes_pre_HSCT
Age
P_Carbamide
P_Calcium_ionfree_pH_7_4
B_reticulocytes_w2
P_Potassium
P_Magnesium_w1
P_Magnesium_pre_HSCT

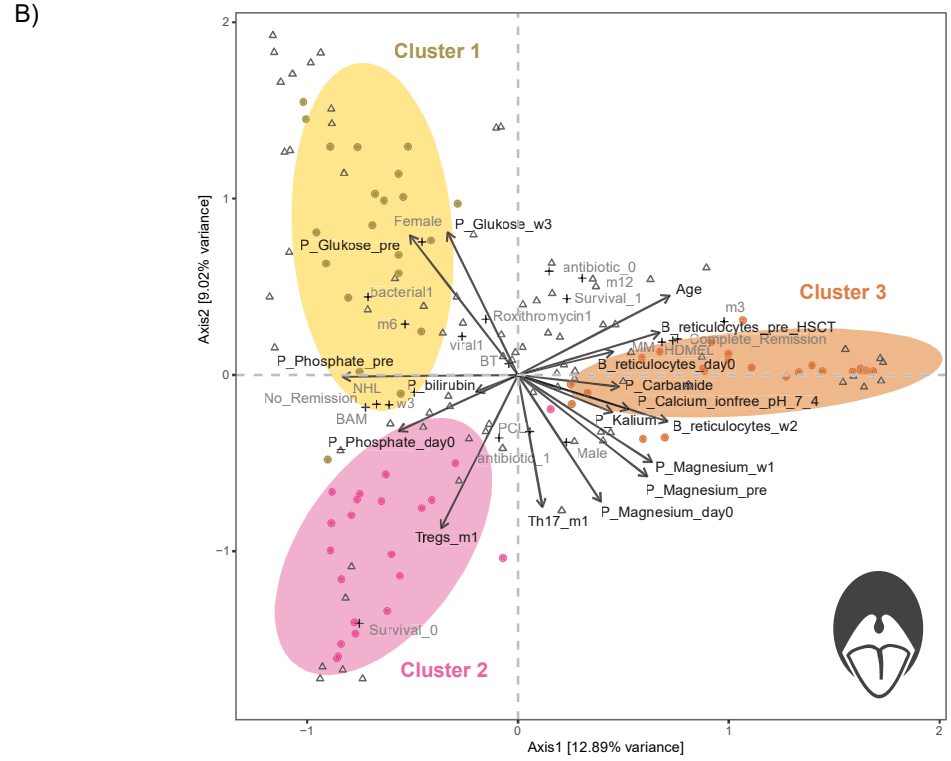

B)
Cluster 1
Cluster 2
Cluster 3
Female
P_Glukose_w3
P_Glukose_pre
antibiotic_0
m12
Survival_1
Age
m3
bacterial1
Roxithromycin1
m6
B_reticulocytes_pre_HSCT
viral1
Complete_Remission
HDMEC
P_Phosphate_pre
BT
B_reticulocytes_day0
P_bilirubin
NHL
P_Carbamide
No_Remission
w3
P_Calcium_ionfree_pH_7_4
BAM
P_Kalium
B_reticulocytes_w2
PCL
P_Phosphate_day0
antibiotic_1
Male
P_Magnesium_w1
P_Magnesium_pre
P_Magnesium_day0
Th17_m1
Tregs_m1
Survival_0
Axis2 [9.02% variance]
Axis1 [12.89% variance]

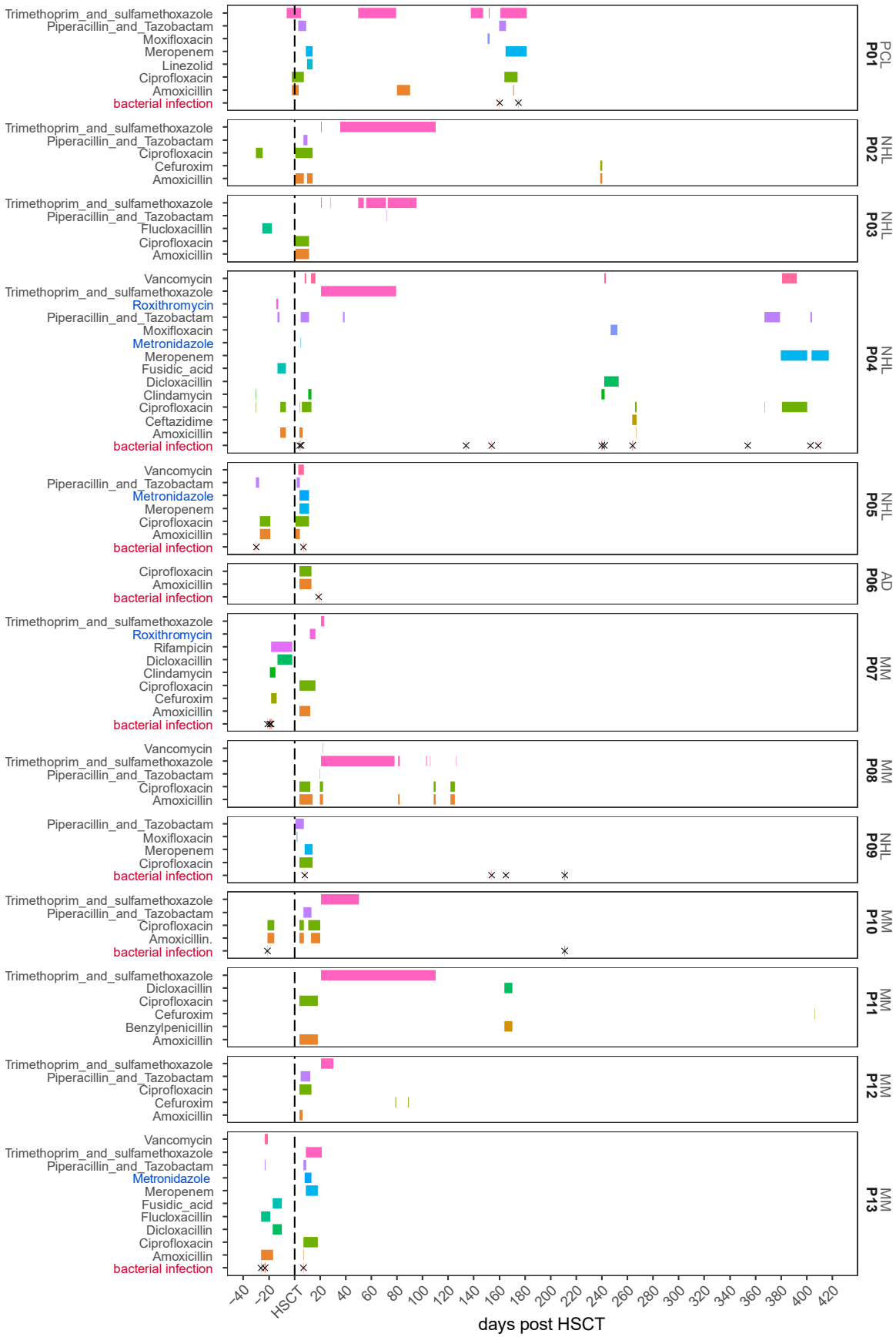
Figure 3. Multivariate host – oral microbial associations in autologous HSCT patients. A) Pairwise correlations (>0.3/<-0.3) of oral ASVs (bottom) and continuous immune and clinical parameters (right) displayed in a hierarchically CIM. Complete linkage clustering was based on the first three dimensions of the sPLS regression model. Red and blue coloring indicates positive and negative Pearson's correlation, respectively. B) Triplot of the CCpnA, showing associations between oral ASV abundances (circles) and continuous (arrows) and categorical (+) immune and clinical parameters. ASVs and variables were included in the CCpnA if they had at least one correlation >0.3/<-0.3 in the sPLS regression analysis. Variables and ASVs with a score >0.3/<-0.3 on at least one the first three CCpnA axes are displayed in dimension 1 versus 2. Samples are shown as triangles. Clusters of ASVs identified by the sPLS-based hierarchical clustering are shown as colored ellipses (with 80% confidence interval). Abbreviations are described in Figure 2. Additional abbreviations: B_, blood; BAM, cytarabine, carmustine, and melphalan treatment; BT, carmustine and thiotepa treatment; HDMEL, high-dose melphalan treatment; PCL, plasma cell leukemia.

Oral and nasal microbiota structure early after HSCT is associated with subsequent T reg and T_H17 cell reconstitution

Positive correlations were found between plasma phosphate levels at the time of HSCT, plasma bilirubin levels, as well as T reg and T_H17 cell counts in month +1 in comparatively younger patients compared to the rest of the cohort and ASVs in cluster 2 in the oral and the nasal cavities, respectively (Figures 3A and S2A). In addition, we observed negative correlations between the same ASVs and plasma glucose levels prior to HSCT at both body sites. Oral cluster 2 comprised mainly *Prevotellaceae,* and *Actinomycetaceae*, as well as a two *Fusobacteriaceae* ASVs (one of them *Fusobacterium nucleatum*) (Figure 3A). In the nasal cavity, we observed mostly *Corynebacteriaceae* in cluster 2 (Figure S2A). It was indicated by the CCpnA that these associations were mainly observed in patients with high mortality (Figure 3B and S2B). Moreover, the CCpnA of both, the oral and the nasal data set, indicated weak positive correlations between the ASVs in these clusters and frequent antibiotic treatment (Figure 3B and S2B). Samples grouped into cluster 2 of the oral and nasal site, respectively, were mainly collected between week +1 and week +3 (Figures 3B and S2B).

Pre- and early post-transplant abundances of distinct gut microbiota members co-vary with innate immune cell counts

A cluster of *Rikenellaceae* and *Tannerellaceae* ASVs in the gut (cluster 1) was associated with high plasma phosphate levels in week +1, high alpha diversity, and high counts of a set of innate immune cell subpopulations (Figure 2A). The CCpnA showed that high innate immune counts and high abundances of ASVs in this cluster were predominantly observed prior to HSCT, at the time of HSCT, and in week +1, and that few antibiotics were given to the concerned patients during that period (Figure 2B).


Trimethoprim_and_sulfamethoxazole
Piperacillin_and_Tazobactam
Moxifloxacin
Meropenem
Linezolid
Ciprofloxacin
Amoxicillin
bacterial infection
PCL
P01
Trimethoprim_and_sulfamethoxazole
Piperacillin_and_Tazobactam
Ciprofloxacin
Cefuroxim
Amoxicillin
NHL
P02
Trimethoprim_and_sulfamethoxazole
Piperacillin_and_Tazobactam
Flucloxacillin
Ciprofloxacin
Amoxicillin
NHL
P03
Vancomycin
Trimethoprim_and_sulfamethoxazole
Roxithromycin
Piperacillin_and_Tazobactam
Moxifloxacin
Metronidazole
Meropenem
Fusidic_acid
Dicloxacillin
Clindamycin
Ciprofloxacin
Ceftazidime
Amoxicillin
bacterial infection
NHL
P04
Vancomycin
Piperacillin_and_Tazobactam
Metronidazole
Meropenem
Ciprofloxacin
Amoxicillin
bacterial infection
NHL
P05
Ciprofloxacin
Amoxicillin
bacterial infection
AD
P06
Trimethoprim_and_sulfamethoxazole
Roxithromycin
Rifampicin
Dicloxacillin
Clindamycin
Ciprofloxacin
Cefuroxim
Amoxicillin
bacterial infection
MM
P07
Vancomycin
Trimethoprim_and_sulfamethoxazole
Piperacillin_and_Tazobactam
Ciprofloxacin
Amoxicillin
MM
P08
Piperacillin_and_Tazobactam
Moxifloxacin
Meropenem
Ciprofloxacin
bacterial infection
NHL
P09
Trimethoprim_and_sulfamethoxazole
Piperacillin_and_Tazobactam
Ciprofloxacin
Amoxicillin.
bacterial infection
MM
P10
Trimethoprim_and_sulfamethoxazole
Dicloxacillin
Ciprofloxacin
Cefuroxim
Benzylpenicillin
Amoxicillin
MM
P11
Trimethoprim_and_sulfamethoxazole
Piperacillin_and_Tazobactam
Ciprofloxacin
Cefuroxim
Amoxicillin
MM
P12
Vancomycin
Trimethoprim_and_sulfamethoxazole
Piperacillin_and_Tazobactam
Metronidazole
Meropenem
Fusidic_acid
Flucloxacillin
Dicloxacillin
Ciprofloxacin
Amoxicillin
bacterial infection
MM
P13
-40
-20
HSCT
20
40
60
80
100
120
140
160
180
200
220
240
260
280
300
320
340
360
380
400
420
days post HSCT

345

Figure 4. Antibiotic treatment and bacterial infections per autologous HSCT patient. Antimicrobial agents and the corresponding duration of treatment per patient (right) is displayed over time in relation to HSCT. Diagnoses are noted for each patient (right). Metronidazole and roxithromycin are noted in blue (left). Dashed lines indicate the day of HSCT. Bacterial infections (noted in red) are depicted as black marks (x). Abbreviations: AD, amyloidosis; NHL, Non-Hodgkin lymphoma; MM, multiple myeloma; PCL, plasma cell leukemia.

Discussion

In the present study, we revealed associations between the microbiota and the human immune system in adult autologous HSCT patients. In comparison to allogeneic HSCT, the microbiota in the autologous HSCT setting has hitherto been insufficiently described. Autologous transplant recipients constitute an important reference to account for donor-related effects in allo-HSCT. Here, we provide for the first time comprehensive insight into gut, oral, and nasal microbial community dynamics prior to, at the time of, and until up to one year post auto-HSCT in relation to immune cell counts, immune markers, and clinical outcomes.

Bacterial alpha diversity was lowest in week +1 or +2 after auto-HSCT at all three body sites. In the gut, diversity increased again in week +2 after the minimum in week +1, and had recovered by month +1. This is in line with a recent report [15]. In addition, we showed here for the first time how diversity changes in the oral and nasal microbiota in autologous HSCT. Interestingly, the decrease in diversity in the oral cavity was delayed compared with the gut. Oral diversity in week +1 was even higher than pre-transplant levels, before a drop in diversity occurred in week +2. One potential explanation might involve the prophylactic amoxicillin treatment in week +1 applied to all patients. A previous study has identified a number of amoxicillin-resistent *Prevotella* and *Neisseria* spp. in dental plaque of children [16]. These species represent taxonomic families that in our study exhibited higher relative abundances in week +1 compared with the preceding and following time point. One could speculate that amoxicillin promotes growth of various resistent oral microbiota members, which could rescue bacterial diversity. However, other studies have also shown a decrease in oral diversity after amoxicillin treatment in healthy adults [17, 18]. Interestingly, in one study the decrease was more pronounced after 7 days compared with after 3 days post amoxicillin-gavage, which could point to a delayed effect in agreement with our finding [18]. Of note, we did not see any direct correlations between oral microbial abundances and amoxicillin treatment in our multivariate analyses. Recovery of oral diversity by month +1 was comparable to the gut microbial diversity trajectory. Oral diversity dropped again in month +3, but sampling coverage at late time points was sparser and diversity levels at those time points cannot currently be statistically evaluated. In the nasal cavity, diversity decreased in week +1, but to a lesser degree than in the gut, and increased again in week +2. Interestingly, nasal bacterial diversity had vastly exceeded pre-transplant levels by month +1. We observed a high proportion of *Moraxellaceae* in the nasal cavity at early time points with low diversity (until and including week +1). From week +2 onwards, *Moraxellaceae* abundances decreased substantially and concurrently diversity increased. This is in agreement with a previous study that described this inverse correlation in children with asthma [19]. Further investigations are required to elucidate any causal relations of changes in bacterial alpha diversity and in the abundance of distinct bacterial families over time in the context of auto-HSCT.

We have assessed host-microbial associations in the present cohort of auto-HSCT patients and found high plasma glucose levels prior to and early after HSCT associated with increased occurrence of bacterial infections. We additionally provide insight into the microbial community structure accompanying this association. In the gut, we observed dominance of *Bacteroides* spp. together with high glucose levels, which is in agreement with the previously described positive correlation of the abundance and growth rate of this genus and glucose levels in adult diabetes mellitus type 2 (T2DM) patients [20, 21]. To our knowledge, we show for the first time a dominance of *Bacteroides* in autologous HSCT patients in relation to high plasma glucose levels. However, it is of utmost importance to note that plasma glucose levels are highly dependent on patients' fasting status, which was not evaluated in connection with the clinical routine measurements of glucose levels in our study. Therefore, findings related to associations with plasma glucose levels cannot be considered conclusive here, but rather warrant further investigations. Generally, with regards to the relevance of plasma glucose levels, T2DM has been established as a condition that can predispose to multiple types of cancer [22]. Apart from organ-specific inflammatory complications and impaired cell signaling pathways, the general state of hyperglycemia, i.e. high levels of circulating glucose, plays a role in diabetic-related cancer risk [22]. Hyperglycemia has consecutively also been investigated in non-diabetic populations of cancer patients and HSCT recipients [23, 24]. High glucose levels have for instance been associated with shorter time in remission, reduced survival, and impaired chemotherapeutic effect e.g. in multiple myeloma patients [23, 25]. In agreement with our findings, previous studies suggested hyperglycemia before and during neutropenia in auto-HSCT as a risk factor for subsequent bacterial infection in adult patients [14, 24]. It should be noted however that patients undergoing HSCT are being treated with prophylactic antibiotics to protect from infections during immunosuppression, and receive further treatment upon the occurrence of infections. Here, our analysis suggested that auto-HSCT patients with high abundances of *Bacteroides* spp. and a high number of infections were treated comparatively often with roxithromycin, an antibiotic of the macrolide class. Therefore, the *Bacteroides* dominance at late time points associated with high glucose levels might also have been reinforced by roxithromycin treatment, which has previously been suggested to increase the proportion of cecal *Bacteroidetes* [26].

In contrast to ASVs associated with high glucose levels, we observed another group of ASVs that was positively correlated with low pre-transplant glucose levels, possibly pointing to positive clinical outcomes [23]. Furthermore, we observed low phosphate levels prior to HSCT in this group. Hypophosphatemia during engraftment post auto-HSCT has previously been reported as a marker of neutrophil recovery since low plasma levels reflect the increased uptake of phosphate by replicating neutrophils [13, 27]. However, plasma phosphate levels might fluctuate drastically in severely ill cancer patients, which strongly limits the conclusive power concerning associations with plasma phosphate in the present study. High reticulocyte counts have been suggested to reflect erythrocyte regeneration and might thereby indirectly point to neutrophil recovery [27]. In agreement, we found high reticulocyte counts prior to HSCT and in week +2 to be associated with low plasma phosphate levels in the present study. Moreover, we related the microbiota to these immune markers and found predominantly *Ruminococcacea* and *Lachnospiraceae* ASVs in the gut. Consistent with the associated immune markers pointing to positive clinical outcomes after HSCT, these taxonomic families are generally associated with health [5, 28]. Of note, neutrophil counts were only weakly positively correlated with two specific

Lachnospiraceae ASVs, predominantly in month +1 and thereafter as indicated by our multivariate analysis.

The low glucose and phosphate levels, as well as high reticulocytes were also associated with distinct groups of oral and nasal ASVs. Interestingly, we observed these host-microbial associations in two different patient groups in the gut compared with the oral and nasal cavities. Our multivariate analysis suggested that these immune markers were correlated with the gut microbiota of Non-Hodgkin lymphoma (NHL) patients, whereas they were correlated with the oral and nasal microbiota of multiple myeloma patients. Of note, the NHL group was associated with adverse outcomes (no remission, death), while the MM group was linked to complete remission. This highlights the importance of accounting for the primary diagnosis of autologous HSCT recipients when attempting to correlate host-microbial associations and clinical outcomes. The adverse outcomes in NHL patients associated with immune markers pointing to rapid neutrophil recovery, and health associated taxonomic families seems counterintuitive. However, taking microbial abundance patterns in the gut, and the oral and nasal cavities into account, gave us more differentiated insights. In addition to *Ruminococcacea* and *Lachnospiraceae*, the concerned gut cluster was the only one that comprised two highly abundant *Pseudomonas* spp. ASVs. One of those ASVs was shared between the gut and the nasal cavity. A previous study has shown that a decrease in intestinal phosphate facilitates virulence activation in, and biofilm formation of *Pseudomonas aeruginosa* [29]. One could speculate that the association between low plasma phosphate levels and high abundances of specific *Pseudomonas* spp. that we detected in the gut might reflect this previously described effect systemically. We have not observed any infections caused by *Pseudomonas* spp. in the present cohort. However, it might be worth investigating in a larger cohort whether plasma phosphate levels play a role in the switch to opportunistic pathogenicity in *Pseudomonas* spp. or other gut microbiota members. In both, the oral and nasal cavities we observed a number of *Actinomyces* spp. in the respective clusters associated with low glucose and phosphate levels, as well as high reticulocyte counts. It has been shown that members of this genus are involved in oral conditions, such as periodontitis, which can also occur after HSCT [30–32]. In this context, it has been suggested that their ability to take up and intracellularly accumulate phosphate might facilitate biofilm formation [31]. We recommend further investigations including oral health assessment to elucidate whether this might be a potential explanation for high *Actinomyces* abundances associated with low phosphate levels in auto-HSCT. In summary, despite the previously described relation of hypophosphatemia and positive outcomes, we highlight the importance of carefully accounting for host-microbial associations when evaluating the potential of these kinds of markers in HSCT.

As opposed to the associations related to low pre-transplant phosphate levels, we observed an association between high plasma phosphate levels at the time of HSCT, and distinct clusters of oral and nasal ASVs. It has previously been suggested that high phosphate levels might indicate impaired neutrophil recovery [13]. In addition, we showed a weak positive correlation with plasma bilirubin levels, pointing to impaired liver function [33]. The respective patients had a higher mortality, but only two patients died during the monitored period, which limits the conclusive power concerning mortality at present. However, the strongest correlation between the oral and nasal microbiota in the respective clusters was with T_{reg} cell counts, and to a lesser degree with T_H17 cell counts in month +1. A major part

of the ASVs from the oral cavity represented in this cluster were affiliated with *Prevotella* spp., including e.g. *P. melaninogenica* and *P. nigrescens*. A role in a number of oral conditions, such as periodontitis, but also in e.g. systemic inflammation, has been attributed to *Prevotella* previously [12]. Interestingly, it has been suggested that *Prevotella* promotes mucosal and systemic inflammation by mediating T_H17 cell induction, in line with our findings of high *Prevotella* abundances and high T_H17 cell counts [12]. Especially *P. nigrescens*, which we also identified here, has previously been shown to facilitate T_H17 cell activity *in vitro* via pro-inflammatory cytokine induction [12]. The concurrent high counts of T reg cells that we observed in the present study might be a compensatory reaction to re-establish immune homeostasis. Furthermore, we observed high abundances of two oral *Fusobacteriaceae* ASVs in this cluster, including one *Fusobacterium nucleatum* ASV. High oral abundances of *F. nucleatum* have been observed previously in HSCT patients with oral mucositis [34]. Therefore, high abundances of this species associated with a possible inflammatory state and adverse outcomes in the present study fits into the framework of previous knowledge. Moreover, our analysis indicated that patients associated with these *Fusobacterium* spp. were treated more often with antibiotics than the rest of the cohort. A recent study found that intestinal *F. nucleatum* expanded after meropenem-gentamicin-vancomycin combination therapy in healthy adults [35]. This effect might be mirrored here in the oral cavity, in line with our finding of associated increased antibiotic use and high *Fusobacterium* abundances. However, further investigations are required to confirm these indications derived from our small cohort.

In addition to host-microbial associations shared between the oral and nasal cavities or all three examined body sites, we detected a number of immune cell subpopulations that were exclusively associated with members of the gut microbiota. We observed that high bacterial alpha diversity was associated with gavage of comparatively few antibiotics in this cluster, a link that is in line with previous findings in HSCT patients [36]. In particular, high abundances of *Parabacteroides* spp. (*Tannerellaceae*) and *Alistipes* spp. (*Rikenellaceae*) prior to, at the time of, and in week +1 after HSCT were observed. A depletion of the *Parabacteroides* genus during post-transplant reduction of alpha diversity has previously been described, in agreement with our findings [37]. Here, we observed high neutrophil, monocyte, and total leukocyte counts concurrent to high abundances of these taxa. *Parabacteroides* spp. produce the short chain fatty acid propionate [37]. The previously described anti-inflammatory and immune regulatory signaling, as well as chemotaxis exerted on neutrophils by propionate via the G-protein-coupled receptor GPR43 might play a role in the correlation we found [38–40].

Conclusions

We show that the microbiota plays a role in autologous HSCT, and encourage further research concerning this patient group. We tracked gut, oral, and nasal microbial dynamics over a year and performed a comprehensive multivariate assessment of microbial abundances in relation to host immune parameters. Our initial findings based on this cohort of 13 adult autologous patients suggest that bacterial alpha diversity decreases after auto-HSCT, and then rapidly recovers. We show that shifts in the community structure are evident early after transplantation, involving an increase in *Streptococcaceae* at all three body sites. In the gut, *Bacteroidaceae* abundances overall decreased after HSCT. However, *Bacteroides* spp. abundances were high in patients with a high number of infections and

antibiotic treatments. This calls attention to the impact antimicrobial treatment can have on the commensal microbiota, and potentially adverse clinical outcomes after auto-HSCT. With regards to the oral cavity, our findings point to an association between *Prevotella* spp. and T_H17 and T_{reg} cell counts. The potential role of this taxon in inflammatory regulation in relation to auto-HSCT might be of worth further investigations. Furthermore, we propose to further examine the dependencies between plasma glucose and phosphate levels and the microbiota in auto-HSCT patients, which might reveal host-microbial interactions relevant for clinical outcomes.

Materials and Methods

Patient recruitment and sample collection

Thirteen adult recipients (age range: 28.7-70.4) of autologous hematopoietic stem cell transplantation were recruited at Copenhagen University Hospital Rigshospitalet (Denmark) between July and November 2016. Patient characteristics are presented in Table S1. Multiple myeloma patients started a conditioning regimen from day -2, and Non-Hodgkin lymphoma patients from day -7 (Table S1). We defined the following sampling time points: pre-HSCT (collected between day -30 and day -2), at the time of HSCT, preferably before graft infusion (collected between day -2 and day +2), and weekly during the first 3 weeks after transplantation (week +1: day +3 to day +10, week +2: day +11 to day +17, week +3: day +18 to day +24). Follow-up time points were divided into broader intervals: Month +1 (between days +25 and +45), month +3 (between days +46 and +120), month +6 (between days +121 and +245), and month +12 (between day +246 and +404). We included bacterial, fungal, viral, and parasitic infection records and antibiotic treatment data from prior to transplantation to month +12 (day -30 to +404).

Analysis of immune cell subpopulations

Leukocytes, thrombocytes, and erythrocyte counts were recorded daily during hospitalization from before transplantation, and later weekly in the outpatient clinic by flow cytometry (Sysmex XN) or in case of very low counts by microscopy (CellaVision DM96 microscope). The following subpopulations were monitored: Monocytes, neutrophils, basophils, eosinophils, blast cells, plasma cells, erythroblasts, erythrocytes, reticulocytes, myelocytes, metamyelocytes, and promyelocytes.

