Microbial metabolites and graft-versus-host disease in patients undergoing allogeneic hematopoietic cell transplant

Robin Cagle

A dissertation
submitted in partial fulfillment of the
requirement for the degree of

Doctor of Philosophy

University of Washington
2026

Reading Committee:
David Fredricks, Chair
Patrick Mitchell
Luke Hoffman

Program Authorized to Offer Degree:
Microbiology

University of Washington

Abstract

Microbial metabolites and graft-versus-host disease in patients undergoing allogeneic hematopoietic cell transplant

Robin Cagle

Chair of Supervisory Committee:
David Fredricks
Department of Microbiology

Allogeneic hematopoietic cell transplantation (HCT) can cure hematologic malignancies, but 30-70% of recipients experience acute graft-versus-host disease (GvHD). GvHD is associated with changes in the gut microbiome, but the mechanistic relationship remains unclear. The gut microbiota produces gut metabolites, which may indicate or modulate GvHD. In a longitudinal case-control study of patients with and without acute gut GvHD, we characterized bile acid concentrations and the gut microbiome in stool. Endogenous secondary bile acid concentrations were associated with gut GvHD (p=0.009). We observed 4.4-fold lower levels of endogenous secondary bile acids in GvHD, particularly lithocholic acid and derivatives ($p=0.004/p_{adjusted} = 0.02$, fold change (FC) = 0.23). There was a 100-fold lower median abundance (p=0.002 and FC < 0.01) and 20-fold lower median diversity of bacterial bile acid 7α-dehydroxylation (*bai*) genes (p=0.0007 and FC < 0.05) in patients with GvHD, indicating that acute gut GvHD patients are deficient in microbial *bai* genes that make secondary bile acids. This work also covers the microbial metabolites short-chain fatty acids (SCFAs) and other compounds associated with GvHD. One SCFA, butyric acid, linked with protection against GvHD pre-HCT (p = 0.02) and several unidentified metabolites were associated with protection against GvHD. These findings suggest that microbial metabolites are markers and potentially mediators of GvHD.

Table of Contents

List of Figures and Tables

Chapter 5:

Acknowledgements

It was impossible for me to complete this thesis on my own. Everyone that has been in my life since I started this PhD in July 2019 has contributed to where I am today. All of my friends, family, mentors, and colleagues deserve endless credit and kudos for their roles in my development as a person and scientist over these seven years.

I am immensely grateful for the opportunity to work in David Fredrick's lab. I learned more than I thought was possible under David's mentorship, foremost of which is the integrity and fidelity that guide great science. My passion for analyzing how systems function and sharing findings with other scientists and the public will guide my future career. David also recognized the importance of having fun in the lab, consistently encouraging and sponsoring lab-wide excursions. I am also thankful for my previous mentors that taught me and supported my endeavors leading up to this PhD program and David's lab: Dr. Andrea Ottesen at the FDA and Dr. Amy Sapkota at the University of Maryland. They treated me with respect even as a young trainee still learning the ropes.

The whole Fredricks lab shuttled me through my thesis, teaching me basic techniques, listening to innumerable practice talks, and offering kindness when it was clear that I was working through thesis challenges. Tina Fiedler for having an answer to every question (no matter how inconvenient), Matt Munch and DJ Valint for discussing the newest episodes of The Last of Us and Severance each week, Susan Strenk for volunteering for a night shift somewhere in Skagit County when I decided to run a Ragnar, Congzhou Liu for teaching me my first technique in the Fredricks lab, Sean Proll for acting as a computational and conceptual project guide, Sujatha Srinivasan for offering a nonjudgemental space and advice on completing a thesis, and Sam Minot for foreseeing the computational needs of the Fred Hutch as a whole and helping me 1:1 to advance my project. Nancy Allbritton and Hao Wang were also integral for collaborating with me during one of the more ambitious sub-projects of my thesis. A huge thank you to all of the clinical coordinators that made my research possible and joined in on Fredricks lab outings. I also appreciate the Shared Resources Core at the Fred Hutch for preparing the whole genome metagenomic libraries, running the library on the sequencer, and teaching me confocal microscopy. I also want to thank the allogeneic hematopoietic stem cell transplant recipients who went out of their way to help others through their contributions of their samples and time participating in this research.

Thank you to the UW's Microbiology department for giving me the chance to learn and work in an environment recognized for its intellectual rigor. Specifically, thank you to my thesis committee for your mentorship and strategic planning of my thesis: Dr. Jason Smith, Dr. Kelly

Smith and my thesis reading committee for going the extra mile, Dr. Patrick Mitchell and Dr. Luke Hoffman. Previous graduate program advisors Amy Gundlach Ritter and Andrea Pardo were amazing in their positions in the department and have set the bar high for supportive advisors. Thank you to any faculty that attended a Journal Club, 1:1 meeting, or Work in Progress talk of mine and gave feedback on my projects- you all are busy and I value your time and consideration highly.

In my (biased) opinion, the graduate students were the best part of the UW Microbiology program. Thank you past and current students for being here; I never felt alone or isolated thanks to you all. Many of us moved to Seattle for the first time and jumped into this program with only each other as company, which turned out to be more than enough. Blowing off steam and goofing around together was integral to my success. Some highlights will be every pre- or post- departmental retreat camping trip, the 2021 Mt. Rainier backpacking trip, the 2022 Utah trip, Ragnar 2024, annual student retreats in Friday Harbor, group runs, Bob's Corn, and more. A specific shoutout is in order for my cohort that started in 2019, quickly got wrapped up in a pandemic and weathered it together: Edward Arnold, Jacob Frick, Ryan Yucha, and Joselyn Landazuri Vinueza. I'm looking forward to seeing what we all get up to in science and life.

I'm so thankful for my community in Seattle. Elliot Lee has been my friend basically every day as we work side by side and chat loudly in the Fredricks lab. Lyndsey Moore was a roommate that made my life better and has remained a great friend. Emma Gomez-Rivas, Alaina Boyle, and Quinn Edwards have taken me and our collective dogs on hikes in beautiful WA nature when I needed a reset. Mattheus Tavares was a compassionate roommate and consistent friend through some difficult trials. The Seattle field hockey community has been immensely valuable as an outlet, accountability for seeing people outside of UW, and exercising while we were at it. Guadalupe Mabry-Torres reminded me why I pursue my PhD and gave care and support when I needed it. Emma Pedraza and Rainey Twohey-Jacobs watched RuPaul with me and helped ensure I had a fun time outside of work/school hours. Devin McDermott, Michael Watson, McKinlee and I found joy in every season from Christmastime gaiety to summer backpacking. Thank you also to my neighbors for showing me what a true community can look like and how much power we have together to care for each other.

Even from across the country, I had tangible support from a consistent group of friends that I have known since high school and college. This includes Cara Walsh, Julia Zhen, Heidi Liu, Caroline Tussing and Angela Park, to name a few, in addition to others that attended Centennial High School in Ellicott City, MD and/or University of Maryland. You all helped me to stay humble and down to earth because you loved me as a friend before AND after the PhD

process. I will treasure our memories from our entire lives, and some more recent memories made while getting out of town and exploring West Seattle, Mt. Rainier National Park, New York City, Shenandoah National Park, Seoul and Tokyo. Another friend that has known and loved me since literally my birth day is Leila Shirvan. I'm so proud of you and what we've accomplished in our lives. Thank you for keeping me around in your life even when I was a blatantly mean child. I think you count more as family than a friend.

Words won't be enough to thank my family for helping me get to this point. My mom, Kathy, saw me through some highs and lows but never wavered in her love and support. She was game every time she came out to visit and I had a minute-to-minute itinerary for us, patient when I traveled to MD and managed to book out every hour for a different activity, and always happy to catch up for hours on the phone. My sister, Veronica, visited me in Seattle and also put up with my stringent itineraries, popped into the occasional phone call, and stayed up to date with my school status. Thank you to my Aunt Barb for joining my mom and sister when traveling out to Seattle to celebrate this milestone in my life and for playing Animal Crossing with me. The Koontz family also helped sustain me since my move to Seattle after a rough 2019. My dad, Jeff, is on my mind daily; although he wasn't there to experience this chapter of my life, he had the best advice for every professional or academic situation, and I still find myself trying to figure out what he would recommend when I need help. I hope he'd be proud of what I have accomplished and what I will do next.

Finally, thank you to my favorites: Cassie the dog and Nook the cat. Just kidding- although they provided emotional wellbeing and cuteness. Thank you to my best friend and partner in everything, Elaina Perry. You have helped get me through every single day since we met, and I can't put into words how thankful I am for you. I think I am the luckiest person ever to have you as my spouse. You accepted unique school/work hours, crises, and questionable eating/meal habits in merging our lives. I'm positive that I wouldn't be here today without you. You're wise, smart, and overall I think you've also earned this PhD. Thank you for forever being by my side.

Chapter 1: Introduction

1.1 Rationale and significance: Why is GvHD important to study?

Allogeneic hematopoietic cell transplantation (allo-HCT) can cure hematological disorders but acute graft-versus-host disease (GvHD) contributes to substantial non-relapse mortality and morbidity.[1–5] It can be difficult to predict which allo-HCT recipients will develop GvHD and which will progress to mild vs. severe GvHD, as even well-matched donors and recipients can lead to GvHD in the recipient. New and more effective therapies may arise from an enhanced understanding of what initiates and modulates GvHD progression and outcomes. I sought to investigate the role of the fecal microbiome in GvHD post-allo-HCT. My thesis work identifies links between GvHD, bile acids, short chain fatty acids (SCFAs), gut microbes and their bile acid metabolic genes in a large cohort of >300 allo-HSCT patients with >2,700 stool samples and robust clinical metadata available. I used metagenomics, metabolomics and 16S rRNA gene sequencing to characterize the gut microbiome post-HCT. This multi-omics analysis helps illuminate connections between acute gut GvHD and the gut microbiome.

1.2 Background

1.2.1 What is allogeneic hematopoietic stem cell transplantation (allo-HCT)?

Allogeneic hematopoietic cell transplantation (allo-HCT) is potentially curative treatment for patients with hematologic malignancies and non-malignant hematologic disorders. The term ‘allogeneic’ refers to the donor or grafted cells being genetically distinct from the recipient’s tissues because the donor/grafted cells are derived from a separate individual other than the recipient. Hematopoietic cells refer to blood cells, including the originating stem cell and differentiated blood cells like platelets/thrombocytes, white blood cells/leukocytes (including granulocytes, monocytes, and lymphocytes) and red blood cells/erythrocytes.[6] The donor cells are either harvested from umbilical cord, peripheral blood, or the bone marrow. Allo-HCT treats disorders of or involving these blood cells, including aplastic anemia, leukemia, myelodysplastic syndromes, multiple myeloma, Hodgkins and non-Hodgkins lymphoma, sickle cell disease, thalassemia and a variety of other hematologic and some autoimmune diseases.[6] The potential for cure and the goal of allo-HCT is to replace the recipient’s native deficient or malignant hematopoietic cells with healthy and functional ones from a donor. Allo-HCT is the central event in the timeline that I focus on in my thesis work.

1.2.1.1 Pre-HCT recipient and donor matching

There are several preparative steps prior to administration of the functional hematopoietic cells via allo-HCT. One of these initial steps is to find a suitable donor with

sufficient genetic similarity to the recipient, also known as a 'match'. Ensuring sufficient genetic similarity also ensures that the antigens (peptides and proteins displayed on the surface of cells), which are encoded by genes, will look and act similarly to the native recipient's antigens. Currently, the best practice to genetically match donors and recipients is through human leukocyte antigens (HLA) matching. The genes encoding the HLA are heritable and therefore siblings to the recipient are prime candidates for providing high HLA matched donor cells. The HLA informs which peptides that major histocompatibility complex (MHC) antigen-presenting cells (APCs) can recognize as foreign, then process and display to $CD4^+$ T cells.[6] This informs activation of other immune cell subsets and guides which peptides and cells should be attacked and destroyed. The minor histocompatibility (mHC) antigens bind the MHC and also distinguish self from non-self peptides, but are much more diverse. This diversity stems from more than 40 loci that encode the minor histocompatibility antigens which are spread throughout the human genome and are only minimally biochemically characterized.[6] Thus, there is mHC mismatching in nearly all allo-HCTs despite the most rigorous MHC matching. This is important because the HCT recipient's mHC antigens are detected by the donor-sourced grafted cells as non-self. Therefore, mHC mismatch may lead to complications arising from donor cells' recognition of self vs. non-self cells/tissues.

1.2.2 What is GvHD?

1.2.2.1 GvHD definition

Recipient and donor cell interactions are an essential part of HCT that must be managed and balanced. Antagonistic interactions between healthy donor immune cells and recipient/host cancerous, malignant or deficient cells that are being treated are beneficial for HCT curative outcomes. The healthy donor cells can recognize the native host's malignant or deficient cells as non-self and destroy them. This is considered an immunologic graft-versus-tumor or graft-versus-leukemia (GvL) effect and is ultimately beneficial for the HCT recipients because it facilitates eradication of the underlying disease, leading to long-term cure.[6] However, this graft cell immune response must be tempered because the detrimental graft-versus-disease (GvHD) complication emerges when allogeneic donor T cells attack healthy host epithelial tissues. GvHD is detrimental to HCT recovery, but complete elimination of mHC mismatch is not necessarily a goal of the field since the GvL effect is critical for transplant success.[6] A goal of my thesis work was to study human- and GvHD-associated microbiological factors that could reduce GvHD incidence and severity.

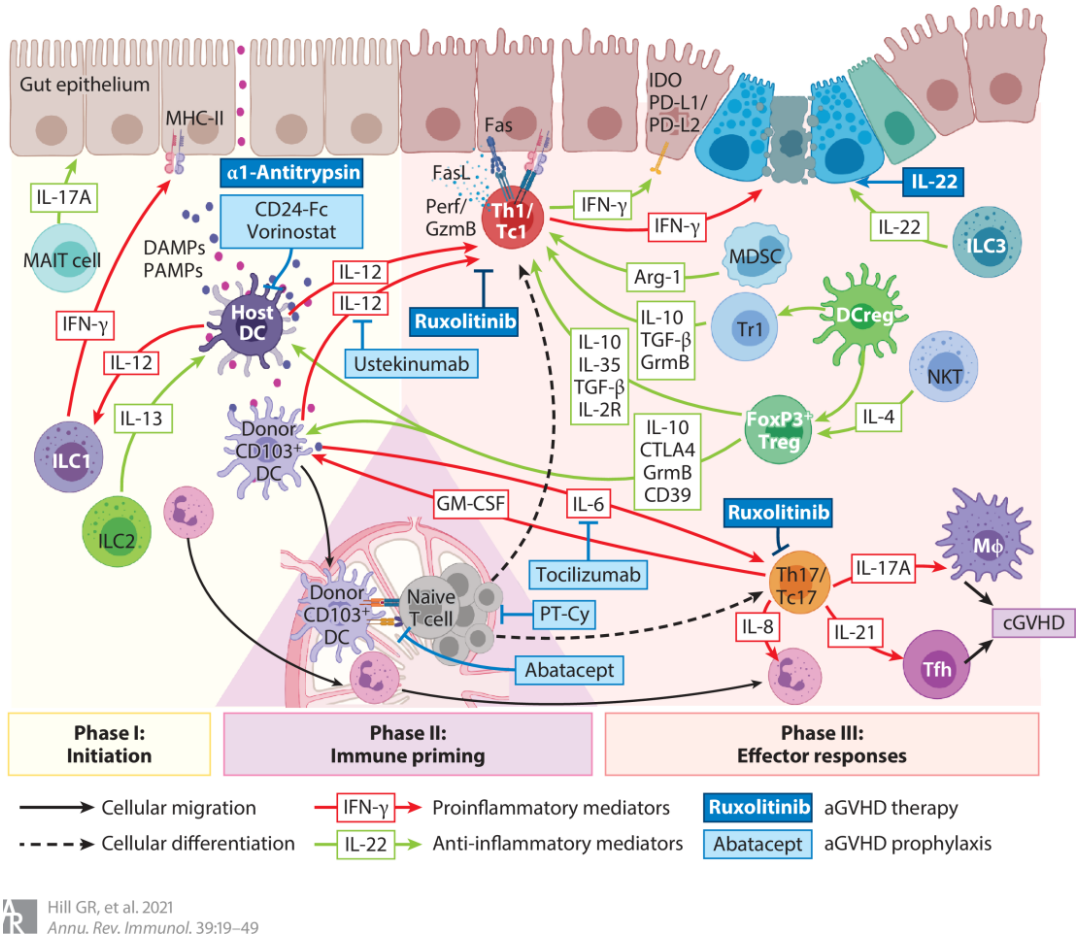
1.2.2.2 GvHD immunology: initiation and progression

Native cancerous cells are a target of donor T cells which will attack via GvL due to mHC mismatch. However, there is often off-target healthy tissue injury at sites of high cell turnover (e.g. the intestinal barrier) because of effects of the preparative chemo-radiotherapy that HCT recipients undergo prior to transplant.[6] The type and extent of chemo-radiotherapy vary based on each individual recipient. Regardless of the intensity, the intent of this preparative conditioning, or the pre-HCT chemo- and/or radiotherapy, is to eradicate the existing native malignancy or deficiency and dampen immune response to the incoming donor cells. This conditioning regimen often succeeds in ablating the native malignant or deficient cells to permit the grafted donor cells to proliferate in their place, but off-target conditioning effects can persist, although the extent of these off-target effects vary with the type and intensity of the conditioning.[6] I considered this in my thesis work by matching/pairing the exact type and intensity of conditioning across the study cohort's cases and controls such that conditioning regimen would not confound my findings.

T cells from the donor graft (since native T cells can be eradicated during preparative conditioning) get recruited to investigate and respond to sites of high immune cell activity. In the case of leukemia or cell malignancy/deficiency, dense immune cell activity can be due to the underlying unregulated cell growth and turnover.[6] However, the preparative conditioning regimen targets highly replicating cells. This intentionally includes cancers and incidentally includes body sites with high cell turnover/growth and replication, such as cells located at mucus membranes where there is increased contact with the environment; for example, epithelial cells in the skin, mouth, liver and intestine (although other healthy host tissue can be implicated). Thus, off-target conditioning injury and recruitment of donor T cells often occurs at the skin, liver and/or gut. These sites are where GvHD typically manifests/occurs, and off-target host tissue damage can be considered the first phase of GvHD.[7,8] On-target suppression or attack of the bone marrow by the cytotoxic conditioning compounds can result in myeloablation, eradication of the key immune and blood cells produced in the bone marrow, or myelosuppression, reduced production of these key immune and blood cells made in the bone marrow. In most conditioning regimens, this myelosuppression/ablation is intentional to eradicate the underlying disease. For some patients, myelosuppression/ablation should be minimized (reduced intensity) or avoided (non-myeloablative) because it may not be well-tolerated due to existing comorbidities (or other factors that would contribute to poor health pre-HCT).[9–11]

Since preparative conditioning causes tissue injury in addition to ablation of malignant or deficient cells, upon transplantation of the grafted hematopoietic cells, donor T cells translocate

to the sites of injury, for instance, the gut epithelium. The mucosal barrier injury in gut tissue contributes towards a pro-inflammatory environment as expected with a normal immune response to tissue damage. This includes release of pro-inflammatory cytokines and alarmins in blood that consistently emerge post-HCT, including IFN-γ, IL-12, GM-CSF, IL-8, IL-17A, and IL-21.[8] This is considered the latter end of the first phase of GvHD (Fig. 1).[8] GvHD can initiate when these donor, undifferentiated T cells encounter the damage-associated molecular patterns (DAMPs) and become activated against host/native cells and antigens (e.g. mHLA antigens) via antigen-presenting cells (APCs). This T cell priming is considered the second phase of GvHD (Fig. 1).[8] Such T cells typically differentiate into T helper (Th) 17, Th1, Th2 or CD8+ cytotoxic T cells that are frequently implicated in GvHD and proceed to further attack the affected tissue, exacerbating tissue damage and inflammation. This produces more pro-inflammatory cytokines and alarmins that recruit even more T cells to the healthy tissue and exacerbate the inappropriate immune response, creating a positive feedback loop that contributes to local and systemic inflammation. T cell differentiation and production of effectors is the third phase of GvHD (Fig. 1).[8] This amplification of excessive inflammation is difficult to disrupt and is characteristic of GvHD.[1,8] It is essential to understand the breakdown of GvHD initiation and progression because it is these phases that contextualize the significance of my thesis projects' findings and that my suggested potential interventions would impact.