Quantification of T_H17 cells and T regulatory cells

We used flow cytometry to determine counts of T cell subtypes, in particular T_H17 cells and T regulatory (T reg) cells, from from frozen peripheral blood mononuclear cells (PBMCs) on a FACS Fortessa III flow cytometer (Becton Dickinson, Albertslund, Denmark) in months +1, +3, and +6. The thawed and washed PBMCs were incubated for 30 minutes on ice with Fixable viability stain 620 (Becton Dickinson) and a set of conjugated monoclonal antibodies: CD3-APC-A750 (Beckmann Coulter), CD4-PE-Cy7 (Beckmann Coulter), CD8-A700 (Becton Dickinson), CD25-PE (Becton Dickinson), CD39-PerCP-Cy5.5 (Beckmann Coulter), CD196-BV510 (Biolegend, San Diego, USA), CD127-BV711 (Biolegend), CD161-BV650 (Becton Dickinson) and CD45RA-BV786 (Becton Dickinson). Subsequently, PBMCs were washed and incubated for 45 minutes on ice with transcription factor buffer set (Becton Dickinson). Next, PBMCs were washed twice and incubated for 45 minutes on ice with intracellular monoclonal antibodies: RORγT-A488 (BD),

553 FOXp3-A647 (BD) and Helios-PB (Beckmann Coulter). T_H17 cells were determined according to the CD4+
554 RORγT+ phenotype, and T_{reg} cells according to the CD4+CD25^{high}FOXp3+ phenotype. We multiplied the
555 frequency of T_H17 and T reg cells with the concurrent total lymphocytes count to obtain absolute cell
556 counts in x10^9/L.
557
558 **Quantification of immune markers and routine clinical assessment tests**
559 Immune marker measurements and routine clinical assessment tests were performed at the
560 Department of Clinical Biochemistry, Copenhagen University Hospital Rigshospitalet, Denmark. Plasma
561 levels of immunoglobulins A, and M in g/L were measured by immunoturbidimetry. As markers of
562 inflammation, we determined C-reactive protein (CRP) in mg/L by latex immunoturbidimetric assays
563 (LIA). We measured the following liver function markers in plasma: alanine transaminase (ALT) and basic
564 phosphatase in U/L were measured by absorption spectroscopy. Bilirubin levels in µmol/L were
565 determined by absorption spectroscopy. Furthermore, the following markers were monitored in plasma
566 (units and detection methods denoted in brackets): ion-free calcium (mmol/L, absorption spectroscopy),
567 potassium (mmol/L, ion selective electrodes measurements), lactate dehydrogenase (U/L, absorbtion
568 photometry), magnesium (mmol/L, absorption photometry), sodium (mmol/L, ion selective electrodes
569 measurements), glucose (mmol/L, absorption photometry), and phosphate (mmol/L, absorption
570 spectroscopy), carbamide (mmol/L, absorption spectroscopy).
571
572 **DNA isolation from fecal, oral, and nasal samples and 16S rRNA gene sequencing**
573 To analyze the gut microbiota, we collected 60 fecal samples from 11 patients at the 9 time points
574 described above. DNA extraction from fecal samples was performed with the QIAamp Fast DNA Stool
575 Mini kit (Qiagen, Venlo, Netherlands), following the manufacturer's instructions with modifications
576 according to [41]. Each extraction comprised a blank control. In total, 14 blank controls were sequenced.
577 In addition, two collection tube controls in total and one mock community sample (Biodefense and
578 Emerging Infectious Research (BEI) Resources of the American Type Culture Collection (ATCC)
579 (Manassas, VA, USA), Catalog No. HM-276D) was included on each sequencing run.
580 Eighty-five buccal and anterior naris swabs were collected each. The QIAamp UCP Pathogen Mini kit
581 (Qiagen, Venlo, Netherlands) was used to isolate DNA from the swab samples, and from one blank
582 control per extraction round (therof sequenced: 28), one mock community sample per run, two
583 collection tube controls, and two sampling swab controls. With the extraction kit, we used the 'Protocol:
584 Pretreatment of Microbial DNA from Eye, Nasal, Pharyngeal, or other Swabs (Protocol without Pre-lysis)'
585 and the 'Protocol: Sample Prep (Spin Protocol)' according to the manufacturer's instructions with
586 modifications. Instead of 500µl Buffer ATL, we used 550µl during pretreatment. Instead of the tubes
587 provided with the kits, DNA was eluted twice with 20µl Buffer AVE into 1.5 ml DNA LoBind tubes
588 (Eppendorf, Hamburg, Germany).
589 Library construction and sequencing, using an Illumina MiSeq instrument (Illumina Inc., San Diego, CA,
590 USA) was performed at the Multi Assay Core facility (DMAC), Technical University of Denmark. A
591 NanoDrop spectrophotometer (Thermo Scientific, Waltham, MA, USA) was used for DNA concentration
592 measurement of each sample. For library construction, we followed the *16S Metagenomic Sequencing*
593 *Library Preparation* protocol by Illumina [42]: The following previously evaluated primers were used to
594 amplify the V3-V4 region of the 16S ribosomal RNA gene in each sample and in the controls in a PCR

[43]: 341F (5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG-3') and 805R (5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC-3'). Quantity and quality of amplicons was assessed in an Agilent 2100 Bioanalyzer with the use of an Agilent RNA 1000 Nano Kit (Agilent Technology, Santa Clara, CA, USA). Then the amplicons were purified on AMPure XP Beads (Beckman Culter, Copenhagen, Denmark) following manufacturer's instructions. In a PCR step, Illumina adapters and dual-index barcodes from the 96 sample Nextera XT Index Kit (Illumina, FC-131–1002) were added to the amplicon target according to Illumina [42]. In another PCR step, AMPure XP Beads (Beckman Culter, Copenhagen, Denmark) were used according to manufacturer's instructions for a final library clean-up. Subsequently, the target size was confirmed in an Agilent 2100 Bioanalyzer (Agilent Technologies). DNA concentration was determined with a Qubit (Life Technologies, Carlsbad, CA, USA) and libraries were pooled and denatured with NaOH, diluted with hybridization buffer, and subsequently heat denatured prior to sequencing. As an internal control for low-diversity libraries, 5% PhiX was included. We performed paired-end sequencing with 2 × 300bp reads using MiSeq v3 reagent kit on an Illumina MiSeq instrument (Illumina Inc., San Diego, CA, USA).

16S rRNA gene sequence pre-processing

For each sample, demultiplexing of raw sequence reads was done based on sample-specific barcodes and 'read 1' and 'read 2' FASTQ files were created using the MiSeq reporter software on a Illumina MiSeq instrument. The cutadapt software (version 1.16) [44] was used for primer removal. We used a tolerated maximum error rate of 15% for matching the primer sequence anchored in the beginning of each read. A read pair was discarded if at least one read of the pair did not contain the primer. Subsequently, ASVs were inferred from the reads by using the R package DADA2 (version 1.8) [45]. Truncation at 280 bp and 200 bp was performed for forward and reverse reads, respectively, to retain a quality score >25 according to MultiQC analysis [46]. This way, an overlap of 480 bp (expected amplicon length of 460 bp + 20 bp) also allowed to merge forward and reverse reads. For the sample inference step (*dada*() function), samples were pooled, otherwise default settings were used for all other quality filtration steps in DADA2. In case of a read count <10,000 after preliminary chimera and contaminant removal, DNA from the concerned sample was re-sequenced. One feces samples with a read count <5,000 was re-extracted. Eventually, sample-wise chimera identification and removal from the whole data set based on a consensus decision (*removeBimeraDenovo*() function, method "consensus") was done. Subsequently, the ASVs were taxonomically assigned based on the Silva reference data base (version 132), formatted for DADA2 [47]. The Silva species-assignment training data base, formatted for DADA2 [47] was used for additional species assignment.

Analysis was continued by integrating the ASV table and associated taxonomy table with the R package phyloseq and its dependencies (version 1.24.0) [48]. In the next step, contaminant removal was performed with the R package decontam [49] for fecal and swab samples separately. Possible technical batch effects were evaluated by ordination (Principal Coordinates Analysis (PCoA)). Sequencing run, 96-well plate, extraction kit, extraction round, experimenter, and extraction date were considered for possible batch effects. Contaminants were removed by sequencing run as a batch effect based on a consensus probability.

We identified contaminants based on both, ASV prevalence in the blank controls, and the relation of ASV frequency in the samples and their post-PCR DNA concentration (method "both", frequency

threshold: 0.2, prevalence threshold: 0.075) [49]. Thereby, and through additional manual evaluation of edge cases, we identified and removed 89 ASVs from the fecal data set as contaminants. Another 7 ASVs were identified as contaminants from the sampling tube controls by the same method and thresholds as stated above.

The same method of contaminant removal was applied for the swab sample data set with different thresholds: Method "both", frequency threshold: 0.1, prevalence threshold: 0.6 [49]. In total, 1137 ASVs were removed from the swab sample data set. Sixteen additional contaminants were identified and removed based on the sampling tube controls and swab controls (method "both", frequency threshold: 0.075, prevalence threshold: 0.5).

After contaminant removal, we took subsets of the ASVs separately at each body site the function *kOverA*() from R package genefilter [50]. We retained 1,030 ASVs with ≥5 reads in ≥2 samples in the fecal data set. From the oral data set, 497 ASVs were retained, and 506 ASVs from the nasal data set. In both cases, these ASVs had ≥5 reads in ≥5 samples. We applied additional manual filtering to the oral and nasal core data sets. A threshold of ≥10 reads in ≥10 samples was applied to retain ASVs of taxonomic families commonly found in both the oral or nasal cavity and the gut. We removed ASVs affiliated with families that are only expected in the gut from the oral and nasal sample sets after manually assessing their abundances. This way, the final oral core set comprised 372 ASVs, and the final nasal core set consisted of 219 ASVs.

Statistical analysis

We performed statistical analyses and generation of graphs in R (version 3.5.1, R Foundation for Statistical Computing, Vienna, Austria) [51]. Experimental and clinical data were integrated with the ASV count data and taxonomy tables in phyloseq [48]. Plots were generated with the packages ggplot2 [52] and mixOmics [53].

We calculated bacterial alpha diversity (inverse Simpson index) from the core ASV sets at each body side. The median inverse Simpson index was compared between time points within each body site. Due to sparse sampling at late follow-up time points, we only compared alpha diversity between time points up to month +1. For this purpose, we used a Friedman test with Benjamini-Hochberg correction for multiple testing, and a post-hoc Conover test.

Subsequently, we agglomerated ASV counts on taxonomic family level with the function *tax_glom()* in phyloseq [48]. For visualizing microbial abundances over time, we displayed the proportions of the 12 most abundant families at each body site for each time point.

In preparation of multivariate analyses, we variance stabilized the ASV count data for each body site using the function *varianceStabilizingTransformation*() in the package DESeq2 [54], based on size factors estimated within the package GMPR [55].

We imputed left- and right-censored measurements (below or above detection level) of immune cell counts and immune markers by using the R package icenReg [56]. First, for each variable containing censored records, we fitted the non-parametric maximum likelihood estimator (NPMLE) for univariate interval censored data by using the function *ic_np*(). Subsequently, we imputed censored records based on the fitted models, using the *imputeCens()* function. For variable that were measure more frequently compared with the characterization of the microbiota, we calculated medians for the time points defined above to obtain comparable data sets. Continuous cell count and immune marker data was split

by time points in case that data was systematically missing for time points. We imputed missing values in continuous cell count and immune marker data in case of ≤ 50% missingness. To do so, we used the R package missForest [57] to achieve simultaneous multivariate non-parametric imputation. In case of >50% missingness, variables were excluded from the analysis.

Subsequently, for each body site, we calculated Manhattan distance matrices of the variance-stabilized community data. We then peformed permutational multivariate analysis of variance using distance matrices (adonis) in the package vegan [58] to examine associations between patients' baseline parameters, immune cell counts, immune markers, and clinical outcomes, and the overall microbial community structure. We accounted for repeated measurement and for keeping the chronological order of those within a patients when setting up the permutation design. The adonis results were used to estimate the number of variables to include in subsequent multivariate multi-table approaches.

In the next step, we employed two multivariate multi-table approaches to assess in detail whether there are association between the fecal, oral, and nasal microbiota and immune cell counts, immune markers, and clinical outcomes in HSCT. Remission status (complete remission, never in remission) and overall survival were included as clinical outcomes. We also included bacterial alpha diversity (inversed Simpson index), antibiotic treatment, infections, and patients' baseline parameters (age, sex, primary disease, conditioning regimen, transplantation number).

We first performed sparse partial least squares regression for the feces, oral, and nasal data set separately by using the function *spls*() in the package mixOmics [53]. We modelled variance stabilized ASV abundances as explanatory variables and continuous clinical and immune parameters as response variables. The number of response variables to be kept for each component (*keepY*) was chosen based on the number of variables that were trending or significant in the adonis analyses. A range of values between 20 and 40 was tested for the number of explanatory variables to be kept on each component (*keepX*) and showed results robust to *keepX*. We set *keepX* to 25 in the sPLS for each body site. We made an informed choice of 3 relevant components by implementing the *perf()* function. Subsequently, we used the *cim()* function (clustering method "complete linkage" and the distance method "Pearson's correlation") for hierarchical clustering based on the sPLS regression models for each body site. Thereby, matrices of coefficients indicating correlations between ASV abundances and continuous clinical and immune parameters were obtained.

Next, we performed canonical correspondence analysis, which allows the inclusion of both categorical and continuous clinical and immune parameters and the assessment of their bidirectional relations to log+1 transformed ASV abundances. We included continuous variables and ASVs that exhibited a correlation of >0.3/<-0.3 in the sPLS analysis for the respective body sites into the CCpnA, and in addition categorical variables were included. We used the the *cca()* function in package vegan [58]. We visualized the CCpnA results as a triplot. Plot dimensions were set to correspond in length to the percentage of variance explained by each axis. We used the function *anova.cca()* in vegan [58] to test in ANOVAs whether the CCpnA models were overall significant, as well as whether the dimensions, and the variables were significant, respectively.

List of abbreviations

ALT: Alanine transaminase
ASV: Amplicon sequence variant
CCpnA: Canonical correspondence analysis
CRP: C-reactive protein
HL: Hodgkin lymphoma
HSCT: Hematopoietic stem cell transplantation
LIA: Latex immunoturbidimetric assay
MM: Multiple myeloma
NHL: Non-Hodgkin lymphomas
PCoA: Principal Coordinates Analysis
sPLS: Sparse partial least squares analysis
T2DM: Diabetes mellitus type 2
T_H17 cell: T helper 17 cell
T_{reg} cell: T regulatory cell

Declarations

Acknowledgements

We would like to thank the patients for participating in this study. We are grateful for the help of Jette Sønderskov Gørløv (Department of Hematology, Rigshospitalet, University Hospital of Copenhagen) with clinical data registration and logistics of sample collection. We also thank Marlene Danner Dalgaard and Neslihan Bicen (Multi Assay Core facility (DMAC), Technical University of Denmark) for library construction and sequencing.

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 will be made available through the European Nucleotide Archive (ENA) at the European Bioinformatics Institute (EBI) at publication. The datasets generated and/or analysed during the current study as well as the R code used to analyze the data will be available from the figshare repository at publication.

Author's contributions

A.C.M., K.K., K.G.M., and S.J.P. designed the research; A.C.M., K.K., H.M., and S.J.P. performed the research; A.C.M., and S.J.P. contributed analytic tools; A.C.M., and S.J.P. analyzed the data; A.C.M. and S.J.P. wrote the manuscript; and K.K., F.M.A., and K.G.M. edited the manuscript.

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-7-2014-016) and the Danish Data Protection Agency.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Passweg JR, Baldomero H, Bader P, Basak GW, Bonini C, Duarte R, et al. Is the use of unrelated donor transplantation leveling off in Europe? The 2016 European Society for Blood and Marrow Transplant activity survey report. Bone Marrow Transplant. 2018;:1. doi:10.1038/s41409-018-0153-1.

2. Appelbaum FR. Hematopoietic-Cell Transplantation at 50. N Engl J Med. 2007;357:1472–5. doi:10.1056/NEJMp078166.

3. 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.

4. Eleutherakis-Papaiakovou E, Kostis E, Migkou M, Christoulas D, Terpos E, Gavriatopoulou M, et al. Prophylactic antibiotics for the prevention of neutropenic fever in patients undergoing autologous stem-cell transplantation: Results of a single institution, randomized phase 2 trial. Am J Hematol. 2010;85:863–7. doi:10.1002/ajh.21855.

5. Masche AC, Kielsen K, Cilieborg MS, Lund O, Holmes S, Aarestrup FM, et al. Specific gut microbiome members are associated with distinct immune markers in allogeneic hematopoietic stem cell transplantation. Under Rev. 2018.

6. Hoe E, Anderson J, Nathanielsz J, Toh ZQ, Marimla R, Balloch A, et al. The contrasting roles of Th17 immunity in human health and disease. Microbiol Immunol. 2017;61:49–56.

7. Arpaia N, Campbell C, Fan X, Dikiy S, Van Der Veeken J, Deroos P, et al. Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation. Nature. 2013;504:451–5. doi:10.1038/nature12726.

8. Atarashi K, Tanoue T, Oshima K, Suda W, Nagano Y, Nishikawa H, et al. Treg induction by a rationally selected mixture of Clostridia strains from the human microbiota. Nature. 2013;500:232–6. doi:10.1038/nature12331.

9. Atarashi K, Tanoue T, Ando M, Kamada N, Nagano Y, Narushima S, et al. Th17 Cell Induction by

799 Adhesion of Microbes to Intestinal Epithelial Cells. Cell. 2015;163:367–80.
800 doi:10.1016/j.cell.2015.08.058.

801 10. Pockley AG, Lindsay JO, Foulds GA, Rutella S, Gribben JG, Alexander T, et al. Immune Reconstitution
802 After Autologous Hematopoietic Stem Cell Transplantation in Crohn's Disease: Current Status and Future
803 Directions. A Review on Behalf of the EBMT Autoimmune Diseases Working Party and the Autologous
804 Stem Cell Transplantation In Refra. Front Immunol. 2018;9:646. doi:10.3389/fimmu.2018.00646.

805 11. Stringer AM, Logan RM. The role of oral flora in the development of chemotherapy-induced oral
806 mucositis. J Oral Pathol Med. 2015;44:81–7.

807 12. Larsen JM. The immune response to *Prevotella* bacteria in chronic inflammatory disease.
808 Immunology. 2017;151:363–74. doi:10.1111/imm.12760.

809 13. Raanani P, Levi I, Holzman F, Grotto I, Brok-Simoni F, Avigdor A, et al. Engraftment-associated
810 hypophosphatemia – the role of cytokine release and steep leukocyte rise post stem cell
811 transplantation. Bone Marrow Transplant. 2001;27:311–7. doi:10.1038/sj.bmt.1702761.

812 14. Derr RL, Hsiao VC, Saudek CD. Antecedent hyperglycemia is associated with an increased risk of
813 neutropenic infections during bone marrow transplantation. Diabetes Care. 2008;31:1972–7.
814 doi:10.2337/dc08-0574.

815 15. Khan N, Peled JU, Gomes AL, Devlin SM, Rondon-Clavo C, Clurman A, et al. Loss of Microbiota
816 Diversity after Autologous Stem Cell Transplant Is Comparable to Injury in Allogeneic Stem Cell
817 Transplant. Blood. 2018;132 Suppl 1:608–608. doi:10.1182/BLOOD-2018-99-110857.

818 16. Ready D, Lancaster H, Qureshi F, Bedi R, Mullany P, Wilson M. Effect of amoxicillin use on oral
819 microbiota in young children. Antimicrob Agents Chemother. 2004;48:2883–7.

820 17. Zaura E, Brandt BW, Teixeira de Mattos MJ, Buijs MJ, Caspers MPM, Rashid M-U, et al. Same
821 Exposure but Two Radically Different Responses to Antibiotics: Resilience of the Salivary Microbiome
822 versus Long-Term Microbial Shifts in Feces. MBio. 2015;6:e01693-15. doi:10.1128/mBio.01693-15.

823 18. Abeles SR, Jones MB, Santiago-Rodriguez TM, Ly M, Klitgord N, Yooseph S, et al. Microbial diversity
824 in individuals and their household contacts following typical antibiotic courses. Microbiome. 2016;4:1–
825 12. doi:10.1186/s40168-016-0187-9.

826 19. Depner M, Ege MJ, Cox MJ, Dwyer S, Walker AW, Birzele LT, et al. Bacterial microbiota of the upper
827 respiratory tract and childhood asthma. J Allergy Clin Immunol. 2017;139:826–834.e13.

828 20. Suez J, Shapiro H, Elinav E. Role of the microbiome in the normal and aberrant glycemic response.
829 Clin Nutr Exp. 2016;6:59–73. doi:10.1016/J.YCLNEX.2016.01.001.

830 21. Larsen N, Vogensen FK, van den Berg FWJ, Nielsen DS, Andreasen AS, Pedersen BK, et al. Gut
831 microbiota in human adults with type 2 diabetes differs from non-diabetic adults. PLoS One.
832 2010;5:e9085. doi:10.1371/journal.pone.0009085.

833 22. Tudzarova S, Osman MA. The double trouble of metabolic diseases: the diabetes–cancer link. Mol
834 Biol Cell. 2015;26:3129–39. doi:10.1091/mbc.e14-11-1550.

835 23. Healy SJ, Dungan KM. Hyperglycemia in Patients with Hematologic Malignancies. Curr Diab Rep.
836 2015;15:8. doi:10.1007/s11892-015-0581-x.

837 24. Hammer MJ, D'Eramo Melkus G, Knobf MT, Casper C, Fletcher J, Cleland CM. Glycemic Status and
838 Infection Risk in Nondiabetic Autologous Hematopoietic Cell Transplantation Recipients. Biol Res Nurs.
839 2016;18:344–50. doi:10.1177/1099800415619227.

840 25. Yu W, Cao D, Zhou H, Hu Y, Guo T. PGC-1α is responsible for survival of multiple myeloma cells under
841 hyperglycemia and chemotherapy. Oncol Rep. 2015;33:2086–92. doi:10.3892/or.2015.3809.

842 26. Zhang C, Li X, Liu L, Gao L, Ou S, Luo J, et al. Roxithromycin regulates intestinal microbiota and alters
843 colonic epithelial gene expression. Appl Microbiol Biotechnol. 2018;:9303–16.

844 27. Grazziutti ML, Dong L, Miceli MH, Cottler-Fox M, Krishna SG, Fassas A, et al. Recovery from
845 neutropenia can be predicted by the immature reticulocyte fraction several days before neutrophil
846 recovery in autologous stem cell transplant recipients. Bone Marrow Transplant. 2006;37:403–9.
847 doi:10.1038/sj.bmt.1705251.

848 28. Weber D, Oefner PJ, Hiergeist a., Koestler J, Gessner a., Weber M, et al. Low urinary indoxyl sulfate
849 levels early after ASCT reflect a disrupted microbiome and are associated with poor outcome. Blood.
850 2015;126:1723–9. doi:10.1182/blood-2015-04-638858.

851 29. Long J, Zaborina O, Holbrook C, Zaborin A, Alverdy J. Depletion of intestinal phosphate after

operative injury activates the virulence of P aeruginosa causing lethal gut-derived sepsis. Surgery. 2008;144:189–97.

30. Flynn KJ, Baxter NT, Schloss PD. Metabolic and Community Synergy of Oral Bacteria in Colorectal Cancer. mSphere. 2016;1. doi:10.1128/mSphere.00102-16.

31. Breiland AA, Flood BE, Nikrad J, Bakarich J, Husman M, Rhee T, et al. Polyphosphate-accumulating Bacteria: Potential Contributors to Mineral Dissolution in the Oral Cavity. Appl Environ Microbiol. 2018;84:e02440-17. doi:10.1128/AEM.02440-17.

32. Doss LM, Dandoy CE, Kramer K, Pate A, Flesch L, El-Bietar J, et al. Oral health and hematopoietic stem cell transplantation: A longitudinal evaluation of the first 28 days. Pediatr Blood Cancer. 2018;65.

33. 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.

34. Laheij AMGA, de Soet JJ, von dem Borne PA, Kuijper EJ, Kraneveld EA, van Loveren C, et al. Oral bacteria and yeasts in relationship to oral ulcerations in hematopoietic stem cell transplant recipients. Support Care Cancer. 2012;20:3231–40. doi:10.1007/s00520-012-1463-2.

35. Palleja A, Mikkelsen KH, Forslund SK, Kashani A, Allin KH, Nielsen T, et al. Recovery of gut microbiota of healthy adults following antibiotic exposure. Nat Microbiol. 2018;3 November. doi:10.1038/s41564-018-0257-9.

36. Taur Y, Coyte K, Schluter J, Robilotti E, Figueroa C, Gjonbalaj M, et al. Reconstitution of the gut microbiota of antibiotic-treated patients by autologous fecal microbiota transplant. Sci Transl Med. 2018;10:eaap9489. doi:10.1126/scitranslmed.aap9489.

37. Biagi E, Zama D, Nastasi C, Consolandi C, Fiori J, Rampelli S, et al. Gut microbiota trajectory in pediatric patients undergoing hematopoietic SCT. Bone Marrow Transplant. 2015;50:992–8. doi:10.1038/bmt.2015.16.

38. Maslowski KM, Vieira AT, Ng A, Kranich J, Sierro F, Yu D, et al. Regulation of inflammatory responses by gut microbiota and chemoattractant receptor GPR43. Nature. 2009;461:1282–6. doi:10.1038/nature08530.

879 39. Vinolo MAR, Rodrigues HG, Hatanaka E, Sato FT, Sampaio SC, Curi R. Suppressive effect of short-
880 chain fatty acids on production of proinflammatory mediators by neutrophils. J Nutr Biochem.
881 2011;22:849–55. doi:10.1016/j.jnutbio.2010.07.009.

882 40. Vinolo MAR, Ferguson GJ, Kulkarni S, Damoulakis G, Anderson K, Bohlooly-Y M, et al. SCFAs induce
883 mouse neutrophil chemotaxis through the GPR43 receptor. PLoS One. 2011;6:e21205.
884 doi:10.1371/journal.pone.0021205.

885 41. Knudsen BE, Bergmark L, Munk P, Lukjancenko O, Prieme A, Aarestrup FM, et al. Impact of Sample
886 Type and DNA Isolation Procedure on Genomic Inference of Microbiome Composition. bioRxiv.
887 2016;1:64394. doi:10.1101/064394.

888 42. 16S Metagenomic Sequencing Library Preparation.
889 https://support.illumina.com/content/dam/illumina-
890 support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-
891 guide-15044223-b.pdf. Accessed 17 Apr 2018.

892 43. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, et al. Evaluation of general 16S
893 ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies.
894 Nucleic Acids Res. 2013;41:e1–e1. doi:10.1093/nar/gks808.

895 44. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
896 EMBnet.journal. 2011;17:10. doi:10.14806/ej.17.1.200.

897 45. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution
898 sample inference from Illumina amplicon data. Nat Methods. 2016;13:581–3. doi:10.1038/nmeth.3869.

899 46. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools
900 and samples in a single report. Bioinformatics. 2016;32:3047–8. doi:10.1093/bioinformatics/btw354.

901 47. Callahan B. Silva taxonomic training data formatted for DADA2 (Silva version 132). 2018.
902 doi:10.5281/ZENODO.1172783.

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

905 49. Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. Simple statistical identification and
906 removal of contaminant sequences in marker-gene and metagenomics data. Microbiome. 2018;6:226.
907 doi:10.1186/s40168-018-0605-2.

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

910 51. R Core Team. R: A Language and Environment for Statistical Computing. 2018. https://www.r-
911 project.org/.

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

913 53. Rohart F, Gautier B, Singh A, Lê Cao K-A. mixOmics: An R package for 'omics feature selection and
914 multiple data integration. PLOS Comput Biol. 2017;13:e1005752. doi:10.1371/journal.pcbi.1005752.

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

917 55. Chen J, Zhang L. GMPR: Geometric Mean of Pairwise Ratios. R package version 0.1.3. 2017.

918 56. Anderson-Bergman C. **icenReg** : Regression Models for Interval Censored Data in *R*. J Stat Softw.
919 2017;81. doi:10.18637/jss.v081.i12.

920 57. Stekhoven DJ, Buhlmann P. MissForest--non-parametric missing value imputation for mixed-type
921 data. Bioinformatics. 2012;28:112–8. doi:10.1093/bioinformatics/btr597.

922 58. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan: Community
923 Ecology Package. R package version 2.5-2. 2018. https://cran.r-project.org/package=vegan.

924

5. Discussion

5.1 Microbiota reconstitution: Temporal dynamics of bacterial diversity and bacterial abundances in HSCT

In all studies included in this thesis, we assessed bacterial alpha diversity over time before, at the time of, and after HSCT. Common to both, allo- and auto-HSCT, as well as to the gut, oral, and nasal sites, was a minimum in diversity within the first three weeks post transplantation, demonstrated by us and others [**Masche I, II, and III**] [150, 151, 184]. The loss of diversity in the gut after autologous HSCT follows a pattern similar to allogeneic HSCT, with minimum diversity occurring within the first two weeks after transplantation [**Masche I, II, and III**] [184]. In adult autologous HSCT patients, gut bacterial diversity returned to levels comparable to before HSCT after one month [**Masche III**] [184], while recovery was first observed after three to six months post transplantation in pediatric allogeneic HSCT patients [**Masche II**] [152]. Besides differences in patient age, underlying primary diseases, and antibiotic regimen in the allogeneic and autologous cohorts, differences in conditioning regimen might also be of importance to explain the disparity in diversity recovery. None of the autologous patients in our third study underwent TBI as part of their conditioning regimen whereas 21% and 33% of allogeneic patients received TBI in our first and second study, respectively [**Masche I, II, and III**]. Studies have shown that radiation diminishes intestinal bacterial count in mice and reduces bacterial diversity in human cancer patients, and indicate that this is independent of chemotherapy [220, 221]. Therefore, one could speculate that the absence of TBI in auto-HSCT patients in our third study might benefit a rapid recovery of gut bacterial diversity. Furthermore, loss of diversity in allo-HSCT is most pronounced in patients with GvHD [154, 156]. Therefore, the occurrence of GvHD and potentially the accompanying immunosuppressive treatment might account in part for the later recovery in allo-HSCT compared with auto-HSCT. In addition, our findings encourage further research elucidating to which extent the mode of transplantation itself (allogeneic versus autologous) might be of relevance to alpha diversity recovery. With regards to understanding immune-microbial interactions on a larger scale, it has been demonstrated that a co-evolved host-specific (human versus mouse) microbiota is crucial for immune system maturation [5]. It could be of interest to further investigate whether host-specificity extents to the individual level to a certain degree. In this light, one could speculate that quicker recovery of diversity indicates a more rapid re-establishment of immune-microbial homeostasis when the host's own cells (auto-HSCT) versus foreign cells (allo-HSCT) are infused. With regards to other body sites, oral bacterial diversity seems to follow a similar temporal sequence of recovery as seen in the gut, with return to pre-transplant like levels after three months (allo-HSCT) and one month (auto-HSCT), respectively [**Masche II and III**]. Differences in nasal bacterial diversity changes in allo- versus auto-HSCT are suggestively less pronounced [**Masche II and III**]. Independent of transplantation mode, the diversity at this site recovered more rapidly compared with the gut, and the oral cavity, within two weeks post-transplant as observed in our second and third studies [**Masche II and III**].

In addition to alpha diversity as an approximate marker of microbial homeostasis, specific taxa have been described by us and others, which typically increase or decrease during the first month after allo-HSCT [**Masche I and II**] [150, 151, 154, 161, 163, 222]. Individual taxa and respective indications for their

expansion or depletion have been discussed and reviewed in **Masche I and II,** and in paragraph 2.5. While the majority of studies have reported changes in bacterial abundances early after allo-HSCT, few studies so far have investigated long-term dynamics exceeding the first month post-transplant [**Masche II**] [163]. Our second study provides insight into microbial abundance trajectories until up to twelve months post allo-HSCT [**Masche II**]. We utilized a novel phylogenetically-informed classification method to identify groups of closely related ASVs with temporal abundance trajectories best discriminating between time points [**Masche II**] [218]. We found that regardless of the directionality of abundance changes (referred to as 'smile'- and 'frown'-shaped trajectories in **Masche II**), reconstitution of pre-transplant like abundance levels consistently occurred starting from three months post HSCT [**Masche II**]. While we have discussed the abundance changes of relevant taxa with regards to differences in immune and clinical markers in our second study, the 'three months'-time point seems to exhibit potential as a consistent temporal threshold for microbiota reconstitution after allo-HSCT [**Masche II**]. In this context, it is interesting to observe the recent focus on fecal microbiota transplantation (FMT) as a therapeutic approach to re-establish microbial homeostasis after allo-HSCT [222–224]. Interestingly, one study demonstrated reconstitution of pre-transplant bacterial alpha diversity and community structure following autologous FMT at day +49 after allo-HSCT in adult patients [222]. The study states that an increase in gut bacterial diversity at this time point is too a lesser degree, and with a compromised functional repertoire of the microbiota, also observed in the control group (without FMT) [222]. In our second study, we illustrated recovery of diversity and abundances of distinct ASVs towards pre-transplant levels in children approximately two months later without FMT intervention ('three months' time point, median day + 92) [**Masche II**]. While FMT might accelerate microbiota reconstitution and restore commensal functionality in allo-HSCT, our study highlights the importance of carefully evaluating intervention-free abundance trajectories, and possibly provides candidates for microbial markers of reconstitution in high taxonomic resolution (i.e. ASVs) [**Masche II**]. Of note, this comparison needs to be interpreted with caution as temporal succession of microbiota recovery might differ in pediatric compared with adult patients.

5.2 Host-gut microbial associations in HSCT

5.2.1 The gut microbiota is associated with immune reconstitution after allogeneic HSCT

In the first two studies, we revealed associations between the gut microbial community structure and immune cell counts, immune markers, and clinical outcomes in children undergoing allogeneic HSCT [**Masche I and II**]. The study populations were comparable (pediatric patients with primarily malignant hematological diseases). However, the two studies differed to a certain extent in the immune cell types and immune markers that were examined. We observed correlations between counts of distinct immune cell types and gut microbial abundances in both studies [**Masche I and II**]. In particular, NK cells and B cells in month +1 and thereafter were correlated with high counts of obligate anaerobes, including *Ruminococcaceae* and *Lachnospiraceae* [**Masche I and II**]. In our first study, we showed that the patients with these characteristics had milder aGvHD and higher survival. Consistently, these positive outcomes are associated with microbial homeostasis and high bacterial alpha diversity, which in turn are linked to high abundances of *Ruminococcaceae* and *Lachnospiraceae* [153, 161]. We argued that the production of

SCFAs, in particular butyrate, by members of these bacterial families could limit inflammation and thereby prevent aGvHD to a certain extent, and that this might entail a direct or indirect effect on NK and B cell reconstitution [**Masche I**]. In agreement, a recent study found decreased fecal butyrate levels 14 days after allo-HSCT in patients with subsequent GvHD (grade I or higher) compared with patients without GvHD [171]. In our second study, we additionally examined immunoglobulin levels and found that high plasma IgA levels twelve months after HSCT was associated with rapid NK and B cell recovery, as well as high *Ruminococcaceae* and *Lachnospiraceae* abundances [**Masche II**]. This led us to suggest that butyrate might directly stimulate B cell differentiation into IgA-producing plasma cells, as previously shown in culture [63]. In line with this, a number of studies have attributed immune cell regulatory effects to butyrate, suggesting histone deacetylase inhibition as a universal mechanism [54, 172, 225]. Interestingly, histone deacetylase inhibition also has a mitigating effect on aGVHD after HSCT, in agreement with our observation of low aGvHD severity in the presence of butyrate-producers [**Masche I**] [226]. Vice versa, the absence of aGvHD and inflammation might also benefit the retention of low oxygen levels in the gut, favoring the growth of obligate anaerobe bacteria [57]. In our second study, the patients in question surprisingly had high mortality [**Masche II**]. However, only three patients (10% of the cohort) died during the monitored period of 21.4 months on average. In comparison, seven patients (19%) died during a longer follow-up period of 62.1 months on average in the first study, providing a slightly more balanced and potentially more reliable source [**Masche I**]. In addition, it has been demonstrated by us and others that antibiotic treatment has a pivotal impact on commensal community structure in HSCT patients, and especially on SCFA-producers, such as *Ruminococcaceae* and *Lachnospiraceae* [**Masche I and II**] [148, 160, 164, 171]. Before the implementation of NGS-based characterization of the microbiota in HSCT, the practice of unspecific antibiotic-induced 'gut decontamination' was suggested to mitigate GvHD risk [146]. However, recent research questions this relation and indicates that the gain of preventing infections by antimicrobial prophylaxis might be compromised by adverse clinical outcomes such as GvHD and TRM that accompany the loss of specific commensal taxa [**Masche I**] [148, 164, 171, 227].

5.2.2 Host-gut microbial associations in autologous HSCT – The relevance of obligate anaerobes across transplantation modes

In the third study of this thesis [**Masche III**], we described host-gut microbial associations in autologous HSCT recipients. As discussed in paragraph 3.1, the autologous cohort differed from the allogeneic group e.g. in age (adult auto-HSCT patients versus pediatric allo-HSCT patients), conditioning regimens, diagnoses, and use of immunosuppressive treatment. Nevertheless, the commonalities and disparities of allogeneic and autologous host-gut microbial associations will be discussed in this paragraph with attention to the underlying differences between the two cohorts.

Concerning the overall gut microbial composition, we observed a slight decrease in *Ruminococcaceae* and *Lachnospiraceae* in week +1 after auto-HSCT [**Masche III**] as well as after allo-HSCT [**Masche II**], which is consistent with what others have previously found [153, 167]. In allogeneic HSCT, it has been shown previously that this decrease is most pronounced in aGvHD patients [162, 164]. The findings of our first study [**Masche I**] were in line with this and potential explanations for the link of these bacterial families and aGvHD severity have been illustrated in paragraph 5.2.1. In this light, it is interesting to observe a likewise reduction of *Ruminococcaceae* and *Lachnospiraceae* in auto-HSCT patients, who have a much

lower GvHD incidence due to absence of immunological disparity between graft- and host-derived cells [**Masche III**] [228]. In the autologous cohort, we found high abundances of *Ruminococcaceae* and *Lachnospiraceae,* predominantly in month +1 and thereafter, to be associated with markers previously related to rapid neutrophil recovery [**Masche III**] [229, 230]. In particular, these were low levels of plasma phosphate, pointing to increased phosphate uptake by replicating neutrophils, and high reticulocyte counts [**Masche III**] [229, 230].

Furthermore, we observed an expansion of *Enterococcaceae* in week +1 after auto-HSCT, similar to what has been described for allo-HSCT by us and others [**Masche II**] [148, 154, 161]. It has been demonstrated that enterococcal expansion in adult patients is primarily driven by antimicrobial treatment, particularly by ciprofloaxacin and metronidazole gavage [148, 154]. Prophylactic antibiotic regimens preceded both, allo- and auto-HSCT in our studies [**Masche I, II, and III**]. Therefore, our findings indicate that the proportion of this taxonomic family increases after both, allogeneic and autologous HSCT, possibly independent of the mode of transplantation. Of note, antibiotic treatment regimens differed between the allogeneic and autologous cohorts, but no antimicrobial agent was directly associated with high abundances of *Enterococcaceae* in any of our studies [**Masche I, II, and III**].

5.3 Beyond the gut – The role of the oral and nasal microbiota in allogeneic and autologous HSCT

The gut is clearly the microbial niche most studied in HSCT patients [149]. Other body sites, such as the oral cavity have been described in few studies via culture- or microarray-based methods in adult allo-HSCT patients [181–183] and recently in pediatric allo- and adult auto-HSCT patients through NGS [**Masche II and III**]. In our second study, we were able to discriminate oral microbiota profiles at different time points prior to and post allo-HSCT by the abundances of distinct *Actinomycetaceae*, *Streptococcaceae*, and *Prevotellaceae* ASVs [**Masche II**]. The temporal trajectories of certain *Prevotellaceae* ASVs showed reduced abundances after allo-HSCT in pediatric patients [**Masche II**]. *Prevotella* spp. have also been described to decrease in adult patients after allo-HSCT [183]. The abundance of a certain *Prevotella* spp. ASVs prior to pediatric allo-HSCT and after twelve months was positively correlated with concurrent high plasma albumin levels, which indicates a good nutritional status [**Masche II**] [183]. However, high pre-transplant abundances of a different oral ASV affiliated with *Prevotella melaninogenica* predicted subsequent moderate to severe aGvHD [**Masche II**]. Increased abundance of *P. melaninogenica* has been reported in adult patients with Crohn's disease (CD), a chronic inflammatory disorder of the intestinal tract, compared with healthy controls [231]. Both, CD and GvHD exhibit impaired gut epithelial barrier function and a high inflammatory burden [99, 178]. A potential explanation for the involvement of *Prevotella* in inflammatory conditions is its ability to promote mucosal and systemic inflammation by stimulating T_H17 cells [**Masche II**] [232, 233]. However, this has been described mainly for *Prevotella nigrescens* [233]. Interestingly, high abundances of both, *P. melaninogenica* and *P. nigrescens* ASVs in the oral cavity were associated with high T_{reg} and T_H17 cell counts in adult autologous-HSCT patients [**Masche III**]. Moreover, this association was accompanied by high abundances of an ASV affiliated with *Fusobacterium nucleatum* [**Masche III**]. In adult allo-HSCT patients, this species was found in lesions caused by oral mucositis, which is a common oral complication in relation to HSCT [234]. One could

speculate that differences in microbiota – immune cell interactions in allogeneic versus autologous HSCT patients are to a certain extent dependent on the origin of the stem cells from which for instance T cells derived (donor- vs. self-derived). Whether the microbiota has a greater modulatory affinity towards recipient-derived immune cells will have to be mechanistically elucidated in future research. Taken together, these observations in distinct groups of patients varying in age and mode of transplantation indicate that *Prevotella* might play a universal role in regulating inflammation in HSCT [**Masche II and III**]. However, 16S rRNA gene profiling might not provide a taxonomic resolution high enough to determine which particular members of the *Prevotellaceae* family are responsible for pro-inflammatory signaling and which might point to microbial homeostasis, but suggest ambiguous influences potentially depending on the particular species, if not the strain.

Interestingly, host-microbial associations of the oral and nasal microbiota in both, allogeneic and autologous HSCT reflected host-gut microbial associations to a large extent. Associations were shared between body sites in the sense that the same sets of specific immune cell types or immune markers exhibited correlations with the microbiota at two or three sites. For instance, CD4+ T cell and T_H17 cell counts in pediatric allogeneic-HSCT patients were associated with a group of *Ruminococcaceae* and *Lachnospiraceae* in the gut, as well as with a group of *Veillonellaceae* in the nasal cavity, and a diverse group of taxa including *Flavobacteriaceae* in the oral cavity [**Masche II**]. This might indicated pro-inflammatory immunomodulatory involvement of the microbiota at multiple body sites as discussed in more detail in our second study [**Masche II**]. In adult autologous patients, we could similarly show shared associations. For instance, high reticulocyte counts and low plasma phosphate levels were associated with *Ruminococcaceae* and *Lachnospiraceae* in the gut (as discussed in paragraph 5.2.2), and with *Moraxellaceae* and *Corynebacteriaceae* in the nasal cavity, as well as e.g. *Prevotellaceae* in the oral cavity [**Masche III**]. This might serve as an example for microbiota in different niches and their shared correlations with immune markers in HSCT. However, in some cases these associations were strongly dependent on the patients' clinical baseline parameters. For instance, the described association in auto-HSCT patients exhibited the mentioned gut microbial pattern primarily in Non-Hodgkin lymphoma patients, while the nasal and oral microbial patterns were dominating in multiple myeloma patients [**Masche III**]. Taken together, we demonstrated similarities of host-microbiota associations between multiple body sites in both, allo- and auto-HSCT, in line with what has recently been shown in CD patients [**Masche II and III**] [235].

5.4 Prediction of aGvHD from pre-transplant *Lactobacillus* sp. and other gut, oral, and nasal microbial taxa

In the two studies comprising allogeneic HSCT patients, we have illustrated changes in the gut microbiota preceding and following aGvHD onset [**Masche I**], and identified gut microbiota members with predictive potential for subsequent aGvHD [**Masche II**]. In the first study, we observed high *Lactobacillaceae* abundances in patients with moderate to severe aGvHD (grade II-IV) [**Masche I**]. When looking in more detail into the temporal succession of this association, we found that *Lactobacillaceae* expanded after the onset of aGvHD in patients who later died due to relapse [**Masche I**]. In contrast, an increase prior to aGvHD onset was observed in patients who were still alive after on average five years [**Masche I**]. In line with this, we were able to show in **Masche II** that high abundances of a specific *Lactobacillus* sp. ASV before HSCT and up to the time of transplantation predicts the development of moderate to severe aGvHD. Because only 3 patients (10%) died during the average follow-up time of 21 months on average in the second study, it was not possible to evaluate the predictive potential for survival. Our findings in the first study indicate that *Lactobacillaceae* expansion might be part of a protective effect in reaction to microbial community disruption and inflammation [**Masche I**]. We argued that the concurrent high levels of the antimicrobial peptide hBD2 might point to both, a high burden of inflammation and a high number of opportunistic pathogen infections in the patients in question [**Masche I**]. It has been suggested previously that specific probiotic *Lactobacillus* spp. can induce the secretion of hBD2 in immune cells [236, 237]. Microbial community disruption after HSCT has been associated with an increase in *Enterococcaceae* in a number of studies [59, 154, 161]. It has been suggested that a concurrent increase in *Lactobacillaceae* limits *Enterococcaceae* expansion [161]. This could be interpreted as a protective reaction to facilitate the re-establishment of microbial homeostasis. Consistently, we have identified temporal trajectories of closely phylogenetically related *Lactobacillaceae* ASV with the use of the tree-based sparse linear discriminant analysis (LDA) in the second study, showing an increase immediately after HSCT that reflects the trajectory of *Enterococcaceae* [**Masche II**]. In summary, the immunomodulatory characteristics attributed to *Lactobacillaceae* and its abundance patterns revealed by us and others suggest this family as a potential marker of a state of inflammation and loss of microbial homeostasis prior to and at the time of allo-HSCT, which might predispose to aGvHD. Moreover, we have for the first time to our knowledge identified a specific ASV that allows us to predict aGvHD from preceding gut microbial abundances, which might serve as a starting point for further specific hypothesis-driven investigations [**Masche II**].

We have identified additional gut ASVs that might have a predictive value in aGvHD, including a *Parabacteroides* spp. ASV and an ASV affiliated with the *Lachnospiraceae* family [**Masche II**]. Furthermore, we demonstrated that pre-transplant abundances of oral and nasal microbiota members could predict subsequent aGvHD.

5.5 Strengths and limitations

The major strength of the work presented in this thesis constitutes the longitudinal study design. In contrast to describing the microbiota at single or few time points, repeated sampling allowed us to gain insight into temporal dynamics of diversity and microbial abundance changes. Another strength of this work is the characterization of the microbiota at non-intestinal sites in HSCT patients, expanding our view on microbial dynamics and host-microbial interactions that are hitherto best described for the gut microbiota. With regards to our analysis strategies, the use of multivariate multi-table approaches robust enough to tolerate a certain degree of collinearity and noise allowed us to integrate and analyze different heterogeneous data types (OTU/ASV count data, continuous and categorical metadata) jointly, and thereby draw a holistic picture of host-microbial interactions in HSCT patients. Moreover, choosing methods that induce sparsity constraints, such as tree-based sparse LDA and sparse partial least squares (sPLS) regression, allowed us to reduce dimensionality and thereby increase interpretability in relation to OTU/ASV count data. At the same time, these methods aim at minimizing the loss of information from the high-dimensional data.

However, the work presented here also has limitations and weaknesses. Despite a reasonable number of fecal and swab samples obtained by longitudinal collection, these samples originated from a limited number of patients, especially in the autologous cohort. Repeated measurements limit statistical power and warrant caution concerning statistical testing, which we accounted for by using appropriate approaches, such as Friedman tests. Another issue is the heterogeneity of the patient populations. HSCT is employed as a curative approach for various conditions, such as hematologic malignancies, severe non-malignant diseases, and primary immunodeficiencies. Conditioning regimens and transplantation modalities differ depending on the diagnosis, and so do risks for certain adverse outcomes after transplantation. Therefore, conclusions about microbial patterns common to all HSCT patients cannot be drawn without carefully taking into account these underlying differences as possible confounders. Even though we did so in our analyses and we have pointed out potential confounders in the presented studies, certain diagnoses and treatment regimens were only represented by few individual patients within our cohorts. Therefore, larger studies with higher patient numbers are necessary to account for heterogeneity in the patient populations more reliably and to validate our findings. Furthermore, it should be noted that our analyses approaches did not by any means serve the purpose or raise the claim to reveal causal host-microbial effects, but rather to assess whether there are associations or correlations between microbial abundances and host parameters or clinical outcomes. This encompasses also the prediction of aGVHD severity from pre-HSCT microbial abundances. However, our studies provide a number of starting points for targeted hypothesis-driven investigations of host-microbial interactions. Moreover, the limited availability of certain data has posed challenges for drawing conclusions. For instance, the investigation of immune cell reconstitution in the autologous cohort was limited by the lack of data on T, B, and NK cell counts. In the allogeneic cohorts, we did not have sufficient data on chronic GvHD at hand. Further research is therefore needed to examine the role of the microbiota in relation to chronic GvHD, especially with regards to B cell reconstitution.

6. Conclusions

Corresponding to the key questions of this PhD project stated in chapter 1, the following general conclusions can be drawn from the presented work:

1. After allogeneic and autologous HSCT, the diversity of the gut, oral, and nasal microbiota decreases. Specific bacterial taxa expand or decrease in abundance after HSCT and successively regain their initial abundance levels. Abundances of these taxa exhibit correlations with adaptive immune cell counts, pointing to specific interactions between the microbiota and the adaptive immune system with regards to maintaining homeostasis.

2. Innate immune components exhibit correlations with gut, oral, and nasal bacterial abundances before, at the time of and after allogeneic and autologous HSCT. This suggests specific interactions between the microbiota and the innate immune system with regards to maintaining homeostasis.

3. The abundances of specific gut, oral, and nasal taxa in both, allogeneic and autologous HSCT correlate with immune markers and HSCT-associated side effects, such as infections and aGvHD before, at the time of and after HSCT. Pre-transplant abundances of distinct taxa at all three body sites are predictive for acute GvHD after allogeneic HSCT.

4. Whether the abundances of specific gut, oral, and nasal taxa in allogeneic and autologous HSCT correlate with clinical outcomes, such as relapse and survival, was difficult to determine conclusively due to a limited number of patients and a limited number of observed adverse events (relapse, death) during the monitored periods.

7. Future perspectives

This thesis includes one of the first comprehensive characterizations of long-term dynamics of microbial abundances in HSCT and suggests that the period between the well described first month post allo-HSCT, and the third month is crucial for microbiota reconstitution [**Masche II**]. To determine in more detail the temporal abundance trajectories from month +1 to month +3, a denser sampling schedule between those time points is recommendable for future research. Whether health-associated taxa, such as *Lachnospiraceae* and *Ruminococcacea*, return to pre-HSCT abundances within the expected time post-transplant could then for instance inform the decision if FMT is necessary to prevent adverse outcomes.
Furthermore, we demonstrated an association between these taxa and time to immune reconstitution after allo-HSCT, which highlights the concern about their antibiotic-induced depletion [**Masche I and II**]. In this light, especially the extensive use of antimicrobial agents, such as vancomycin, ciprofloxacin, and metronidazole, which target obligate anaerobes might have to be reevaluated in the future [**Masche I**] [148, 160].
For autologous HSCT, our findings suggest that *Ruminococcaceae* and *Lachnospiraceae* could possibly serve as a marker for neutrophil recovery, in addition to aGvHD in allogeneic HSCT [**Masche III**]. However, directionality and temporal succession of these associations remain to be investigated in more detail, preferably in a larger cohort. In addition to comparing host-microbial associations in allogeneic and autologous HSCT, future studies should aim at disentangling transplantation- and conditioning-related effects. Investigating host-microbial associations in a control group of patients who receive chemotherapy and/or TBI, but do not undergo HSCT is therefore recommendable and originally was part of the project plan for this thesis.
We showed that the microbiota at different body sites exhibited shared associations with immune markers in both, allo- and auto-HSCT [**Masche II and III**]. Collection of buccal and anterior naris swabs has the advantage of increased feasibility compared with fecal samples, especially during clinical routine, since it does not depend on patients' bowel movements, and patient compliance for sample collection might be higher. This might provide new opportunities for establishing regular screening of the microbiota to use microbial markers for informed preventive management of adverse clinical outcomes in HSCT. However, inferences between body sites need to be interpreted with caution and under consideration of clinical baseline parameters, e.g. primary disease. The inclusion of oral mucositis scoring data would be of great interest for further assessment of the oral microbiota in HSCT, since our findings suggest the importance of a number of taxa with potential involvement in this condition [**Masche II and III**].
Other studies have related gut microbial abundances to aGvHD severity before [156, 159, 162, 163, 238]. However, as we argued in **Masche II**, previous findings were mostly based on regression or correlation analyses and/or considered microbial abundances close to, concurrent to or after aGvHD onset, limiting the options for early interventions in the clinic. Certainly, our results, especially the machine learning models, need to be validated in an additional independent, preferably larger cohort. However, our findings might open up opportunities to use specific bacterial taxa, such as intestinal *Lactobacillus,* as early microbial predictive markers for GvHD and potentially also other clinical outcomes [**Masche I and II**]. This could facilitate early interventions in the clinic for patients at increased risk of adverse clinical outcomes. We might in the future be able to make use of the predictive potential of especially oral and nasal taxa since swab collection might facilitate frequent and rapid microbiota screening in the clinic.

References

1. Sender R, Fuchs S, Milo R. Are We Really Vastly Outnumbered? Revisiting the Ratio of Bacterial to Host Cells in Humans. Cell. 2016;164:337–40. doi:10.1016/j.cell.2016.01.013.

2. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The human microbiome project. Nature. 2007;449:804–10. doi:10.1038/nature06244.

3. Gilbert JA, Blaser MJ, Caporaso JG, Jansson JK, Lynch S V., Knight R. Current understanding of the human microbiome. Nat Med. 2018;24:392–400.

4. Heintz-Buschart A, Wilmes P. Human Gut Microbiome: Function Matters. Trends Microbiol. 2018;26:563–74. doi:10.1016/J.TIM.2017.11.002.

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

6. Milani C, Duranti S, Bottacini F, Casey E, Turroni F, Mahony J, et al. The First Microbial Colonizers of the Human Gut: Composition, Activities, and Health Implications of the Infant Gut Microbiota. Microbiol Mol Biol Rev. 2017;81:e00036-17. doi:10.1128/MMBR.00036-17.

7. Jovel J, Dieleman LA, Kao D, Mason AL, Wine E. The Human Gut Microbiome in Health and Disease. Metagenomics. 2018;:197–213. doi:10.1016/B978-0-08-102268-9.00010-0.

8. Korpela K, de Vos WM. Early life colonization of the human gut: microbes matter everywhere. Curr Opin Microbiol. 2018;44:70–8. doi:10.1016/J.MIB.2018.06.003.

9. Wampach L, Heintz-Buschart A, Hogan A, Muller EEL, Narayanasamy S, Laczny CC, et al. Colonization and Succession within the Human Gut Microbiome by Archaea, Bacteria, and Microeukaryotes during the First Year of Life. Front Microbiol. 2017;8:738. doi:10.3389/fmicb.2017.00738.

10. Bergström A, Skov TH, Bahl MI, Roager HM, Christensen LB, Ejlerskov KT, et al. Establishment of intestinal microbiota during early life: a longitudinal, explorative study of a large cohort of Danish infants. Appl Environ Microbiol. 2014;80:2889–900. doi:10.1128/AEM.00342-14.

11. Lin A, Bik EM, Costello EK, Dethlefsen L, Haque R, Relman DA, et al. Distinct distal gut microbiome diversity and composition in healthy children from Bangladesh and the United States. PLoS One. 2013;8:e53838. doi:10.1371/journal.pone.0053838.

12. Hollister EB, Riehle K, Luna RA, Weidler EM, Rubio-Gonzales M, Mistretta T-A, et al. Structure and function of the healthy pre-adolescent pediatric gut microbiome. Microbiome. 2015;3:36. doi:10.1186/s40168-015-0101-x.