Hill GR, et al. 2021
Annu. Rev. Immunol. 39:19–49

Figure 1. Phases of GvHD initiation and progression. Figure is from Hill et al., 2021 *Annu Rev Immunol* 39:19-49. Caption is also from the original publication: "Phase I: GVHD initiation. Pretransplantation conditioning causes barrier integrity loss, gut lumen pathogen-associated molecular pattern (PAMP) translocation, and dying host cell release of damage-associated molecular patterns (DAMPs) that activate host antigen-presenting cells (APCs). Host mucosal-associated invariant T (MAIT) cells produce IL-17A, which prevents intestinal dysbiosis. Activated APCs secrete IL-12, inducing host type 1 innate lymphoid cells (ILC1s) and T cells to release IFN-γ, which upregulates MHC-II on nonprofessional APCs (epithelial and stromal cells). APC activation is suppressed by IL-13-producing ILC2s, or by regulatory T cells (Tregs) utilizing secreted (IL-10) or contact-dependent (CTLA4, granzyme B, CD39) mechanisms, and modified by Siglec G–CD24 pathway agonists (e.g., CD24-Fc) or IL-12 inhibition (ustekinumab). Histone deacetylase inhibitor vorinostat and α-1-antitrypsin inhibit PAMP/DAMP-induced APC proinflammatory cytokine production and costimulation. Phase II: Donor T cell priming. Donor T cell activation follows interaction with recipient hematopoietic APCs [e.g., dendritic cells (DCs)] or nonhematopoietic APCs activated during phase I. Reconstituting donor CD103+ DCs migrate to draining lymph nodes, where they further prime donor T cells. Abatacept (CTLA4-Ig) disrupts Tcell:APC costimulation to inhibit aGVHD. Alloreactive donor T cell expansion is curtailed by early posttransplantation cyclophosphamide (PT-Cy). Phase III: effector responses. Primed T cells differentiate to type 1 T helper (Th1)/type 1 CD8+ T (Tc1) or Th17/Tc17 cells, directed by IL-12 or IL-6, respectively, and are susceptible to the Jak1/2 inhibitor ruxolitinib, IL-12/23p40 (ustekinumab), or IL-6R (tocilizumab) blocking monoclonal antibodies. Th1/Tc1 cells secrete IFN-γ, which is intestinal stem cell cytotoxic, while augmenting the negative regulatory indoleamine-2,3-dioxygenase (IDO) and PD-L1/PD-L2 pathways to restrain aGVHD. Th17/Tc17 cells secrete proinflammatory cytokines that mediate tissue pathology both directly and by recruitment of secondary effectors [e.g., neutrophils, Mϕs (macrophages)]. IL-22-secreting ILC3s or exogenous IL-22 supplementation support intestinal stem cell regeneration. Myeloid-derived suppressor cells (MDSCs) suppress T effector functions via an arginase-1-dependent mechanism. Tregs [supported by IL-4-producing natural killer T (NKT) cells and regulatory DC cells (DCregs)] and regulatory type 1 T cells exert multiple pathways to control T effector cells. Abbreviations: cGVHD, chronic GVHD; GM-CSF, granulocyte-macrophage colony-stimulating factor; Tfh, T follicular helper; Tr1, regulatory type 1 T cell. Figure adapted from images created with BioRender.com." Figure and caption are from Geoffrey R. Hill, Brian C. Betts, Victor Tkachev, Leslie S. Kean, Bruce R. Blazar. 2021. Current Concepts and Advances in Graft-Versus-Host Disease Immunology. *Annual Review Immunology*. 39:19-49. https://doi.org/10.1146/annurev-immunol-102119-073227

1.2.2.3 GvHD: overview from the clinical perspective

1.2.2.3.1 Preparation for allo-HCT

Broadly, conditioning regimens can be considered either high-dose (myeloablative), reduced-intensity or non-myeloablative. Each recipient's age, comorbidities, history of transplant, etc. informs which conditioning regimen will be tolerated and appropriate. Initially, high-dose conditioning entailed total body irradiation (TBI) and the cytotoxic chemotherapeutic cyclophosphamide. Over time, preference shifted towards cyclophosphamide in combination with another chemotherapeutic, busulfan, without TBI. Currently, a combination of cytotoxic compounds and approaches is the best practice; the exact combination varies depending on the individual recipient's genotype, strength of HLA match between donor and recipient, treatment history, and underlying illness, although the goal is to preserve GvL effects but restrict tissue injury and GvHD adverse outcomes. It is also common to combine agents to span multiple modes of cytotoxic action. Types of cytotoxic conditioning drugs include DNA-modifying agents (e.g. topoisomerase inhibitors like etoposide and teniposide, radiation therapy (or radiotherapy), alkylating agents (e.g. cyclophosphamide, busulfan, melphelan), and nucleoside analogs (e.g. cytarabine, fludarabine, clofarabine). Many of these treatments are myelosuppressive or myeloablative, meaning they suppress or halt the production of blood and immune cells in the bone marrow. Depending on the context, this can be helpful in eradicating the underlying disease or this can be detrimental to blood and immune cell reconstitution.[9,10] As was mentioned earlier in this text, off-target toxicity to healthy tissues restricts use of high-dose conditioning. Reduced intensity conditioning (RIC) is frequently used when the anticipated tolerance to HCT is low; RIC aims to strategically decrease dosage of particular cytotoxic compounds/agents to minimize off-target tissue injury/toxicity yet maintain beneficial immunotherapeutic GvL effects. Another conditioning strategy is T cell depletion of the graft before transplanting to the recipient. This is considered in cases where there is a poor match between donor and recipient and results in a lower GvHD risk; however, it also reduces the ability of the recipient to fight off infection (e.g. cytomegalovirus) that could lead to complications post-HCT.[11] Non-myeloablative conditioning relies more on immunosuppression and entirely avoids targeting the bone marrow, making it better tolerated and used to treat individuals who may not otherwise qualify for HCT (due to high risk of myeloablative conditioning intolerance)[12]. The best conditioning/chemotherapy regimen is chosen based on which strategy would facilitate each individual's hematopoietic cell reconstitution in the long-term.[9–11] These types of preparations

are described in detail here because each is employed to prepare individuals within this thesis project's study cohort for allo-HCT.

1.2.2.3.1.1 Antibiotics

Anti-anaerobic antibiotics like piperacillin-tazobactam or meropenem are used sparingly when there is no suitable alternative, such as cases of sepsis, because use of these antibiotics post-HCT is associated with GvHD.[13] For adult HCT-recipients, antibiotics can be administered prophylactically, empirically, or in a directed fashion for documented infections. The identification of an underlying infectious agent contributing to fever may trigger a shorter course of a targeted antibiotic vs. a longer duration of a broader-spectrum antibiotic in situations where the underlying infectious agent remains unknown. Prophylactic antibiotics include those administered based on low absolute neutrophil count (ANCs) and empiric antibiotics administered based on fever. Prophylactic antibiotics typically include levofloxacin, although alternatives (e.g. cefpodoxime, amoxicillin/clavulanic, ceftriaxone) are used with intolerance or allergy to levofloxacin. Empiric antibiotics are determined for fever and neutropenia; fever with low ANC generally entails antipseudomonal beta-lactams such as ceftazidime or cefepime. Cases of suspected bacteremia/sepsis also trigger a distinct antibiotic therapy of meropenem and linezolid, although alternatives like vancomycin and aminoglycoside antibiotics may be employed. It is important to understand that the antibiotics administered impact the gut microbiomes that I characterized in my thesis work. My analysis of microbial community composition and microbially-mediated bile acid metabolism in the gut microbiome considers the impact of any antibiotics that were administered.

1.2.2.3.2 GvHD pathophysiology: Diagnosis of GvHD

GvHD can affect a range of organs/tissues and the symptoms can vary. To guide treatment, a staging and grading scheme is used to assess the degree of GvHD based on clinical symptoms/parameters. In the gastrointestinal tract, this is the daily volume of diarrhea: Stage 1 gut GvHD describes 500 - 1,000 mL/day of diarrhea, Stage 2 is 1,000 - 1,500 mL/day of diarrhea, Stage 3 is 1,500 – 2,000 mL/day of diarrhea, and Stage 4 is > 2,000 mL/day of diarrhea or ileal pain. Stool samples used in this thesis research are likely impacted by the Stage of gut GvHD/degree of diarrhea, although it is not possible to control for this since diarrhea and Stage of gut GvHD are inherently related. More symptoms of gut GvHD include intestinal bleeding, cramping, poor appetite, nausea, and indigestion. The liver GvHD Stage is informed by concentration/quantity of bilirubin, and skin GvHD by degree and type of rash. The Stage and number of organs affected by GvHD determine the overall Grade of GvHD. According to the Glucksberg Grade scale, Grade I means a Stage 1 or 2 skin involvement

without liver or gut involvement. Grade II indicates Stage 1 - 3 skin involvement and Stage 1 liver or gut involvement. Grade III includes Stage 2 or 3 skin, liver, or gut involvement. Finally, Grade IV entails Stage 1 - 4 skin involvement and Stage 2 - 4 liver or gut involvement. Although GvHD is diagnosed symptomatically, gut GvHD pathology is characterized by increased apoptotic cells, tissue fibrosis, edema, mucosal sloughing, lesions, and crypt destruction in a gut biopsy.[14–16] However it remains challenging to determine if these pathologic signatures are due to GvHD, pre-HCT conditioning, and/or something else (e.g. infection). Further, identification of biomarkers for prediction of GvHD occurrence, prognosis and outcomes is under active investigation in the field. Biomarkers for GvHD are not currently used as standard of care, however several candidates have been evaluated: suppression of tumorgenicity 2 (ST2), TNF-alpha, IL-13, IL-6, soluble IL-2 receptor alpha chain, citrulline, elafin, amphiregulin, T-cell immunoglobulin and mucin domain-containing protein 3 (TIM-3), long non-coding RNAs, high-density lipoprotein cholesterol (HDL-C), and Reg3alpha (for greater detail, see section 1.3.1.1: GvHD biomarkers.[17,18] Depending on the timeline, gut GvHD can be considered acute or chronic based on whether GvHD is diagnosed within the first 100 days post-HCT vs. after the first 100 days post-HCT, although there can be exceptions where acute or chronic GvHD occur outside of this timeline.[19,20] Acute vs. chronic GvHD have distinct clinical features and courses of treatment.

1.2.2.3.3 Treatment and prevention of acute GvHD

1.2.2.3.3.1 Immunosuppression: steroids, other agents

The findings of my thesis work concern acute gut GvHD and could impact treatment of acute gut GvHD, so treatment of acute GvHD is the focus. The incidence of acute GvHD post-allogeneic HCT varies based on GvHD prophylaxis and conditioning regimen, strength of the HLA match, source type of donor cells, specific center that HCT was performed, recipient and donor characteristics/traits (e.g. age, sex, blood type, etc.), recipient/host microbial community and more, but ranges from 20-50% of allogeneic HCT recipients. Approximately half of these HCT recipients develop acute gut GvHD, specifically. Pre-HCT, the FDA has recently approved use of abatacept (Orencia) to prevent GvHD.[21] However, the improvement to rates of survival in HCT-recipients is approximately 20% and abatacept did not improve survival for severe acute GvHD[21], so new therapeutic interventions could still help improve outcomes post-HCT. Calcineurin inhibitors like cyclosporine, mycophenolate, and methotrexate are often used post-HCT to suppress the immune system to prevent GvHD.[22,23] Once GvHD is diagnosed, the primary treatment is corticosteroids administered orally, topically, or intravenously, depending on the severity and organ involvement. The corticosteroids can inhibit T cell activation, which

mediates GvHD, and in turn reduce production of pro-inflammatory cytokines. If successful, this line of treatment continues one to two weeks after symptoms resolve.[22,23] Approximately 50% of allogeneic HCT recipients respond to corticosteroids (e.g. prednisone, methylprednisolone), and those who do not respond receive salvage treatments for steroid-resistant GvHD. Notably, a recent intervention, post-HCT cyclophosphamide, has been observed to reduces the incidence and severity of GvHD.[24] One of the first salvage or secondary treatments considered is antithymocyte globulin (ATG) which can raise the response rate to be up to 80%, although weaker response rates of 30-50% have also been reported.[22] Ruxolitinib can also improve response to corticosteroid treatment.[25] Targeted T cell depletion is another strategy to treat steroid-resistant GvHD; this includes monoclonal anti-CD3 antibodies (e.g. OKT3, visilizumab), anti-CD52 antibodies (e.g. alemtuzumab), anti-CD162 antibodies (e.g. neihulizuma), anti-CD25 antibodies (e.g. basiliximab), anti-cytokine (e.g. TNF-α, IL-5, IL-6, IL-12/23) molecules (e.g. infliximab, etanercept), tissue-protective cytokines (e.g. Il-22),[26] anti-T cell proliferation compounds (e.g. pentostatin, mycophenolate mofetil, sirolimus), fecal microbiota transplant,[26] protease inhibitors (e.g. alpha-1 antitrypsin),[27] extracorporeal photopheresis that immunosuppresses T cells, and/or mesenchymal stem cells to immunosuppress T cells and repair tissue damage. Octreotide is used to control symptoms of gut GvHD.[20,22] Although these second-line treatments for steroid-refractory acute GvHD are promising, none beyond ruxolitinib are FDA-approved, leaving options limited.[28]

1.2.2.3.3.2 Ursodiol/ursodeoxycholic acid (UDCA)

The secondary bile acid ursodeoxycholic acid (UDCA) is available as an FDA-approved prescription therapeutic. It is known as ursodiol, although it is chemically identical to UDCA.[23,29,30] Ursodiol/UDCA is administered orally to all allogeneic HCT recipients regardless of GvHD status to mitigate post-transplant liver complications, such as sinusoidal obstruction syndrome (SOS) also known as hepatic veno-occlusive disease (VOD). Ursodiol is administered from before conditioning through 90 days post-HCT. Although the intent behind UDCA clinical administration is prevention of liver damage, protection against acute gut GvHD has also been observed.[31,32] Bile acid levels in stool were measured as a part of the study and so UDCA (a bile acid) treatment is relevant.

1.3 Overview of the field: What is known vs. unknown about GvHD?

As summarized in section 1.1 (Rationale and significance: Why is GvHD important to study), I examined how GvHD links with different aspects of the gut microbiome, including the intestinal barrier, bile acids, SCFAs, diet, gut microbes and their bile acid metabolic genes. Thus

in this section, the current state of the association between GvHD and each aspect of the gut microbiome is described in detail.

1.3.1 Role of intestinal barrier integrity in GvHD

New therapies to treat or prevent GvHD may arise from a greater understanding of how intestinal barrier disruption propagates GvHD. The healthy intestine supports a diverse commensal microbial community; while Paneth cells release antimicrobial peptides (AMPs) in response to microbe-associated molecular patterns (MAMPs), concentrated local immune cells use numerous effectors to limit growth of pathogens. Goblet cells produce mucin to maintain separation between the gut epithelium and microbes. Cells that typically heal intestinal tissue damage (fibroblasts) and promote IEC turnover (intestinal stem cells) are also located within the gut barrier. Each of these cell types are subject to damage from pre-transplant conditioning and GvHD. Loss of barrier integrity has been shown to contribute to GvHD initiation and severity.[2,33–37] In humans, more intensive pre-HCT conditioning causes more pronounced intestinal barrier damage and acute gut GvHD therapy.[38] Murine model experiments revealed that intestinal barrier loss is necessary and sufficient to exacerbate minor histocompatibility mismatches to drive GvHD.[39] Colonic intestinal barrier degradation was observed in the presence/production of inflammatory TNF-α, which is implicated in acute gut GvHD development.[40] Neutralization of TNF-α reduced intestinal barrier damage and delayed onset of acute gut GvHD, supporting the importance of barrier integrity in GvHD.[41]

TNF-α can be induced by bacterial production of lipopolysaccharide (LPS), providing a pathway for bacteria to impact barrier integrity and acute gut GvHD.[39] Microbial LPS binds and signals via Toll-like Receptor 4 (TLR4) on donor cells, triggering MyD88-mediated inflammation and disease in mice.[39] Tight junction degradation has also been implicated in GvHD.[42] Tight junctions are the scaffolding that hold the epithelium together by regulating paracellular permeability between individual cells within the epithelial barrier.[42] They are a structured mesh of proteins and filaments, such as occludin, zonula occludens-1, myosin, and more, that form a cytoskeleton which can be modulated by environmental stimuli and cell signaling.[42] Microbes that produce LPS or otherwise induce TNF-α reduce the integrity and disrupt the regulation of tight junctions, providing another mechanism for microbially-mediated barrier disruption.[42] Additional microbial metabolites can impact gut barrier integrity via tight junctions, such as SCFAs. Further analysis of how microbial SCFAs impact tight junctions and barrier integrity can be found in Chapter 4's Introduction section.

Fecal microbial dysbiosis, a surrogate measure for intra-intestinal microbial dysbiosis, can be observed pre- and post-HCT, and is exacerbated with gut GvHD. Microbial dysbiosis is observed in several different ways, including by measuring the diversity of the gut bacterial community by comparing the number of bacterial taxa represented as well as the evenness of each taxon's relative abundance (alpha-diversity). Individuals with GvHD exhibit lower bacterial community alpha-diversity vs. HCT recipients without GvHD.[43] Both the types and abundances of bacterial species within the gut differ in the GvHD disease state.[43] In particular, pathogenic bacteria like *Enterococcus* or *Escherichia coli* often mono-dominate or bloom within an HCT-recipient's gut microbial community, contributing more damage through production and release of virulence factors to promote infection.[44] Finally, through infection, damage from microbial effectors, or pre-HCT conditioning, epithelial cell cytotoxicity is observed. This can dampen the strength of the immune response to pathogenic bacterial blooms, particularly if Goblet or Paneth cells are affected, as their mucin and AMP production are a crucial part of the immune response. Cell death also opens the intestinal barrier for transcellular transport of microbial effectors to the bloodstream (compared to paracellular permeability when tight junctions are weakened). Intestinal stem cells (ISCs) must compensate for epithelial cytotoxicity, although ISC regeneration is impacted by microbial metabolites as well. For instance, microbially produced butyrate impairs ISC recovery and butyrogenic bacteria have been associated with GvHD that does not respond to steroid treatment in one study, though butyrate has been associated with reduced risk of GvHD in other studies.[45–49]

1.3.1.1 GvHD biomarkers

Some of the thesis work's results indicate a possible role as biomarkers for GvHD, and so the current knowledge on such biomarkers is described. There have been substantial research efforts to identify a biomarker of GvHD before disease onset to begin GvHD prophylaxis as early as possible and ideally, prevent GvHD entirely or mitigate GvHD severity. Attention to biomarkers in stool samples revealed statistically significant separation of fecal calprotectin and α1 antitrypsin levels by GvHD status, corticosteroid response and response to first vs. second line therapies, however some of the median levels were similar.[50] While α1 antitrypsin specifically indicates poor mucosal barrier integrity, fecal calprotectin is nonspecific and indicates general inflammation.[51,52] Within the previous approximate decade, two promising serum biomarkers were identified: suppressor of tumorigenicity-2 (ST2) and regenerating islet-derived protein 3-α (REG3α). The levels of ST2 and REG3α in serum collected from HCT recipients seven days post-HCT and one week after GvHD onset or treatment associated with non-relapse mortality (NRM), overall survival (OS) and steroid-response (SR).[53–56] An algorithm

incorporates the ST2 and REG3α level to classify the individual as high or low probability for NRM, SR and OS post-HCT.[53–56] These outcomes post-HCT did correlate with the two-biomarker probability classification, supporting the use of this classification system to inform tailored treatment or prophylaxis post-HCT. This was validated in multiple studies and cohorts.[53,55,56] These findings are promising and undergoing validation in clinical trials, but none have undergone every step recommended by the FDA to be considered clinically reliable and used in clinical practice.[57] Thus, there is still a need for biomarker(s) to identify GvHD that can be used by physicians.

1.3.2 Role of gut microbiome in GvHD

1.3.2.1 Microbial taxa and GvHD

As described above, tissue injury instigated by pre-HCT conditioning can lead to mucosal barrier injury which compromises integrity. A more permeable gastrointestinal tract permits much greater crosstalk between the gut, bloodstream and body. Interactions between the host and gut microbes are even higher stakes when the gut barrier is has low integrity, as microbial metabolites can further prime APCs and contribute to immunity, systemic inflammation and GvHD. Low gut microbial diversity has been linked to poor outcomes post-HCT and GvHD in numerous studies.[2–4,58–69] GvHD is marked by loss of intestinal barrier integrity and inflammation.[3,34,35,70–72] Studies in humans and mice have linked presence and/or high abundance particular microbial taxa, such as *Akkermansia muciniphila,*[73] *Clostridiales* order,[70,74] *Clostridiaceae* family,[72] *Clostridia* class,[75] *Clostridiales leptum*,[72] *Oscillospiraceae* family,[65,72,76] *Coprococcus* genus,[77] *Eubacterium* like *E. rectale and E. limosum,*[66,72,75,77,78] *Lachnospiraceae*,[65,66,68,72,76] *Blautia* genus like *B. obeum,*[45,66,71–73,76,77,79] *Ruminococcus* genus,[77] and *Bacteroides* genus such as *B. thetaiotaomicron, B. ovatus,* and *B. caccae*[73,79] with better GvHD outcomes. Other taxa, such as *Enterobacteriaceae* family,[65,66,72] *Enterococcus* genus like *E. faecium* and *E. faecalis,*[44,63,64,67,70,72,73,75,80–82] *Streptococcus* genus,[63,67,82] *Staphylococcaceae* family,[68] *Staphylococcus* genus,[67] *Klebsiella* genus,[67] Veillonella genus,[83] *Escherichia* genus,[63,67] *Enterococcaceae* family,[68] *Lactobacillales* order,[63,70] *Lactobacillus* genera like *L. gasseri,*[63,70,73,75] *Pasteurellaceae* family,[69] and *Bacteroides doeri,*[79] presence and/or high abundance are linked with worse GvHD outcomes.[1,58] The gut microbiota and their metabolites may influence GvHD risk through impacts on the intestinal barrier and immune cell populations.[2,33–35,84,85] For instance, germ-free mice have lower incidence of acute gut GvHD vs. conventional mice. But when decontamination of the human gastrointestinal (GI) tract is attempted, the off-target effects (e.g. fungal overgrowth) outweigh any potential benefit from reducing the microbial load. The preparative chemoradiotherapy can also induce oxidative

stress in the HCT recipient. The gut microbiota can mitigate or exacerbate such oxidative stress depending on the exact composition. *Enterococcus faecalis*, which has been associated with proliferation post-HCT and during GvHD, can induce free radicals to facilitate DNA damage whereas *Clostridium* species can produce H2 that is able to detoxify reactive oxygen species and downregulate inflammatory genes.[86,87] Gastrointestinal tissue damage also recruits leukocytes, like macrophages and neutrophils, to the site of injury. Mucosal damage itself produces nitric oxide (NO) and leukocyte infiltration contributes to ROS production. At the intestinal barrier, free radicals can damage the epithelial cells and contribute to barrier damage that permits microbial metabolites, such as LPS, to translocate and cause serious complications post-HCT.[88] LPS triggers an inflammatory response from endothelial cells, exacerbating the pro-inflammation feedback cycle and enabling the bacterial translocation that leads to sepsis.[89,90] The level of tissue damage caused by free radicals and ROS in the intestine is even more relevant for GvHD because they cause tissue damage to the recipient, upregulating DAMPs, worsening inflammation and feeding into GvHD progression. Further, levels of ROS impact hematopoietic stem cell characteristics. Greater ROS levels can reduce stem cell quiescence, reduce pluripotency and increase differentiation, which can impede HCT recovery.[91,92]

1.3.2.2 Microbial metabolites and GvHD

Post-allogeneic HCT, recipients with GvHD tend to have perturbed microbiomes vs. recipients without GvHD, including characteristic microbial metabolites. Trimethylamine N-oxide, a microbially-modulated metabolite, is linked with GvHD.[93] On the other hand, high concentrations of dietary and fecal short-chain fatty acids (SCFAs) butyrate and propionate are protective against GvHD.[76,94,95] The exact mechanism of this protection may be multifold. Butyrate has been observed to modulate immune cell populations; for example butyrate is a histone deacetylase inhibitor (HDACi) which promotes hyper-acetylation of follicular regulatory T cells (T_{FR}) genes *Bcl6*, *Cxcr5*, and *Tcf7*.[96] This could impact GvHD because it is a T-cell mediated disease and T_{FR} cells maintain immune homeostasis by regulating self-targeting autoantibodies and self-antigens, which are key for GvHD initiation.[96] High levels of other microbial metabolites associated with health/protection against GvHD, including urinary indoxyl sulfate,[97] dietary indole-3-carboxaldehyde,[98] dietary tyrosine,[99] dietary, fecal, and intestinal taurine,[100] and colonic intestinal epithelial cell indoleamine 2,3-dioxygenase.[101] Most microbial metabolites are produced in the large intestine (e.g. the cecum, colon and rectum) because that is where the microbial load is greatest. Microbial load increases along the length of the intestines, with up to nine orders of magnitude separating the microbial content of the duodenum (approximately 10^3 colony forming units (CFU)/mL) from the colon (approximately

10^{12} CFU/mL).[102–106] This is despite the majority of carbohydrate (like starch), protein or lipid breakdown and absorption of these reaction byproducts (vitamins, amino acids, bile acids) occurring in the small intestine (e.g. duodenum, jejunum and ileum). Low levels of these dietary and/or host-biosynthesized compounds progress to the large intestine, where they encounter the microbes that modify/metabolize tryptophan, bile acids, fiber and more.[102–104,106] The stool samples used in this study do not reflect these nuanced alterations in microbial community composition and function throughout the GI tract; rather, stool more closely resembles the nearest portion: the large intestine and, in particular, the colon.[105] Although imperfect, stool more closely reflects this portion of the GI tract (large intestine/colon[105]) that I am interested in capturing: where microbial load is highest and where microbial metabolism of diet- and host-derived products (such as fiber and bile acids) occur. It remains challenging to determine whether microbial and microbially-mediated metabolic shifts in the gut may be causes, correlations, and/or consequences of gut GvHD.[20]

1.3.2.2.1 Tryptophan metabolites

One type of microbial metabolite that has been studied extensively is microbial tryptophan catabolites. This includes proteolytic degradation products such as indole, tryptamine, indoleacetate, indolepropionate, and indoleethanol. Each of these are primarily found in the gut. One reason that these metabolites receive research attention is because they uniquely have been linked to beneficial health effects, whereas microbial protein degradation products typically have an adverse impact on health.[107] The exact mechanisms behind the protective effect remain unclear, but may include modulation of the redox environment, improving gut barrier function, cell receptor binding (e.g. pregnane X receptor), host metabolism, and reduction of inflammation.[108,109] Metabolites like indole are examples of microbially produced signaling compounds that uniquely activate human vs. murine receptors, such that conclusions of indole bioactivity derived from mouse models should not be used as a proxy for human activity.[110] A producer of the tryptophan metabolite tryptamine, *Ruminococcus gnavus* resides in the gut although the impacts in humans require further investigation.[111] Some associations between tryptophan and tryptophan metabolites in human stool during disease have been observed; for instance, lower levels of tryptophan and indoleacetic acid were detected in stool samples of individuals with inflammatory bowel disease.[112,113] Although not microbially-modulated, the upregulation of host-expressed indoleamine 2,3-dioxygenase (IDO), the enzyme responsible for tryptophan catabolism, was reported to suppress GvHD in mice whereas its over-expression was associated with severe GvHD in humans. This underlies the importance of human studies for associating metabolites with disease states.[114,115]

1.3.2.2.2 Role of bile acids in GvHD

Recent studies have linked microbially-metabolized bile acids with intestinal health, including GvHD.[116] Bile acids function as lipid emulsifiers in the intestinal tract, facilitating absorption of dietary fats. Further research has elucidated their role as signaling molecules in the intestine, binding host cell membrane receptors like G-protein coupled receptor TGR5 and sphingosine-1 phosphate receptor 2 (S1PR2) and nuclear receptors such as farnesoid X receptor (FXR), pregnane X receptor (PXR), and vitamin D receptor (VDR) in intestine-resident cells, like intestinal epithelial cells, or immune cells, such as T cells.[116–122]. There are a variety of chemical modifications that the human host and gut microbiota can perform on bile acids, but transformations between primary vs. secondary and conjugated vs. unconjugated are crucial because each unique form binds and activates various cell receptors differently.[117–119] Some bile acids including isolithocholic acid, isodeoxycholic acid, and 3-oxolithocholic acid can suppress T cell differentiation into Th_{17} cells and promote Treg differentiation.[120–122] Isolithocholic acid and 3-oxolithocholic acid are both produced by human gut microbial commensals *Eggerthella lenta* and *Clostridium citroniae,*[121] isolithocholic acid by *Bacteroides fragilis* and *Mediterraneibacter gnavus*,[121] and isodeoxycholic acid by *Mediterraneibacter gnavus* and *Clostridium scindens.*[120] The relationship between bile acids and the microbes that modulate them is significant because each can be manipulated to influence the other.[123] Tissue-associated microbial communities can be curated to influence the bile acid pool in specific organs such as the liver, kidneys, heart and intestines to ultimately modulate disease.[123,124]

A previous study showed that T cell-induced inflammation alters microbial bile acid metabolism, contributing to bile acid dysbiosis that exacerbates GvHD.[116] VDR is specifically activated by lithocholic acid (LCA), and VDR expression varies based on the composition of the microbial community.[125] Knocking out VDR and FXR impacts immunity, as was observed in VDR and FXR knockout mice supplemented with bile acids that reduced the pool of colonic RORγ+ Treg cells. These cells were most impacted by VDR knockout vs. other bile acid receptor knockouts (including FXR).[125] Further, colonic bile acids influenced only the colonic RORγ+ Treg cells, without affecting other biological compartments (e.g. the spleen or mesenteric lymph node) and without impacting abundance of other T cells (e.g. TH17 and FOXP3+ Treg).[125] Colonic RORγ+ Treg cells promote intestinal homeostasis and minimize intestinal colitis, so this reduction in bile acid signaling and RORγ+ Treg cells could contribute to GvHD.[125]

1.3.2.2.2.1 Microbial bile acid metabolism genes: deconjugation via bile salt hydrolase/cholylglycine hydrolase (*bsh* genes)

Bile acids are enterohepatically circulated; they are produced in the liver from dietary and host-biosynthesized cholesterol before undergoing several host-mediated biotransformations. They are excreted in bile ducts and stored in the gall bladder for transport to the intestine as needed (e.g. after a meal). Bile acids enter the intestine as primary bile acids cholic acid (CA) or chenodeoxycholic acid (CDCA) conjugated to a taurine or glycine. These conjugated, primary bile acids encounter a range of gut microbes, some of which possess a choloylglycine hydrolase enzyme, also known as a bile salt hydrolase enzyme (encoded by *bsh* genes), such as found in *Roseburia intestinalis* and *Blautia obeum*. Canonically, bile salt hydrolases deconjugate bile acids in one step, freeing amino acids and creating unconjugated primary bile acids. This can occur in the small or large intestine.[105] Recent findings identified a novel ability of bile salt hydrolases to re-conjugate taurine or glycine back onto amino acids[126] Furthermore, bile salt hydrolases can re-conjugate bile acids to a diverse range of compounds, including all amino acids, polyamines, 5-aminovalerate, aminobutyrate.[127] The *bsh* gene is highly diverse; exact *bsh* gene sequences vary by strain, and evolutionary relationships do not correlate with *bsh* sequence identity[128]. Further, bile acid metabolic ability is not consistent within a bacterial genus or species.

1.3.2.2.2.2 Microbial bile acid metabolism genes: 7α-dehydroxylation (*bai* genes)

A large portion of bile acids are actively reabsorbed across the small intestinal barrier as unconjugated, primary bile acids where they are recycled via the portal vein back to the liver. The remaining bile acid pool continues to migrate along the intestine, encountering more gut bacteria in the colon, including microbes like *Clostridium scindens, Mediterraneibacter gnavus* and others that possess the *bile acid inducible* (*bai)* gene operon that encodes a series of enzymes required for the multi-step conversion of unconjugated, primary bile acids into secondary bile acids by removing a hydroxyl group from the seventh carbon (7α-dehydroxylation).[105,129] The majority of these unconjugated, secondary bile acids are passively reabsorbed across the intestinal barrier for recirculation back to the liver whereas a smaller portion are excreted into stool. In total, approximately 95% of bile acids are recirculated while 5% are excreted via stool.[130,131]

1.3.2.2.3 SCFAs

Ch 4 focuses on SCFAs in detail. Here, their known and hypothesized functions with respect to GvHD are summarized, but refer to Chapter 4.1.1 for an in-depth review. Short- and branched-chain fatty acids (SCFAs, BCFAs) are the products of microbial fermentation of nondigestible carbohydrates and proteins, respectively. For this thesis, both BCFAs and SCFAs are referred to as SCFAs.[132,133] They have a variety of physiological health impacts due to their

signaling function in the gut and due to their transport throughout the body via the bloodstream and downstream signaling in additional organs. A host of microbial taxa can produce SCFAs. These metabolites are present at levels ranging from 10-100 mM in the intestine; the precise level varies with the location along the gastrointestinal tract.[134,135] SCFAs are often associated with health, although the number of mechanisms underlying this beneficial effect remain ambiguous. SCFAs may trigger enhanced transcription of tight junction proteins that increase the integrity of the intestinal barrier, reducing risk of inflammation and infection.[42,136] SCFAs are also hypothesized to promote anti-inflammatory gene transcription by inhibiting histone deacetylation. SCFAs have been observed to bind G protein-coupled receptors (GPRs), including GPR109A and GPR43, to promote intestinal homeostasis and prevent inflammation.[137,138] There may also be an immune priming effect of propionate in which the inflammasome can more rapidly and robustly respond to pathogenic challenge.[94] When available to colonocytes, SCFAs serve as an energy source that permits faster turnover of damaged cells to reduce intestinal permeability.[49]

1.3.3 Role of diet/nutrition in GvHD

Diet and nutrition broadly impact human health and disease; in particular, diseases and health of the gastrointestinal tract (e.g. celiac disease). This also holds true for gastrointestinal GvHD. Dietary components can modulate the gut directly or indirectly by changing the intestinal environment and gut microbiome.[139,140] It is known that changes in the gut microbial community, such as reduced diversity, are associated with poor outcomes post-HCT.[63,71,79] Thus, diet and nutrition are viable pathways to control or mitigate GvHD.

1.3.3.1 Fiber

One dietary component, fiber, has been linked with human health, hematological malignancies, and gut GvHD.[139–141] Dietary fibers consist of non-digestible carbohydrates, although within this category are several subtypes defined by the solubility, source, and chemical structure of the fiber compound. These subtypes have nuanced effects on the gut environment; for instance, most insoluble fibers primarily contribute to fecal bulking because they are not digestible by humans or bacteria. On the other hand, soluble fibers are able to be fermented by some gut bacteria into microbial metabolites relevant for gut and overall health, like short-chain-fatty acids.[139] A high fat diet can induce gut microbial dysbiosis and intestinal mucosa degradation, but dietary fermentable fiber protects against these changes by promoting colonic epithelial cell turnover and growth of beneficial gut microbes.[142] There is also early evidence that fiber may impact the immune cell population, namely T cell differentiation.[143]

Post-HCT and during acute GvHD, supplemental dietary fiber reduced diarrheal and oral mucositis symptoms.[144] Using a murine model, dietary fiber was found to promote overall survival post-HCT. This outcome may be attributed to the observed protective effect of dietary fiber against gut microbial dysbiosis and pathogen outgrowth, or the promotion of regulatory T cells vs. conventional T cells.[141]

1.3.3.2 Vitamin D

Vitamin D deficiency has been extensively studied in the context of GvHD and linked to poor overall survival post-HCT in humans.[145,146] Vitamin D can be biosynthesized by humans in the skin with exposure to sunlight/ultraviolet radiation, but dietary supplementation is often needed to ensure sufficient levels.[147] Human hydroxylases metabolize dietary and biosynthesized vitamin D into the compound that is clinically assayed to evaluate vitamin D levels, 25-hydroxyvitamin D (will be referred to as vitamin D).[147] Canonically, vitamin D ensures adequate calcium and phosphorous absorption for bone health, although vitamin D is needed for a greater breadth of physiological processes like metabolism, muscle function, innate immunity, cell growth, and immunomodulation.[147] Such a widespread impact may be due to the presence of Vitamin D Receptor (VDR) on nearly all tissues and cell types in humans.[147] Case reports describe resolution of chronic GvHD symptoms and reduced relapse rates post-HCT in human recipients supplemented with Vitamin D.[148] Vitamin D receptor (VDR) binds both Vitamin D and LCA, but not other bile acids.[149] VDR is present on intestinal epithelial cells and T cells, which are crucial cells for GvHD progression.[150] VDR activity, like that which can be triggered by LCA binding and activation, can promote tight junction protein gene expression to preserve intestinal barrier integrity, abrogate pro-inflammatory IL-12, NFκB, and TNFα gene expression, and impact differentiation of T cells.[151–154]

1.4 Experimental conceptualization/design

Before beginning my thesis project, the Fredricks lab was interested in how the gut microbiome may impact GvHD. Longitudinal stool samples from ten healthy control individuals and eight HCT-recipients with no GvHD, eight with low severity GvHD, and four high severity GvHD (and ten control blank swabs) were analyzed for metabolites that may link with GvHD severity. The timepoints examined for each participant were pre-HCT and pre-conditioning vs. post-HCT and pre-GvHD diagnosis. This analysis included over 650 metabolites and was performed by the company Metabolon. One striking observation was a higher level of the conjugated bile acids, taurocholic acid, glycocholic acid, and taurodeoxycholic acid post-HCT in HCT recipients with severe GvHD vs. all other participant groups examined. This was intriguing

to the Fredricks lab, yet the sample size was small, prompting the targeted and untargeted metabolomics approaches that built the foundation for my thesis project.

From an existing GvHD project in the Fredricks lab which enrolled more than 300 allogeneic HCT recipients, 49 control HCT recipients without GvHD were matched and paired to 49 HCT recipients with stage 2-4 intestinal or gut GvHD (Fig. 2). HCT recipients with stage 1 GvHD were excluded from this study to focus on the two poles of GvHD. The matching and pairing of control and case HCT recipients was based on the strength of the HLA match, HCT transplant type (e.g. umbilical cord vs. bone marrow vs. peripheral blood), donor type (e.g. sibling, unrelated, etc.), conditioning regimen (e.g. degree of chemo-radiotherapy, such as total body irradiation vs. reduced intensity conditioning), and GvHD prophylaxis regimen. This reduced many confounding factors attributed to HCT besides GvHD status. From each of the 98 participants in the study, three longitudinal stool samples were selected for analysis (when available): pre-transplant to represent a baseline, day 30 post-transplant to represent the impact of the HCT, and day 60 post-transplant to represent recovery and engraftment post-HCT. Many participants had stool samples available on the exact day of interest, but for those without stool samples from precisely day 30 or day 60 post-HCT, a window of variability of +/- 7 days was acceptable for the purpose of our study. Even with the window of variability, a few participants did not have a corresponding stool sample. There were 247 total samples across cases, controls, for all timepoints.

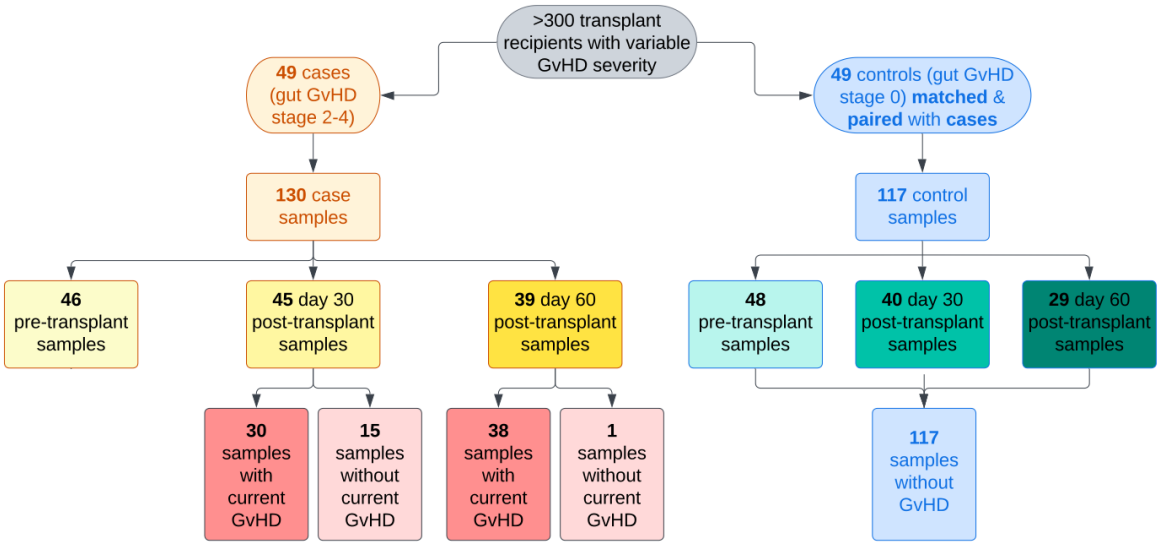


Figure 2. Study population and sample selection. From an existing GvHD project in the Fredricks lab which enrolled more than 300 allogeneic HCT recipients, 49 control HCT recipients without GvHD were matched and paired to 49 HCT recipients with stage 2-4 intestinal or gut GvHD. This produced 130 case samples and 117 control samples across all three timepoints. The GvHD case samples can be further subdivided by normalizing to GvHD diagnosis (e.g. whether GvHD had been diagnosed by the point of sample).

Chapter 2: Materials and Methods

Several of the Methods in this chapter were adapted from published work: Cagle et al., Acute gastrointestinal graft-versus-host disease is associated with reductions of secondary bile acids after allogeneic hematopoietic cell transplantation. *Frontiers in Microbiology (accepted 3/31/26).*

2.1 Study design and population

Prior to my joining the Fredricks lab in 2020, a cohort of 329 allo-HCT recipients undergoing allogeneic HCT at the Fred Hutchinson Cancer Center in Seattle, WA between April 2013-March 2020 was recruited with prospective collection of weekly stool, blood and/or tissue samples and dietary questionnaires under IRB#2608 with written informed consent. For the stool samples, weekly sample collection kits were provided to patients that contained polyurethane foam swabs (Puritan Medical, Guilford ME) to self-collect stool samples at home; nurses collected swabs from inpatients. Swabs were returned to our laboratory by courier on the same day or by patients using insulated containers and ice packs if more than a day. Upon arrival, samples were tagged and stored at −80°C until used for DNA extraction and metabolomics.

2.2 Experimental design

The goal of this study is to test the hypothesis that microbially-metabolized bile acids play a role in gut GvHD initiation and progression to better understand why particular HCT-recipients develop severe gut GvHD whereas others do not. The study also seeks to identify bile acid targets associated with GvHD to investigate for diagnostic or therapeutic application to protect against gut GvHD. To accomplish this goal, we performed targeted bile acid metabolomics to measure bile acid concentrations, metagenomics to capture relative levels of microbial bile acid metabolic genes, and 16S rRNA sequencing to characterize the bacterial communities of stool samples from HCT-recipients. Stool samples were collected approximately weekly, but this study's sample sizes were informed by key timepoints in the HCT timeline (pre-HCT, 30 days post-HCT, and 60 days post-HCT) and the number of HCT-recipients that developed GvHD and suitably matched to an HCT-recipient that did not develop GvHD along clinical and demographic criteria. No data from these analyses were excluded in this study.

2.3 Case-control design

From the greater study population of 329 allogeneic HCT recipients at the Fred Hutchinson Cancer Center, participants with stage 2-4 acute gut GvHD were selected and matched 1:1 to participants without gut GvHD (stage 0). 49 participants with stage 2-4 acute gut GvHD were matched to 49 control participants without gut GvHD. Although GvHD is a time-dependent variable, we treated GvHD status as binary variable for the purpose of this study because there is no time-dependency in the no-GvHD group to match upon. The median time to GvHD diagnosis within the GvHD case group was 25 days (Fig. 3). The criteria for matching/pairing participants included degree of human leukocyte (HLA) matching, conditioning regimen, stem cell source, and GvHD prophylaxis/immunosuppression. Stool samples from the selected participants were chosen from three timepoints, plus or minus seven days from that timepoint.

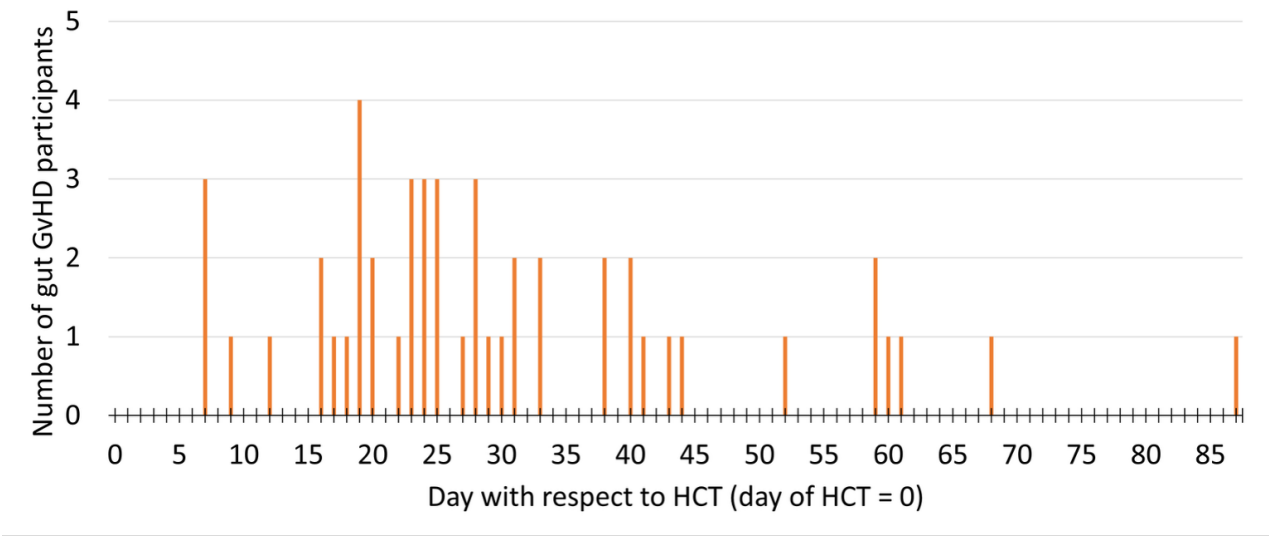


Figure 3. Histogram describing the distribution of time to GvHD diagnosis in days relative to HCT (day 0) for all 49 gut GvHD participants. Median time (days) to GvHD is 25.

2.4 DNA extraction

Bacterial DNA from all 247 samples in the 98-participant cohort was extracted using the BiOstic Bacteremia DNA Isolation Kit (Qiagen, Germantown, MD) as described previously.[155] Blank swabs were included as DNA extraction (negative) controls to monitor for contamination from extraction reagents. This work was performed by technicians Martha DeMeules and Congzhou Liu.

2.5 Broad-range 16S rRNA gene PCR sequencing and taxonomic assignment

Bacterial DNA concentrations were measured as described previously.[156] Broad-range 16S rRNA gene PCR with Illumina sequencing targeting the V3-V4 region of 16S rRNA gene was performed on all samples, as previously described.[155] A mock bacterial community from ATCC was also sequenced as a positive control to ensure performance of laboratory methods and the bioinformatics pipeline. Negative controls for sequencing included both no-template PCR and DNA extraction controls to monitor for contamination. Raw sequence reads were processed using the DADA2 package[157] and a list of unique sequence variants (SVs) was generated. Sequence reads are available from the NCBI Short Read Archive (Bioproject PRJNA1281039). Sequences were classified using a reference set tailored for the human gut microbiota.[79] Taxonomy was assigned to each unique SV by phylogenetic placement (*pplacer*). Bacterial taxa represented by fewer than 25 reads in a sample were removed to minimize environmental

contaminant sequences from being included in the final dataset. This work was done by Sujatha Srinivasan, PhD.

2.6 Metagenomic sequencing and analysis

2.6.1 Rationale for metagenomic approach

I was first interested in quantifying the abundance of *bsh*- and *bai*-possessing microbial genes in stool samples and connecting these data with the bile acid metabolomics. A review of the literature revealed that a wide variety of gut microbes possess *bsh* genes and possession is variable within each microbial taxon. Thus, a gene-forward approach was most appropriate to best capture *bsh* quantity in each sample. Ideally, a quantitative polymerase chain reaction (qPCR) assay would most comprehensively assess the relationship between gene concentrations and metabolite. The first step of creating a *bsh*-specific qPCR was to determine the sequence identity of a representative set of *bsh* genes to assess feasibility of designing probes and primers. The set of *bsh* genes was chosen by selecting public/published *bsh* gene sequences from microbial species and genera that were detected by 16S rRNA sequencing in my sample set. I performed a multiple sequence alignment of these representative, diverse *bsh* genes and found very low sequence identity and no broad regions of similarity across the representative genes (Fig. 4). In the absence of regions of similarity, it was impossible to design primers or probes to cover diverse *bsh* genes and therefore not feasible to create a qPCR approach to quantify *bsh* genes. Thus, I pivoted to a metagenomic approach since this did not require any sequence similarity/identity of the target genes (*bsh, bai*).