13. Verma D, Garg PK, Dubey AK. Insights into the human oral microbiome. Arch Microbiol. 2018;200:525–40. doi:10.1007/s00203-018-1505-3.

14. Xu X, He J, Xue J, Wang Y, Li K, Zhang K, et al. Oral cavity contains distinct niches with dynamic microbial communities. Environ Microbiol. 2015;17:699–710. doi:10.1111/1462-2920.12502.

15. Bik EM, Long CD, Armitage GC, Loomer P, Emerson J, Mongodin EF, et al. Bacterial diversity in the oral cavity of 10 healthy individuals. ISME J. 2010;4:962–74. doi:10.1038/ismej.2010.30.

16. Yan M, Pamp SJ, Fukuyama J, Hwang PH, Cho D-Y, Holmes S, et al. Nasal Microenvironments and Interspecific Interactions Influence Nasal Microbiota Complexity and S. aureus Carriage. Cell Host

Microbe. 2013;14:631–40. doi:10.1016/j.chom.2013.11.005.

17. Proctor DM, Relman DA. The Landscape Ecology and Microbiota of the Human Nose, Mouth, and Throat. Cell Host Microbe. 2017;21:421–32. doi:10.1016/J.CHOM.2017.03.011.

18. Krismer B, Weidenmaier C, Zipperer A, Peschel A. The commensal lifestyle of Staphylococcus aureus and its interactions with the nasal microbiota. Nat Rev Microbiol. 2017;15:675–87. doi:10.1038/nrmicro.2017.104.

19. Liu CM, Price LB, Hungate BA, Abraham AG, Larsen LA, Christensen K, et al. Staphylococcus aureus and the ecology of the nasal microbiome. Sci Adv. 2015;1:1–8.

20. Hermansen GMM, Sazinas P, Kofod D, Millard A, Andersen PS, Jelsbak L. Transcriptomic profiling of interacting nasal staphylococci species reveals global changes in gene and non-coding RNA expression. FEMS Microbiol Lett. 2018;365:1–9.

21. Mika M, Mack I, Korten I, Qi W, Aebi S, Frey U, et al. Dynamics of the nasal microbiota in infancy: a prospective cohort study. J Allergy Clin Immunol. 2015;135:905–12.e11. doi:10.1016/j.jaci.2014.12.1909.

22. Ta LDH, Yap GC, Tay CJX, Lim ASM, Huang C-H, Chu CW, et al. Establishment of the nasal microbiota in the first 18 months of life: Correlation with early-onset rhinitis and wheezing. J Allergy Clin Immunol. 2018;142:86–95. doi:10.1016/j.jaci.2018.01.032.

23. Oh J, Conlan S, Polley EC, Segre JA, Kong HH. Shifts in human skin and nares microbiota of healthy children and adults. Genome Med. 2012;4:77. doi:10.1186/gm378.

24. Shanks N, Greek R, Greek J. Are animal models predictive for humans? Philos Ethics Humanit Med. 2009;4:2. doi:10.1186/1747-5341-4-2.

25. Gensollen T, Iyer SS, Kasper DL, Blumberg RS. How colonization by microbiota in early life shapes the immune system. Science (80-). 2016;352:539–44.

26. Honda K, Littman DR. The microbiota in adaptive immune homeostasis and disease. Nature. 2016;535:75–84. doi:10.1038/nature18848.

27. Khosravi A, Yáñez A, Price JG, Chow A, Merad M, Goodridge HS, et al. Gut microbiota promote hematopoiesis to control bacterial infection. Cell Host Microbe. 2014;15:374–81.

28. Deshmukh HS, Liu Y, Menkiti OR, Mei J, Dai N, O'Leary CE, et al. The microbiota regulates neutrophil homeostasis and host resistance to Escherichia coli K1 sepsis in neonatal mice. Nat Med. 2014;20:524–30. doi:10.1038/nm.3542.

29. Round JL, Mazmanian SK. The gut microbiome shapes intestinal immune responses during health and disease. Nat Rev Immunol. 2009;9:313–23.

30. Bain CC, Bravo-Blas A, Scott CL, Gomez Perdiguero E, Geissmann F, Henri S, et al. Constant replenishment from circulating monocytes maintains the macrophage pool in the intestine of adult mice. Nat Immunol. 2014;15:929–37.

31. Belkaid Y, Harrison OJ. Homeostatic Immunity and the Microbiota. Immunity. 2017;46:562–76. doi:10.1016/j.immuni.2017.04.008.

32. Lysenko ES, Clarke TB, Shchepetov M, Ratner AJ, Roper DI, Dowson CG, et al. Nod1 signaling overcomes resistance of S. pneumoniae to opsonophagocytic killing. PLoS Pathog. 2007;3:1073–81.

33. Clarke TB, Davis KM, Lysenko ES, Zhou AY, Yu Y, Weiser JN. Recognition of peptidoglycan from the microbiota by Nod1 enhances systemic innate immunity. Nat Med. 2010;16:228–31. doi:10.1038/nm.2087.

34. Zhang D, Chen G, Manwani D, Mortha A, Xu C, Faith JJ, et al. Neutrophil ageing is regulated by the microbiome. Nature. 2015;525:528–32. doi:10.1038/nature15367.

35. Shi C, Jia T, Mendez-Ferrer S, Hohl TM, Serbina N V., Lipuma L, et al. Bone Marrow Mesenchymal Stem and Progenitor Cells Induce Monocyte Emigration in Response to Circulating Toll-like Receptor Ligands. Immunity. 2011;34:590–601. doi:10.1016/j.immuni.2011.02.016.

36. Shi C, Pamer EG. Monocyte recruitment during infection and inflammation. Nat Rev Immunol. 2011;11:762–74. doi:10.1038/nri3070.

37. Hill DA, Siracusa MC, Abt MC, Kim BS, Kobuley D, Kubo M, et al. Commensal bacteria-derived signals regulate basophil hematopoiesis and allergic inflammation. Nat Med. 2012;18:538–46. doi:10.1038/nm.2657.

38. Sommer F, Bäckhed F. The gut microbiota-masters of host development and physiology. Nat Rev Microbiol. 2013;11:227–38. doi:10.1038/nrmicro2974.

39. Ganal SC, Sanos SL, Kallfass C, Oberle K, Johner C, Kirschning C, et al. Priming of Natural Killer Cells by Nonmucosal Mononuclear Phagocytes Requires Instructive Signals from Commensal Microbiota. Immunity. 2012;37:171–86.

40. Abt MC, Osborne LC, Monticelli LA, Doering TA, Alenghat T, Sonnenberg GF, et al. Commensal Bacteria Calibrate the Activation Threshold of Innate Antiviral Immunity. Immunity. 2012;37:158–70.

41. Harder J, Gläser R, Schröder J-M. Review: Human antimicrobial proteins — effectors of innate immunity. J Endotoxin Res. 2007;13:317–38. doi:10.1177/0968051907088275.

42. Boman HG. Antibacterial peptides: basic facts and emerging concepts. J Intern Med. 2003;254:197–215. doi:10.1046/j.1365-2796.2003.01228.x.

43. Izadpanah A, Gallo RL. Antimicrobial peptides. J Am Acad Dermatol. 2005;52:381–90. doi:10.1016/j.jaad.2004.08.026.

44. Gupta S, Bhatia G, Sharma A, Saxena S. Host defense peptides: An insight into the antimicrobial world. J Oral Maxillofac Pathol. 2018;22:239–44. doi:10.4103/jomfp.JOMFP_113_16.

45. Sankaran-Walters S, Hart R, Dills C. Guardians of the Gut: Enteric Defensins. Front Microbiol. 2017;8:647. doi:10.3389/fmicb.2017.00647.

46. Ostaff MJ, Stange EF, Wehkamp J. Antimicrobial peptides and gut microbiota in homeostasis and pathology. EMBO Mol Med. 2013;5:1465–83. doi:10.1002/emmm.201201773.

47. Ostaff MJ, Stange EF, Wehkamp J. Antimicrobial peptides and gut microbiota in homeostasis and pathology. EMBO Mol Med. 2013;5:1465–83.

48. Duits LA, Ravensbergen B, Rademaker M, Hiemstra PS, Nibbering PH. Expression of beta-defensin 1 and 2 mRNA by human monocytes, macrophages and dendritic cells. Immunology. 2002;106:517–25. doi:10.1046/J.1365-2567.2002.01430.X.

49. Günther C, Josenhans C, Wehkamp J. Crosstalk between microbiota, pathogens and the innate

immune responses. Int J Med Microbiol. 2016;306:257–65. doi:10.1016/j.ijmm.2016.03.003.

50. Belkaid Y, Hand TW. Role of the Microbiota in Immunity and Inflammation. Cell. 2014;157:121–41. doi:10.1016/J.CELL.2014.03.011.

51. Perez-Lopez A, Behnsen J, Nuccio S-P, Raffatellu M. Mucosal immunity to pathogenic intestinal bacteria. Nat Rev Immunol. 2016;16:135–48. doi:10.1038/nri.2015.17.

52. Hooper L V., Littman DR, Macpherson AJ. Interactions between the microbiota and the immune system. Science. 2012;336:1268–73. doi:10.1126/science.1223490.

53. Postler TS, Ghosh S. Understanding the Holobiont: How Microbial Metabolites Affect Human Health and Shape the Immune System. Cell Metab. 2017;26:110–30. doi:10.1016/J.CMET.2017.05.008.

54. Chang P V, Hao L, Offermanns S, Medzhitov R. The microbial metabolite butyrate regulates intestinal macrophage function via histone deacetylase inhibition. Proc Natl Acad Sci U S A. 2014;111:2247–52. doi:10.1073/pnas.1322269111.

55. Vinolo MAR, Ferguson GJ, Kulkarni S, Damoulakis G, Anderson K, Bohlooly-Y M, et al. SCFAs induce mouse neutrophil chemotaxis through the GPR43 receptor. PLoS One. 2011;6:e21205. doi:10.1371/journal.pone.0021205.

56. Vinolo MAR, Rodrigues HG, Hatanaka E, Sato FT, Sampaio SC, Curi R. Suppressive effect of short-chain fatty acids on production of proinflammatory mediators by neutrophils. J Nutr Biochem. 2011;22:849–55. doi:10.1016/j.jnutbio.2010.07.009.

57. 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.

58. Donohoe DR, Wali A, Brylawski BP, Bultman SJ. Microbial Regulation of Glucose Metabolism and Cell-Cycle Progression in Mammalian Colonocytes. PLoS One. 2012;7.

59. 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.

60. Roche AM, Richard AL, Rahkola JT, Janoff EN, Weiser JN. Antibody blocks acquisition of bacterial colonization through agglutination. Mucosal Immunol. 2015;8:176–85.

61. 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.

62. Peterson DA, McNulty NP, Guruge JL, Gordon JI. IgA Response to Symbiotic Bacteria as a Mediator of Gut Homeostasis. Cell Host Microbe. 2007;2:328–39.

63. Kim M, Qie Y, Park J, Kim CH. Gut Microbial Metabolites Fuel Host Antibody Responses. Cell Host Microbe. 2016;20:202–14. doi:10.1016/j.chom.2016.07.001.

64. Kawamoto S, Maruya M, Kato LM, Suda W, Atarashi K, Doi Y, et al. Foxp3+T Cells Regulate Immunoglobulin A Selection and Facilitate Diversification of Bacterial Species Responsible for Immune Homeostasis. Immunity. 2014;41:152–65. doi:10.1016/j.immuni.2014.05.016.

65. Pamer EG. Microbial Tuning of the Mammalian Immune System. Trends Mol Med. 2017;23:379–80. doi:10.1016/J.MOLMED.2017.03.006.

66. Atarashi K, Tanoue T, Ando M, Kamada N, Nagano Y, Narushima S, et al. Th17 Cell Induction by Adhesion of Microbes to Intestinal Epithelial Cells. Cell. 2015;163:367–80. doi:10.1016/j.cell.2015.08.058.

67. Sano T, Huang W, Hall JA, Yang Y, Chen A, Gavzy SJ, et al. An IL-23R/IL-22 Circuit Regulates Epithelial Serum Amyloid A to Promote Local Effector Th17 Responses. Cell. 2015;163:381–93. doi:10.1016/j.cell.2015.08.061.

68. Bel S, Hooper L V. A bacterial nudge to T-cell function. Nature. 2015;526:328–30. doi:10.1038/526328a.

69. Wacleche VS, Landay A, Routy JP, Ancuta P. The Th17 lineage: From barrier surfaces homeostasis to autoimmunity, cancer, and HIV-1 pathogenesis. Viruses. 2017;9:1–33.

70. Hoe E, Anderson J, Nathanielsz J, Toh ZQ, Marimla R, Balloch A, et al. The contrasting roles of Th17 immunity in human health and disease. Microbiol Immunol. 2017;61:49–56.

71. Chen X, Oppenheim JJ. Th17 cells and Tregs: unlikely allies. J Leukoc Biol. 2014;95:723–31. doi:10.1189/jlb.1213633.

72. Atarashi K, Tanoue T, Oshima K, Suda W, Nagano Y, Nishikawa H, et al. Treg induction by a rationally selected mixture of Clostridia strains from the human microbiota. Nature. 2013;500:232–6.

73. Faith JJ, Ahern PP, Ridaura VK, Cheng J, Gordon JI. Identifying Gut Microbe – Host Phenotype Relationships Using Combinatorial Communities in Gnotobiotic Mice. 2014;5. doi:10.1126/scitranslmed.3008051.

74. Ding T, Schloss PD. Dynamics and associations of microbial community types across the human body. Nature. 2014;509:357–60. doi:10.1038/nature13178.

75. Tsukamoto Y, Usui M, Yamamoto G, Takagi Y, Tachikawa T, Yamamoto M, et al. Role of the junctional epithelium in periodontal innate defense and homeostasis. J Periodontal Res. 2012;47:750–7. doi:10.1111/j.1600-0765.2012.01490.x.

76. Dixon DR, Reife RA, Cebra JJ, Darveau RP. Commensal Bacteria Influence Innate Status Within Gingival Tissues: A Pilot Study. J Periodontol. 2004;75:1486–92. doi:10.1902/jop.2004.75.11.1486.

77. Jain R, Waldvogel-Thurlow S, Darveau R, Douglas R. Differences in the paranasal sinuses between germ-free and pathogen-free mice. Int Forum Allergy Rhinol. 2016;6:631–7. doi:10.1002/alr.21712.

78. Costalonga M, Cleary PP, Fischer LA, Zhao Z. Intranasal bacteria induce Th1 but not Treg or Th2. Mucosal Immunol. 2009;2:85–95.

79. Thaiss CA, Zmora N, Levy M, Elinav E. The microbiome and innate immunity. Nature. 2016;535:65–74.

80. Charlie G. Buffie, Pamer EG. Microbiota-mediated colonization resistance against intestinal pathogens. Nat Rev Immunol. 2013;13:790–801.

81. Grainger JR, Wohlfert EA, Fuss IJ, Bouladoux N, Askenase MH, Legrand F, et al. Inflammatory monocytes regulate pathologic responses to commensals during acute gastrointestinal infection. Nat Med. 2013;19:713–21. doi:10.1038/nm.3189.

82. Duvallet C, Gibbons SM, Gurry T, Irizarry RA, Alm EJ. Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. Nat Commun. 2017;8:1784. doi:10.1038/s41467-017-01973-8.

83. Jackson MA, Verdi S, Maxan M-E, Shin CM, Zierer J, Bowyer RCE, et al. Gut microbiota associations

with common diseases and prescription medications in a population-based cohort. Nat Commun. 2018;9:2655. doi:10.1038/s41467-018-05184-7.

84. Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward D V, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol. 2012;13:R79. doi:10.1186/gb-2012-13-9-r79.

85. Peters BA, Shapiro JA, Church TR, Miller G, Trinh-Shevrin C, Yuen E, et al. A taxonomic signature of obesity in a large study of American adults. Sci Rep. 2018;8:1–13. doi:10.1038/s41598-018-28126-1.

86. Hand TW, Vujkovic-Cvijin I, Ridaura VK, Belkaid Y. Linking the Microbiota, Chronic Disease, and the Immune System. Trends Endocrinol Metab. 2016;27:831–43. doi:10.1016/j.tem.2016.08.003.

87. Maslowski KM, Vieira AT, Ng A, Kranich J, Sierro F, Yu D, et al. Regulation of inflammatory responses by gut microbiota and chemoattractant receptor GPR43. Nature. 2009;461:1282–6. doi:10.1038/nature08530.

88. Rivera-Chávez F, Zhang LF, Faber F, Lopez CA, Byndloss MX, Olsan EE, et al. Depletion of Butyrate-Producing Clostridia from the Gut Microbiota Drives an Aerobic Luminal Expansion of Salmonella. Cell Host Microbe. 2016;19:443–54. doi:10.1016/J.CHOM.2016.03.004.

89. Litvak Y, Byndloss MX, Tsolis RM, Bäumler AJ. Dysbiotic Proteobacteria expansion: a microbial signature of epithelial dysfunction. Curr Opin Microbiol. 2017;39:1–6. doi:10.1016/J.MIB.2017.07.003.

90. Chabannon C, Kuball J, Bondanza A, Dazzi F, Pedrazzoli P, Toubert A, et al. Hematopoietic stem cell transplantation in its 60s: A platform for cellular therapies. Sci Transl Med. 2018;10:eaap9630. doi:10.1126/scitranslmed.aap9630.

91. Passweg JR, Baldomero H, Bader P, Basak GW, Bonini C, Duarte R, et al. Is the use of unrelated donor transplantation leveling off in Europe? The 2016 European Society for Blood and Marrow Transplant activity survey report. Bone Marrow Transplant. 2018;:1. doi:10.1038/s41409-018-0153-1.

92. 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.

93. Passweg JR, Baldomero H, Bader P, Bonini C, Duarte RF, Dufour C, et al. Use of haploidentical stem cell transplantation continues to increase: the 2015 European Society for Blood and Marrow Transplant activity survey report. Bone Marrow Transplant. 2017;52:811–7. doi:10.1038/bmt.2017.34.

94. Gooptu M, Kim HT, Ho VT, Alyea EP, Koreth J, Armand P, et al. A Comparison of the Myeloablative Conditioning Regimen Fludarabine/Busulfan with Cyclophosphamide/Total Body Irradiation, for Allogeneic Stem Cell Transplantation in the Modern Era: A Cohort Analysis. Biol Blood Marrow Transplant. 2018. doi:10.1016/J.BBMT.2018.03.011.

95. Arai Y, Jo T, Matsui H, Kondo T, Takaori-Kondo A. Efficacy of antithymocyte globulin for allogeneic hematopoietic cell transplantation: a systematic review and meta-analysis. Leuk Lymphoma. 2017;58:1840–8. doi:10.1080/10428194.2016.1266624.

96. Urbano-Ispizua A. Risk assessment in haematopoietic stem cell transplantation: Stem cell source. Best Pract Res Clin Haematol. 2007;20:265–80. doi:10.1016/J.BEHA.2006.09.006.

97. 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.

98. Bassan R, Hoelzer D. Modern therapy of acute lymphoblastic leukemia. J Clin Oncol. 2011;29:532–43.

99. Ghimire S, Weber D, Mavin E, Wang X nong, Dickinson AM, Holler E. Pathophysiology of GvHD and Other HSCT-Related Major Complications. Front Immunol. 2017;8:79. doi:10.3389/fimmu.2017.00079.

100. Nassereddine S, Rafei H, Elbahesh E, Tabbara I. Acute Graft Versus Host Disease: A Comprehensive Review. Anticancer Res. 2017;37:1547–55. doi:10.21873/anticanres.11483.

101. Daguindau E, Gautier T, Chagué C, Pais de Barros J-P, Deckert V, Lagrost L, et al. Is It Time to Reconsider the Lipopolysaccharide Paradigm in Acute Graft-Versus-Host Disease? Front Immunol. 2017;8:952. doi:10.3389/fimmu.2017.00952.

102. Staffas A, Burgos da Silva M, van den Brink MRM. The intestinal microbiota in allogeneic hematopoietic cell transplant and graft-versus-host disease. Blood. 2017;129:927–33. doi:10.1182/blood-2016-09-691394.

103. Verneris MR, Lee SJ, Ahn KW, Wang H-L, Battiwalla M, Inamoto Y, et al. HLA Mismatch Is Associated with Worse Outcomes after Unrelated Donor Reduced-Intensity Conditioning Hematopoietic Cell Transplantation: An Analysis from the Center for International Blood and Marrow Transplant Research. Biol Blood Marrow Transplant. 2015;21:1783–9. doi:10.1016/J.BBMT.2015.05.028.

104. Kanda J. Effect of HLA mismatch on acute graft-versus-host disease. Int J Hematol. 2013;98:300–8. doi:10.1007/s12185-013-1405-x.

105. Zeiser R, Socié G, Blazar BR. Pathogenesis of acute graft-versus-host disease: from intestinal microbiota alterations to donor T cell activation. British Journal of Haematology. 2016;175.

106. Garnett C, Apperley JF, Pavlů J. Treatment and management of graft-versus-host disease: improving response and survival. Ther Adv Hematol. 2013;4:366–78. doi:10.1177/2040620713489842.

107. 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.

108. Dickinson AM, Holler E. Polymorphisms of cytokine and innate immunity genes and GVHD. Best Pract Res Clin Haematol. 2008;21:149–64.

109. Hanash AM, Kappel LW, Yim NL, Nejat RA, Goldberg GL, Smith OM, et al. Abrogation of donor T-cell IL-21 signaling leads to tissue-specific modulation of immunity and separation of GVHD from GVL. Blood. 2011;118:446–55.

110. Hanash AM, Dudakov JA, Hua G, O'Connor MH, Young LF, Singer NV, et al. Interleukin-22 Protects Intestinal Stem Cells from Immune-Mediated Tissue Damage and Regulates Sensitivity to Graft versus Host Disease. Immunity. 2012;37:339–50. doi:10.1016/J.IMMUNI.2012.05.028.

111. Cowen EW. Cutaneous Graft-Versus-Host Disease. In: Clinical and Basic Immunodermatology: Second Edition. 2017. p. 665–83.

112. Eriguchi Y, Takashima S, Oka H, Shimoji S, Nakamura K, Uryu H, et al. Graft-versus-host disease disrupts intestinal microbial ecology by inhibiting Paneth cell production of α-defensins. Blood. 2012;120:223–31. doi:10.1182/blood-2011-12-401166.

113. Levine JE, Huber E, Hammer STG, Harris AC, Greenson JK, Braun TM, et al. Low Paneth cell numbers at onset of gastrointestinal graft-versus-host disease identify patients at high risk for nonrelapse mortality. Blood. 2013;122:1505–9. doi:10.1182/blood-2013-02-485813.

114. Weber D, Frauenschläger K, Ghimire S, Peter K, Panzer I, Hiergeist A, et al. The association between acute graft-versus-host disease and antimicrobial peptide expression in the gastrointestinal tract after allogeneic stem cell transplantation. PLoS One. 2017;12:e0185265. doi:10.1371/journal.pone.0185265.

115. Eriguchi Y, Nakamura K, Hashimoto D, Shimoda S, Shimono N, Akashi K, et al. Decreased secretion of Paneth cell α-defensins in graft-versus-host disease. Transpl Infect Dis. 2015;17:702–6. doi:10.1111/tid.12423.

116. Zhao D, Kim Y-H, Jeong S, Greenson JK, Chaudhry MS, Hoepting M, et al. Survival signal REG3α prevents crypt apoptosis to control acute gastrointestinal graft-versus-host disease. J Clin Invest. 2018;128. doi:10.1172/JCI99261.

117. Glucksberg H, Storb R, Fefer A, Buckner CD, Neiman PE, Clift RA, et al. Clinical manifestations of graft-versus-host disease in human recipients of marrow from HL-A-matched sibling donors. Transplantation. 1974;18:295–304. http://www.ncbi.nlm.nih.gov/pubmed/4153799. Accessed 19 Oct 2018.

118. Appelbaum FR. Hematopoietic-Cell Transplantation at 50. N Engl J Med. 2007;357:1472–5. doi:10.1056/NEJMp078166.

119. Jones JA, Qazilbash MH, Shih Y-CT, Cantor SB, Cooksley CD, Elting LS. In-hospital complications of autologous hematopoietic stem cell transplantation for lymphoid malignancies. Cancer. 2008;112:1096–105. doi:10.1002/cncr.23281.

120. Salar A, Sierra J, Gandarillas M, Caballero M, Marín J, Lahuerta J, et al. Autologous stem cell transplantation for clinically aggressive non-Hodgkin's lymphoma: the role of preparative regimens. Bone Marrow Transplant. 2001;27:405–12. doi:10.1038/sj.bmt.1702795.

121. Byun JM, Lee J, Shin S-J, Kang M, Yoon S-S, Koh Y. Busulfan plus melphalan versus high-dose melphalan as conditioning regimens in autologous stem cell transplantation for newly diagnosed multiple myeloma. Blood Res. 2018;53:105–9. doi:10.5045/br.2018.53.2.105.

122. Patir P, Soyer N, Durusoy IR, Sahin F, Saydam G, Tobu M, et al. A Retrospective Comparison of TECAM and BEAM Conditioning Regimens Before Autologous Hematopoietic Stem Cell Transplant in Lymphoma Patients: Efficacy and Toxicity. Exp Clin Transplant. 2018. doi:10.6002/ect.2017.0274.

123. Sahin U, Toprak SK, Atilla PA, Atilla E, Demirer T. An overview of infectious complications after allogeneic hematopoietic stem cell transplantation. J Infect Chemother. 2016;22:505–14. doi:10.1016/J.JIAC.2016.05.006.

124. Dandoy CE, Ardura MI, Papanicolaou GA, Auletta JJ. Bacterial bloodstream infections in the allogeneic hematopoietic cell transplant patient: new considerations for a persistent nemesis. Bone Marrow Transplant. 2017;52:1091–106. doi:10.1038/bmt.2017.14.

125. Dandoy CE, Haslam D, Lane A, Jodele S, Demmel K, El-Bietar J, et al. Healthcare Burden, Risk Factors, and Outcomes of Mucosal Barrier Injury Laboratory-Confirmed Bloodstream Infections after Stem Cell Transplantation. Biol Blood Marrow Transplant. 2016;22:1671–7. doi:10.1016/j.bbmt.2016.06.002.

126. Satwani P, Freedman JL, Chaudhury S, Jin Z, Levinson A, Foca MD, et al. A Multicenter Study of Bacterial Blood Stream Infections in Pediatric Allogeneic Hematopoietic Cell Transplantation Recipients:

The Role of Acute Gastrointestinal Graft-versus-Host Disease. Biol Blood Marrow Transplant. 2017;23:642–7. doi:10.1016/J.BBMT.2017.01.073.

127. Hong J, Moon SM, Ahn HK, Sym SJ, Park YS, Park J, et al. Comparison of Characteristics of Bacterial Bloodstream Infection between Adult Patients with Allogeneic and Autologous Hematopoietic Stem Cell Transplantation. Biol Blood Marrow Transplant. 2013;19:994–9. doi:10.1016/J.BBMT.2013.03.019.

128. Kimura S, Akahoshi Y, Nakano H, Ugai T, Wada H, Yamasaki R, et al. Antibiotic prophylaxis in hematopoietic stem cell transplantation. A meta-analysis of randomized controlled trials. J Infect. 2014;69:13–25. doi:10.1016/J.JINF.2014.02.013.

129. Rungoe C, Malchau EL, Larsen LN, Schroeder H. Infections during induction therapy for children with acute lymphoblastic leukemia. The role of sulfamethoxazole-trimethoprim (SMX-TMP) prophylaxis. Pediatr Blood Cancer. 2010;55:304–8. doi:10.1002/pbc.22423.

130. Calitri C, Ruberto E, Castagnola E. Antibiotic prophylaxis in neutropenic children with acute leukemia: do the presently available data really support this practice? Eur J Haematol. 2018. doi:10.1111/ejh.13162.

131. Freifeld AG, Bow EJ, Sepkowitz KA, Boeckh MJ, Ito JI, Mullen CA, et al. Clinical Practice Guideline for the Use of Antimicrobial Agents in Neutropenic Patients with Cancer: 2010 Update by the Infectious Diseases Society of America. Clin Infect Dis. 2011;52:e56–93. doi:10.1093/cid/cir073.