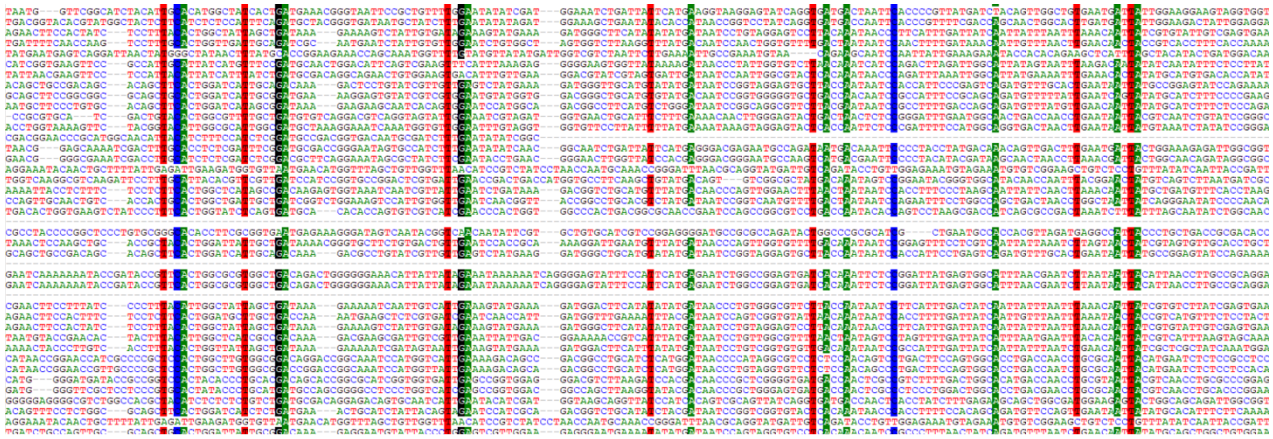


Figure 4. No broad regions of similarity in diverse *bsh* genes by multiple sequence alignment. Across diverse *bsh* sequences from representative bacterial taxa detected in the case-control cohort stool samples there was very poor sequence identity and no broad regions of similarity across the representative genes. Columns of colors correspond to nucleotides that are consistent/conserved across the *bsh* sequences and letters denote the nucleotide, such that red = T/thymine, blue = C/cytosine, green = A/adenine, black = G/guanine and blanks are unknown.

2.6.2 Whole genome metagenomic sequencing

This workflow was entirely performed on all 247 stool samples within the 98-participant case-control cohort by Martha DeMeules and Congzhou Liu. Genomic DNA was quantified using the Quant-iT™ dsDNA Assay Kit, high sensitivity (HS) (Thermofisher Scientific; Waltham, MA). Sequencing libraries were prepared using 625 pg genomic DNA and the quarter reaction workflow using the Illumina Nextera-XT library prep (Illumina; SanDiego, CA). The Illumina DNA/ RNA UD indices and 12 cycle indexing was performed. Indexed libraries were pooled by volume and clean up was performed using the 0.8X Agencourt AMPure XP beads (Beckman Coulter; Indianapolis, IN). Size distribution within the library pool was validated using an Agilent High Sensitivity D5000 Screen Tape on the Agilent 4200 TapeStation (Agilent Technologies Inc; Santa Clara, CA). Additional library QC and cluster optimization was done using an Invitrogen Qubit® 2.0 Fluorometer (Life Technologies; Carlsbad, CA). Sequencing was performed on a NovaSeq 6000 instrument using an S2 flowcell employing paired-end, 150 bp read lengths.

2.7 Metabolomics

The targeted and untargeted LC-MS metabolomics analyses (including sample preparation and LC-MS processing) described in this section (2.7) were performed by Dr. Dan Raftery and members of his lab: Danijel Djukovic, Dr. Wentao Zhu, and Hayley Purcell.

2.7.1 Targeted bile acid analysis by liquid chromatography-mass spectrometry (LC-MS)

The MS based approach used here was adapted from previously reported methods.[158,159] Human stool swab samples were moved from -80 °C storage and thawed at 4 °C in 250 µL of cold MeOH/PBS (4:1, v/v), then vortexed for 60 sec. A portion (200 µL) of sample was transferred into a separate 2 mL Eppendorf vial, and 10 µL of isotope labeled internal standard stock solution (composition of the internal standard stock solution can be found in Table 1) was added, along with 600 µL methanol. These sample vials were vortexed again for 10 sec and stored at -20 °C for 20 min. Vials were centrifuged at 18,000 x g (at 4 °C) for 15 min. The supernatant was collected (650 µL) and placed in another 2 mL Eppendorf vial, then dried at 30 °C for approximately 2.5 hrs in a Speed Vac. Samples were reconstituted with 100 µL MeOH/water (1:1, v/v), vortexed for 10 sec, then centrifuged for 5 min at 18,000 x g (at 4 °C). The supernatants (80 µL) were transferred into LC vials for MS analysis.

Concentration (µM)	Compound name	Compound abbreviation	Source
9.94	Deuterated glycochenodeoxycholic acid	d4-GCDCA	Cambridge Isotope Labs (Tewksbury, MA)
9.94	Deuterated lithocholic acid	d4-LCA	Cambridge Isotope Labs (Tewksbury, MA)
9.94	Deuterated deoxycholic acid	d4-DCA	Cambridge Isotope Labs (Tewksbury, MA)
9.94	Deuterated glycocholic acid	d4-GCA	Cambridge Isotope Labs (Tewksbury, MA)
10.08	Deuterated cholic acid	d4-CA	Cambridge Isotope Labs (Tewksbury, MA)

Table 1. Composition of Internal Standard stock solution used in targeted bile acid liquid chromatography-mass spectrometry. These compounds are in solvent composed of MeOH and water in a 1:1 ratio.

A Waters Acquity I-Class UPLC TQS-micro MS (Waters, Milford, MA) was used for targeted MS analysis under negative ionization mode and using a Waters XSelect HSS T3 column kept at 40 °C for chromatographic separation. The mobile phase was composed of two solvents: 5 mM ammonium acetate in H_2O with 0.1% acetic acid (A) and acetonitrile with 0.1% acetic acid (B). After 1 min of isocratic elution at 75% of solvent A, it was decreased to 5% at t=15 min. Run solvent composition was then maintained at 5% for 10 min, followed by an increase to 75% at t=25 min. The total experimental time for each injection was 40 min. A total of 55 bile acids were targeted, of which 32 bile acids had measurable signals and were analyzed. The five stable isotope (deuterium)-labeled internal standards were used to determine the bile acid molar amounts.[160,161]

2.7.2 Targeted short chain fatty acid (SCFA) metabolomic analysis

This targeted LC-MS metabolomic analysis uses isotope-labeled derivatization to quantify ten SCFA metabolites in fecal samples. Samples were prepared by transferring approximately 10-20 mg of dried fecal sample into a 1.5 mL Eppendorf vial. The precise weight was measured and recorded to use for normalization. To the same vial, 100 µL Milli-Q water and zirconium oxide beads (proportional to sample size) were added. This mixture was homogenized in a Bullet Blender, cooled to 4 °C. Next, 300 µL methanol was added before a 30 second vortex, after which the sample was stored at -20 °C for 20 min. The suspension was sonicated in an ice bath for 15 minutes, centrifuged at 14000 rpm for 30 min at 4 °C, and all liquid was collected into a new 2 mL Eppendorf vial. The solids were carefully rinsed with methanol:H_2O (3:1, v/v) if needed. Samples were dried at 30°C in an Eppendorf Vacufuge drier (~1 hour).

When ready for derivatization, samples were reconstituted with 225 µL 50% ACN/50% H_2O. After a 10 second vortex and shake for 2 mins, samples were centrifuged at 14000 rpm for 10 min at 4 °C. Then 40 µL supernatant were transferred into 2 mL LC glass vials. The solvents needed for derivatization were also prepared: 500 mL 10ACN/90H2O (2 mL for dilution of each sample; 200 ml for Internal Standards [IS]), 100 mL 1:1 ACN/H2O and 2-3 mL 1:1 ACN/H2O with 6% pyridine. The solid reactants were also weighed according to Table 2*,* precise weights were recorded and the measured reactants were stored at -20 °C until the solutions are prepared for dissolving the solids.

	MW	**grams for 2 mL, 100 samples**	**grams for 3 mL, 150 samples**
200 mM 3NPH HCL in 1:1 ACN/H2O	**189.6**	**0.07584**	**0.11376**
120 mM EDC HCL in 1:1 ACN/H2O with 6% pyridine	**191.7**	**0.07668**	**0.11502**

Table 2. Weighing solid reactants for internal standards to perform targeted SCFA metabolomics analysis. Different quantities of reactants are required based on the number of samples analyzed. 3NPH = 3-nitrophenylhydrazine, HCl = hydrochloric acid, ACN = acetonitrile, EDC = N′-ethylcarbodiimide HCl

The mobile phase solvents were prepared**:** mobile phase A consisted of 0.01 % HFA in 100% H2O 2L and mobile phase B of 0.01 % HFA in 100% ACN 1L. The LC column was 5 cm long, 2.1 mm id with 1.7 micron beads (BEH C18). The negative (-) electrospray ionization (ESI) mode was selected and the LC Column Chamber Temp set to 40 °C. A 20 µL auto Wash (10%ACN+90%H2O) was injected for 10 seconds. Flow rate (ml/min) was set to 0.40.

1684 µL was added into a 2 mL glass LC vial and the acids added one by one according to Table 3. From the solution with mixed acids 75 mM (1.5 M HAc), 10 µL was transferred into 1865 µL 1:1 ACN/H2O in a 2 mL glass LC vial for 400 µM SCFAs (8 mM HAc; solution 0). This was used for serial dilution of Light Standards (L) and for making isotopic labeled Heavy Standards (H). The following were diluted from stock solution 0 using rinsed 2 mL glass vials or Eppendorf vials:

- 200 uM (500 µL 0 + 500 µL 1:1 ACN/H_2O)
- 100 uM (500 µL a + 500 µL 1:1 ACN/H_2O)
- 50 uM (500 µL b + 500 µL 1:1 ACN/H_2O)
- 10 uM (200 µL c + 800 µL 1:1 ACN/H_2O)
- 2 uM (200 µL d + 800 µL 1:1 ACN/H_2O)
- 500 nM (250 µL e + 750 µL 1:1 ACN/H_2O)
- 200 nM (400 µL f + 600 µL 1:1 ACN/H_2O)
- 100 nM (500 µL g + 500 µL 1:1 ACN/H_2O)
- 50 nM (500 µL h + 500 µL 1:1 ACN/H_2O)
- 25 nM (500 µL i + 500 µL 1:1 ACN/H_2O)

Acids	MW	Density	Volume (µL) 150 µM (3 mM for Acetic acid)	Diluted to 2mL for 75 mM 1.5 M HAc
acetic acid	60.05	1.05	171.5714286	2000-315.94=
propionic acid	74.08	0.99	11.22424242	1684 µL
butyric acid	88.11	0.96	13.7671875	Add acids to 1:1 ACN/H2O
isobutyric acid, 2-methylpropanoic acid	88.11	0.97	13.62525773	
2-methylbutyric acid	102.13	0.936	16.36698718	
isovaleric acid, 3-methylbutyric acid	102.13	0.925	16.56162162	
valeric acid, pentanoic acid	102.13	0.93	16.47258065	
caproic acid, hexanoic acid	116.16	0.93	18.73520968	
3-methyl-valeric acid, 3-methylpentanoic acid	116.16	0.93	18.73548387	
isocaproic acid, 4-methylvaleric acid	116.16	0.923	18.87757313	
Total			315.9375724	

Table 3. SCFA Standards. These standards are needed for calibration of the targeted SCFA metabolomic analysis. HAc = acetic acid, ACN = acetonitrile

For both L and real samples, 40 µL of each acid was mixed with 20 µL 200 mM 3NPH in 1:1 ACN/H2O, and 20 µL 120 mM EDC in 1:1 ACN/H2O with 6% pyridine. This was vortexed first for 10 seconds and then again for 30 min at 40 °C in glass vials. MRM parameters were tuned using SmartX, use derivatives of 400 uM FAs (8 mM HAc). Each product was diluted by adding 1.92 mL 10ACN/90H2O and stored at -20 °C, as needed, after removing 200 uL.

For H, 1 mg 13C6-3NPH HCL was weighed in a 2 mL glass LC vial and 50 µL of 400 uM FAs (8 mM HAc; solution 0) then 25 µL of 120 mM EDC HCL with 6% pyridine was added. The solution was vortexed for 10 seconds and then again at 40 °C for 30 min, followed by dilution with 10ACN/90H2O to 200 ml. The final solution was stored at -20 °C.

For LC-MS/MS, 500 µL H and 500 µL L/sample were mixed as 1:1, vortexed, and 20 µL was injected.

2.7.3 Untargeted/global metabolomic analysis

The LC system is composed of an Agilent 1290 high speed binary pump, an Agilent 1290 multisampler, an Agilent 1290 column compartment and an Agilent 1260 iso pump (Agilent Technologies, Santa Clara, CA). The LC modules are controlled by the Agilent Masshunter Workstation B.11.0.203 software (Agilent Technologies, Santa Clara, CA). The chromatographic separation is performed in hydrophilic interaction liquid chromatography (HILIC) mode on an XBridge BEH Amide column (150 x 2.1 mm, 2.5 µm particle size, Waters Corporation, Milford, MA, Part No. 186009930) and in reversed-phase chromatography on an ACQUITY BEH C18 column (150 x 2.1 mm, 1.7 µm particle size, Waters Corporation, Milford, MA, Part No. 188002353). Each sample is injected twice in each chromatography, 10 µL for analysis using negative ionization mode and 5 µL for positive ionization mode. The flow rate is 0.300 mL/min, the auto-sampler temperature is kept at 4 °C and the column compartment is set at 45 °C in each chromatography. The total separation time for both ionization modes is 30 min in HILIC mode and 25 min in reversed-phase chromatography mode. The mobile phase is composed of Solvent A (95% H20 [dispensed by Millipore Synergy UV water purifier, EMD Millipore, Billerica, MA]/3% ACN [MS-grade, Sigma-Aldrich, Saint Louis, MO]/2% MeOH [Sigma-Aldrich, Saint Louis, MO] with 0.2% [volume] acetic acid [Sigma-Aldrich, Saint Louis, MO] and 10 mM ammonium acetate [Sigma-Aldrich, Saint Louis, MO]) and Solvent B (5% H20/93% ACN/2% MeOH with 0.2% acetic acid and 10 mM ammonium acetate) in HILIC mode, Solvent A (100% H20 with 0.1% [volume] formic acid) and Solvent B (100% ACN with 0.1% formic acid) in reversed-phase chromatography. The gradient conditions for both ionization modes are identical in each chromatographic separation, and are shown in Tables 4 and 5.

Time Segment, min.	Solvent A, %	Solvent B, %
0	5	95
3	5	95
8	50	50
12	50	50
13	5	95
30	5	95

Table 4. LC Gradient Conditions in hydrophilic interaction liquid chromatography (HILIC) mode.

Time Segment, min.	Solvent A, %	Solvent B, %
0	99	1
1	99	1
10	1	99
16	1	99
16.1	99	1
25	99	1

Table 5. LC Gradient Conditions in reversed-phase chromatography mode.

After chromatographic separation, MS ionization and data acquisition were performed using an Agilent 6546 Q-TOF mass spectrometer (Agilent Technologies, Santa Clara, CA) equipped with Dual Agilent Jet Stream Electrospray Ionization (Dual AJS ESI) source. The instrument is controlled by the Agilent Masshunter Workstation B.11.0.203 software. Aqueous global profiling data acquisition is performed in the full scan profile and centroid mode in the m/z range of 60-1,000. The drying and nebulizer gas for ESI source is N_2 (99.999% purity). The drying gas flow rate is 10 L/min, and the nebulizer gas pressure is 35 psi. The sheath gas flow rate is 10 L/min, and the sheath gas temperature is 350 °C. The ion source temperature is 325 °C. The ionization voltage is 3.0 kV.

Data processing was performed using Progenesis Qi software v. 2.2.5826.42898 (Nonlinear Dynamics; Newcastle; UK). Peak alignment was carried out taking samples as reference. Peak-picking was performed using sensitivity and chromatographic peak width at 1 and 0.05 min; respectively. The retention time limit was set 1.0-15.0 min. Possible adduct ions were defined as follows: [M+H]+; [M+Na]+; [M+NH4]+; [M-H2O+H]+; and [2M+H]+ in (+) ESI mode; and [M–H]–; [M+Cl]–; and [2M+CH3COO–H]– in (-) ESI mode. The adduct ions were grouped into mass

features through peak deconvolution. Putative peak annotation was performed by searching metabolites from MONA Database (https://mona.fiehnlab.ucdavis.edu/) using accurate m/z measurements from the full scan data and MS2 spectra. We set the m/z tolerance of 20 ppm (MS1) and 30 ppm (MS2). Metabolomics methods were developed and applied by Danijel Djukovic, Dr. Wentao Zhu, Hayley Purcell, and Dr. Dan Raftery.

2.8 Data handling, transformations, statistics

2.8.1 Targeted metagenomic gene analysis

We curated a database for the microbial bile acid metabolic genes of interest: *bile salt hydrolase (bsh)* encoding cholylglycine hydrolase enzymes that can re-/deconjugate bile acids, and the *bile acid inducible (bai)* operon encoding the enzymes needed to 7α-dehydroxylate bile acids. Two distinct approaches were adopted for the custom *bsh* vs. *bai* gene databases, using different search criteria for the metagenomic data because *bsh* gene sequences are highly diverse compared to *bai* gene sequences. The *bsh* gene database/catalog: Since *bsh* gene sequences are diverse, a sequence-based search would be inherently limited in breadth of *bsh* genes included. Thus, we performed a keyword search, "choloylglycine hydrolase," in February 2025 on the NCBI Protein database (https://www.ncbi.nlm.nih.gov/protein/). We applied criteria restricting the search space to 'Bacteria' and 'RefSeq' as the source database to promote inclusion of on-target, non-redundant, data-validated protein sequences. At the time of the search, this yielded 9,768 sequences which were downloaded and used as the *bsh* gene database/catalog for metagenomic analyses. This work was done with the assistance of Dr. Sam Minot at Fred Hutch.

2.8.2 The bai gene database/catalog

The *bai* genes are more similar in sequence identity and more conserved within taxa. This operon is composed of at least 7 *bai* genes which are well defined in *Clostridium scindens*.[162] The *baiA2, baiB, baiCD, baiE, baiF, baiG* and *baiH* genes from *C. scindens* were used as search criteria in April 2025 on NCBI BLAST (blast.ncbi.nlm.nih.gov/Blast.cgi) blastp program[163,164] and all sequences with > 49% sequence identity were selected and included in the *bai* gene database/catalog. Using the microbial genes described above, the *gig-map* bioinformatics utility (github.com/FredHutch/gig-map) was used to generate a microbial gene catalog indexed for rapid alignment as follows:

- Build Gene Catalog:
    - Summary: Deduplicate the genes' protein-coding sequences based on amino acid similarity clustering using CD-HIT.

    - Nextflow workflow: FredHutch/gig-map/deduplicate.nf.
    - Analysis parameters:
        - cluster_coverage:0.9,
        - min_gene_length:50,
        - cluster_similarity:0.9
- Align Reads to Genes:
    - Summary: Quantify the number of reads which align to each gene in a precomputed catalog using the DIAMOND short read aligner.
    - Nextflow workflow: FredHutch/gig-map/align_reads.nf
    - Analysis parameters:
        - max_evalue: 0.001
        - min_identity: 90
        - min_score_reads: 50

This work was done with the assistance of Dr. Sam Minot at Fred Hutch.

2.8.3 Statistical analysis and software

The concentrations of bile acids, the sums or diversity of bile acid metabolism genes, and abundance of microbial taxa by 16S rRNA gene sequencing within stool samples at each time point were analyzed for statistical significance using the Mann-Whitney U test since the means are non-parametric and non-evenly distributed. A 10% prevalence filter of all samples was applied to the microbial taxa abundance data before analyzing by prevalence or abundance. Odds ratios and prevalence comparisons were analyzed for statistical significance using the Fisher's Exact test since the prevalence counts are non-parametric. Correlative relationships' strength was evaluated by Kendall-Tau correlation (more conservative vs. Spearman) given that the sample populations were non-parametric and non-evenly distributed and the data could be sparse with small values. A 10% prevalence filter of all samples was applied to the microbial taxa abundance data and to the bile acid concentration data before calculating correlations. Where p-value adjustment for multiple comparisons was appropriate, the Benjamini-Hochberg false discovery rate (FDR) adjustment was used. Adjusted p-values ≥ 0.05 were considered significant. The Python (version 3.12.6)[165] pandas (2.2.3), numpy (2.1.1), and openpyxl (3.1.5) packages were used for data organization.[166–168] The Python statsmodels (0.14.4), skbio (version 0.7.2), sklearn (1.5.2), and scipy (1.14.1) packages were used for all statistical analyses.[169–172] The following formula was used to calculate the R^2 values in: $R^2 = 1 / (1+((n-g)/((g-1)*F)))$, where n = number of samples, g = number of groups, and F = the PERMANOVA F-prime statistic.[173] Figures were created using matplotlib (3.9.2) and seaborn

(0.13.2).[174,175] Comparisons between relative abundance levels of microbial taxa were made in R (4.4.1) using Physloseq (1.50.0), microViz (0.12.6), ggplot2 (3.5.1), plyr (1.8.9), dplyr (1.1.4), RColorBrewer (1.1-3), and ggiraph (0.8.12) packages.[176–183] Additional R packages used for data handling and analysis include grid (4.4.1), scimpute (0.0.8), tidygraph (1.3.1), ggraph (2.2.1), purr (1.0.4), readxl (1.4.3), tidyverse (2.0.0), compositions (2.0-8), writexl (1.5.4), tidyr (1.3.1), and pheatmap (1.0.13).[183–193]

2.8.4 Bile acid dataset from a similar Cancer Center

Publicly available data were downloaded from the Lindner et al. research article published in 2024 from Memorial Sloan Kettering Cancer Center.[116] The data analyzed were from, 'Source Data Fig. 4: Information on the cohort (GVHD vs controls), the concentrations/AUC of BAs, BA family.' Metadata from sheet, 'cohort_BAS,' columns A-D was used with bile acid concentration data from sheet, 'filtered_combined_table,' column AK, 'lithocholic_acid,' and column AL, 'lithocholic_acid_3_sulfate.' Mann-Whitney U test was used to compare the distribution of each bile acid's concentrations in GvHD cases vs. controls at timepoints peri-engraftment and post-engraftment for each participant.

2.8.5 Steroid (prednisone/methylprednisolone) analysis

Active prednisone/methylprednisolone therapy was defined as receiving any dose of prednisone or methylprednisolone on any day in the seven days prior to sample collection date. Samples that were collected while the participant was receiving steroid treatment were excluded in a separate sensitivity analysis to determine if prednisone may bias the measured bile acid levels.

2.9 Study approval

The study was Institutional Review Board (IRB) approved and all patients provided written informed consent prior to participation (under IRB#2608).

2.10 Data availability

The original contributions presented in the study are publicly available. The 16S rRNA gene sequencing read data can be found in the NCBI Short Read Archive: Bioproject PRJNA1281039. The metagenomic sequencing read data will be made available/accessible in the NCBI Short Read Archive. The metabolomics data will be made available/accessible on the National Metabolomics Data Repository (NMDR)/Metabolomics Workbench.