132. Ogonek J, Kralj Juric M, Ghimire S, Varanasi PR, Holler E, Greinix H, et al. Immune Reconstitution after Allogeneic Hematopoietic Stem Cell Transplantation. Front Immunol. 2016;7:507. doi:10.3389/fimmu.2016.00507.

133. Camargo JF, Komanduri K V. Emerging concepts in cytomegalovirus infection following hematopoietic stem cell transplantation. Hematol Oncol Stem Cell Ther. 2017;10:233–8. doi:10.1016/J.HEMONC.2017.05.001.

134. Wang C-H, Chang F-Y, Chao T-Y, Kao W-Y, Ho C-L, Chen Y-C, et al. Characteristics comparisons of bacteremia in allogeneic and autologous hematopoietic stem cell-transplant recipients with levofloxacin prophylaxis and influence on resistant bacteria emergence. J Microbiol Immunol Infect. 2018;51:123–31. doi:10.1016/J.JMII.2016.02.003.

135. Chaudhry MS, Velardi E, Malard F, van den Brink MRM. Immune Reconstitution after Allogeneic Hematopoietic Stem Cell Transplantation: Time To T Up the Thymus. J Immunol. 2017;198:40–6. doi:10.4049/jimmunol.1601100.

136. de Koning C, Plantinga M, Besseling P, Boelens JJ, Nierkens S. Immune Reconstitution after Allogeneic Hematopoietic Cell Transplantation in Children. Biol Blood Marrow Transplant. 2016;22:195–206. doi:10.1016/j.bbmt.2015.08.028.

137. Storek J, Geddes M, Khan F, Huard B, Helg C, Chalandon Y, et al. Reconstitution of the immune system after hematopoietic stem cell transplantation in humans. Semin Immunopathol. 2008;30:425–37. doi:10.1007/s00281-008-0132-5.

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

139. Leung W, Iyengar R, Turner V, Lang P, Bader P, Conn P, et al. Determinants of Antileukemia Effects of Allogeneic NK Cells. J Immunol. 2004;172:644–50. doi:10.4049/jimmunol.172.1.644.

140. van den Brink MRM, Velardi E, Perales M-A. Immune reconstitution following stem cell transplantation. Hematol Am Soc Hematol Educ Progr. 2015;2015:215–9. doi:10.1182/asheducation-2015.1.215.

141. Storek J, Wells D, Dawson MA, Storer B MD. Factors influencing B lymphopoiesis after allogenic hematopoietic cell transplantation. Blood. 2001;98: 489-49:489–92.

142. Wiegering V, Eyrich M, Winkler B, Schlegel PG. Comparison of immune reconstitution after allogeneic vs. autologous stem cell transplantation in 182 pediatric recipients. Pediatr Hematol Oncol J. 2017;2:2–6. doi:10.1016/j.phoj.2017.05.002.

143. Porrata LF, Litzow MR, Markovic SN. Immune reconstitution after hematopoietic stem cell transplantation. Hematop Stem Cell Transplant Pediatr Hematol. 2017;76 April:371–83.

144. Noor F, Kaysen A, Wilmes P, Schneider JG. The Gut Microbiota and Hematopoietic Stem Cell Transplantation: Challenges and Potentials. J Innate Immun. 2018;:1–11. doi:10.1159/000492943.

145. Beelen DW, Elmaagacli A, Müller KD, Hirche H, Schaefer UW. Influence of intestinal bacterial decontamination using metronidazole and ciprofloxacin or ciprofloxacin alone on the development of acute graft- versus-host disease after marrow transplantation in patients with hematologic malignancies: Final results and. Blood. 1999;93:3267–75. https://www.scopus.com/record/display.uri?eid=2-s2.0-0033562823&origin=inward&txGid=6a3d3b9a39c3259863121fa126ee624e. Accessed 16 Feb 2018.

146. Whangbo J, Ritz J, Bhatt A. Antibiotic-mediated modification of the intestinal microbiome in allogeneic hematopoietic stem cell transplantation. Bone Marrow Transplant. 2017;52:183–90. doi:10.1038/bmt.2016.206.

147. Horton LE, Haste NM, Taplitz RA. Rethinking Antimicrobial Prophylaxis in the Transplant Patient in the World of Emerging Resistant Organisms—Where Are We Today? Curr Hematol Malig Rep. 2018;13:59–67. doi:10.1007/s11899-018-0435-0.

148. 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.

149. Andermann TM, Peled JU, Ho C, Reddy P, Riches M, Storb R, et al. The Microbiome and Hematopoietic Cell Transplantation: Past, Present, and Future. Biol Blood Marrow Transplant. 2018. doi:10.1016/J.BBMT.2018.02.009.

150. 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.

151. Kaysen A, Heintz-Buschart A, Muller EEL, Narayanasamy S, Wampach L, Laczny CC, et al. Integrated meta-omic analyses of the gastrointestinal tract microbiome in patients undergoing allogeneic hematopoietic stem cell transplantation. Transl Res. 2017;186:79–94.e1. doi:10.1016/j.trsl.2017.06.008.

152. Lähteenmäki K, Wacklin P, Taskinen M, Tuovinen E, Lohi O, Partanen J, et al. Haematopoietic stem cell transplantation induces severe dysbiosis in intestinal microbiota of paediatric ALL patients. Bone Marrow Transplant. 2017;52:1479–82. doi:10.1038/bmt.2017.168.

153. 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.

154. 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.

155. Peled JU, Gomes A, Buchan ML, Stein-Thoeringer CK, Slingerland JB, Slingerland AE, et al. Intestinal Microbiota Injury during Allo-Hsct Is Generalizable across Transplantation Centers and Is Associated with Increased Mortality, Broad-Spectrum Antibiotics, and Decreased Calorie Intake. Blood. 2017;130 Suppl 1. http://www.bloodjournal.org/content/130/Suppl_1/750?sso-checked=true. Accessed 20 Oct 2018.

156. Mancini N, Greco R, Pasciuta R, Barbanti MC, Pini G, Morrow OB, et al. Enteric Microbiome Markers as Early Predictors of Clinical Outcome in Allogeneic Hematopoietic Stem Cell Transplant: Results of a Prospective Study in Adult Patients. Open Forum Infect Dis. 2017;4. doi:10.1093/ofid/ofx215.

157. Lane JP, Stewart CJ, Cummings SP, Gennery AR. Gut microbiome variations during hematopoietic stem cell transplant in severe combined immunodeficiency. J Allergy Clin Immunol. 2015;135:1654–1656.e2. doi:10.1016/j.jaci.2015.01.024.

158. Ubeda C, Taur Y, Jenq RR, Equinda MJ, Son T, Samstein M, et al. Vancomycin-resistant Enterococcus domination of intestinal microbiota is enabled by antibiotic treatment in mice and precedes bloodstream invasion in humans. J Clin Invest. 2010;120:4332–41. doi:10.1172/JCI43918.

159. Han L, Jin H, Zhou L, Zhang X, Fan Z, Dai M, et al. Intestinal Microbiota at Engraftment Influence Acute Graft-Versus-Host Disease via the Treg/Th17 Balance in Allo-HSCT Recipients. Front Immunol. 2018;9:669. doi:10.3389/fimmu.2018.00669.

160. 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.

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

162. Golob JL, Pergam SA, Srinivasan S, Fiedler TL, Liu C, Garcia K, et al. Stool Microbiota at Neutrophil Recovery Is Predictive for Severe Acute Graft vs Host Disease After Hematopoietic Cell Transplantation. Clin Infect Dis. 2017;65:1984–91. doi:10.1093/cid/cix699.

163. Biagi E, Zama D, Nastasi C, Consolandi C, Fiori J, Rampelli S, et al. Gut microbiota trajectory in pediatric patients undergoing hematopoietic SCT. Bone Marrow Transplant. 2015;50:992–8. doi:10.1038/bmt.2015.16.

164. 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.

165. Weber D, Oefner PJ, Hiergeist a., Koestler J, Gessner a., Weber M, et al. Low urinary indoxyl sulfate levels early after ASCT reflect a disrupted microbiome and are associated with poor outcome. Blood. 2015;126:1723–9. doi:10.1182/blood-2015-04-638858.

166. Taur Y, Pamer EG. The intestinal microbiota and susceptibility to infection in immunocompromised patients. Curr Opin Infect Dis. 2013;26:332–7. doi:10.1097/QCO.0b013e3283630dd3.

167. 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.

168. Montassier E, Al-Ghalith GA, Ward T, Corvec S, Gastinne T, Potel G, et al. Pretreatment gut microbiome predicts chemotherapy-related bloodstream infection. Genome Med. 2016;8:49. doi:10.1186/s13073-016-0301-4.

169. 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.

170. Woo V, Alenghat T. Host-microbiota interactions: epigenomic regulation. Curr Opin Immunol. 2017;44:52–60. doi:10.1016/j.coi.2016.12.001.

171. Romick-Rosendale LE, Haslam DB, Lane A, Denson L, Lake K, Wilkey A, et al. Antibiotic Exposure and Reduced Short Chain Fatty Acid Production after Hematopoietic Stem Cell Transplant. Biol Blood Marrow Transplant. 2018;24:2418–24. doi:10.1016/j.bbmt.2018.07.030.

172. Arpaia N, Campbell C, Fan X, Dikiy S, Van Der Veeken J, Deroos P, et al. Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation. Nature. 2013;504:451–5. doi:10.1038/nature12726.

173. Atarashi K, Tanoue T, Oshima K, Suda W, Nagano Y, Nishikawa H, et al. Treg induction by a rationally selected mixture of Clostridia strains from the human microbiota. Nature. 2013;500:232–6. doi:10.1038/nature12331.

174. Michonneau D, Latis E, Dubouchet L, Latour RP De, Robin M, Fontbrune FS De, et al. Metabolomics Profiling after Allogeneic Hematopoietic Stem Cell Transplantation Unravels a Specific Signature in Human Acute GVHD. Blood. 2018;132 Suppl 1:69–69. doi:10.1182/BLOOD-2018-99-110441.

175. Ghimire S, Matos C, Caioni M, Weber D, Peter K, Holler E, et al. Indoxyl 3-sulfate inhibits maturation and activation of human monocyte-derived dendritic cells. Immunobiology. 2018;223:239–45. doi:10.1016/J.IMBIO.2017.10.014.

176. Gao J, Xu K, Liu H, Liu G, Bai M, Peng C, et al. Impact of the Gut Microbiota on Intestinal Immunity Mediated by Tryptophan Metabolism. Front Cell Infect Microbiol. 2018;8:13. doi:10.3389/fcimb.2018.00013.

177. Bansal T, Alaniz RC, Wood TK, Jayaraman A. The bacterial signal indole increases epithelial-cell tight-junction resistance and attenuates indicators of inflammation. Proc Natl Acad Sci U S A. 2010;107:228–33. doi:10.1073/pnas.0906112107.

178. Gersemann M, Wehkamp J, Stange EF. Innate immune dysfunction in inflammatory bowel disease. J Intern Med. 2012;271:421–8. doi:10.1111/j.1365-2796.2012.02515.x.

179. Meisch JP, Nishimura M, Vogel RM, Sung HC, Bednarchik BA, Ghosh SK, et al. Human β-defensin 3 peptide is increased and redistributed in Crohn's ileitis. Inflamm Bowel Dis. 2013;19:942–53. doi:10.1097/MIB.0b013e318280b11a.

180. Stebe-Frick S, Ostaff MJ, Stange EF, Malek NP, Wehkamp J. Histone deacetylase-mediated regulation of the antimicrobial peptide hBD2 differs in intestinal cell lines and cultured tissue. Sci Rep. 2018;8:1–11. doi:10.1038/s41598-018-31125-x.

181. Soga Y, Maeda Y, Ishimaru F, Tanimoto M, Maeda H, Nishimura F, et al. Bacterial substitution of coagulase-negative staphylococci for streptococci on the oral mucosa after hematopoietic cell transplantation. Support Care Cancer. 2011;19:995–1000.

182. Ames NJ, Sulima P, Ngo T, Barb J, Munson PJ, Paster BJ, et al. A Characterization of the Oral Microbiome in Allogeneic Stem Cell Transplant Patients. PLoS One. 2012;7:e47628. doi:10.1371/journal.pone.0047628.

183. Osakabe L, Utsumi A, Saito B, Okamatsu Y, Kinouchi H, Nakamaki T, et al. Influence of Oral Anaerobic Bacteria on Hematopoietic Stem Cell Transplantation Patients: Oral Mucositis and General Condition. Transplant Proc. 2017;49:2176–82. doi:10.1016/j.transproceed.2017.09.012.

184. Khan N, Peled JU, Gomes AL, Devlin SM, Rondon-Clavo C, Clurman A, et al. Loss of Microbiota Diversity after Autologous Stem Cell Transplant Is Comparable to Injury in Allogeneic Stem Cell Transplant. Blood. 2018;132 Suppl 1:608–608. doi:10.1182/BLOOD-2018-99-110857.

185. Doki N, Suyama M, Sasajima S, Ota J, Igarashi A, Mimura I, et al. Clinical impact of pre-transplant gut microbial diversity on outcomes of allogeneic hematopoietic stem cell transplantation. Ann Hematol. 2017;96:1517–23. doi:10.1007/s00277-017-3069-8.

186. Forcina A, Lorentino F, Marasco V, Oltolini C, Marcatti M, Greco R, et al. Clinical Impact of Pretransplant Multidrug-Resistant Gram-Negative Colonization in Autologous and Allogeneic Hematopoietic Stem Cell Transplantation. Biol Blood Marrow Transplant. 2018. doi:10.1016/J.BBMT.2018.02.021.

187. Bik EM. The Hoops, Hopes, and Hypes of Human Microbiome Research. Yale J Biol Med. 2016;89:363–73. http://www.ncbi.nlm.nih.gov/pubmed/27698620. Accessed 15 Aug 2018.

188. Knudsen BE, Bergmark L, Munk P, Lukjancenko O, Prieme A, Aarestrup FM, et al. Impact of Sample Type and DNA Isolation Procedure on Genomic Inference of Microbiome Composition. bioRxiv. 2016;1:64394. doi:10.1101/064394.

189. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol. 2014;12:87. doi:10.1186/s12915-014-0087-z.

190. Glassing A, Dowd SE, Galandiuk S, Davis B, Chiodini RJ. Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples. Gut Pathog. 2016;8:1–12.

191. Nelson MC, Morrison HG, Benjamino J, Grim SL, Graf J. Analysis, Optimization and Verification of Illumina-Generated 16S rRNA Gene Amplicon Surveys. PLoS One. 2014;9:e94249. doi:10.1371/journal.pone.0094249.

192. Jovel J, Patterson J, Wang W, Hotte N, O'Keefe S, Mitchel T, et al. Characterization of the Gut Microbiome Using 16S or Shotgun Metagenomics. Front Microbiol. 2016;7:459. doi:10.3389/fmicb.2016.00459.

193. Salipante SJ, Kawashima T, Rosenthal C, Hoogestraat DR, Cummings LA, Sengupta DJ, et al.

Performance Comparison of Illumina and Ion Torrent Next-Generation Sequencing Platforms for 16S rRNA-Based Bacterial Community Profiling. Appl Environ Microbiol. 2014;80:7583–91. doi:10.1128/AEM.02206-14.

194. Tremblay J, Singh K, Fern A, Kirton ES, He S, Woyke T, et al. Primer and platform effects on 16S rRNA tag sequencing. Front Microbiol. 2015;6:771. doi:10.3389/fmicb.2015.00771.

195. Nguyen N-P, Warnow T, Pop M, White B. A perspective on 16S rRNA operational taxonomic unit clustering using sequence similarity. npj Biofilms Microbiomes. 2016;2:16004. doi:10.1038/npjbiofilms.2016.4.

196. 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.

197. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13:581–3. doi:10.1038/nmeth.3869.

198. Eren AM, Maignien L, Sul WJ, Murphy LG, Grim SL, Morrison HG, et al. Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data. Methods Ecol Evol. 2013;4:1111–9. doi:10.1111/2041-210X.12114.

199. Callahan B. Silva taxonomic training data formatted for DADA2 (Silva version 132). 2018. doi:10.5281/ZENODO.1172783.

200. Callahan BJ, McMurdie PJ, Holmes SP. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J. 2017;11:2639–43. doi:10.1038/ismej.2017.119.

201. Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome. 2018;6:226. doi:10.1186/s40168-018-0605-2.

202. 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.

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

204. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan: Community Ecology Package. R package version 2.5-2. 2018. https://cran.r-project.org/package=vegan.

205. Rohart F, Gautier B, Singh A, Lê Cao K-A. mixOmics: An R package for 'omics feature selection and multiple data integration. PLOS Comput Biol. 2017;13:e1005752. doi:10.1371/journal.pcbi.1005752.

206. Fukuyama J. treeDA: Tree-Based Discriminant Analysis. 2017. https://github.com/jfukuyama/treeda.

207. Kuhn M, Wing J, Weston S, Williams A, Keefer C, Engelhardt A, et al. caret: Classification and Regression Training. R package version 6.0-80. 2018. https://cran.r-project.org/package=caret.

208. Callahan B, Proctor D, Relman D, Fukuyama J, Holmes S. Reproducible Research Workflow in R for the Analysis of Personalized Human Microbiome Data. Biocomput 2016. 2016;:183–94. doi:10.1142/9789814749411_0018.

209. 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.

210. Huttenhower C, Knight R, Brown CT, Caporaso JG, Clemente JC, Gevers D, et al. Advancing the Microbiome Research Community. Cell. 2014;159:227–30. doi:10.1016/j.cell.2014.09.022.

211. Ravel J, Wommack K. All hail reproducibility in microbiome research. Microbiome. 2014;2:8. doi:10.1186/2049-2618-2-8.

212. Xie Y. Dynamic Documents with R and knitr. Chapman and Hall/CRC; 2016. doi:10.1201/b15166.

213. R Core Team. R: A Language and Environment for Statistical Computing. 2018. https://www.r-project.org/.

214. McMurdie PJ. Normalization of Microbiome Profiling Data. Methods Mol Biol. 2018;1849:143–68.

215. 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.

216. Faust K, Lahti L, Gonze D, de Vos WM, Raes J. Metagenomics meets time series analysis: unraveling microbial community dynamics. Curr Opin Microbiol. 2015;25:56–66.

217. Ridenhour BJ, Brooker SL, Williams JE, Van Leuven JT, Miller AW, Dearing MD, et al. Modeling time-series data from microbial communities. ISME J. 2017;11:2526–37. doi:10.1038/ismej.2017.107.

218. Fukuyama J, Rumker L, Sankaran K, Jeganathan P, Dethlefsen L, Relman DA, et al. Multidomain analyses of a longitudinal human microbiome intestinal cleanout perturbation experiment. PLoS Comput Biol. 2017;13:e1005706. doi:10.1371/journal.pcbi.1005706.

219. Mallick H, Ma S, Franzosa EA, Vatanen T, Morgan XC, Huttenhower C. Experimental design and quantitative analysis of microbial community multiomics. Genome Biol. 2017;18:228. doi:10.1186/s13059-017-1359-z.

220. Johnson LB, Riaz AA, Adawi D, Wittgren L, Bäck S, Thornberg C, et al. Radiation enteropathy and leucocyte-endothelial cell reactions in a refined small bowel model. BMC Surg. 2004;4:10. doi:10.1186/1471-2482-4-10.

221. Nam Y-D, Kim HJ, Seo J-G, Kang SW, Bae J-W. Impact of Pelvic Radiotherapy on Gut Microbiota of Gynecological Cancer Patients Revealed by Massive Pyrosequencing. PLoS One. 2013;8:e82659. doi:10.1371/journal.pone.0082659.

222. Taur Y, Coyte K, Schluter J, Robilotti E, Figueroa C, Gjonbalaj M, et al. Reconstitution of the gut microbiota of antibiotic-treated patients by autologous fecal microbiota transplant. Sci Transl Med. 2018;10:eaap9489. doi:10.1126/scitranslmed.aap9489.

223. DeFilipp Z, Peled JU, Li S, Mahabamunuge J, Dagher Z, Slingerland AE, et al. Third-party fecal microbiota transplantation following allo-HCT reconstitutes microbiome diversity. Blood Adv. 2018;2:745–53. doi:10.1182/bloodadvances.2018017731.

224. DeFilipp Z, Hohmann E, Jenq RR, Chen Y-B. Fecal microbiota transplantation: restoring the injured microbiome after allogeneic hematopoietic cell transplantation. Biol Blood Marrow Transplant. 2018;0:1–6. doi:10.1016/j.bbmt.2018.10.022.

225. Park J, Kim M, Kang SG, Jannasch AH, Cooper B, Patterson J, et al. Short-chain fatty acids induce both effector and regulatory T cells by suppression of histone deacetylases and regulation of the mTOR-S6K

pathway. Mucosal Immunol. 2015;8:80–93. doi:10.1038/mi.2014.44.

226. Choi SW, Gatza E, Hou G, Sun Y, Whitfield J, Song Y, et al. Histone deacetylase inhibition regulates inflammation and enhances Tregs after allogeneic hematopoietic cell transplantation in humans. Blood. 2015;125:815–9. doi:10.1182/blood-2014-10-605238.

227. Hidaka D, Hayase E, Shiratori S, Hasegawa Y, Ishio T, Tateno T, et al. The association between the incidence of intestinal graft-versus-host disease and antibiotic use after allogeneic hematopoietic stem cell transplantation. Clin Transplant. 2018;:e13361. doi:10.1111/ctr.13361.

228. Hammami MB, Talkin R, Al-Taee AM, Schoen MW, Goyal SD, Lai J-P. Autologous Graft-Versus-Host Disease of the Gastrointestinal Tract in Patients With Multiple Myeloma and Hematopoietic Stem Cell Transplantation. Gastroenterol Res. 2018;11:52–7. doi:10.14740/gr925w.

229. Raanani P, Levi I, Holzman F, Grotto I, Brok-Simoni F, Avigdor A, et al. Engraftment-associated hypophosphatemia – the role of cytokine release and steep leukocyte rise post stem cell transplantation. Bone Marrow Transplant. 2001;27:311–7. doi:10.1038/sj.bmt.1702761.

230. Grazziutti ML, Dong L, Miceli MH, Cottler-Fox M, Krishna SG, Fassas A, et al. Recovery from neutropenia can be predicted by the immature reticulocyte fraction several days before neutrophil recovery in autologous stem cell transplant recipients. Bone Marrow Transplant. 2006;37:403–9. doi:10.1038/sj.bmt.1705251.

231. Said HS, Suda W, Nakagome S, Chinen H, Oshima K, Kim S, et al. Dysbiosis of salivary microbiota in inflammatory bowel disease and its association with oral immunological biomarkers. DNA Res. 2014;21:15–25. doi:10.1093/dnares/dst037.

232. Larsen JM. The immune response to *Prevotella* bacteria in chronic inflammatory disease. Immunology. 2017;151:363–74. doi:10.1111/imm.12760.

233. de Aquino SG, Abdollahi-Roodsaz S, Koenders MI, van de Loo FAJ, Pruijn GJM, Marijnissen RJ, et al. Periodontal Pathogens Directly Promote Autoimmune Experimental Arthritis by Inducing a TLR2- and IL-1-Driven Th17 Response. J Immunol. 2014;192:4103–11. doi:10.4049/jimmunol.1301970.

234. Laheij AMGA, de Soet JJ, von dem Borne PA, Kuijper EJ, Kraneveld EA, van Loveren C, et al. Oral bacteria and yeasts in relationship to oral ulcerations in hematopoietic stem cell transplant recipients. Support Care Cancer. 2012;20:3231–40. doi:10.1007/s00520-012-1463-2.

235. Xun Z, Zhang Q, Xu T, Chen N, Chen F. Dysbiosis and Ecotypes of the Salivary Microbiome Associated With Inflammatory Bowel Diseases and the Assistance in Diagnosis of Diseases Using Oral Bacterial Profiles. Front Microbiol. 2018;9:1136. doi:10.3389/fmicb.2018.01136.

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

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

238. 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.

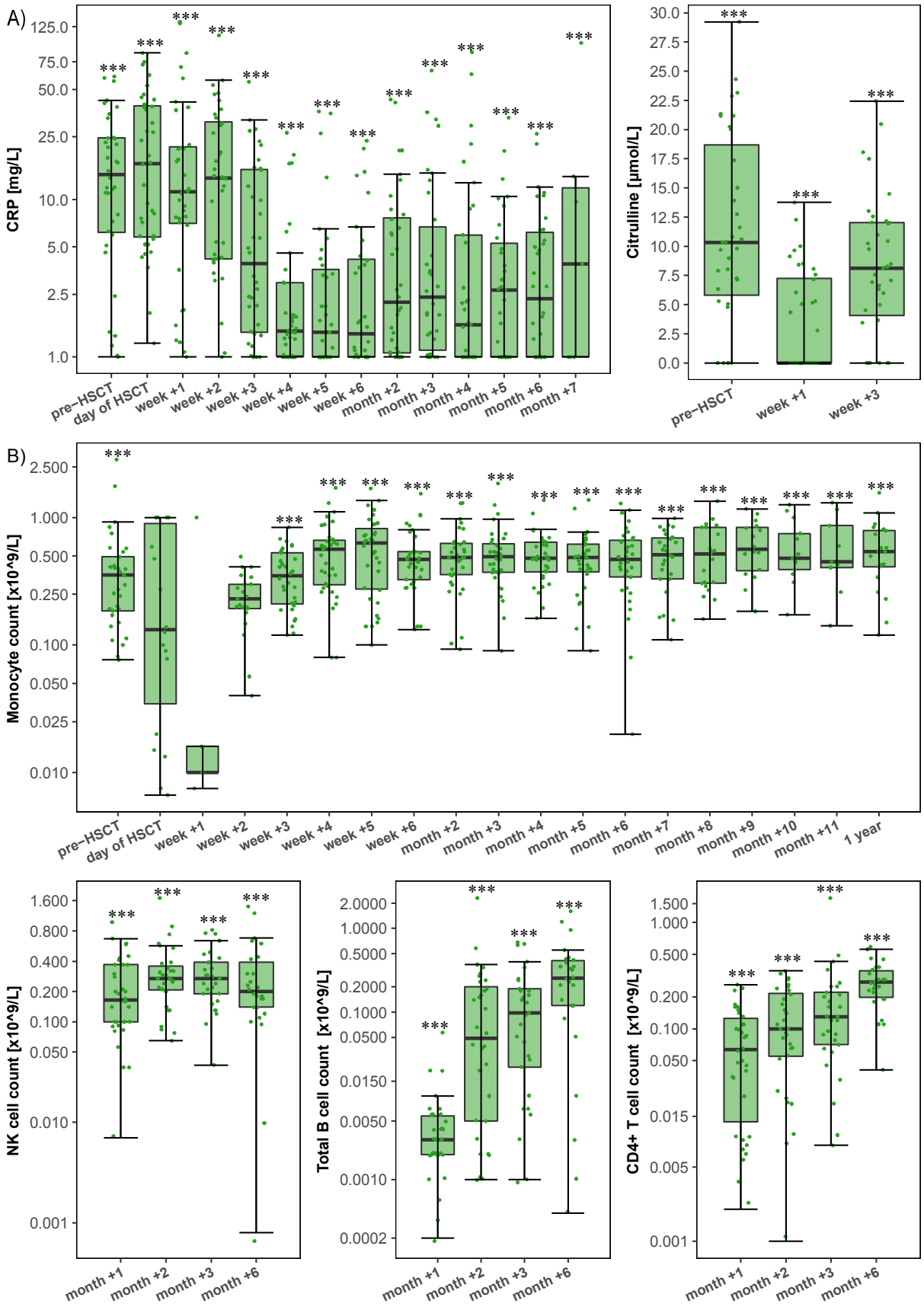
Appendix

Supplementary Material – Manuscript I

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.

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


A)
CRP [mg/L]
125.0
75.0
50.0
25.0
10.0
5.0
2.5
1.0
pre−HSCT
day of HSCT
week +1
week +2
week +3
week +4
week +5
week +6
month +2
month +3
month +4
month +5
month +6
month +7
Citrulline [µmol/L]
30.0
27.5
25.0
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
pre−HSCT
week +1
week +3
B)
Monocyte count [x10^9/L]
2.500
1.000
0.500
0.250
0.100
0.050
0.025
0.010
pre−HSCT
day of HSCT
week +1
week +2
week +3
week +4
week +5
week +6
month +2
month +3
month +4
month +5
month +6
month +7
month +8
month +9
month +10
month +11
1 year
NK cell count [x10^9/L]
1.600
0.800
0.400
0.200
0.100
0.050
0.010
0.001
month +1
month +2
month +3
month +6
Total B cell count [x10^9/L]
2.0000
1.0000
0.5000
0.2000
0.1000
0.0500
0.0150
0.0050
0.0010
0.0002
month +1
month +2
month +3
month +6
CD4+ T cell count [x10^9/L]
1.500
1.000
0.500
0.200
0.100
0.050
0.015
0.005
0.001
month +1
month +2
month +3
month +6

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 *** $P < 0.001$.

Additional file 3: **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.

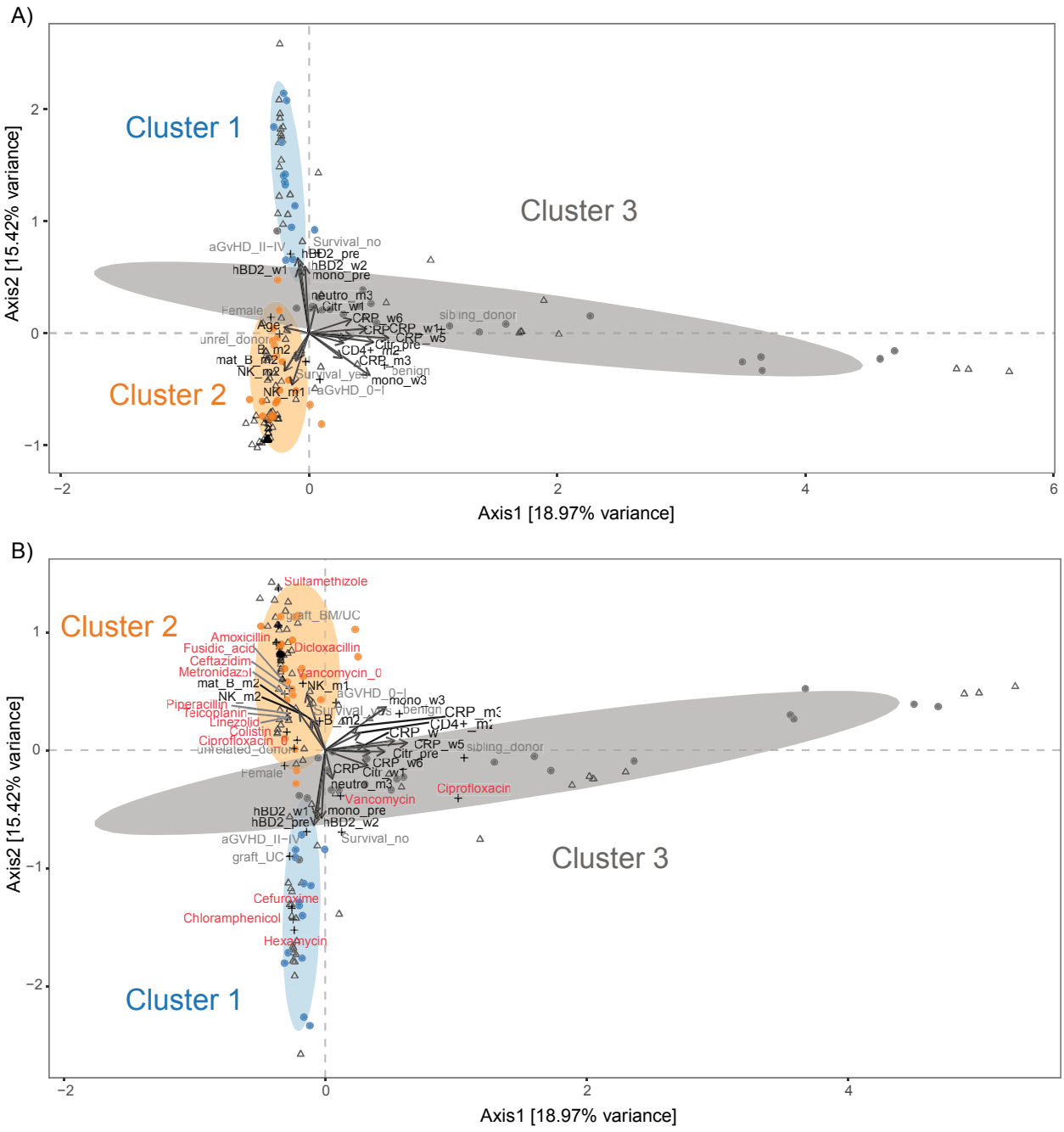
Variable	**SumOfSqs**	**F**	***P*-value**
hBD2_sim	9763	0.4635	0.915
hbd2_pre	89958	4.3373	**0.001**
hbd2_w0	39432	1.9012	**0.037**
hbd2_w1	62425	3.0098	**0.007**
hbd2_w2	70010	3.3755	**0.004**
hbd2_w3	32106	1.5480	0.088
hbd2_m2	33806	1.6300	0.104
CRP_sim	28707	1.3629	**0.015**
CRP_pre	15714	0.8102	0.631
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

Additional file 4: **Table S3: Taxonomy and cluster affiliation of OTUs significantly 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.

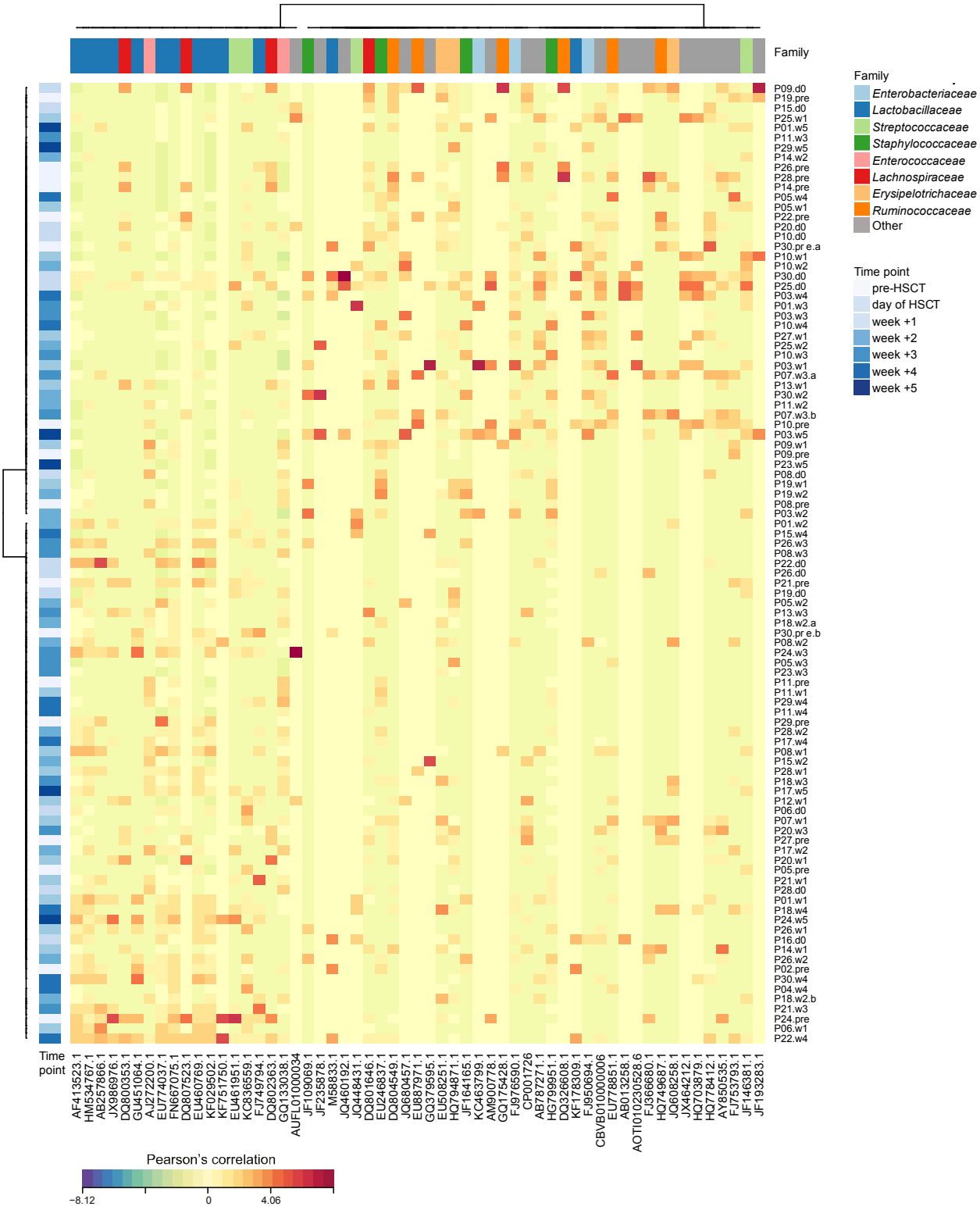
OTU number	Phylum	Order	Family	Genus/Species/Description	Cluster	
					sPLS	CST
HM534767.1	*F*	*Lactobacillales*	*Lactobacillaceae*	*Lactobacillus* sp. Akpobro1	1	1
AF413523.1	*F*	*Lactobacillales*	*Lactobacillaceae*	*Lactobacillus pantheris*	1	1
KF751750.1	*F*	*Lactobacillales*	*Lactobacillaceae*	*Lactobacillus* 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*	*Corynebacteriales*	*Corynebacteriaceae*	*Corynebacterium* sp. DYS15	3	3
GQ379595.1	*P*	*Xanthomonadales*	*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*	*Enterobacteriales*	*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*	*Enterobacteriales*	*Enterobacteriaceae*	*Escherichia coli*	3	3
JF164165.1	*F*	*Bacillales*	*Staphylococcaceae*	*Staphylococcus* sp.	3	3
KC463799.1	*P*	*Enterobacteriales*	*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



Additional file 5: **Figure S2. 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

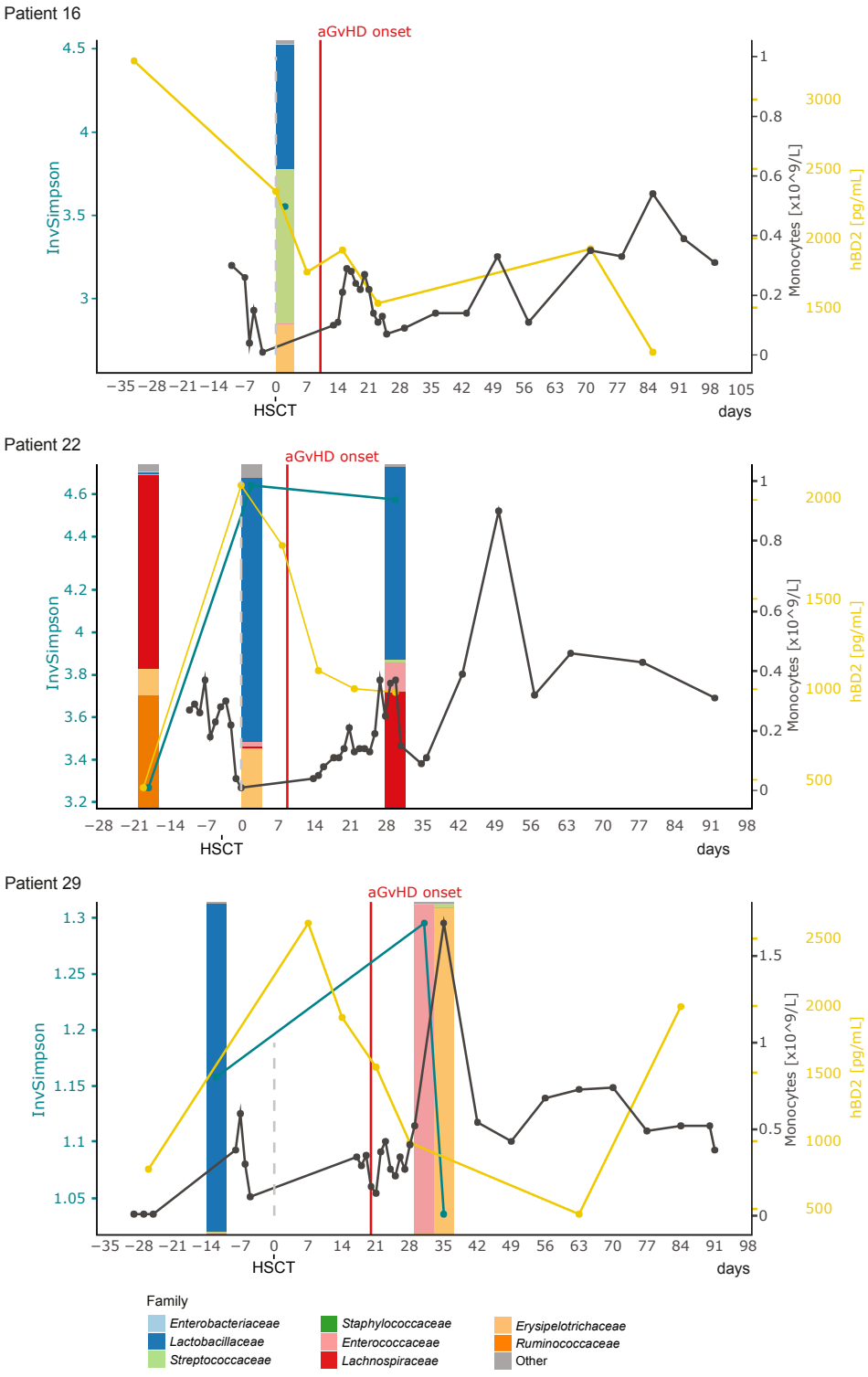
antibiotic treatment at time points simultaneous to microbiome characterization. Antibiotics were added as categorical variables. Depiction of the antibiotic's name (in red) indicates administration of the particular antibiotic, and the extension "_0" indicates no administration of the respective antibiotic. 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.



Additional file 6: **Figure S3. 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.

Additional file 7: **Figure S4. 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.



Additional file 8: **Figure S5. 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.

Additional file 9: **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.

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.

Additional files 10 – 16: **R data analyses reports.** Available online at: https://figshare.com/projects/Specific_gut_microbiome_members_are_associated_with_distinct_immune_markers_in_allogeneic_hematopoietic_stem_cell_transplantation/35201

Additional file 10: **R data analysis report 1**. Data preparation, filtering and transformation.

Additional file 11: **R data analysis report 2**. Bacterial alpha-diversity over time and rank abundance curve for the gut microbiome of HSCT patients.

Additional file 12: **R data analysis report 3**. Temporal patterns of immune markers and immune cells in HSCT patients.

Additional file 13: **R data analysis report 4**. Correlations between immune markers, immune cell counts, and outcomes in HSCT patients.

Additional file 14: **R data analysis report 5**. Variable selection and multivariate analyses of immune parameters and intestinal bacterial taxa in HSCT patients.

Additional file 15: **R data analysis report 6**. Clustering of samples into Community State Types (CSTs) based on Jenson-Shannon divergence.

Additional file 16: **R data analysis report 7**. Longitudinal profiles of microbial community composition and immune markers.

Supplementary Material – Manuscript II

Supplementary Table 1. Patient characteristics. Abbreviations: HLA, human leukocyte antigen; TBI, total body irradiation; CY, Cyclophosphamide; VP16, Etoposide; BU, Busulfan; MEL, Melphalan; GvHD, graft-versus-host disease.

Characteristics		**Number of patients**	**Percentage of all patients (%)**
Transplant recipients		29	100
Average recipient age in years		9.5 (Range: 2.5-16.4)	NA
Gut microbiota characterized	**At ≤ 6 timepoints**	9	31
	At 7-8 timepoints	13	44.8
	At 9-10 timepoints	7	24.1
Oral microbiota chracterized	**At ≤ 6 timepoints**	3	10.3
	At 7-8 timepoints	11	37.9
	At 9-10 timepoints	15	51.7
Nasal microbiota chracterized	**At ≤ 6 timepoints**	3	10.3
	At 7-8 timepoints	9	31
	At 9-10 timepoints	17	58.6
Patient sex	**Female**	13	44.8
	Male	16	55.2
Disease at transplantation	**malignant hematologic diseases**	20	69
	Severe aplastic anemia	1	3.4
	Other benign disorders including immunodeficiencies	8	27.6
Donor type	**HLA-matched sibling**	14	48.3
	HLA-matched unrelated donor (9/10 or 10/10 match)	12	41.4
	Haplo-identical related donor (≤8/10 match)	3	10.3
Stem cell source	**Bone marrow**	23	79.3
	Umbilical cord blood	2	6.9
	Peripheral blood	4	13.8
Conditioning regimen	**TBI + CY / TBI + VP16**	6	20.7
	Combinations of BU, CY, VP16 and MEL	6	20.7
	Fludarabine-thiothepa combinations	12	41.4
	Other fludarabine-based conditioning	5	17.2
Anti-thymocyte globulin treatment		16	55.2

Antibiotics pre- and post-HSCT		29	100
Sex mismatch (female donor to male recipient)		7	24.1
Acute GvHD	**Grade 0-I**	20	69
	Grade II-IV	9	31
Infections	**At least one registered bacterial infection**	25	86.2
	At least one registered viral infection	29	100
	At least one registered fungal infection	14	48.3
Overall Survival*	**Alive**	26	89.7
	Dead	3	10.3
Relapse of primary disease*		2	6.9
Non-relapse mortality		2	6.9
Re-transplantation		1	3.4

*Mean follow-up time after HSCT: 21.4 months (range: 10.1 – 32.7 months)

Supplementary Table 2. Taxonomy of LDA clade members and corresponding LDA-scores in the gut, oral cavity, and nasal cavity.

ASV number	Body site	Phylum	Family	Genus/Species/ Description	LDA score	Number of patients with presence of this ASV at ≥ 1 time point
ASV_1	gut	*Firmicutes*	*Enterococcaceae*	*Enterococcus* sp.	0.082	29
ASV_30	gut	*Firmicutes*	*Enterococcaceae*	*Enterococcus* sp.	0.006	26
ASV_158	gut	*Firmicutes*	*Enterococcaceae*	*Enterococcus* sp.	0.006	13
ASV_178	gut	*Firmicutes*	*Enterococcaceae*	*Enterococcus* sp.	0.006	12
ASV_344	gut	*Firmicutes*	*Enterococcaceae*	*Enterococcus* sp.	0.006	8
ASV_395	gut	*Firmicutes*	*Enterococcaceae*	*Enterococcus* sp.	0.012	14
ASV_424	gut	*Firmicutes*	*Enterococcaceae*	*Enterococcus* sp.	0.012	9
ASV_543	gut	*Firmicutes*	*Enterococcaceae*	*Enterococcus* sp.	0.006	7
ASV_552	gut	*Firmicutes*	*Enterococcaceae*	*Enterococcus avium*	0.006	7
ASV_720	gut	*Firmicutes*	*Enterococcaceae*	*Enterococcus* sp.	0.078	15
ASV_730	gut	*Firmicutes*	*Enterococcaceae*	*Enterococcus* sp.	0.006	12
ASV_784	gut	*Firmicutes*	*Enterococcaceae*	*Enterococcus* sp.	0.006	8
ASV_951	gut	*Firmicutes*	*Enterococcaceae*	*Enterococcus* sp.	0.006	12
ASV_1186	gut	*Firmicutes*	*Enterococcaceae*	*Melissococcus* sp.	0.006	9
ASV_3	gut	*Firmicutes*	*Lactobacillaceae*	*Lactobacillus* sp.	0.011	26
ASV_31	gut	*Firmicutes*	*Lactobacillaceae*	*Lactobacillus* sp.	0.006	11
ASV_67	gut	*Firmicutes*	*Lactobacillaceae*	*Lactobacillus sakei*	0.006	15
ASV_74	gut	*Firmicutes*	*Lactobacillaceae*	*Lactobacillus* sp.	0.006	16
ASV_144	gut	*Firmicutes*	*Lactobacillaceae*	*Pediococcus* sp.	0.006	10
ASV_151	gut	*Firmicutes*	*Lactobacillaceae*	*Pediococcus pentosaceus*	0.006	12
ASV_189	gut	*Firmicutes*	*Lactobacillaceae*	*Lactobacillus* sp.	0.006	10
ASV_222	gut	*Firmicutes*	*Lactobacillaceae*	*Lactobacillus rhamnosus*	0.006	8
ASV_586	gut	*Firmicutes*	*Lactobacillaceae*	*Lactobacillus* sp.	0.006	9
ASV_78	gut	*Firmicutes*	*Lachnospiraceae*	*Blautia* sp.	-0.039	26
ASV_237	gut	*Firmicutes*	*Lachnospiraceae*	*Blautia faecis*	-0.026	23
ASV_265	gut	*Firmicutes*	*Lachnospiraceae*	*Blautia* sp.	-0.026	24
ASV_271	gut	*Firmicutes*	*Lachnospiraceae*	*Blautia obeum*	-0.026	12
ASV_375	gut	*Firmicutes*	*Lachnospiraceae*	*Blautia* sp.	-0.026	15
ASV_554	gut	*Firmicutes*	*Lachnospiraceae*	*Blautia* sp.	-0.026	16
ASV_1291	gut	*Firmicutes*	*Lachnospiraceae*	*Blautia* sp.	-0.026	9
ASV_132	gut	*Firmicutes*	*Ruminococcaceae*	*Ruminococcus bromii*	-0.007	16
ASV_192	gut	*Firmicutes*	*Ruminococcaceae*	*Ruminococcus bromii*	-0.007	24
ASV_206	gut	*Firmicutes*	*Ruminococcaceae*	DTU089	-0.024	23
ASV_306	gut	*Firmicutes*	*Ruminococcaceae*	*Ruminococcus* sp.	-0.007	13
ASV_348	gut	*Firmicutes*	*Ruminococcaceae*	*Ruminococcus* sp.	-0.007	11

ASV_556	gut	*Firmicutes*	*Ruminococcaceae*	*Caproiciproducens* sp.	-0.024	9
ASV_566	gut	*Firmicutes*	*Ruminococcaceae*	*Ruminoclostridium sp.*	-0.024	13
ASV_583	gut	*Firmicutes*	*Ruminococcaceae*	*Ruminoclostridium sp.*	-0.024	23
ASV_682	gut	*Firmicutes*	*Ruminococcaceae*	*Caproiciproducens* sp.	-0.024	9
ASV_750	gut	*Firmicutes*	*Ruminococcaceae*	*Caproiciproducens* sp.	-0.024	13
ASV_792	gut	*Firmicutes*	*Ruminococcaceae*	*Ruminoclostridium sp.*	-0.024	15
ASV_18	oral	*Actinobacteria*	*Actinomycetaceae*	*Actinomyces* sp.	0.021	27
ASV_66	oral	*Actinobacteria*	*Actinomycetaceae*	*Actinomyces* sp.	0.003	26
ASV_117	oral	*Actinobacteria*	*Actinomycetaceae*	*Actinomyces naeslundii*	0.016	27
ASV_126	oral	*Actinobacteria*	*Actinomycetaceae*	*Actinomyces naeslundii*	0.016	19
ASV_138	oral	*Actinobacteria*	*Actinomycetaceae*	*Actinomyces* sp.	0.003	22
ASV_155	oral	*Actinobacteria*	*Actinomycetaceae*	*Actinomyces odontolyticus*	0.003	20
ASV_227	oral	*Actinobacteria*	*Actinomycetaceae*	*Actinomyces* sp.	0.016	7
ASV_235	oral	*Actinobacteria*	*Actinomycetaceae*	F0332	0.003	18
ASV_262	oral	*Actinobacteria*	*Actinomycetaceae*	*Actinomyces* sp.	0.003	18
ASV_345	oral	*Actinobacteria*	*Actinomycetaceae*	*Actinomyces* sp.	0.016	9
ASV_389	oral	*Actinobacteria*	*Actinomycetaceae*	*Actinomyces odontolyticus*	0.003	7
ASV_403	oral	*Actinobacteria*	*Actinomycetaceae*	*Actinomyces* sp.	0.003	15
ASV_407	oral	*Actinobacteria*	*Actinomycetaceae*	*Actinomyces* sp.	0.016	7
ASV_2693	oral	*Actinobacteria*	*Actinomycetaceae*	*Actinomyces* sp.	0.021	5
ASV_422	oral	*Actinobacteria*	*Actinomycetaceae*	*Actinomyces odontolyticus*	0.003	15
ASV_431	oral	*Actinobacteria*	*Actinomycetaceae*	*Actinomyces* sp.	0.016	12
ASV_2697	oral	*Actinobacteria*	*Actinomycetaceae*	*Actinomyces naeslundii*	0.016	9
ASV_461	oral	*Actinobacteria*	*Actinomycetaceae*	*Actinomyces* sp.	0.003	11
ASV_475	oral	*Actinobacteria*	*Actinomycetaceae*	*Actinomyces gerencseriae*	0.003	6
ASV_484	oral	*Actinobacteria*	*Actinomycetaceae*	*Actinomyces* sp.	0.003	4
ASV_501	oral	*Actinobacteria*	*Actinomycetaceae*	*Actinomyces odontolyticus*	0.003	13
ASV_516	oral	*Actinobacteria*	*Actinomycetaceae*	*Actinomyces odontolyticus*	0.003	12
ASV_2700	oral	*Actinobacteria*	*Actinomycetaceae*	*Actinomyces gerencseriae*	0.003	14
ASV_568	oral	*Actinobacteria*	*Actinomycetaceae*	*Actinomyces* sp.	0.003	12
ASV_600	oral	*Actinobacteria*	*Actinomycetaceae*	*Actinomyces* sp.	0.003	7
ASV_642	oral	*Actinobacteria*	*Actinomycetaceae*	*Actinomyces* sp.	0.003	7
ASV_798	oral	*Actinobacteria*	*Actinomycetaceae*	F0332	0.003	10
ASV_871	oral	*Actinobacteria*	*Actinomycetaceae*	*Actinomyces massiliensis*	0.003	13
ASV_2729	oral	*Actinobacteria*	*Actinomycetaceae*	*Actinomyces graevenitzii*	0.003	9
ASV_1055	oral	*Actinobacteria*	*Actinomycetaceae*	*Actinomyces* sp.	0.003	6
ASV_1172	oral	*Actinobacteria*	*Actinomycetaceae*	*Actinomyces* sp.	0.003	7
ASV_2751	oral	*Actinobacteria*	*Actinomycetaceae*	*Actinomyces* sp.	0.016	8

ASV_2664	oral	*Firmicutes*	*Streptococcaceae*	*Streptococcus* sp.	0.010	29
ASV_10	oral	*Firmicutes*	*Streptococcaceae*	*Streptococcus* sp.	0.010	29
ASV_16	oral	*Firmicutes*	*Streptococcaceae*	*Streptococcus* sp.	0.010	29
ASV_27	oral	*Firmicutes*	*Streptococcaceae*	*Streptococcus* sp.	0.010	29
ASV_28	oral	*Firmicutes*	*Streptococcaceae*	*Streptococcus* sp.	0.040	28
ASV_37	oral	*Firmicutes*	*Streptococcaceae*	*Streptococcus* sp.	0.010	19
ASV_48	oral	*Firmicutes*	*Streptococcaceae*	*Streptococcus* sp.	0.010	27
ASV_173	oral	*Firmicutes*	*Streptococcaceae*	*Streptococcus salivarius*	0.010	15
ASV_183	oral	*Firmicutes*	*Streptococcaceae*	*Streptococcus* sp.	0.010	19
ASV_188	oral	*Firmicutes*	*Streptococcaceae*	*Streptococcus* sp.	0.010	14
ASV_2674	oral	*Firmicutes*	*Streptococcaceae*	*Streptococcus* sp.	0.010	14
ASV_230	oral	*Firmicutes*	*Streptococcaceae*	*Streptococcus* sp.	0.010	8
ASV_269	oral	*Firmicutes*	*Streptococcaceae*	*Streptococcus cristatus*	0.010	29
ASV_282	oral	*Firmicutes*	*Streptococcaceae*	*Streptococcus parasanguinis*	0.010	13
ASV_2683	oral	*Firmicutes*	*Streptococcaceae*	*Streptococcus mitis*	0.010	11
ASV_480	oral	*Firmicutes*	*Streptococcaceae*	*Streptococcus* sp.	0.010	18
ASV_481	oral	*Firmicutes*	*Streptococcaceae*	*Streptococcus peroris*	0.010	12
ASV_802	oral	*Firmicutes*	*Streptococcaceae*	*Streptococcus* sp.	0.010	24
ASV_1531	oral	*Firmicutes*	*Streptococcaceae*	*Streptococcus* sp.	0.010	11
ASV_1599	oral	*Firmicutes*	*Streptococcaceae*	*Streptococcus* sp.	0.010	13
ASV_42	oral	*Bacteroidetes*	*Prevotellaceae*	*Prevotella melaninogenica*	0.028	27
ASV_226	oral	*Bacteroidetes*	*Prevotellaceae*	*Prevotella melaninogenica*	0.028	15
ASV_800	oral	*Bacteroidetes*	*Prevotellaceae*	*Prevotella* sp.	0.028	6
ASV_2665	oral	*Firmicutes*	Family XI	*Gemella* sp.	0.009	29
ASV_208	oral	*Firmicutes*	Family XI	*Gemella sanguinis*	0.009	28
ASV_2701	oral	*Firmicutes*	Family XI	*Gemella* sp.	0.009	5
ASV_14	nasal	*Actinobacteria*	*Corynebacteriaceae*	*Corynebacterium* sp.	0.117	24
ASV_354	nasal	*Actinobacteria*	*Corynebacteriaceae*	*Corynebacterium durum*	0.054	15
ASV_360	nasal	*Actinobacteria*	*Corynebacteriaceae*	*Corynebacterium* sp.	0.054	17
ASV_2704	nasal	*Actinobacteria*	*Corynebacteriaceae*	*Corynebacterium* sp.	0.022	12
ASV_2707	nasal	*Actinobacteria*	*Corynebacteriaceae*	*Corynebacterium durum*	0.054	14
ASV_1206	nasal	*Actinobacteria*	*Corynebacteriaceae*	*Lawsonella* sp.	0.007	16

Supplentary Table 3. Taxonomy of aGvHD predictors within the fecal, oral, and nasal microbiota.