2.11 Summary of work performed by others

Wherever work/methods were performed by others, the exact workflow(s) and individual(s) responsible are specified and credited. In summary, 16S rRNA gene sequence analysis and construction and maintenance of the taxonomic analysis/pipeline was performed by Dr. Sujatha Srinivasan. Martha DeMeules and Congzhou Liu aggregated stool samples, extracted the microbial DNA from the samples and prepared the samples' DNA as libraries for sequencing. Danijel Djukovic, Dr. Wentao Zhu, Hayley Purcell and Dr. Dan Raftery performed all LC-MS preparation, methods and techniques, including SCFA targeted metabolomics, bile acid targeted metabolomics and untargeted/global metabolomics. This group also performed the processing of the resulting LC-MS prior to my analyses. Dr. Sam Minot constructed the pipelines that I used for targeted metagenomic gene analysis, namely the *gig-map* bioinformatics utility (github.com/FredHutch/gig-map), including Nextflow workflows that deduplicated genes' protein-coding sequences (FredHutch/gig-map/deduplicate.nf) and that aligned then quantified number of metagenomic reads which aligned to each gene in the constructed *bai* and *bsh* gene catalog (FredHutch/gig-map/align_reads.nf). Analysis of study participants' diet questionnaires was facilitated by the Fred Hutch Nutrition Assessment shared resource. Integration, analysis and comparison of these diet data with metabolomic and metagenomic data was initially investigated by interns Supraja Kadagandla and Annie Frohlich in the lab.

Chapter 3: Bile acids and GvHD

Most of this chapter was adapted from published work: Cagle et al., Acute gastrointestinal graft-versus-host disease is associated with reductions of secondary bile acids after allogeneic hematopoietic cell transplantation. *Frontiers in Microbiology (accepted).*

3.1 Introduction

3.1.1 What is allo-HCT and GvHD?

Allogeneic hematopoietic cell transplantation (HCT) is a potentially curative treatment for patients with hematologic malignancies and non-malignant hematologic disorders. However, significant morbidity and mortality is associated with acute graft-versus-host disease (GvHD) that affects 30-70% of HCT recipients. While the graft-versus-leukemia effect may help eradicate malignancy, detrimental GvHD occurs when allogeneic donor T cells attack healthy host epithelial tissues. Acute GvHD most commonly affects the skin, liver or gut, though other tissues can be involved. Staging of acute gut GvHD is primarily based on volume of diarrhea per day, but is also characterized by increased apoptotic cells and intestinal crypt destruction in a

gut biopsy. Overall grade of GvHD is evaluated by the organs involved and stage of GvHD at each site.[14–16]

3.1.2 Effects of microbially-modulated bile acids

Low gut microbial diversity has been linked to GvHD in numerous studies.[4,33,58,59,62,75,194] GvHD is marked by loss of intestinal barrier integrity and inflammation.[2,33–37] The gut microbiota and their metabolites may influence GvHD risk through impacts on the intestinal barrier and immune cell populations.[2,33–35,84,85] Recent studies have linked microbially-metabolized bile acids with intestinal health, including GvHD.[116] Bile acids function as lipid emulsifiers in the intestinal tract, but research has elucidated their role as signaling molecules in the intestine, binding host cell membrane receptors like G-protein coupled receptor TGR5 and sphingosine-1 phosphate receptor 2 (S1PR2) and nuclear receptors such as farnesoid X receptor (FXR), pregnane X receptor (PXR), and vitamin D receptor (VDR).[195] Particular bile acids including isolithocholic acid, isodeoxycholic acid, and 3-oxolithocholic acid can suppress T cell differentiation into Th_{17} cells and promote Treg differentiation.[120–122] Isolithocholic acid and 3-oxolithocholic acid are both produced by human gut microbial commensals *Eggerthella lenta* and *Clostridium citroniae,*[121] isolithocholic acid by *Bacteroides fragilis* and *Mediterraneibacter gnavus*,[121] and isodeoxycholic acid by *Mediterraneibacter gnavus* and *Clostridium scindens*.[120]

3.1.3 Enterohepatic circulation of bile acids

Bile acids are enterohepatically circulated; they are produced in the liver from dietary and host-biosynthesized cholesterol before undergoing several host-mediated biotransformations. They are excreted in bile ducts and stored in the gall bladder for transport to the intestine as needed (e.g. after a meal) (Fig. 5). Bile acids enter the intestine as primary bile acids cholic acid (CA) or chenodeoxycholic acid (CDCA) conjugated to a taurine or glycine. Most bile acids are reabsorbed in the small intestine (ileum) for recycling back to the liver, but some progress to the large intestine.[105,106] In the small or large intestine, these conjugated, primary bile acids encounter a range of gut microbes, most of which possess a choloylglycine hydrolase enzyme, also known as a bile salt hydrolase enzyme (encoded by *bsh* genes). Canonically, bile salt hydrolases deconjugate bile acids in one step, freeing amino acids and creating unconjugated primary bile acids. Recent findings identified a novel ability of bile salt hydrolases to re-conjugate taurine or glycine back onto bile acids.[126] Furthermore, bile salt hydrolases can re-conjugate bile acids to a diverse range of compounds, including all amino acids, polyamines, 5-aminovalerate, aminobutyrate, and more.[127] The *bsh* gene is highly diverse; exact *bsh* gene sequences vary by strain, and evolutionary relationships do not

correlate with *bsh* sequence identity.[128] Further, capacity for bile acid metabolism is not consistent within a bacterial genus or species.[128]

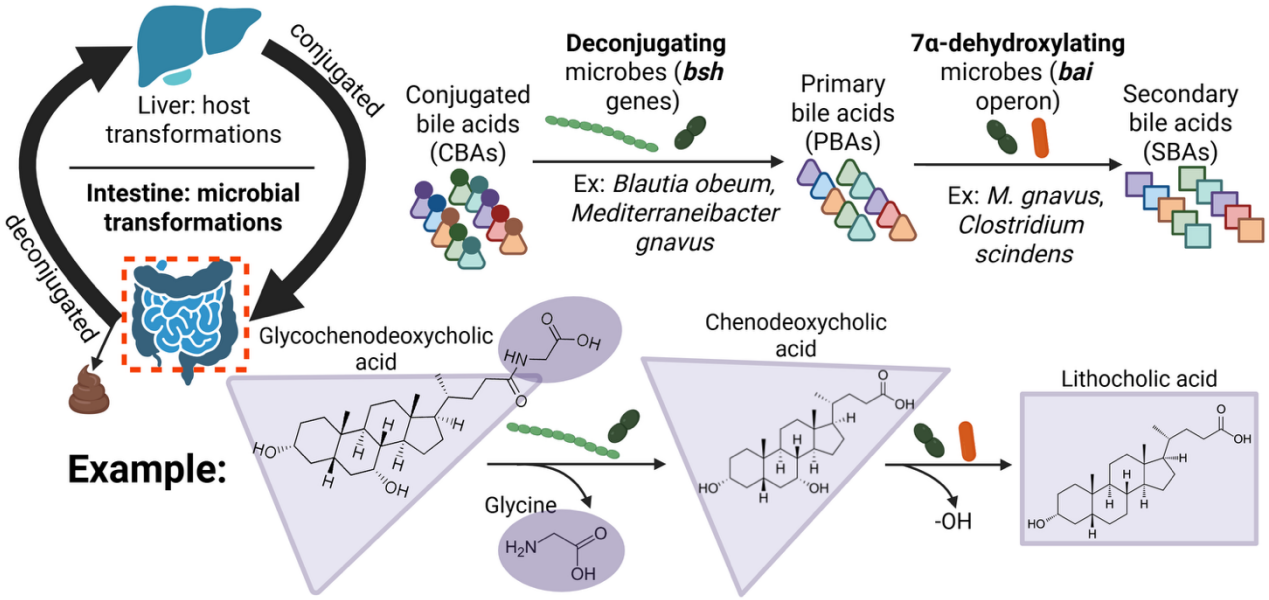


Figure 5. Enterohepatic circulation of bile acids. Bile acids are produced from cholesterol in the liver by the host. They undergo several host-mediated biotransformations in the liver before being transported to the intestine as primary bile acids conjugated to taurine or glycine, such as glycochenodeoxycholic acid. Most bile acids are reabsorbed in the small intestine (ileum), but some progress to the large intestine. In the small or large intestine, the taurine or glycine are removed by microbes that possess a *bile salt hydrolase (bsh)* gene encoding a choloylglycine hydrolase enzyme. This process is called deconjugation, after which the bile acids are in primary form, for example, chenodeoxycholic acid. These primary bile acids travel along the intestine to the descending colon and encounter a different subset of microbes that possess the *bile acid inducible (bai)* operon. These microbes can perform 7α-dehydroxylation, removing the hydroxyl group from the seventh carbon within the bile acid structure. This multi-step reaction results in production of secondary bile acids, like lithocholic acid. The majority of intestinal bile acids are reabsorbed and recirculated back to the liver via the portal vein for enterohepatic circulation. A small portion of bile acids are excreted in stool. Created in BioRender. Cagle, R. (2026) https://BioRender.com/wkbvayg.

After reabsorption for circulation back to the liver, the remaining bile acid pool migrates along the intestine, encountering more gut bacteria in the large intestine (namely the colon), including microbes like *Clostridium scindens* and others that possess the *bile acid inducible* (*bai)* gene operon that encodes a series of enzymes required for the multi-step conversion of unconjugated, primary bile acids into secondary bile acids by removing a hydroxyl group from the seventh carbon (7α-dehydroxylation).[105,106,129,195,196] The majority of these unconjugated,

secondary bile acids are passively reabsorbed across the intestinal barrier (95%) for recirculation back to the liver whereas a smaller portion are excreted into stool (5%).[131,197,198]

3.1.4 Bile acids impact GvHD

The human host and gut microbiota chemically modify bile acids, but transformations between primary vs. secondary and conjugated vs. unconjugated are key because these unique forms activate cell receptors differently (see section 3.3.4 for detail).[117–119,199] Receptors include those on cell membranes (e.g. TGR5) and in the nucleus (e.g. FXR) in intestine-resident cells, like intestinal epithelial cells (IECs), or immune cells, such as T cells.[116–122] A previous study showed that T cell-induced inflammation alters microbial bile acid metabolism, contributing to bile acid dysbiosis that exacerbates GvHD.[116] This work investigates the relationships between bile acid disturbances, the gut microbiota, and GvHD in humans and considers how diverse bile acids may modulate or serve as markers of GvHD. An enhanced understanding of gut microbiota dynamics and bile acid transformations/flux may pave the way to therapies that reduce GvHD and improve HCT outcomes.

3.2 Results

3.2.1 Study population

Demographic and clinical characteristics of the cases and controls were well matched (by design) except for the presence of acute gut GvHD (Table 6). Further details on the distribution of overall grade and individual organ stages of GvHD in case and control participants can be found in Fig. 3 and Table 7.

		Cases (n=49), n (%)	Controls (n=49), n (%)
Age	Median	56 (20-75)	55 (26-72)
Race	Asian	2 (4.08)	2 (4.08)
	Black or African American	3 (6.12)	0 (0)
	White	41 (83.67)	43 (87.76)
	More than one race	1 (2.04)	3 (6.12)
	Unknown or not reported	2 (4.08)	1 (2.04)
Ethnicity	Hispanic or Latino	2 (4.08)	0 (0)
	Not Hispanic or Latino	46 (93.88)	48 (97.96)
	Unknown	1 (2.04)	1 (2.04)
Sex	Female	20 (40.82)	14 (28.57)
	Male	29 (59.18)	35 (71.43)
Donor relationship	Related	12 (24.49)	10 (20.41)
	Not Related	37 (75.51)	39 (79.59)
HLA	Haploidentical	5 (10.2)	3 (6.12)
	Matched	36 (73.47)	37 (75.51)
	Cord	6 (12.24)	5 (10.2)
	Mismatch	2 (4.08)	4 (8.16)
Underlying diagnosis	Leukemia	27 (55.1)	27 (55.1)
	Myelofibrosis	4 (8.16)	5 (10.2)
	Aplastic anemia	9 (18.37)	8 (16.327)
	Lymphoma	4 (8.16)	6 (12.24)
	Other	5 (10.2)	3 (6.12)
Stem cell source	Bone marrow	3 (6.12)	3 (6.12)
	Cord	6 (12.24)	5 (10.2)
	Peripheral blood stem cell	40 (81.63)	41 (83.67)
Conditioning regimen	Myeloablative	23 (46.94)	21 (42.86)
	Reduced intensity	25 (51.02)	22 (44.90)
	Non-myeloablative	1 (2.04)	6 (12.24)
GvHD prophylaxis	CNI + methotrexate	13 (26.53)	17 (34.69)
	CNI + MMF	11 (22.45)	13 (26.53)
	CNI + MMF + rapamycin	8 (16.33)	4 (8.16)
	Other	17 (34.69)	15 (30.61)

Table 6. Demographics of the study population. The demographic and clinical characteristics of the cases and controls are well matched except for the presence of acute gut GvHD. The numbers in parentheses after the median age reflect the age range of participants in this study. CNI = Calcineurin-inhibitor, MMF = mycophenolate mofetil.

		Cases (n = 49) (%)	Controls (n = 49) (%)
Grade acute GvHD	**0**	0 (0)	34 (69.39)
	1	0 (0)	4 (8.16)
	2	0 (0)	11 (22.45)
	3	35 (71.43)	0 (0)
	4	14 (28.57)	0 (0)
Stage acute skin GvHD	**0**	28 (57.14)	34 (69.39)
	1	1 (2.04)	2 (4.08)
	2	8 (16.33)	2 (4.08)
	3	9 (18.37)	11 (22.45)
	4	2 (4.08)	0 (0)
	NA	1 (2.04)	0 (0)
Stage acute liver GvHD	**0**	39 (79.59)	49 (100)
	1	1 (2.04)	0 (0)
	2	5 (10.2)	0 (0)
	3	3 (6.12)	0 (0)
	4	1 (2.04)	0 (0)
Stage acute gut GvHD	**0**	0 (0)	49 (100)
	1	0* (0*)	0* (0*)
	2	22 (44.9)	0 (0)
	3	15 (30.61)	0 (0)
	4	12 (24.49)	0 (0)

Table 7. GvHD demographics of the patient population. *Participants with stage 1 acute gut GvHD were excluded from study cohort to focus on differences between no acute gut GvHD (stage 0) and moderate-severe acute gut GvHD (stages 2-4)

3.2.2 Global bile acid concentrations post-HCT

To examine global differences in bile acid concentrations between GvHD cases and no-GvHD controls, we first used principal component analyses (PCAs) based on the concentrations of each of the 32 bile acids measured in this study. We observed statistically significant differences between GvHD case and control samples on days 30 and 60 post-transplant ($p = 0.008$ and 0.05, $R^2 = 0.042$ and 0.041, respectively; Fig. 6B-C). Nevertheless, we observed overlap in covariance ellipses (Fig. 6A-C), with similar median case and control concentrations of summed bile acids at each timepoint (Fig. 6D). Examination of the change in bile acid levels over time in GvHD cases revealed that summed bile acid levels decreased 1.5-fold from 5.6 mM pre-transplant to 3.8 mM ($p = 0.02$, FC = 0.68) at day 30 post-transplant whereas levels remained similar in no-GvHD controls (5.5 mM vs. 4.3 mM, $p = 0.09$; Fig. 6D, Fig. 7A). We next sought to determine which bile acid groups contribute to the variability seen in the PCA plots.

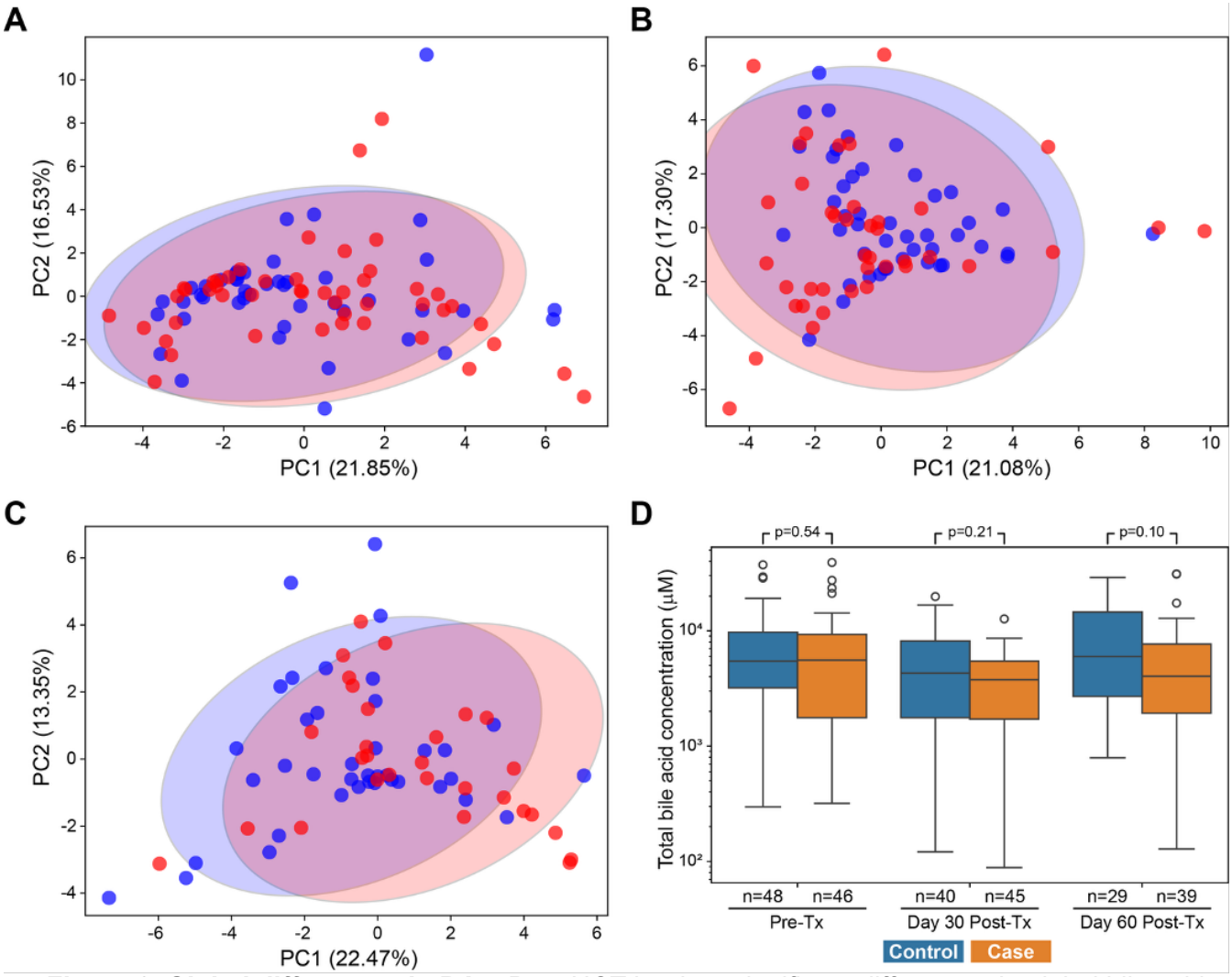


Figure 6. Global differences in BAs. Post-HCT is when significant differences in global bile acid levels emerge. There were no global differences in BAs (reflecting the population of 32 measured BAs) between GvHD cases vs controls as observed by principal component analysis (PCA) (A) pre-transplant. PCA visualized the statistically significant differences between GvHD cases' and controls' bile acid concentrations at (B) 30 days post-transplant ($p = 0.008$, $R^2 = 0.042$) and (C) 60 days post-transplant ($p = 0.05$, $R^2 = 0.041$). Values next to the component on each axis display the amount of variation attributed that component. Cases are in blue, controls in red. (D) There were lower levels of summed BAs in cases at 60 days, although this was not statistically significant. The box represents the interquartile range (IQR): the lower bound represents the first quartile, the center bound represents the median, and the upper bound the third quartile. The lowest whisker represents the minimum and the highest whisker the maximum value, where observations outside of 1.5 * IQR are considered outlying values and displayed independently. PERMANOVA test was performed to generate R^2 and p-values. PC = principal component. Tx = hematopoietic cell transplant.

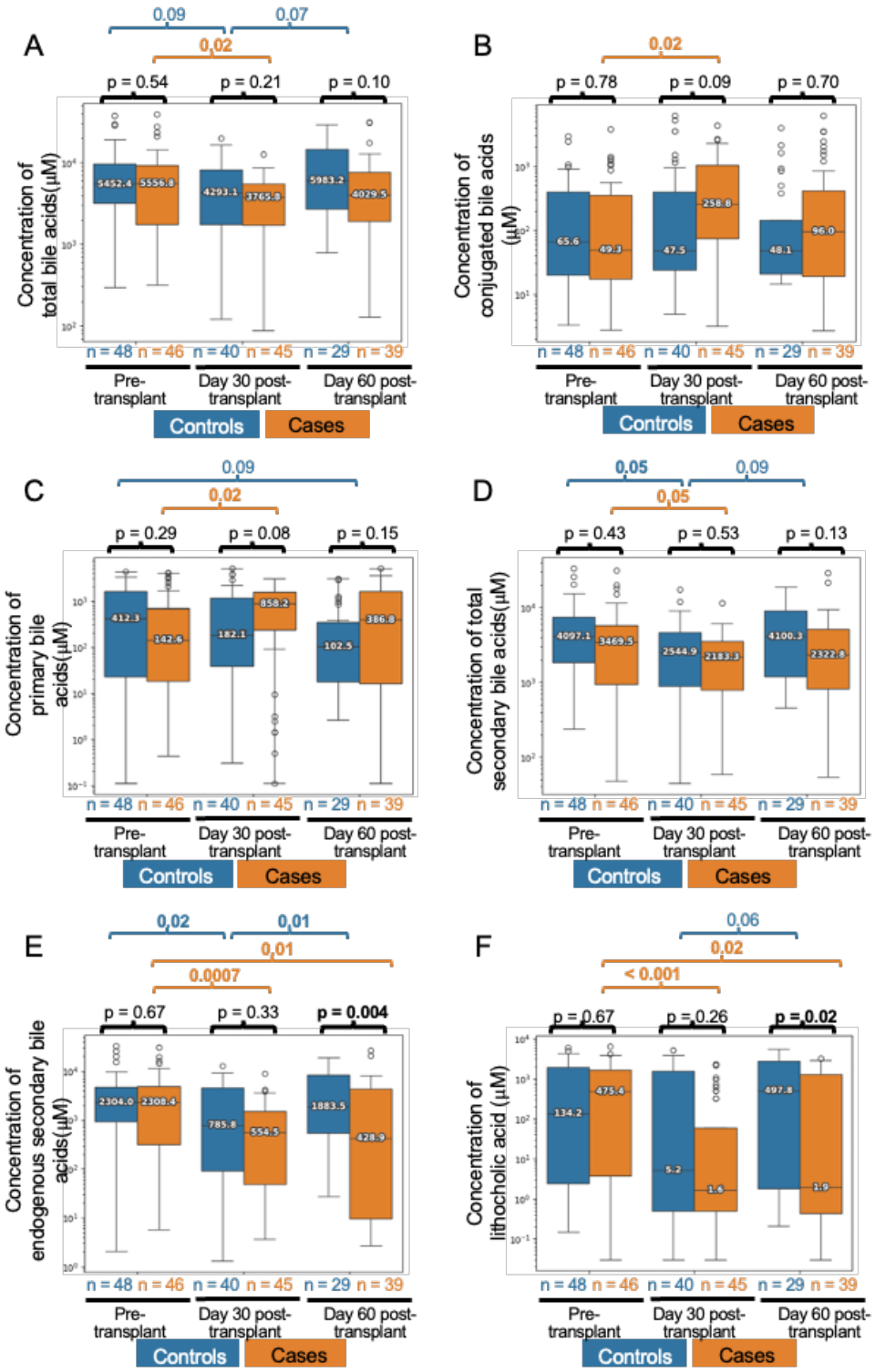


Figure 7. Differences in bile acid levels of within-group changes over time. (A) Total bile acid levels, (B) conjugated bile acid levels, (C) primary bile acid levels, (D) secondary bile acid levels, (E) endogenous secondary bile acids (excluding UDCA, TUDCA, GUDCA), (F) lithocholic acid. Mann-Whitney U test was performed to generate p-values. Only p-values < 0.1 are displayed. All p-values shown are raw/unadjusted. When adjusted, in panels E and F, the day 60 post-HCT FDR-adjusted p-value is 0.02 and 0.16, respectively.