ASV number	Phylum	Family	Genus/Species/ Description	Body Site	Importance in svmLinear model	Mean Importance in Boruta
ASV_3	***Firmicutes***	***Lactobacillaceae***	***Lactobacillus sp.***	**gut**	**0.74**	**8.85**
ASV_7	*Bacteroidetes*	*Tannerellaceae*	*Parabacteroides merdae*	gut	0.63	4.62
ASV_8	*Proteobacteria*	*Enterobacteriaceae*	*Escherichia/Shigella* sp.	gut	0.70	4.16
ASV_12	*Firmicutes*	*Ruminococcaceae*	UBA1819 sp.	gut	0.68	5.52
ASV_35	*Bacteroidetes*	*Bacteroidaceae*	*Bacteroides uniformis*	gut	0.68	4.66
ASV_50	*Firmicutes*	*Ruminococcaceae*	*Subdoligranulum* sp.	gut	0.60	3.78
ASV_128	***Bacteroidetes***	***Tannerellaceae***	***Parabacteroides distasonis***	**gut**	**0.68**	**7.42**
ASV_131	*Bacteroidetes*	*Tannerellaceae*	*Parabacteroides distasonis*	gut	0.66	6.49
ASV_166	*Fusobacteria*	*Fusobacteriaceae*	*Fusobacterium* sp.	gut	0.60	4.08
ASV_268	***Firmicutes***	***Lachnospiraceae***	***Lachnospiraceae_NK4A136_group* sp.**	**gut**	**0.61**	**5.02**
ASV_361	*Bacteroidetes*	*Tannerellaceae*	*Parabacteroides* sp.	gut	0.61	4.40
ASV_477	*Firmicutes*	*Lachnospiraceae*	*Eisenbergiella tayi*	gut	0.62	3.64
ASV_507	*Firmicutes*	*Lachnospiraceae*	*Lachnospira pectinoschiza*	gut	0.60	3.91
ASV_563	*Firmicutes*	*Erysipelotrichaceae*	*Coprobacillus cateniformis*	gut	0.63	5.38
ASV_567	*Firmicutes*	*Ruminococcaceae*	*Flavonifractor* sp.	gut	0.60	2.81
ASV_585	*Bacteroidetes*	*Tannerellaceae*	*Parabacteroides* sp.	gut	0.62	4.53
ASV_687	*Firmicutes*	*Lachnospiraceae*	NA	gut	0.59	3.83
ASV_15	*Firmicutes*	*Streptococcaceae*	*Lactococcus* sp.	oral	0.61	7.31
ASV_34	*Proteobacteria*	*Neisseriaceae*	*Neisseria* sp.	oral	0.62	5.08
ASV_66	*Actinobacteria*	*Actinomycetaceae*	*Actinomyces* sp.	oral	0.65	4.84
ASV_72	*Actinobacteria*	*Micrococcaceae*	*Rothia* sp.	oral	0.63	4.49
ASV_92	*Bacteroidetes*	*Bacteroidaceae*	*Bacteroides* sp.	oral	0.65	5.36
ASV_98	*Actinobacteria*	*Micrococcaceae*	*Rothia mucilaginosa*	oral	0.58	5.87
ASV_135	*Firmicutes*	*Streptococcaceae*	*Streptococcus mutans*	oral	0.60	4.24
ASV_205	*Firmicutes*	*Ruminococcaceae*	*Ruminococcaceae_UCG-002* sp.	oral	0.63	4.38
ASV_226	***Bacteroidetes***	***Prevotellaceae***	***Prevotella_7 melaninogenica***	**oral**	**0.63**	**8.19**
ASV_247	*Firmicutes*	*Ruminococcaceae*	*Ruminococcaceae_UCG-013* sp.	oral	0.66	4.80
ASV_270	*Firmicutes*	*Veillonellaceae*	*Veillonella* sp.	oral	0.59	3.47
ASV_338	*Bacteroidetes*	*Prevotellaceae*	*Alloprevotella* sp.	oral	0.61	4.26

ASV_360	*Actinobacteria*	*Corynebacteriaceae*	*Corynebacterium* sp.	oral	0.64	6.50
ASV_374	*Fusobacteria*	*Leptotrichiaceae*	*Leptotrichia wadei*	oral	0.60	4.39
ASV_378	*Firmicutes*	*Lachnospiraceae*	*Oribacterium sinus*	oral	0.65	7.17
ASV_500	***Actinobacteria***	***Propionibacteriaceae***	***Pseudopropionibacterium propionicum***	**oral**	**0.67**	**8.51**
ASV_518	*Bacteroidetes*	*Bacteroidaceae*	*Bacteroides* sp.	oral	0.64	4.67
ASV_568	***Actinobacteria***	***Actinomycetaceae***	***Actinomyces* sp.**	**oral**	**0.71**	**8.79**
ASV_640	*Bacteroidetes*	*Tannerellaceae*	*Parabacteroides* sp.	oral	0.62	2.81
ASV_871	*Actinobacteria*	*Actinomycetaceae*	*Actinomyces massiliensis*	oral	0.61	4.65
ASV_894	*Firmicutes*	*Ruminococcaceae*	*Ruminococcaceae_UCG-014* sp.	oral	0.62	2.74
ASV_1403	*Firmicutes*	*Veillonellaceae*	*Selenomonas_3* sp.	oral	0.59	2.82
ASV_2664	*Firmicutes*	*Streptococcaceae*	*Streptococcus* sp.	oral	0.68	6.11
ASV_2666	*Actinobacteria*	*Micrococcaceae*	*Rothia dentocariosa*	oral	0.61	4.47
ASV_2676	*Firmicutes*	*Aerococcaceae*	*Abiotrophia defectiva*	oral	0.61	5.74
ASV_2692	*Actinobacteria*	*Micrococcaceae*	*Rothia* sp.	oral	0.62	6.30
ASV_2664	*Firmicutes*	*Streptococcaceae*	*Streptococcus* sp.	nasal	0.61	11.49
ASV_2666	*Actinobacteria*	*Micrococcaceae*	*Rothia dentocariosa*	nasal	0.56	5.80
ASV_47	***Actinobacteria***	***Micrococcaceae***	***Rothia* sp.**	**nasal**	**0.59**	**6.41**
ASV_51	*Proteobacteria*	*Pasteurellaceae*	*Haemophilus* sp.	nasal	0.64	5.67
ASV_52	*Actinobacteria*	*Micrococcaceae*	*Rothia mucilaginosa*	nasal	0.61	8.58
ASV_66	***Actinobacteria***	***Actinomycetaceae***	***Actinomyces* sp.**	**nasal**	**0.67**	**9.05**
ASV_2670	*Proteobacteria*	*Neisseriaceae*	NA	nasal	0.61	6.69
ASV_117	*Actinobacteria*	*Actinomycetaceae*	*Actinomyces naeslundii*	nasal	0.58	4.73
ASV_125	*Actinobacteria*	*Corynebacteriaceae*	*Corynebacterium_1 accolens*	nasal	0.57	5.43
ASV_252	*Firmicutes*	*Veillonellaceae*	*Veillonella rogosae*	nasal	0.58	5.54
ASV_270	*Firmicutes*	*Veillonellaceae*	*Veillonella* sp.	nasal	0.57	5.41
ASV_2694	*Actinobacteria*	*Corynebacteriaceae*	*Lawsonella* sp.	nasal	0.65	8.42

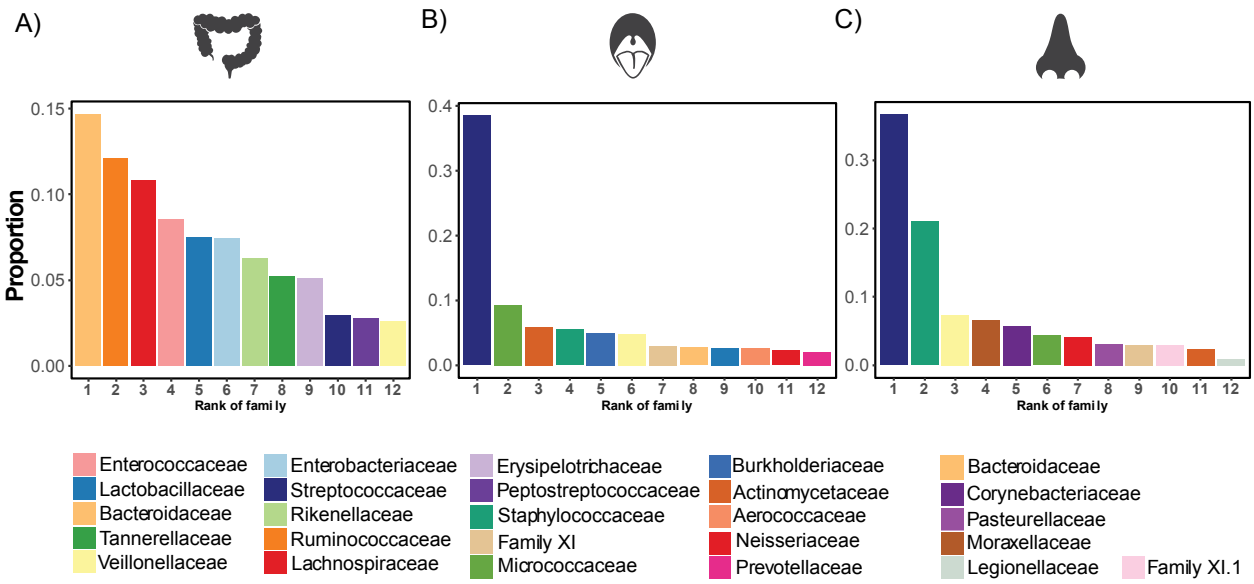


Figure S1. Most abundant taxonomic families in the gut, oral cavity, and nasal cavity in allo-HSCT patients. Rank abundance curves displaying the proportions of the 12 most abundant taxonomic families at each body site (gut, oral cavity, and nasal cavity).

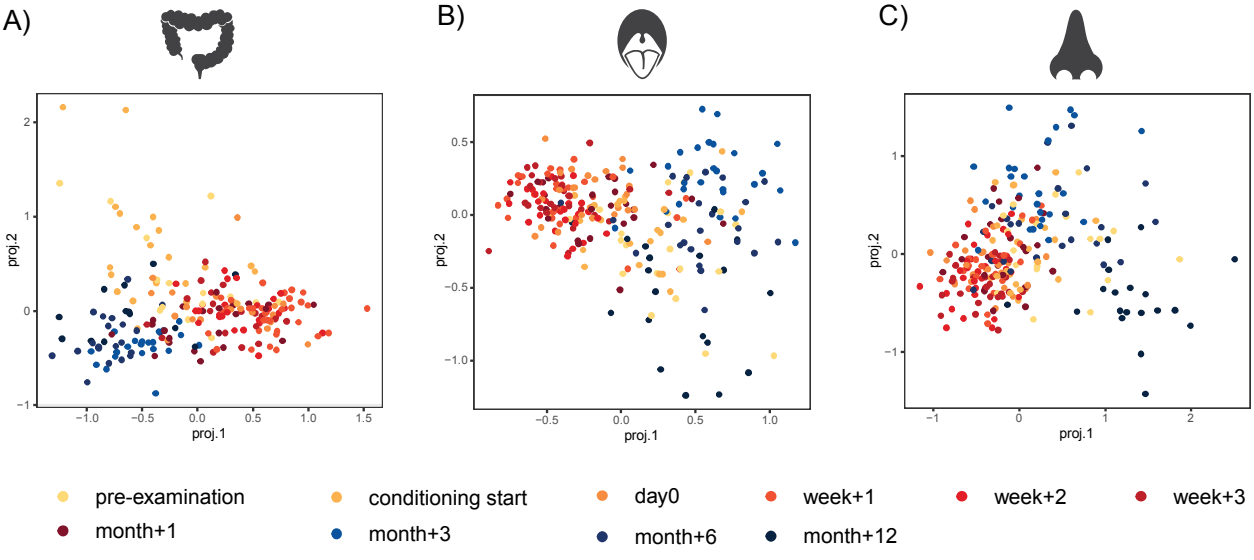


Figure S2. Sample scores on the first two axes of tree-based sparse linear discriminant (LDA) analyses by time point in relation to HSCT. Tree-based sparse LDA performs phylogeny-informed selection of microbial predictors. The projection maximizes differences between samples from different time points. A) Fecal sample scores of LDA axis 2 versus 1 showing separation of time points on LDA axis 1 with the highest scores assigned to samples collected immediately post HSCT. B) Oral sample scores of LDA axis 2 versus 1 showing separation of time points on LDA axis 1 with the highest scores assigned to samples from before HSCT and from late follow up-time points. C) Nasal sample scores of LDA axis 2 versus 1 showing separation of time points on LDA axis 1 with the highest scores assigned to from before HSCT and from late follow up-time points.

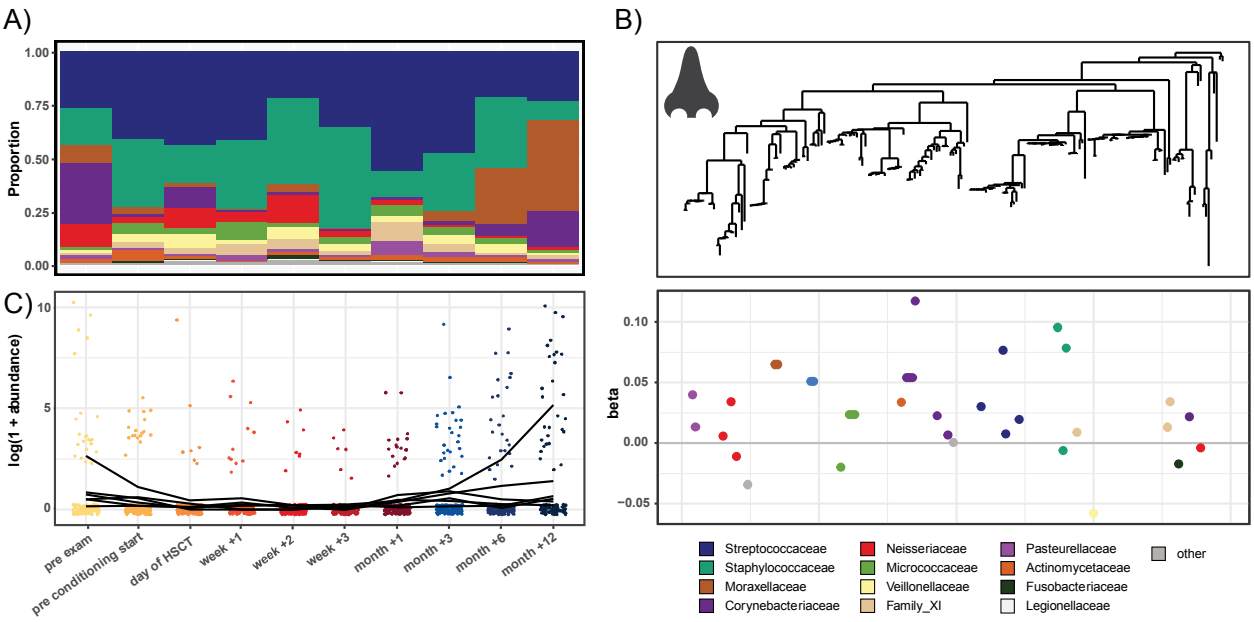


Figure S3. Tree-based sparse linear discriminant analysis revealing nasal ASVs that distinguish time points in relation to HSCT. A) Relative abundances over time of the 12 most abundant families in the oral cavity. B) Coefficients of discriminating clades of ASVs on the first LDA axis, colored by taxonomic family, and plotted along the phylogenetic tree. C) Trajectories of ASVs in one discriminating group, affiliated with the families *Corynebacteriaceae,* with decreasing abundances after HSCT and recovery at late follow-up time points. Detailed taxonomic information and LDA-coefficients of the displayed ASVs are listed in Table S2.

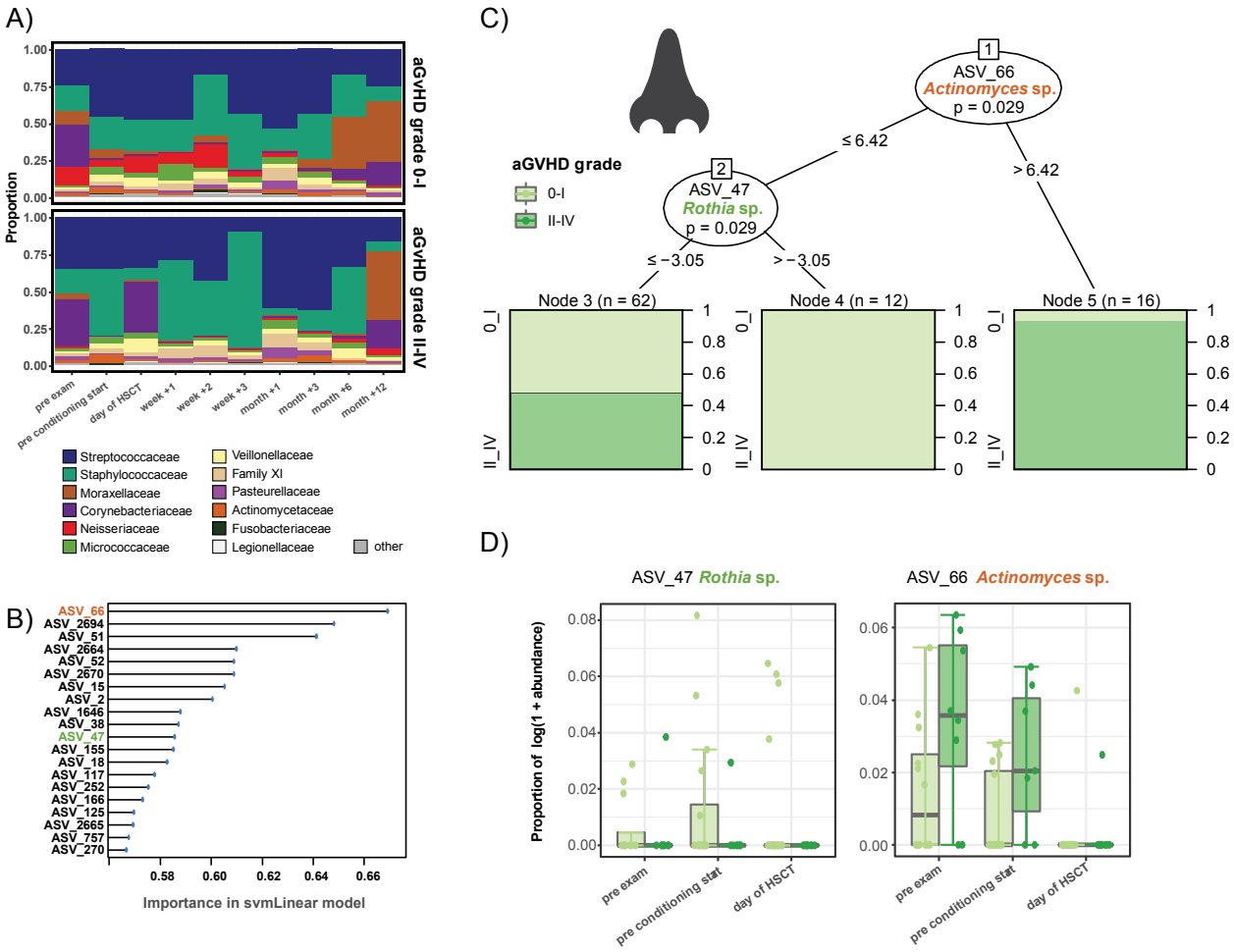
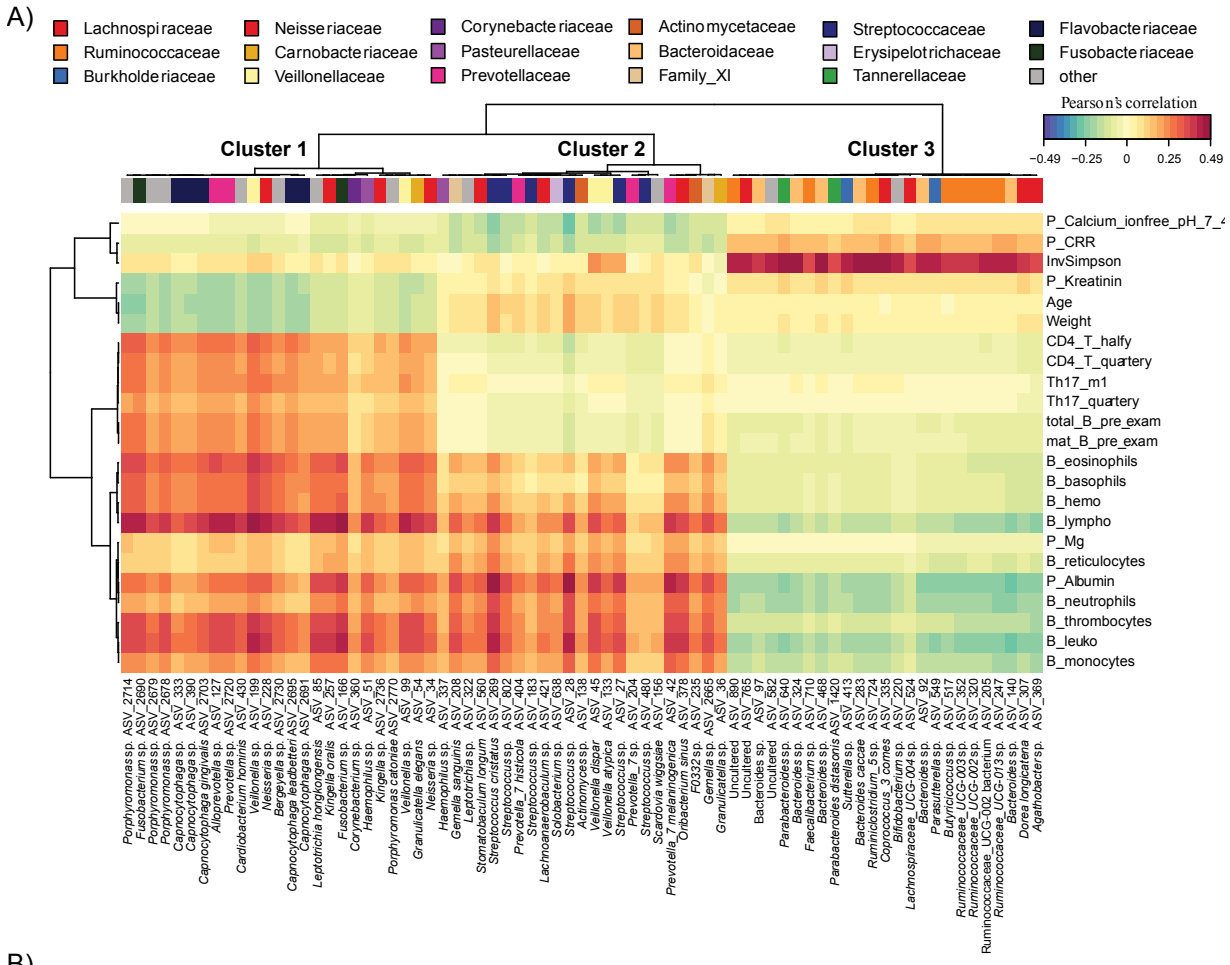


Figure S4. Machine learning-based prediction of aGvHD severity from nasal microbial abundances pre-HSCT. A) Relative abundances over time of the 12 most abundant families in the nasal cavity in patients with aGvHD grade 0-I versus II-IV. B) Importance plot of top 20 predictive nasal ASVs identified by the svmLinear model with importance scores indicating the mean decrease in prediction accuracy in case the respective ASV would be excluded from the model. C) Conditional inference tree (CTREE) displaying ASVs identified as significant split nodes by nonparametric regression for prediction of aGvHD. Numbers along the branches indicate split values of variance stabilized bacterial abundances. The terminal nodes show the proportion of samples originating from patients with aGvHD grade 0-I vs II-IV. D) Boxplots depict the log transformed relative abundances of the predictive ASVs at time points up to the transplantation in aGvHD grade 0-I compared with grade II-IV patients.