3.2.3 Lower concentrations of endogenous secondary bile acids associate with GvHD post-HCT

We examined the summed concentrations of three key groups of bile acids: conjugated, primary, and secondary bile acids (Fig. 5) that are influenced by bacterial metabolism. At first, it appeared that there were no significant differences in bile acid concentrations between GvHD cases vs. controls at any timepoint for the 12 conjugated, six primary and 14 secondary bile acids measured (Fig. 8A-C). Individual primary bile acids cholic acid and chenodeoxycholic acid also showed no difference between GvHD cases vs. controls at any timepoint. However, levels of both conjugated and primary bile acids significantly increased in GvHD case samples from pre-transplant to day 30 post-transplant ($p=0.02$, FC = 5.2 and 6.0, respectively; Fig. 7B-C), whereas secondary bile acid concentrations decreased 1.6-fold from pre-transplant to day 30 post-transplant in GvHD cases and no-GvHD controls ($p = 0.05$ and 0.05; FC = 0.63 and 0.62, respectively; Fig. 7D). Additionally, grouping all secondary bile acids together ignores a key distinguishing factor in HCT recipients that we note below.

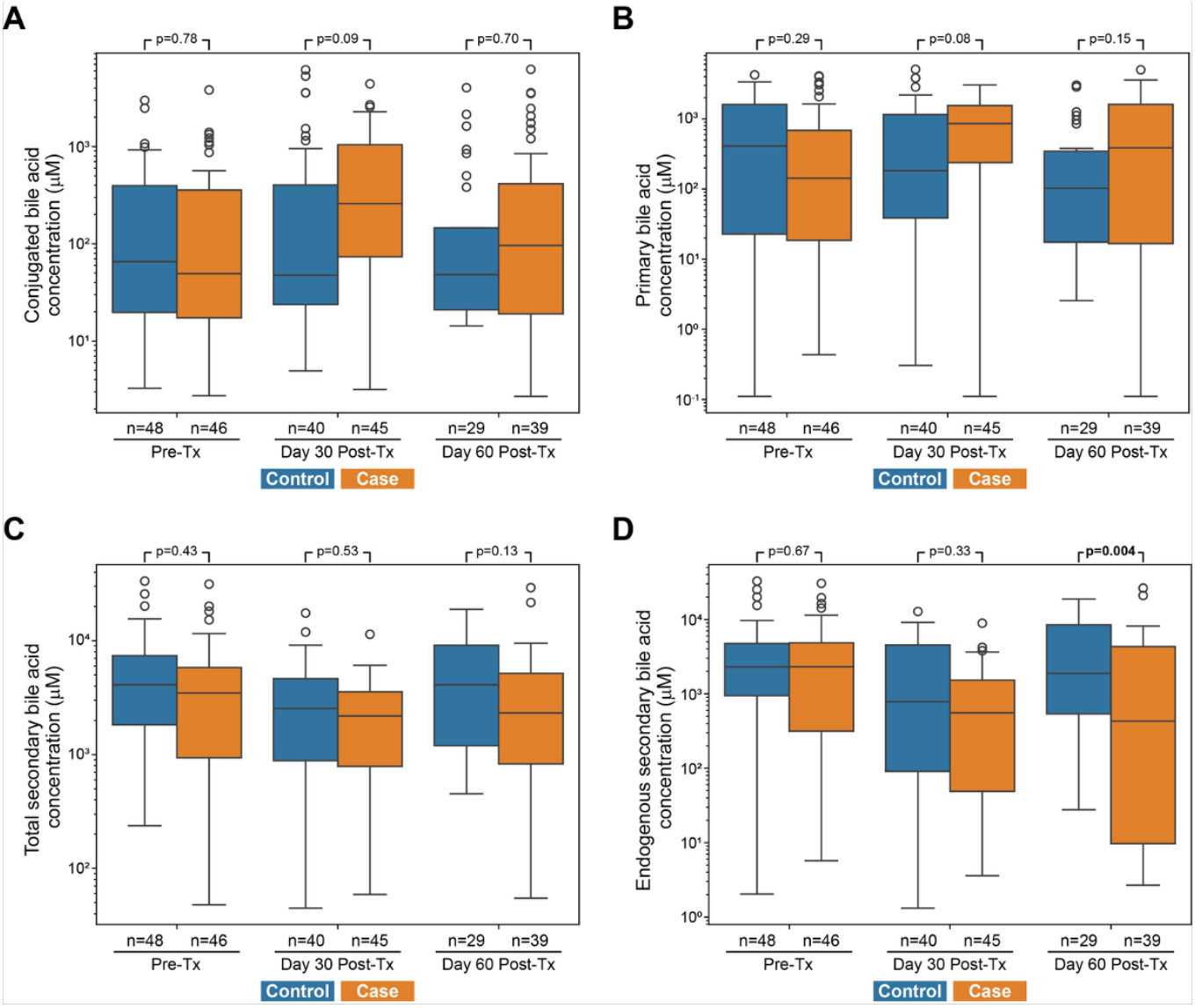


Figure 8. Group-level differences in BAs. Endogenous secondary bile acids are present at lower levels in GvHD cases vs. controls 60 days post-transplant. (A) There are no significant differences in conjugated BAs (12) between GvHD cases vs controls at the 3 assessed time points. (B) There are no significant differences in primary BAs (6) between GvHD cases vs controls. (C) There are no significant differences in secondary BAs (14) between GvHD cases vs controls. (D) There are lower levels of endogenous secondary BAs (11) in GvHD cases vs controls 60 days post-transplant. The number shown on each box represents the sample group's median. Mann-Whitney U test was performed to generate p-values. Raw/unadjusted p-values are shown. The FDR-adjusted p-value for panel D day 60 post-HCT is 0.02. The bounds of the boxes, the lines within the boxes, the whiskers, and the consideration of outlying values are the same as those defined in the Figure 6 legend. Tx = hematopoietic cell transplant.

The exogenous secondary bile acid ursodeoxycholic acid (UDCA, also known as ursodiol) is administered to all allogeneic HCT recipients regardless of GvHD status to mitigate post-transplant liver complications, such as sinusoidal obstruction syndrome (SOS) also known

as hepatic veno-occlusive disease (VOD). Ursodiol is administered from before conditioning through 90 days post-HCT. Although the intent behind UDCA clinical administration is prevention of liver damage, protection against acute gut GvHD has also been observed [31,32]. Since all recipients in the study received exogenous UDCA, we excluded UDCA from the endogenous secondary bile acid group (Fig. 9). When excluding UDCA, and metabolites tauro-UDCA and glyco-UDCA, statistically significant differences in the 11 endogenous secondary bile acids were observed on days 30 and 60 post-HCT by PCA ($p = 0.009$ and 0.05, respectively; Fig. 10B-C). There was a 4.4-fold decrease in levels of endogenous secondary bile acids in GvHD case samples (median 430 μM) vs. no-GvHD samples (median 1,900 μM) on day 60 post-transplant ($p = 0.004$, $p_{adjusted} = 0.02$, FC = 0.23; Fig. 8D, Table 8). Endogenous secondary bile acids levels also significantly change over time within cases or controls. GvHD case median endogenous secondary bile acid levels were 2,300 μM pre-transplant and decreased 4.2-fold to 550 μM by 30 days post-transplant ($p = 0.0007$, FC = 0.24; Fig. 7E). Control (no-GvHD) median endogenous secondary bile acid levels were 2,300 μM pre-transplant and also decreased, but only 2.9-fold to 790 μM by 30 days post-transplant ($p = 0.02$, FC = 0.34; Fig. 7E). Although these endogenous secondary bile acid levels decreased in both cases and controls from pre-transplant to 30 days post-HCT, only the control endogenous secondary bile acid levels recovered by day 60 post-transplant to a median of 1,900 μM ($p = 0.01$, FC = 2.4; Fig. 7E). This contrasted with GvHD case endogenous secondary bile acid levels on day 60 post-HCT, median 430 μM, which remained 5.3-times lower than pre-transplant levels ($p = 0.01$, FC = 0.19). These data provided evidence of persistent deficiency of endogenous secondary bile acids during GvHD post-HCT (Fig. 7E).

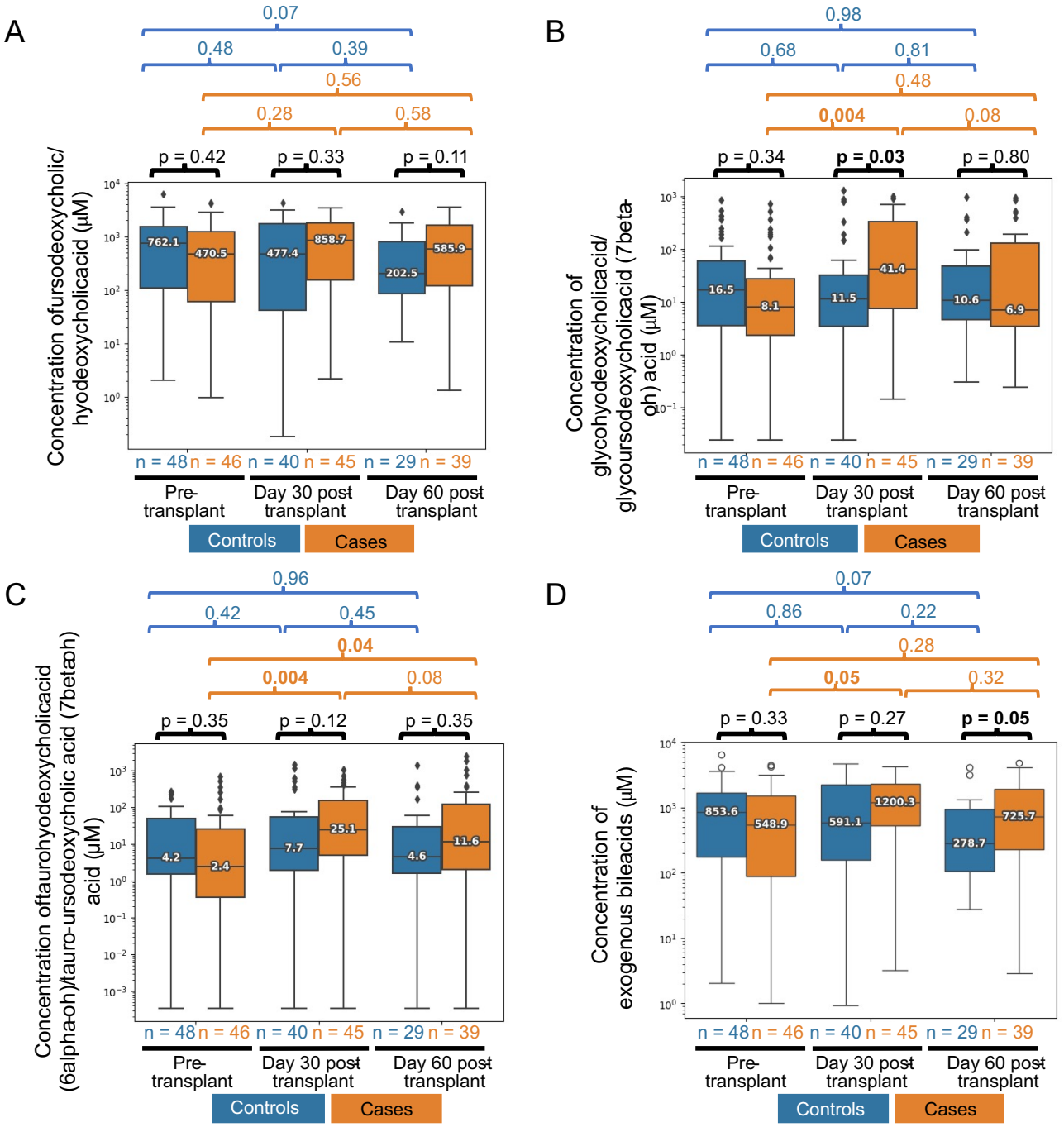


Figure 9. Levels of exogenous bile acids. (A) Levels of ursodeoxycholic acid and hyodeoxycholic acid, which are indistinguishable from each other by the LC-MS method. (B) Levels of glycohyodeoxycholic acid and glycoursodeoxycholic acid (7beta-oh), which are indistinguishable from each other by the LC-MS method. (C) Levels of taurohyodeoxycholic acid (6alpha-oh) and tauro-ursodeoxycholic acid (7beta-oh), which are indistinguishable from each other by the LC-MS method. (D) Combined levels of ursodeoxycholic acid/hyodeoxycholic acid, glycohyodeoxycholic acid/glycoursodeoxycholic acid (7beta-oh) and taurohyodeoxycholic acid (6alpha-oh)/tauro-ursodeoxycholic acid (7beta-oh). All p-values shown are raw/unadjusted. Mann-Whitney U test was performed to generate p-values.

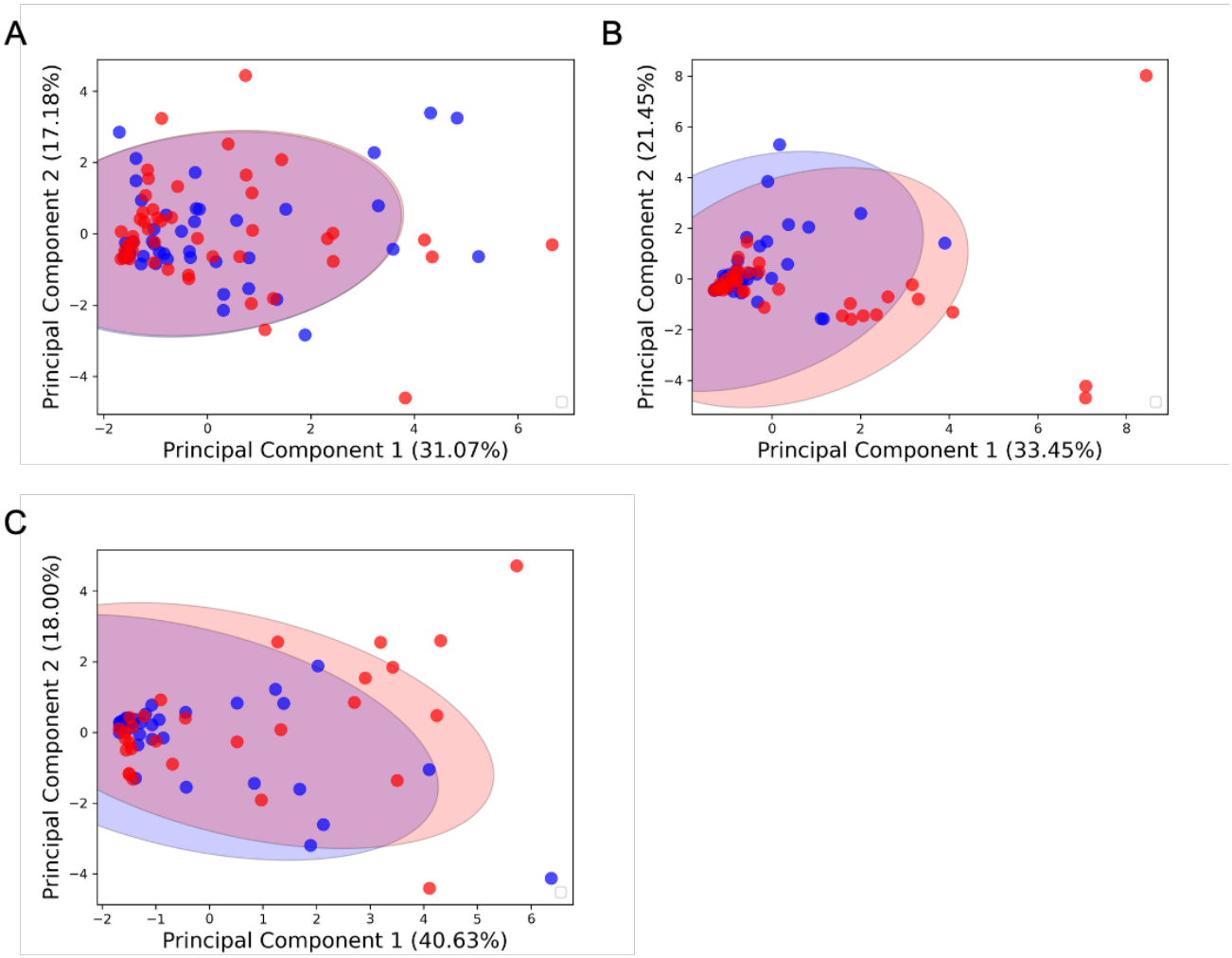


Figure 10. Differences in endogenous secondary bile acids. Principal component analysis (PCA) was used to assess differences in the population of 11 endogenous secondary bile acid levels between GvHD cases and controls. (A) Pre-transplant, there were no differences between GvHD cases vs controls. However, PCA revealed statistically significant differences between GvHD cases and controls along Principal Component 2 at (B) 30 days post-transplant ($p = 0.009$) and along Principal Component 1 at (C) 60 days post-transplant ($p = 0.05$). Values next to the component on each axis display the amount of variation attributed that component. Cases are in blue, controls in red. All p-values shown are raw/unadjusted. Mann-Whitney U test was performed to generate p-values.

	Pre-Tx				Day 30				Day 60			
	median (μM)			adj.	median (μM)			adj.	median (μM)			adj.
	control	case	p-value	p-value	control	case	p-value	p-value	control	case	p-value	p-value
12KLCA814 (3alpha-hydroxy-12 ketolithocholic acid)	182.7	139.8	0.483	0.693	0.2	0.2	0.124	0.37	54.3	0.2	0.029	0.157
7KLCAolone (3alpha-hydroxy-7-ketolithocholic acid)	419.1	137.1	0.213	0.503	175.9	187.7	0.478	0.693	169.4	49.7	0.035	0.167
5beta-cholanic acid-3beta, 12alpha-diol	25.2	11.7	0.107	0.354	8.3	7.2	0.824	0.938	19.5	4.4	0.011	0.157
allolithocholic acid	0	0.6	0.46	0.693	0	0	0.021	0.157	0.4	0	0.101	0.354
deoxycholic acid	142.9	75.3	0.96	0.997	0.1	0.1	0.301	0.552	15.7	0.1	0.146	0.37
glycodeoxycholic acid	0.4	0.5	0.997	0.997	0	0	0.973	0.997	0.5	0	0.238	0.507
glycolithocholic acid	3	3.9	0.342	0.595	2.7	1.9	0.027	0.157	1.7	1.9	0.76	0.896
isolithocholic acid	8.9	35.4	0.363	0.6	3	1.6	0.249	0.507	237.4	4.8	0.14	0.37
lithocholic acid	134.2	475.4	0.669	0.849	5.2	1.6	0.261	0.507	497.8	1.9	0.02	0.157
taurodeoxycholic acid	0.3	0.1	0.604	0.797	0	0	0.547	0.753	0.1	0.5	0.867	0.954
taurolithocholic acid	0.4	0.4	0.743	0.896	0.6	0.2	0.013	0.157	0.4	0.2	0.054	0.222
Total endogenous secondary bile acids	1891.1	1521.3	0.568	0.789	696.5	622.9	0.158	0.439	1848	429.6	0.004	0.019

Table 8. Concentrations of endogenous secondary bile acids. Median concentrations (μM), raw p-values and adjusted p-values from endogenous secondary bile acids and key comparisons by GvHD status and timepoints throughout the HCT process. Mann-Whitney U test was performed to generate p-values.

We gained even greater insight by focusing on the 45 GvHD case participants' stool samples from day 30 post-HCT. These samples were divided into two groups: those diagnosed with GvHD by the time of the day 30 sample collection time (n = 30, current GvHD) and those not yet diagnosed with GvHD at that time (n = 15, pre-GvHD). As a group, the summed levels of endogenous secondary bile acids on day 30 in the pre-GvHD cases were statistically indistinguishable from the no-GvHD group (p = 0.64; Fig. 11), whereas the current GvHD cases had 4-fold lower levels of endogenous secondary bile acids compared to the pre-GVHD group (p = 0.03/p_{adj} = 0.22, FC = 0.25; Fig. 11, Table 9). These data provided more evidence that endogenous secondary bile acid levels are linked to GvHD status.

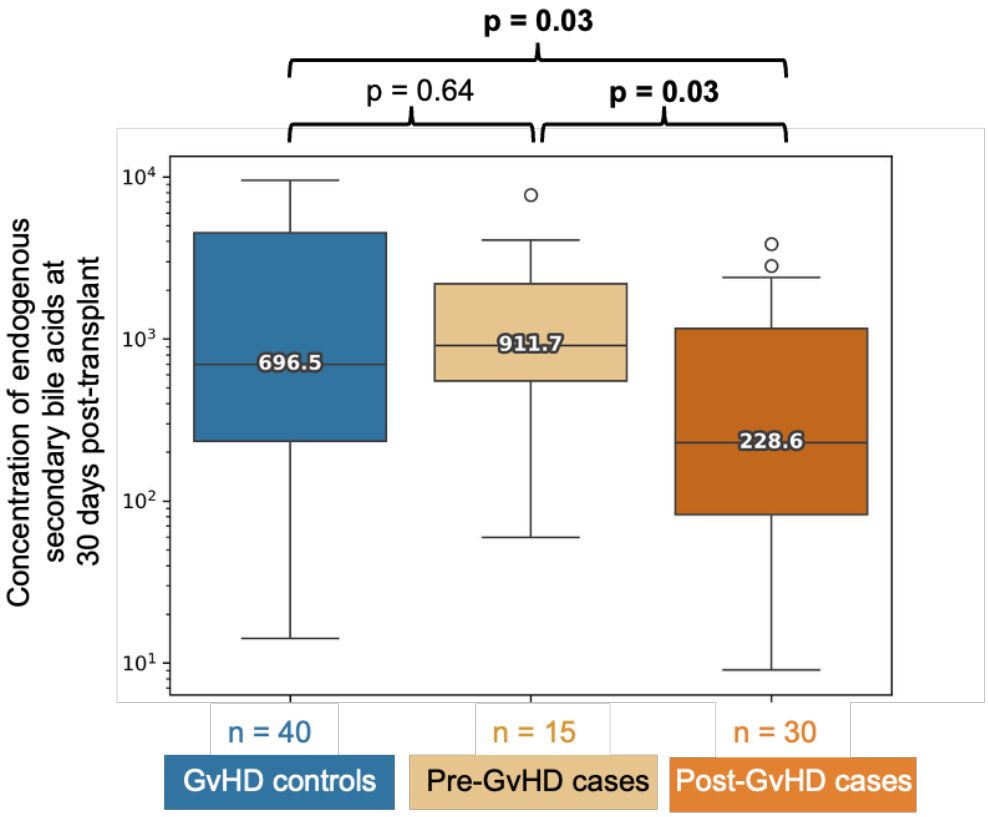
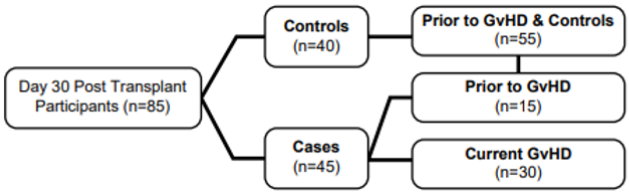


Figure 11. Endogenous secondary bile acid levels when adjusted for time to GvHD. At day 30 post-transplant, most of the GvHD case recipients had been diagnosed with GvHD (n = 30), but a minority of the GvHD case recipients had not yet been diagnosed with GvHD (n = 15). Recipient GvHD cases with GvHD already diagnosed at the time of day 30 post-transplant sample had significantly lower levels of endogenous secondary bile acids vs. recipient GvHD cases without GvHD yet and vs. GvHD controls. Recipient GvHD cases without GvHD diagnosed yet had similar levels of endogenous secondary bile acids as GvHD controls. The p-values shown are raw, and when FDR-adjusted, are no longer statistically significant: GvHD controls vs. pre-GvHD cases adjusted p = 0.82, pre-GvHD cases vs. post-GvHD cases adjusted p = 0.22, GvHD controls vs. post-GvHD cases adjusted p = 0.22. Mann-Whitney U test was performed to generate p-values.



	Median (μM)			Comparisons							
				Prior to GvHD vs. Current GvHD		Prior to GvHD vs. Controls		Current GvHD vs. Controls		Current GvHD vs. Prior to GvHD & Controls	
	Prior to GvHD (n=15)	Current GvHD (n=30)	Prior to GvHD & Controls (n=55)	p-value	adj. p-value	p-value	adj. p-value	p-value	adj. p-value	p-value	adj. p-value
12KLCA814 (3alpha-hydroxy-12 ketolithocholic acid)	0.2	0.2	0.2	0.509	0.784	0.549	0.784	0.093	0.388	0.121	0.403
7KLCAolone (3alpha-hydroxy-7-ketolithocholic acid)	538.5	108.3	230.4	0.148	0.435	0.101	0.388	0.948	0.995	0.543	0.784
5beta-cholanic acid-3beta, 12alpha-diol	10.6	3.6	9.4	0.112	0.4	0.209	0.523	0.627	0.824	0.323	0.66
allolithocholic acid	0	0	0	0.068	0.378	0.462	0.745	0.008	0.113	0.009	0.113
deoxycholic acid	0.1	0.1	0.1	0.963	0.995	0.533	0.784	0.33	0.66	0.433	0.745
glycodeoxycholic acid	0	0	0	0.957	0.995	0.958	0.995	0.995	0.995	0.926	0.995
glycolithocholic acid	1.6	2.1	2.3	0.462	0.745	0.048	0.3	0.083	0.388	0.29	0.631
isolithocholic acid	0.1	1.6	2	0.721	0.901	0.241	0.574	0.418	0.745	0.623	0.824
lithocholic acid	1.9	1.6	3.9	0.263	0.598	0.813	0.945	0.174	0.458	0.137	0.429
taurodeoxycholic acid	0	0	0	0.746	0.91	0.41	0.745	0.782	0.931	0.992	0.995
taurolithocholic acid	0.4	0.1	0.6	0.095	0.388	0.356	0.685	0.006	0.113	0.006	0.113
Total endogenous secondary bile acids	911.7	228.6	713.3	0.031	0.221	0.643	0.824	0.028	0.221	0.011	0.113

Table 9. Concentrations and comparisons of endogenous secondary bile acids on day 30 post-HCT. Median concentrations (μM), raw p-values and adjusted p-values from endogenous secondary bile acids and key comparisons by GvHD status on day 30 post-transplant. Sample sizes (n) are displayed in parentheses in the column headers. At this day 30 post-HCT timepoint, there were 40 GvHD control participants and 45 GvHD case participants. Within the 45 cases, there were 15 GvHD case participants without GvHD diagnosed at the time of sample ("Prior to GvHD") and 30 GvHD control participants with GvHD diagnosed at the time of sample ("Current GvHD"). There were 55 total participants without GvHD at the time of sample (15 cases Prior to GvHD + 40 Controls). Mann-Whitney U test was performed to generate p-values.

3.2.4 Low concentrations of lithocholic acid and similarly structured secondary bile acids associate with GvHD

Since significant differences in summed endogenous secondary bile acid levels were observed between GvHD cases vs. controls post-HCT, we also examined each individual endogenous secondary bile acid (Table 10, Fig. 12, Table 8-9). We noted 263-fold lower levels of the endogenous secondary bile acid lithocholic acid (LCA) in GvHD cases (median 1.9 μM) vs. no-GvHD controls (500 μM) on day 60 post-transplant ($p = 0.02/p_{adj} = 0.16$, FC = 0.004; Fig. 12, Table 8). Similarly, LCA levels in pre-transplant GvHD cases (median 480 μM) decreased 300-fold to 1.6 μM by day 30 post-transplant ($p < 0.001$, FC = 0.003; Fig. 7F). Consistent with

summed endogenous secondary bile acids as a group, day 60 post-transplant LCA levels in GvHD cases were 250-times reduced from pre-transplant levels (p = 0.02, FC = 0.004; Fig. 7F), which showed the persistent loss of LCA was unique to GvHD (vs. no-GvHD).

Bile Acid	Timepoint	Controls	Cases	Difference	p-value	Adjusted p-value
Taurolithocholic acid	Day 30	0.6	0.2	0.4	0.013	0.157
Taurolithocholic acid	Day 60	0.4	0.2	0.2	0.053	0.222
5beta-cholanic acid-3beta, 12alpha-diol	Day 60	19.5	4.4	15.1	0.011	0.157
Glycolithocholic acid	Day 30	2.7	1.9	0.8	0.027	0.157
Lithocholic acid	Day 60	497.8	1.9	495.9	0.019	0.157
7KLCAolone*	Day 60	169.4	49.7	119.7	0.035	0.167
12KLCA814*	Day 60	54.3	0.2	54.1	0.029	0.157

Table 10. Differences in concentrations of individual endogenous secondary BAs. Table summarizing significant differences in concentrations of lithocholic acid isomers and derivatives between GvHD cases vs controls. Difference in medians are calculated by subtracting the case from the control median concentration. Statistical significance and p-values for comparing the distribution of concentrations from the Mann-Whitney U test. P-values were adjusted for false discovery rate (FDR). Median concentrations in μM are shown. *7KLCAolone is an abbreviation for the bile acids 3alpha-hydroxy-7-ketolithocholic acid and 5alpha-cholanic acid-3alpha-ol-6-one, which are indistinguishable from each other in our LC-MS analysis. 12KLCA814 is an abbreviation for the bile acids 3alpha-hydroxy-12 ketolithocholic acid and 8(14),(5beta)-cholenic acid-3alpha, 12alpha-diol, which are indistinguishable from each other in our LC-MS analysis.

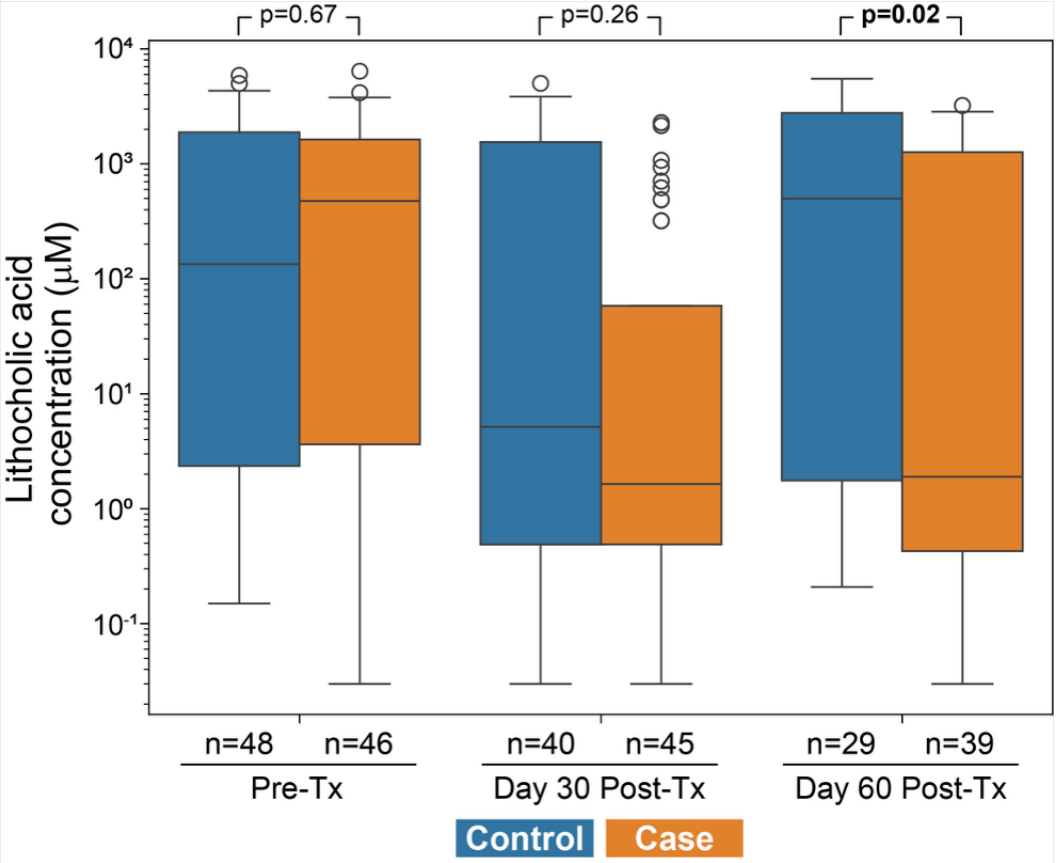


Figure 12. Differences in the endogenous secondary BA, lithocholic acid. There were differences in endogenous secondary BAs (11) between GvHD cases vs. controls, such as seen with lithocholic acid. The bounds of the boxes, the lines within the boxes, the whiskers, and the consideration of outlying values are the same as those defined in the Figure 6 legend. Mann-Whitney U test was performed to generate p-values. Tx = hematopoietic cell transplant.

Six additional endogenous secondary bile acids were less prevalent and/or at lower concentration in GvHD cases vs. controls post-transplant (Tables 10-11). Concentrations of tauroLCA and glycoLCA (conjugated forms of LCA) were significantly lower (3 and 1.4-fold, respectively) in GvHD cases vs. controls at 30 days post-transplant ($p = 0.01/p_{adj} = 0.16$ and 0.03/0.16, respectively; FC = 0.33 and 0.70, respectively; Table 10). Concentrations of tauroLCA and LCA were also lower (2 and 263-fold, respectively) in GvHD cases at 60 days post-transplant ($p = 0.05/p_{adj} = 0.22$ and 0.02/0.16, respectively; FC = 0.50 and 0.004, respectively; Table 10). At 30 days post-transplant, alloLCA (LCA conformer) was less prevalent in GvHD cases than controls ($p = 0.05/p_{adj} = 0.35$; Table 11). Bile acids tauroLCA and LCA were also less prevalent in GvHD cases at 60 days post-transplant ($p = 0.05/p_{adj} = 0.35$ and 0.02/0.29, respectively; Table 11). It was striking that five of the significantly different endogenous secondary bile acids were highly structurally similar to LCA.

Bile Acid	Time-point	Controls	Cases	Differ-ence	p-value	Adjusted p-value	Odds ratio	Odds ratio 95% CI
Allolithocholic acid	Day 30	37.50 (15)	17.78 (8)	19.72	0.052	0.354	2.74	0.93 - 8.68
Taurolithocholic acid	Day 60	86.21 (25)	64.10 (25)	22.11	0.054	0.354	3.44	0.91 - 16.36
Lithocholic acid	Day 60	100.00 (29)	82.05 (32)	17.5	0.017	0.288	Inf**	1.17 – inf**
12KLCA814*	Day 60	68.97 (20)	43.59 (17)	25.38	0.05	0.354	2.83	0.94 - 9.02
5beta-cholanic acid-3beta, 12alpha-diol	Day 60	96.55 (28)	66.67 (26)	29.88	0.002	0.078	13.58	1.8 - 613.81

Table 11. Differences in prevalence of individual endogenous secondary BAs. Table summarizing significant differences in prevalence of lithocholic acid isomers and derivatives between GvHD cases vs controls. Statistical significance and p-values for comparing prevalence derived from Fisher's Exact test. P-values were adjusted for false discovery rate (FDR). Prevalence data are shown in % (count, #). *7KLCAolone is an abbreviation for the bile acids 3alpha-hydroxy-7-ketolithocholic acid and 5alpha-cholanic acid-3alpha-ol-6-one, which are indistinguishable from each other in our LC-MS analysis. 12KLCA814 is an abbreviation for the bile acids 3alpha-hydroxy-12 ketolithocholic acid and 8(14),(5beta)-cholenic acid-3alpha, 12alpha-diol, which are indistinguishable from each other in our LC-MS analysis. **The reported "inf" or "infinite" odds ratio and corresponding 95% confidence intervals are unable to be estimated due to counts of 0 and division by 0 errors. CI = confidence interval.

To delve deeper into the LCA findings, we revisited the day 30 post-HCT GvHD sample classifications (15 cases pre-GvHD, 30 cases with current GvHD, 40 no-GvHD controls). Adjustment or normalization by time to GvHD diagnosis was performed at this timepoint by comparing 55 samples without GvHD (ever or yet) vs. 30 samples with GvHD. This analysis revealed that concentrations of two LCA-like bile acids, alloLCA and tauroLCA, as well as the sum of the 11 endogenous secondary bile acids were significantly lower during current GvHD vs. the combined samples without GvHD ($p = 0.009/p_{adj} = 0.11$, 0.006/0.11 and 0.01/0.11, respectively; Table 9).

Next, we sought to quantify the effect size of the six bile acids significantly associated with GvHD by calculating the odds ratio for GvHD and corresponding 95% confidence intervals. The odds ratios indicate that presence of each bile acid in the gastrointestinal tract was associated with a reduced chance of having GvHD. The odds ratios for 5beta-cholanic acid-3beta, 12alpha-diol was the largest at 14 (95% confidence interval [CI] = 1.8 – 610; Table 11). The odds ratio of LCA is incalculable ("inf") due to 100% prevalence in 29 control patients

without GvHD at 60 days post-HCT (95% CI = 1.17 – inf; Table 11). Nonetheless, upon adjusting p-values for multiple testing (FDR), the adjusted differential concentration and prevalence p-values were greater than 0.07 and no longer statistically significant (Tables 10-11). Thus, we sought to query an analogous dataset to determine if our findings (particularly the LCA findings) were reproducible.

3.2.5 Analysis of the association between bile acid deficiency and GvHD at another cancer center

We analyzed a publicly available dataset from a comparable cancer center to assess the reproducibility of our findings regarding secondary bile acids and GvHD. These data describe bile acid levels in a patient population that received allogeneic HCT at Memorial Sloan Kettering (MSK) Cancer Center. The data were published by Lindner et al. in 2024 and represent a different patient population and Center than this study. We delved into the supplemental data file and brought the reported concentrations and associated metadata into the Python environment for analysis by differential abundance (including Mann Whitney U test). Several findings were concordant in this dataset; namely, LCA was depleted in GvHD cases vs. controls overall ($p = 0.0007$) and at both timepoints examined in the Lindner study, peri-engraftment ($p = 0.02$) and peri-GvHD ($p = 0.02$; Fig. 13A-C). Also of interest was lithocholic acid-3-sulfate (not included in this study's bile acid metabolomic panel), which was lower in GvHD cases vs. controls overall ($p = 0.007$) and at peri-engraftment ($p = 0.0004$; Fig. 13D-F).[116] Indeed, the MSK group's LCA findings were consistent with those observed in this study.

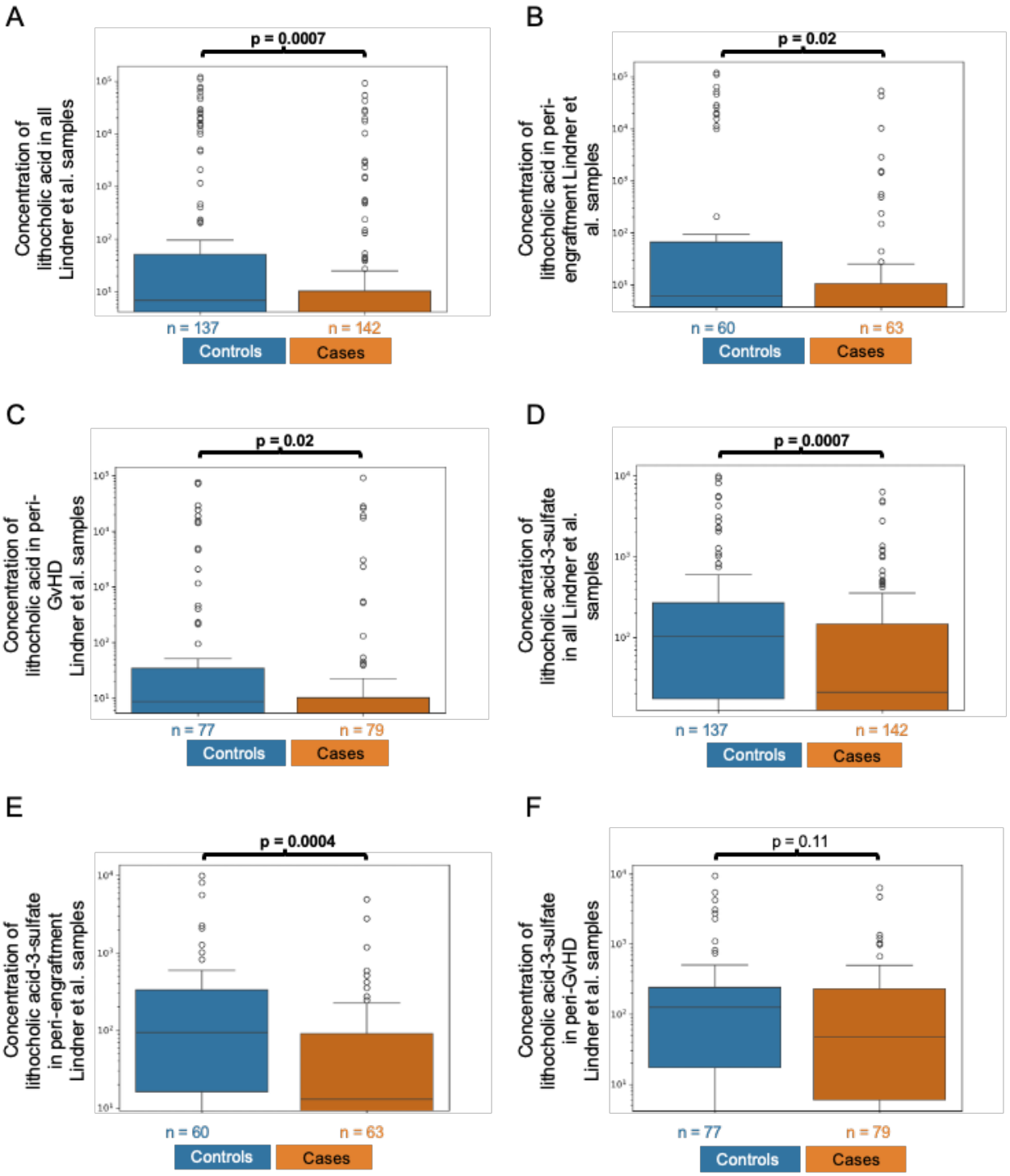


Figure 13. Findings from a similar study are concordant with our observations. A publicly available dataset from Lindner et al. (2024) in human HCT-recipients with and without GvHD validate our study's findings that lithocholic acid and similarly structured bile acids are lower in GvHD vs. no GvHD. Lithocholic acid concentrations from Lindner et al. (2024) samples (A) overall, (B) peri-engraftment, and (C) peri-GvHD. Lithocholic acid-3-sulfate concentrations from Lindner et al. (2024) samples (D) overall, (E) peri-engraftment, and (F) peri-GvHD. All p-values shown are raw/unadjusted. Mann-Whitney U test was performed to generate p-values.

3.2.6 Endogenous bile acid levels remain low in GvHD even when prednisone therapy is considered

We investigated if prednisone/methylprednisolone therapy could have biased the measured bile acid concentrations in stool samples by excluding the 14 samples that were collected from patients while they were receiving this steroid treatment. Bile acid levels were then re-analyzed in a sub-cohort that was off active prednisone/methylprednisolone therapy. Lithocholic acid and endogenous secondary bile acid levels remained significantly reduced in the participants with vs. without GvHD 60 days post-HCT ($p = 0.02$ and 0.002, respectively; Fig. 14).

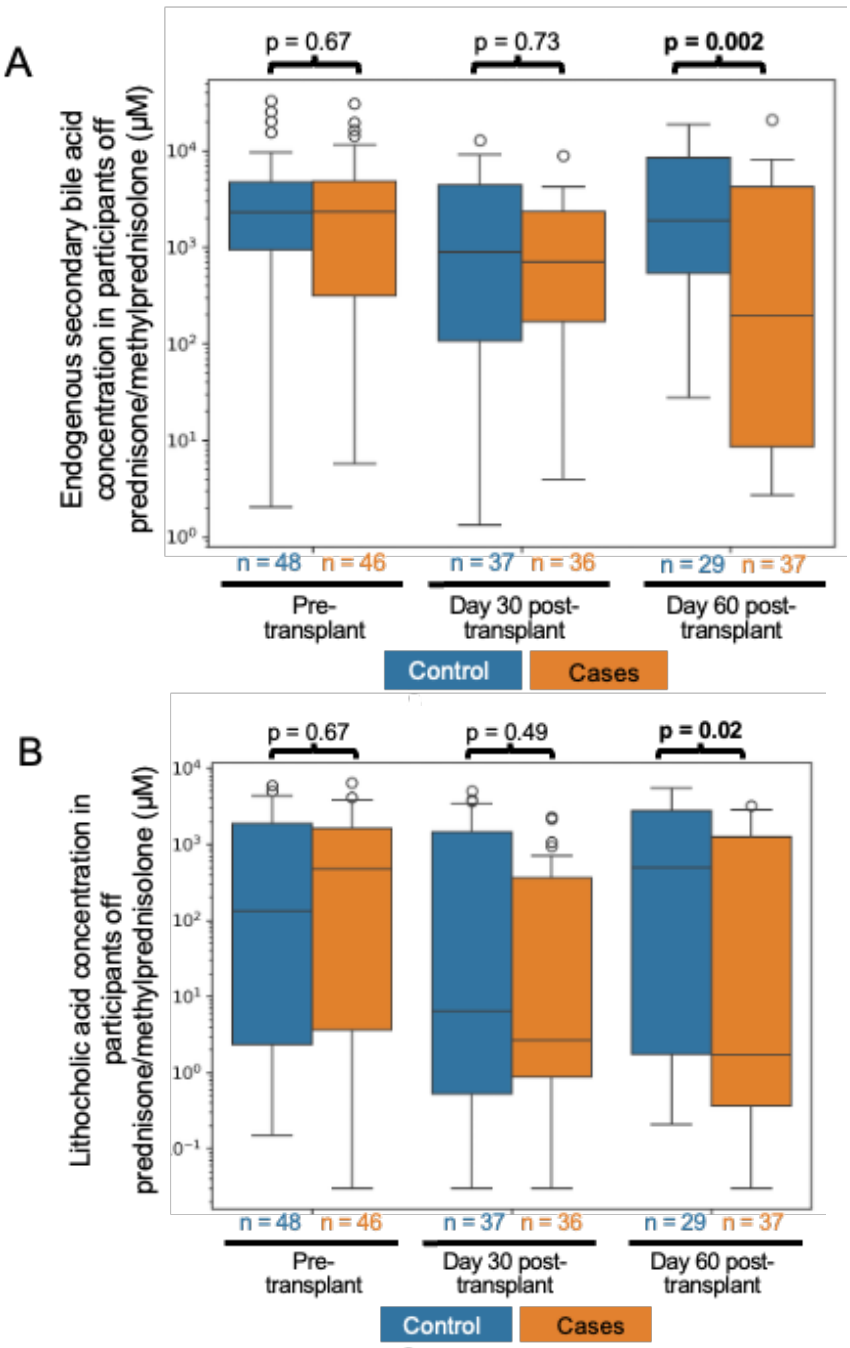


Figure 14. Prednisone/methylprednisolone therapy does not bias bile acid levels. To investigate if prednisone/methylprednisolone treatment biases the observed bile acid levels in stool, we removed the 14 samples collected from patients that were actively on this steroid therapy. We defined this as having received prednisone or methylprednisolone in the seven or fewer days prior to sample date. Key findings were replicated without the effect of this steroid therapy. (A) Endogenous secondary bile acid and (B) lithocholic acid, an endogenous secondary bile acid, concentrations were statistically significantly reduced in GvHD cases vs. controls at day 60 post-HCT. Statistical significance of these differences was not impacted when considering steroid use. This supports that prednisone/methylprednisolone steroid therapy does not bias bile acid levels. All p-values shown are raw/unadjusted. Mann-Whitney U test was performed to generate p-values.