A)
Lachnospiraceae
Ruminococcaceae
Burkholderiaceae
Neisseriaceae
Carnobacteriaceae
Veillonellaceae
Corynebacteriaceae
Pasteurellaceae
Prevotellaceae
Actinomycetaceae
Bacteroidaceae
Family_XI
Streptococcaceae
Erysipelotrichaceae
Tannerellaceae
Flavobacteriaceae
Fusobacteriaceae
other
Pearson's correlation
−0.49 −0.25 0 0.25 0.49
Cluster 1
Cluster 2
Cluster 3
P_Calcium_ionfree_pH_7_4
P_CRR
InvSimpson
P_Kreatinin
Age
Weight
CD4_T_halfy
CD4_T_quartery
Th17_m1
Th17_quartery
total_B_pre_exam
mat_B_pre_exam
B_eosinophils
B_basophils
B_hemo
B_lympho
P_Mg
B_reticulocytes
P_Albumin
B_neutrophils
B_thrombocytes
B_leuko
B_monocytes

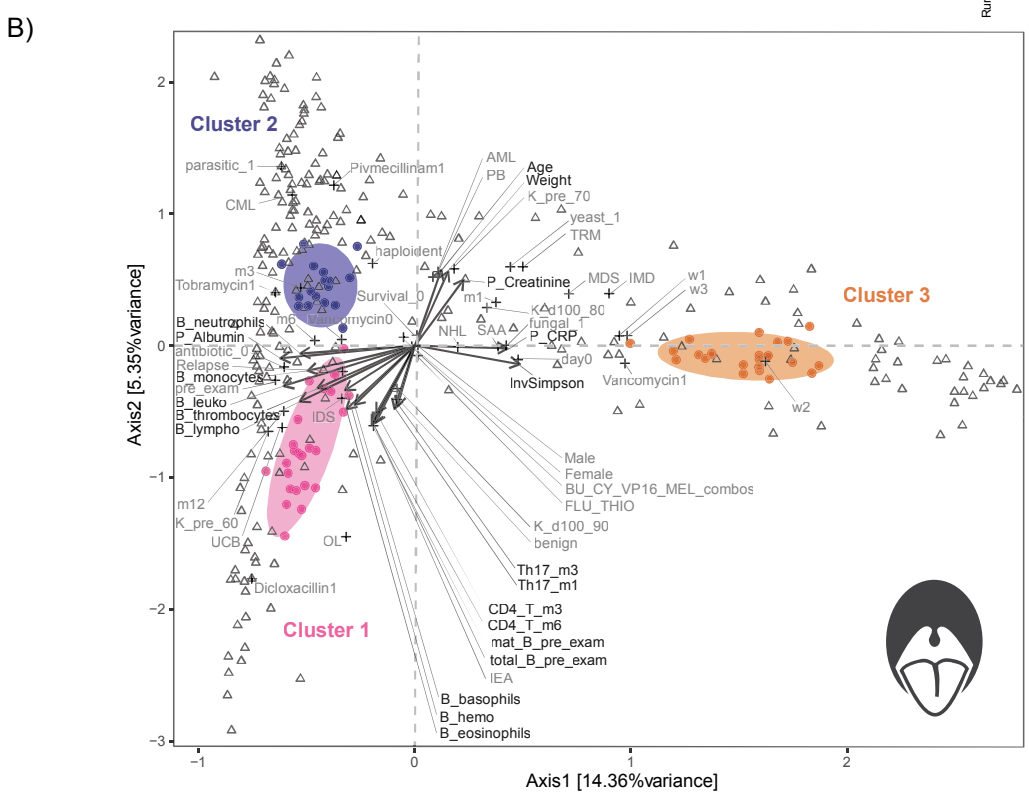

B)
Cluster 2
Cluster 1
Cluster 3
parasitic_1
CML
Pivmecillinam1
AML
PB
Age
Weight
K_pre_70
yeast_1
TRM
haploident
m3
Tobramycin1
Survival_0
m1
P_Creatinine
MDS
IMD
w1
w3
K_d100_80
fungal_1
NHL
SAA
P_CRP
day0
InvSimpson
Vancomycin1
w2
B_neutrophils
P_Albumin
antibiotic_0
Relapse
B_monocytes
pre_exam
B_leuko
B_thrombocytes
B_lympho
IDS
Male
Female
BU_CY_VP16_MEL_combos
FLU_THIO
K_d100_90
benign
m12
K_pre_60
UCB
OL
Dicloxacillin1
Th17_m3
Th17_m1
CD4_T_m3
CD4_T_m6
mat_B_pre_exam
total_B_pre_exam
IEA
B_basophils
B_hemo
B_eosinophils
Axis2 [5.35%variance]
Axis1 [14.36%variance]

Figure S5. Multivariate associations of the oral microbiota with immune and clinical parameters in HSCT. A) Clustered image map (CIM) based on sparse partial least squares (sPLS) regression analysis dimensions 1, 2, and 3, displaying pairwise correlations >0.2/<-0.2 between oral ASVs (bottom), and continuous immune and clinical parameters (right). Red indicated positive correlation, and blue indicates negative correlation, respectively. Based on the sPLS regression model, hierarchical clustering (clustering method: complete linkage, distance method: Pearson's correlation) was performed resulting in the three depicted clusters. B) Canonical correspondence analysis (CCpnA) relating oral microbial abundances (circles) to continuous (arrows) and categorical (+) immune and clinical parameters. ASVs and variables with at least one correlation >0.2/<-0.2 in the sPLS analysis were included in the CCpnA. The triplot shows variables and ASVs with a score >0.3/<-0.3 on at least one the first three CCpnA axes, displayed on axis 1 versus 2 with samples depicted as triangles. The colored ellipses (depicted with 80% confidence interval) correspond to the clusters of ASVs identified by the sPLS-based hierarchical clustering. For visualization purposes, a focused section of the CCpnA triplot is shown. Abbreviations are described in Figure 6. Additional abbreviations: fungal, fungal infection; haploident, haploidentical donor; hemo, hemoglobin; leuko, leukocytes; lympho, lymphocytes; w1, week+1; w2, week+2; w3, week+3.

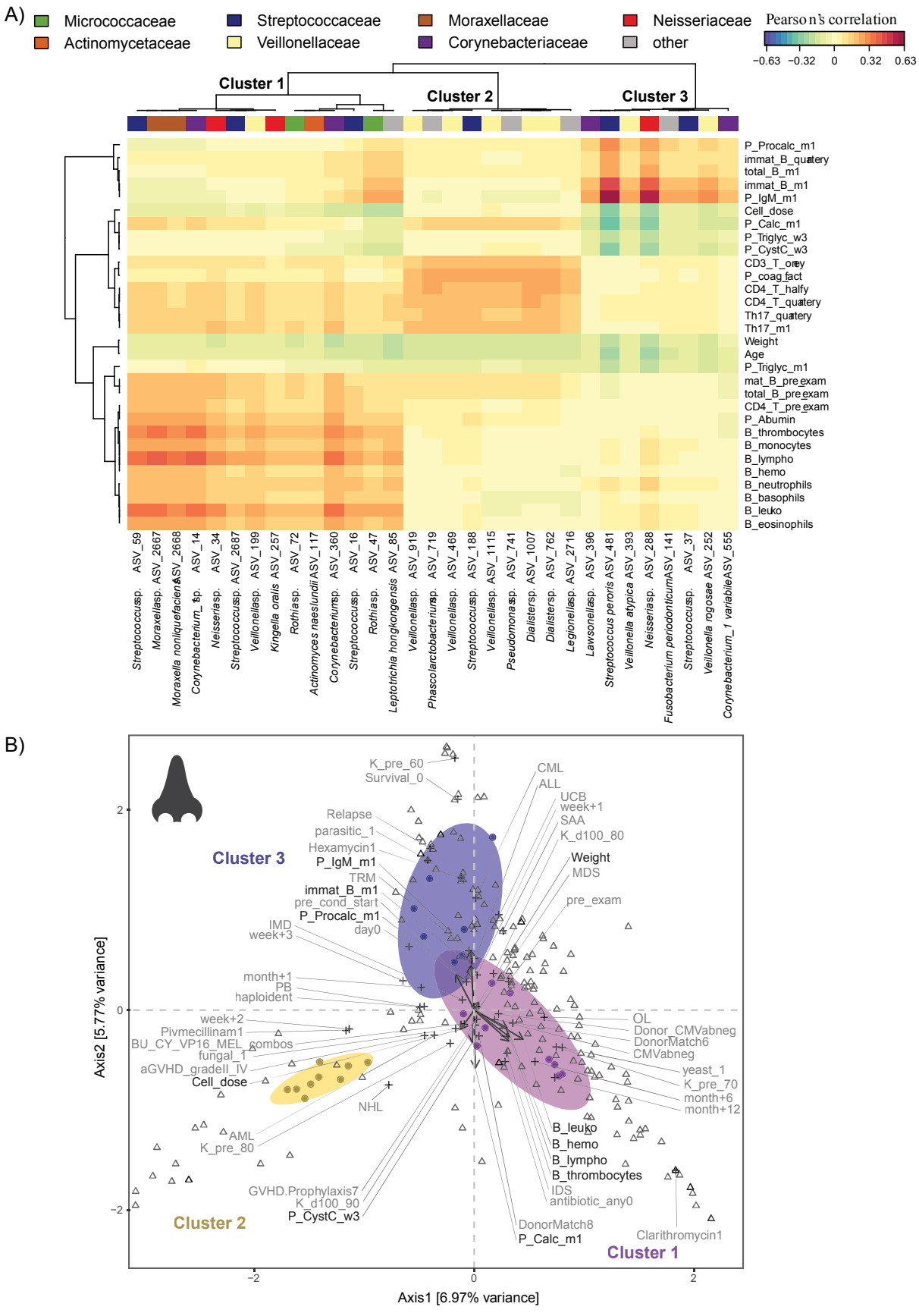

A)
Micrococcaceae
Actinomycetaceae
Streptococcaceae
Veillonellaceae
Moraxellaceae
Corynebacteriaceae
Neisseriaceae
other
Pearson's correlation
−0.63 −0.32 0 0.32 0.63
Cluster 1
Cluster 2
Cluster 3
P_Procalc_m1
immat_B_quatery
total_B_m1
immat_B_m1
P_IgM_m1
Cell_dose
P_Calc_m1
P_Triglyc_w3
P_CystC_w3
CD3_T_orey
P_coag_fact
CD4_T_halfy
CD4_T_quatery
Th17_quatery
Th17_m1
Weight
Age
P_Triglyc_m1
mat_B_preexam
total_B_preexam
CD4_T_preexam
P_Abumin
B_thrombocytes
B_monocytes
B_lympho
B_hemo
B_neutrophils
B_basophils
B_leuko
B_eosinophils
ASV_59 Streptococcus sp.
ASV_2667 Moraxella sp.
ASV_2668 Moraxella nonliquefaciens
ASV_14 Corynebacterium_1 sp.
ASV_34 Neisseria sp.
ASV_2687 Streptococcus sp.
ASV_199 Veillonella sp.
ASV_257 Kingella oralis
ASV_72 Rothia sp.
ASV_117 Actinomyces naeslundii
ASV_360 Corynebacterium sp.
ASV_16 Streptococcus sp.
ASV_47 Rothia sp.
ASV_85 Leptotrichia hongkongensis
ASV_919 Veillonella sp.
ASV_719 Phascolarctobacterium sp.
ASV_469 Veillonella sp.
ASV_188 Streptococcus sp.
ASV_1115 Veillonella sp.
ASV_741 Pseudomonas sp.
ASV_1007 Dialister sp.
ASV_762 Dialister sp.
ASV_2716 Legionella sp.
ASV_396 Lawsonella sp.
ASV_481 Streptococcus peroris
ASV_393 Veillonella atypica
ASV_288 Neisseria sp.
ASV_141 Fusobacterium periodonticum
ASV_37 Streptococcus sp.
ASV_252 Veillonella rogosae
ASV_555 Corynebacterium_1 variabile
B)
K_pre_60
Survival_0
CML
ALL
UCB
week+1
SAA
K_d100_80
Relapse
parasitic_1
Hexamycin1
P_IgM_m1
TRM
immat_B_m1
pre_cond_start
P_Procalc_m1
day0
IMD
week+3
Weight
MDS
pre_exam
Cluster 3
month+1
PB
haploident
week+2
Pivmecillinam1
BU_CY_VP16_MEL_combos
fungal_1
aGVHD_gradeII_IV
Cell_dose
NHL
AML
K_pre_80
GVHD.Prophylaxis7
K_d100_90
P_CystC_w3
Cluster 2
OL
Donor_CMVabneg
DonorMatch6
CMVabneg
yeast_1
K_pre_70
month+6
month+12
B_leuko
B_hemo
B_lympho
B_thrombocytes
IDS
antibiotic_any0
DonorMatch8
P_Calc_m1
Clarithromycin1
Cluster 1
Axis2 [5.77% variance]
Axis1 [6.97% variance]
2
0
−2
−2
0
2

Figure S6. Multivariate associations of the nasal microbiota with immune and clinical parameters in HSCT. A) Clustered image map (CIM) based on sparse partial least squares (sPLS) regression analysis dimensions 1, 2, and 3, displaying pairwise correlations >0.2/<-0.2 between nasal ASVs (bottom), and continuous immune and clinical parameters (right). Red indicated positive correlation, and blue indicates negative correlation, respectively. Based on the sPLS regression model, hierarchical clustering (clustering method: complete linkage, distance method: Pearson's correlation) was performed resulting in the three depicted clusters. B) Canonical correspondence analysis (CCpnA) relating nasal microbial abundances (circles) to continuous (arrows) and categorical (+) immune and clinical parameters. ASVs and variables with at least one correlation >0.2/<-0.2 in the sPLS analysis were included in the CCpnA. The triplot shows variables and ASVs with a score >0.3/<-0.3 on at least one the first three CCpnA axes, displayed on axis 1 versus 2 with samples depicted as triangles. The colored ellipses (depicted with 80% confidence interval) correspond to the clusters of ASVs identified by the sPLS-based hierarchical clustering. For visualization purposes, a focused section of the CCpnA triplot is shown. Abbreviations are described in Figures 6 and S5. Additional abbreviations: Calc, Calcium; CMVabneg, Recipient is CMV seronegative; CystC, cystacin C; Donor_CMVabneg, Donor is CMV seronegative; DonorMatch8, unrelated donor with 1 HLA mismatch; PB, peripheral blood;

Supplementary Material – Manuscript III

Table S1. Clinical patient characteristics. General patient characteristics, conditioning regimens, complications, and clinical outcomes for the autologous HSCT cohort (*n=13*). Abbreviations: VP16, Etoposide.

Characteristics		**Number of autologous HSCT patients**	**Percentage of all patients [%]**
Transplant recipients		13	100
Average recipient age in years		54.2 (Range: 28.7 - 70.4)	NA
Gut microbiota characterized	**not characterized**	2	15.4
	At ≤ 5 timepoints	5	38.5
	At 6-7 timepoints	6	46.2
	At 8-9 timepoints	0	0
Oral and nasal microbiota chracterized	**At ≤ 5 timepoints**	3	23.1
	At 6-7 timepoints	6	46.2
	At 8-9 timepoints	4	30.8
Patient sex	**Female**	5	38.5
	Male	8	61.5
Disease at transplantation	**Multiple myeloma**	6	46.2
	Non-Hodgkin lymphoma	5	38.5
	Amyloidosis	1	7.7
	Plasma cell leukemia	1	7.7
Stem cell source	**Bone marrow**	0	0
	Umbilical cord blood	0	0
	Peripheral blood	13	100
Conditioning regimen	**Cytarabine + Carmustine + Melphalan**	4	30.8
	High-dose Melphalan	7	53.8
	Cytarabine + Cyclophosphamide + VP16 + Thiotepa + Bortezomib + Lenalidomide	1	7.7
	Carmustine + Thiotepa	1	7.7
Antibiotics pre-HSCT		8	61.5
Antibiotics post-HSCT*		13	100

Infections*	**At least one registered bacterial infection**	8	61.5
	At least one registered viral infection	2	15.4
	At least one registered fungal infection	0	0
Overall Survival**	**Alive**	11	84.6
	Dead	2	15.4
Relapse of primary malignancy**		2	15.4
Status of remission**	**Complete remission**	6	46.2
	Never in complete remission	6	46.2
	Not evaluated	1	7.7
Non-relapse mortality		0	0
Re-transplantation		0	0

*Follow-up time after HSCT: until day +404

**Mean follow-up time after HSCT: 8 months (range: 5.6 – 10.1 months)

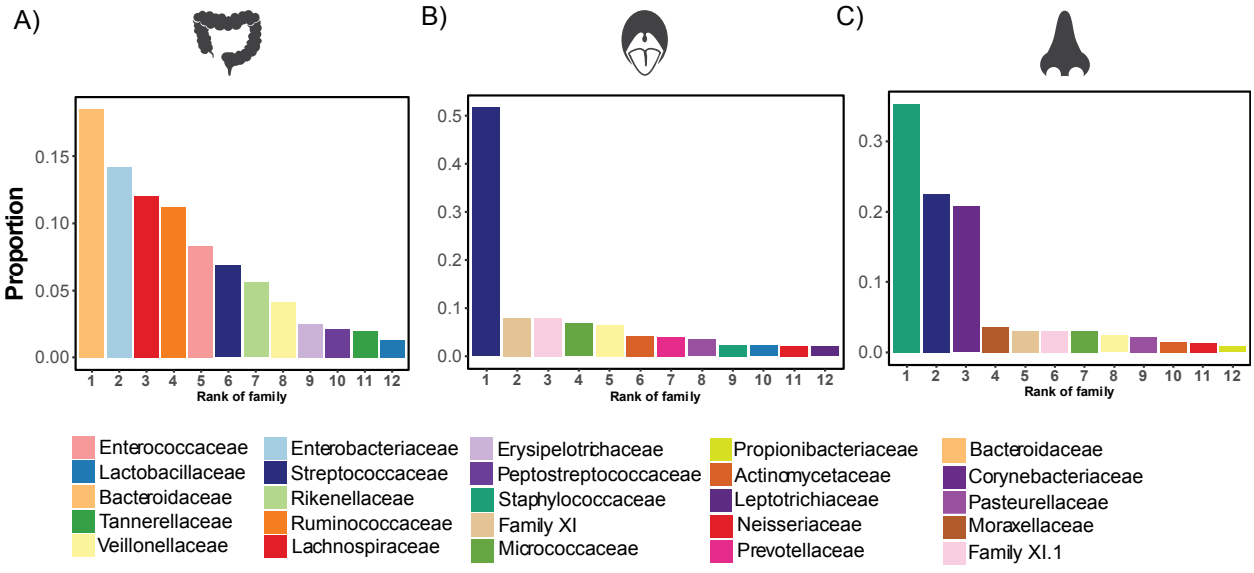
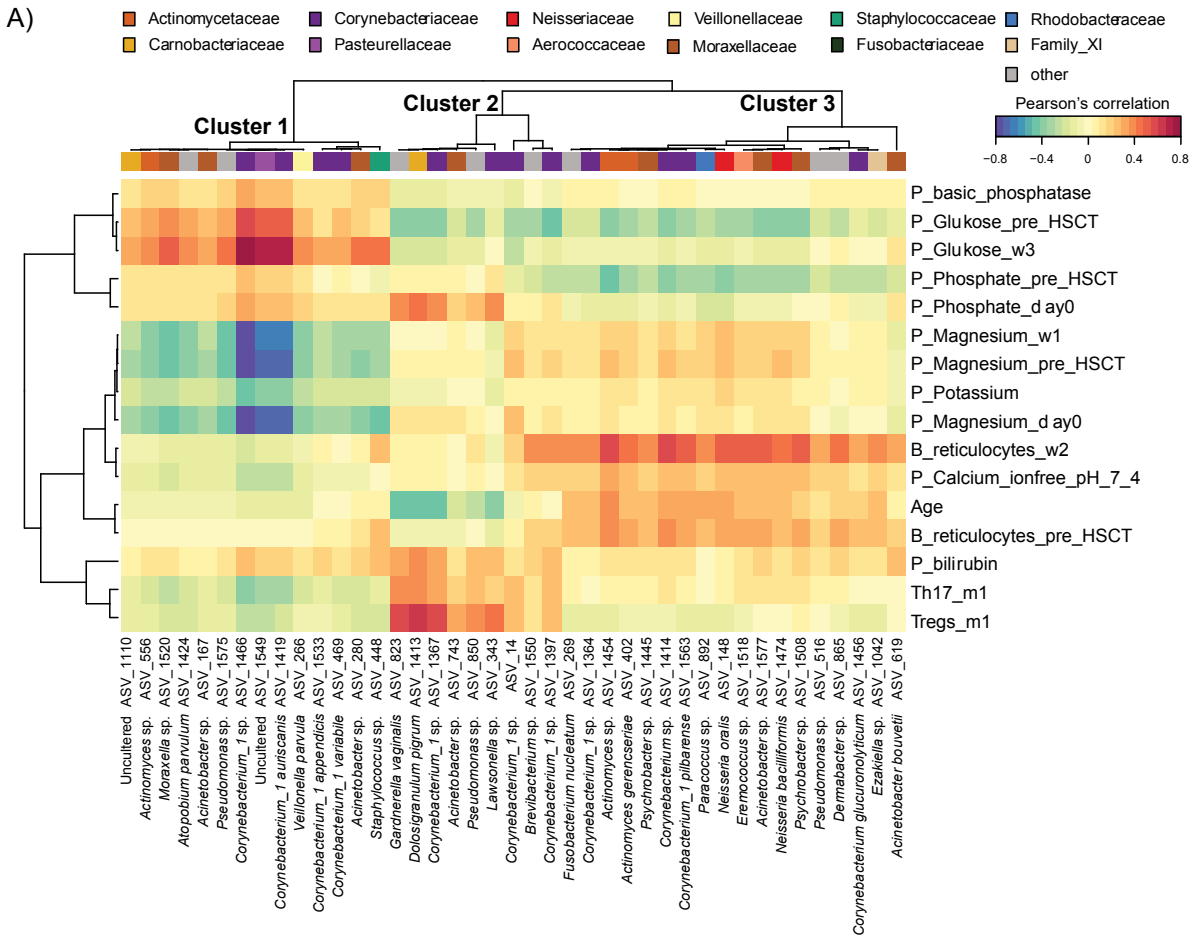


Figure S1. Most abundant taxonomic families in the gut, oral cavity, and nasal cavity in autologous HSCT patients. A) Proportions of the 12 most abundant taxonomic families displayed as rank abundance curves, in the gut, B) the oral cavity, and C) the nasal cavity.


A)
Actinomycetaceae
Carnobacteriaceae
Corynebacteriaceae
Pasteurellaceae
Neisseriaceae
Aerococcaceae
Veillonellaceae
Moraxellaceae
Staphylococcaceae
Fusobacteriaceae
Rhodobacteraceae
Family_XI
other
Pearson's correlation
−0.8 −0.4 0 0.4 0.8
Cluster 1
Cluster 2
Cluster 3
P_basic_phosphatase
P_Glukose_pre_HSCT
P_Glukose_w3
P_Phosphate_pre_HSCT
P_Phosphate_d ay0
P_Magnesium_w1
P_Magnesium_pre_HSCT
P_Potassium
P_Magnesium_d ay0
B_reticulocytes_w2
P_Calcium_ionfree_pH_7_4
Age
B_reticulocytes_pre_HSCT
P_bilirubin
Th17_m1
Tregs_m1
ASV_1110 Uncultured
ASV_556 Actinomyces sp.
ASV_1520 Moraxella sp.
ASV_1424 Atopobium parvulum
ASV_167 Acinetobacter sp.
ASV_1575 Pseudomonas sp.
ASV_1466 Corynebacterium_1 sp.
ASV_1549 Uncultured
ASV_1419 Corynebacterium_1 auriscanis
ASV_266 Veillonella parvula
ASV_1533 Corynebacterium_1 appendicis
ASV_469 Corynebacterium_1 variabile
ASV_280 Acinetobacter sp.
ASV_448 Staphylococcus sp.
ASV_823 Gardnerella vaginalis
ASV_1413 Dolosigranulum pigrum
ASV_1367 Corynebacterium_1 sp.
ASV_743 Acinetobacter sp.
ASV_850 Pseudomonas sp.
ASV_343 Lawsonella sp.
ASV_14 Corynebacterium_1 sp.
ASV_1550 Brevibacterium sp.
ASV_1397 Corynebacterium_1 sp.
ASV_269 Fusobacterium nucleatum
ASV_1364 Corynebacterium_1 sp.
ASV_1454 Actinomyces sp.
ASV_402 Actinomyces gerencseriae
ASV_1445 Psychrobacter sp.
ASV_1414 Corynebacterium sp.
ASV_1563 Corynebacterium_1 pilbarense
ASV_892 Paracoccus sp.
ASV_148 Neisseria oralis
ASV_1518 Eremococcus sp.
ASV_1577 Acinetobacter sp.
ASV_1474 Neisseria bacilliformis
ASV_1508 Psychrobacter sp.
ASV_516 Pseudomonas sp.
ASV_865 Dermabacter sp.
ASV_1456 Corynebacterium glucuronolyticum
ASV_1042 Ezakiella sp.
ASV_619 Acinetobacter bouvetii

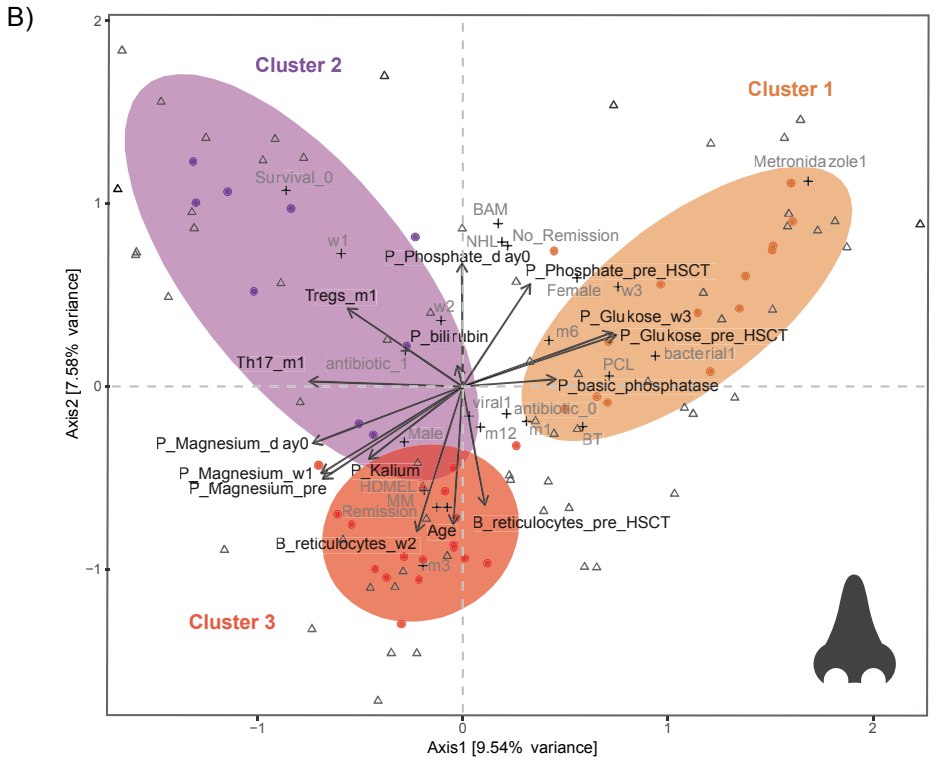

B)
Cluster 2
Cluster 1
Cluster 3
Survival_0
Metronidazole1
BAM
NHL
No_Remission
w1
P_Phosphate_d ay0
P_Phosphate_pre_HSCT
Female
w3
Tregs_m1
w2
P_Glukose_w3
P_bilirubin
m6
P_Glukose_pre_HSCT
bacterial1
Th17_m1
antibiotic_1
PCL
P_basic_phosphatase
viral1
antibiotic_0
m1
m12
BT
Male
P_Magnesium_d ay0
P_Magnesium_w1
P_Magnesium_pre
P_Kalium
HDMEL
MM
Remission
Age
B_reticulocytes_pre_HSCT
B_reticulocytes_w2
m3
Axis2 [7.58% variance]
Axis1 [9.54% variance]
2
1
0
−1
−1
0
1
2

Figure S2. Multivariate host – nasal microbial associations in autologous HSCT patients. A) Pairwise correlations (>0.3/<-0.3) of nasal ASVs (bottom) and continuous immune and clinical parameters (right) displayed in a hierarchically CIM. Complete linkage clustering was based on the first three dimensions of the sPLS regression model. Red and blue coloring indicates positive and negative Pearson's correlation, respectively. B) Triplot of the CCpnA, showing associations between nasal ASV abundances (circles) and continuous (arrows) and categorical (+) immune and clinical parameters. ASVs and variables were included in the CCpnA if they had at least one correlation >0.3/<-0.3 in the sPLS regression analysis. Variables and ASVs with a score >0.3/<-0.3 on at least one the first three CCpnA axes are displayed in dimension 1 versus 2. Samples are shown as triangles. Clusters of ASVs identified by the sPLS-based hierarchical clustering are shown as colored ellipses (with 80% confidence interval). Abbreviations are explained in Figures 2 and 3.