3.2.7 Bile acid-modulating microbes associate with endogenous secondary bile acids post-HCT

With evidence that our study's LCA findings were reproducible, the next question was: why are LCA and LCA-like endogenous secondary bile acids deficient during GvHD? It is known that certain gut microbes metabolize bile acids (Fig. 5). Therefore, we hypothesized that differences in GvHD case vs. control gut microbial communities may be responsible for differences in levels of secondary bile acids. We first investigated bacterial community-level differences between GvHD cases vs. controls by comparing the alpha-diversity as measured by Shannon Diversity Index using 16S rRNA gene sequencing data in stool samples (Fig. 15). SDI in this context is a metric of bacterial species richness (number of species) and evenness in stool. By 30 days post-transplant, median Shannon Diversity Index (SDI) in cases was 2.05, which was significantly lower than that of no-GvHD samples, 2.49 ($p = 0.002$; Fig. 15). At 60 days post-transplant the median SDI was 2.50 in GvHD cases and 2.72 in controls ($p = 0.074$; Fig. 15). These data provide evidence that gut bacterial diversity is lower post-HCT in patients with GvHD in our study.

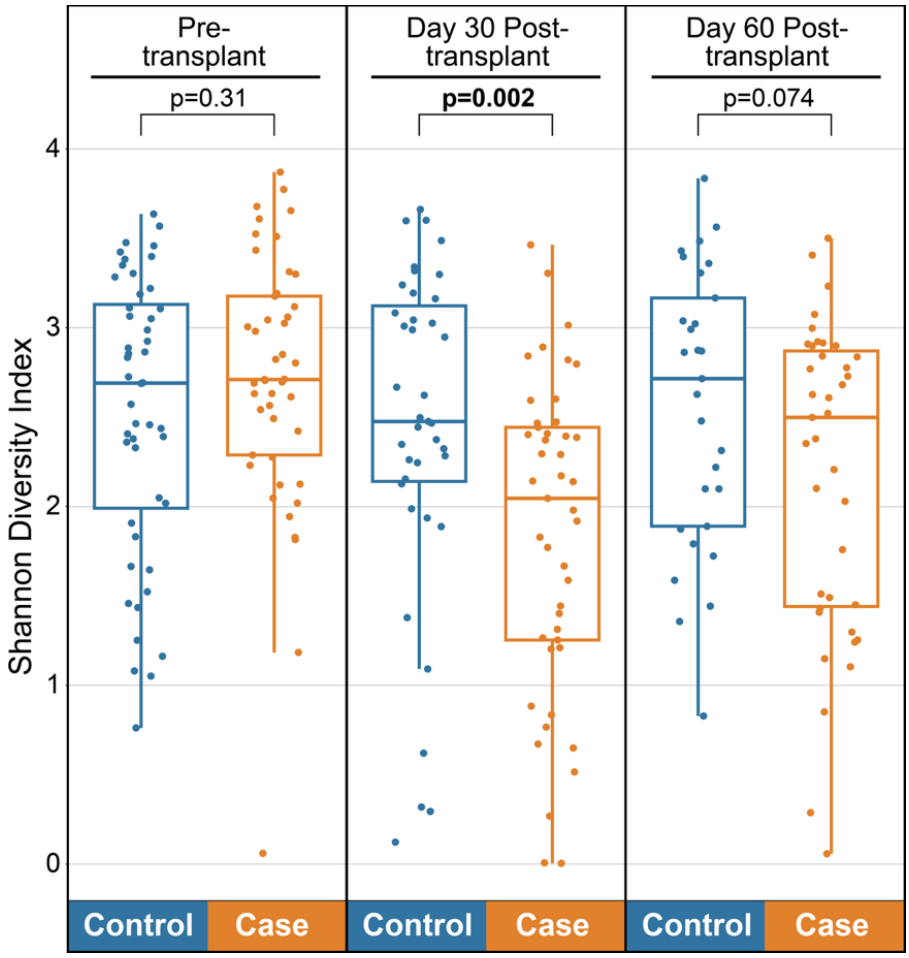


Figure 15. Community-level differences in gut microbial diversity. Microbial alpha-diversity was associated with GvHD at the 30 day time point post HCT. The bounds of the boxes, the lines within the boxes, the whiskers, and the consideration of outlying values are the same as those defined in the Figure 6 legend. Mann-Whitney U test was performed to generate p-values.

We next sought to determine how this difference in alpha-diversity and stool bacterial composition might lead to differences in bile acid concentrations. We measured bacterial relative abundance and prevalence at each timepoint using 16S rRNA gene PCR with sequencing, but no specific taxa at the family, genus or species level was statistically significantly different by GvHD status when using the Mann Whitney U test and when adjusting for multiple comparison (Fig. 16). Despite this, we found that numerous microbial taxa correlated with concentrations of endogenous secondary bile acids (Table 12). Notably, *Clostridium scindens* (a *bai*-possessing microbe) associated with the LCA-like endogenous secondary bile acid 3alpha-hydroxy-12 ketolithocholic acid/8(14),(5beta)-cholenic acid-3alpha, 12alpha-diol (12KLCA814) at 30 days post-HCT (correlation coefficient rho = 0.57, p_{adj} = 1.87E-6; Table 12). *Dysosmobacter welbionis* associated with the LCA conformer, alloLCA (rho = 0.57,

p_{adj} = 4.34E-6; Table 12), at 60 days post-HCT. AlloLCA also correlated with multiple microbial taxa: *Ruminococcus torques* at 60 days post-HCT (rho = 0.61, p_{adj} = 6.49E-7; Table 12), *Lachnospiraceae* on day 60 post-HCT (rho = 0.59, p_{adj} = 1.28E-6; Table 12) and *Oscillospiraceae* on day 30 post-HCT (rho = 0.58, p_{adj} = 2.51E-6; Table 12). Further, we noticed that these taxa have a key commonality.

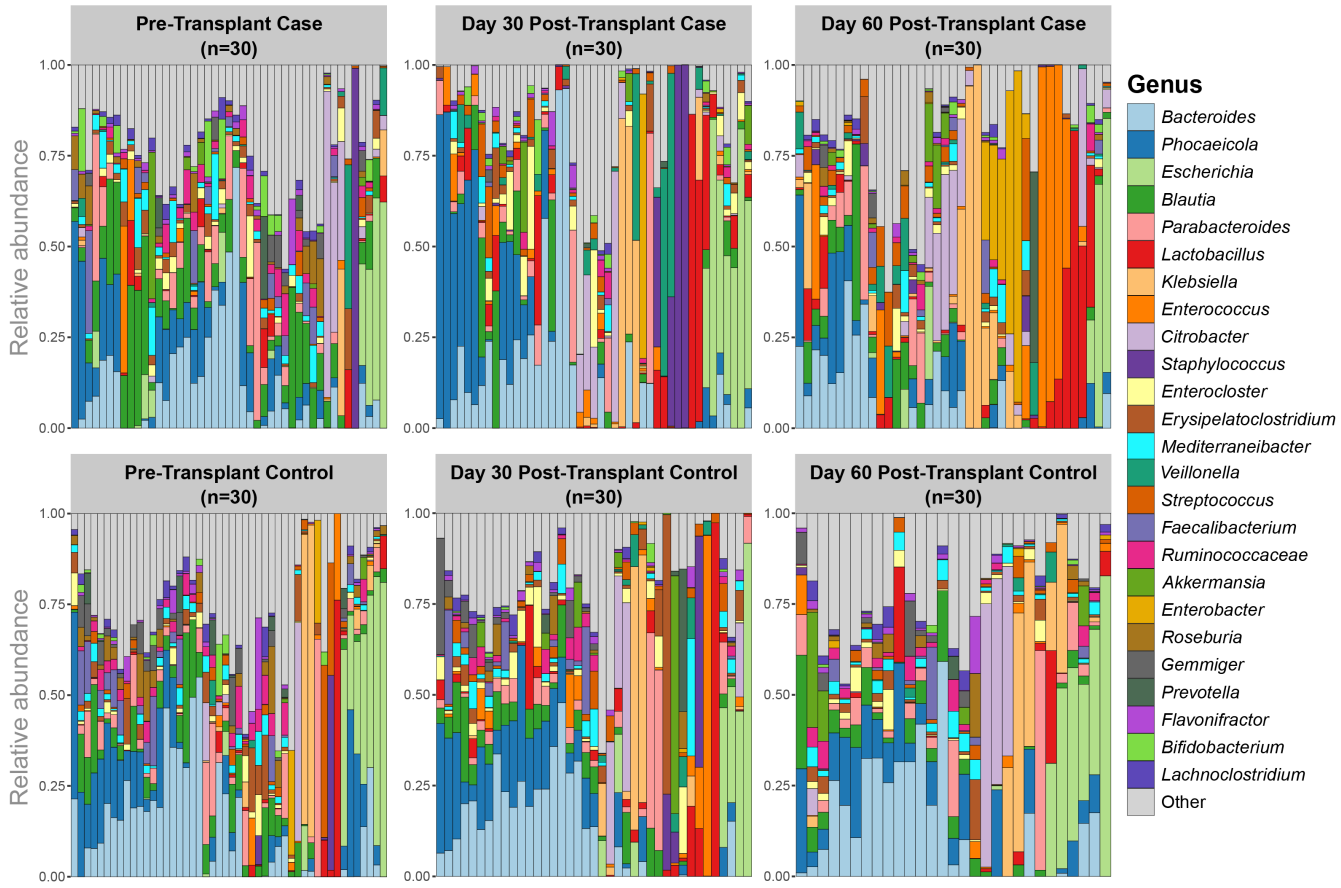


Figure 16. Relative abundance of top 20 bacterial genera calculated from 16S rRNA gene sequencing data. There are global similarities between GvHD case and control stool samples. We observed no specific taxa at the family, genus or species level that was statistically significantly different by GvHD status when using the Mann Whitney U test and when adjusting for multiple comparisons.

Bile acid	Microbial taxa	rho	p-value	Adjusted p-value	Timepoint
12KLCA814	*Clostridium scindens*	0.57	2.53E-09	1.87E-06	Day 30
Deoxycholic acid	*Clostridium scindens*	0.55	4.70E-08	1.67E-05	Day 60
3-ketocholanic acid	*Ruminococcus torques*	0.59	2.46E-10	3.39E-07	Day 60
Allolithocholic acid	*Ruminococcus torques*	0.61	7.22E-10	6.49E-07	Day 60
Allolithocholic acid	*Oscillospiraceae*	0.58	3.96E-09	2.51E-06	Day 30
3-ketocholanic acid	*Dysosmobacter welbionis*	0.57	8.26E-10	6.49E-07	Day 60
Allolithocholic acid	*Dysosmobacter welbionis*	0.57	7.89E-09	4.34E-06	Day 60
3-ketocholanic acid	*Lachnospiraceae*	0.56	6.42E-10	6.49E-07	Day 60
Allolithocholic acid	*Lachnospiraceae*	0.59	1.86E-09	1.28E-06	Day 60
Allolithocholic acid	*Neglecta*	0.65	1.77E-10	3.25E-07	Day 60
Allolithocholic acid	*Neglecta*	0.55	1.20E-08	3.54E-06	Day 30
3-ketocholanic acid	*Eubacteriales*	0.63	7.30E-12	2.01E-08	Day 60
3-ketocholanic acid	*Eubacteriales*	0.55	7.35E-11	3.27E-07	Day 30
Allolithocholic acid	*Eubacteriales*	0.76	5.15E-15	2.83E-11	Day 60
Allolithocholic acid	*Eubacteriales*	0.57	6.80E-10	1.01E-06	Day 30

Table 12. Bile acids and microbial taxa correlate post-HCT. Correlations between abundance of microbial taxa and bile acids that were moderately strong (rho ≥ 0.55). Several bile acids link with multiple microbial taxa, and vice versa. These strong correlations emerge post-HCT. Kendall Tau correlation was used to evaluate the relationships between bile acids and 16S microbial taxa. 'Timepoint' refers to the timepoint post-transplant.

Several microbial taxa in Table 12 that moderately strongly correlated with a bile acid (in this study, a moderately strong correlation coefficient rho was considered ≥ 0.55) either possess or include members that possess *bsh* (e.g. *Oscillospiraceae, Lachnospiraceae, D. welbionis, R. torques*) and/or *bai* genes (e.g. *C. scindens*), but it is challenging to assess bile acid metabolism by taxonomy alone.[200–205] This is because *bsh* genes are highly diverse in sequence, making it difficult to use sequence-based methods (such as PCR) to characterize them. *Bsh* gene possession is also highly inconsistent and yet abundant, with *bsh* genes present at zero to four copies per genome even within a single taxon, such as a given species or strain.[126,128,206–209] Gene orthologs of *bsh* within the same genome or taxon may not have high sequence similarity, similar functions, or target specificity/efficiency.[128,210–214] In contrast, the presence of the *bai* gene is more conserved within a species. Altogether, this informed our decision to adopt a gene-forward approach to further probe microbial bile acid metabolism.

3.2.8 Loss of secondary bile acid-producing bai genes associates with GvHD post-HCT

The gene-forward approach involved a targeted *bsh* and *bai* gene survey in stool that was performed for GvHD case and control samples using shotgun metagenomic sequencing data. This method quantified capacity for bile acid metabolism in gut microbial communities without using taxonomy as a proxy. Surprisingly, the normalized sum of metagenomic reads that aligned to a *bsh* gene ortholog did not associate with GvHD status at any timepoint (p values > 0.30; Fig. 17A). Rather, a greater number of unique *bsh* gene orthologs associated with absence of GvHD at day 60 post-HCT ($p = 0.02$, FC = 3.2; Fig. 17B). This was concordant with the observed association between a diverse microbial community and absence of GvHD (Fig. 15) and suggests that quantity of *bsh* gene orthologs may be less important than type(s) or diversity of *bsh* gene orthologs. For the *bai* genes, a 100-fold lower normalized sum of reads that aligned to *bai* gene orthologs associated with GvHD 30 days post-allo-HCT ($p = 0.002$, FC < 0.01, the day 30 post-HCT median sum of *bai* reads in GvHD cases is the pseudocount 0.1; Fig. 18A). Diversity of *bai* genes was also important as a 20-fold lower number of unique *bai* gene orthologs associated with GvHD 30 days post-allo-HCT ($p = 0.0007$, FC < 0.05, the day 30 post-HCT median number of *bai* genes in GvHD cases is the pseudocount 0.01; Fig. 18B). Finally, detection of at least one *bai* gene ortholog was associated with absence of GvHD 30 days post -HCST ($p = 0.004/p_{adj} = 0.01$, odds ratio: 4.1; Table 13), further highlighting the importance of *bai* genes.

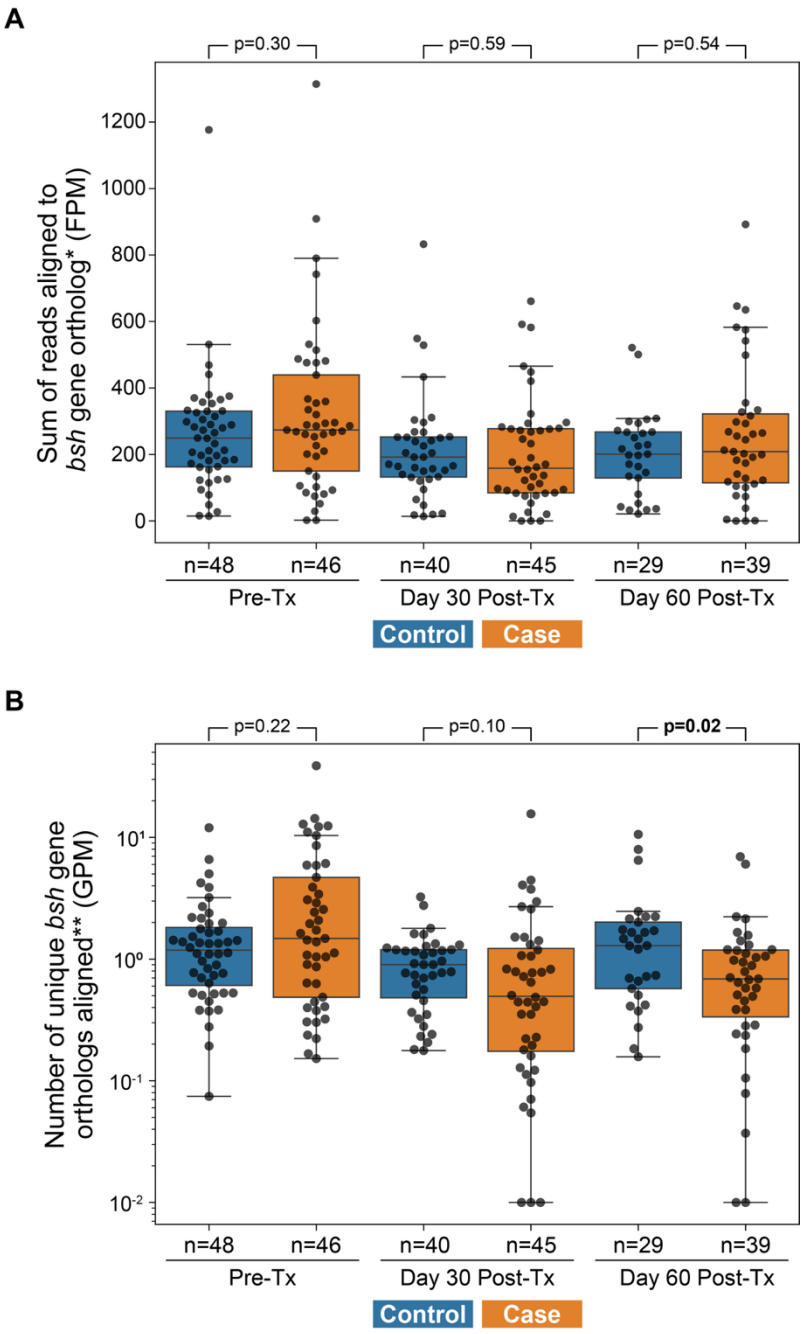


Figure 17. Abundance and diversity of deconjugation *bsh* genes. (A) The distributions of sum of reads that aligned to a *bsh* gene ortholog (normalized to the sum of bacterial reads per sample) in a sample were statistically the same between GvHD cases and controls at each timepoint. **Bsh* reads were normalized to the sum of bacterial reads per sample, reported in fragments per million (FPM). (B) Metagenomic reads aligned to statistically fewer *bsh* gene orthologs (normalized to the sum of bacterial reads per sample) in GvHD cases vs. controls at day 60 post-transplant. Pre-transplant and 30 days post-transplant, GvHD cases and controls had statistically the same number of *bsh* genes hit by the metagenomic reads. **Number of *bsh* genes were normalized to the sum of bacterial reads per sample, reported in genes per million (GPM). A read cutoff of 5 was implemented to consider any genes a true hit. The bounds of the boxes, the lines within the boxes, the whiskers, and the consideration of outlying values are the same as those defined in the Figure 6 legend. Mann-Whitney U test was performed to generate p-values.

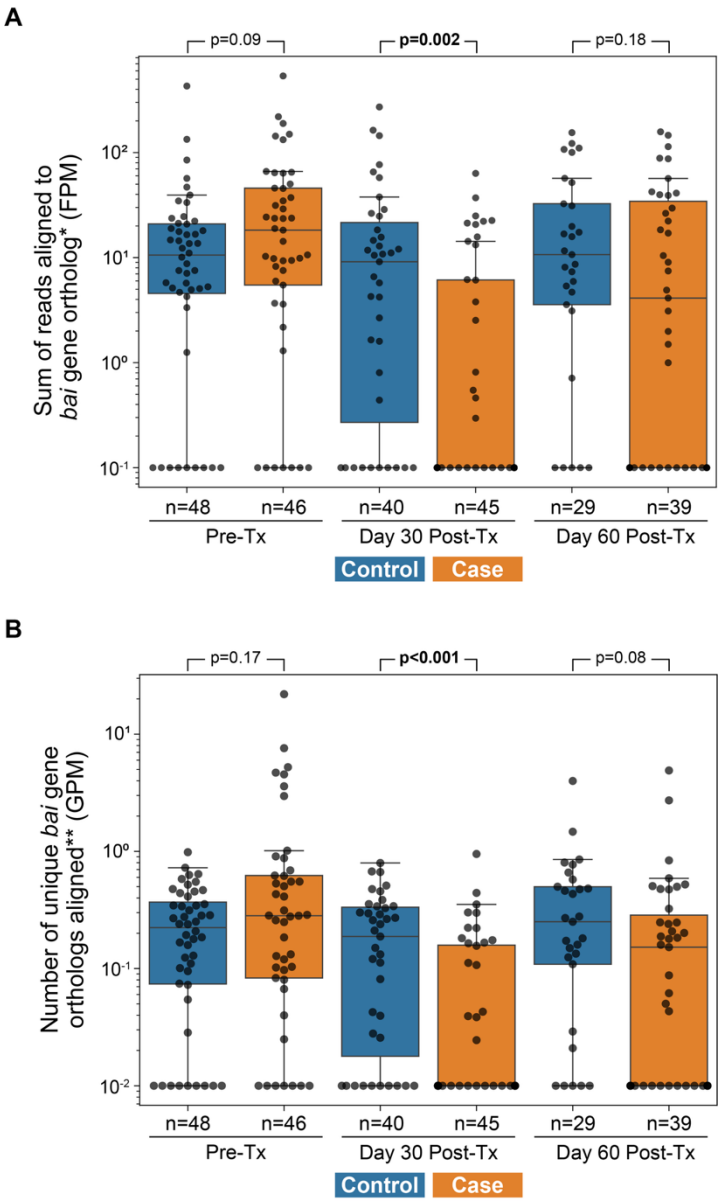


Figure 18. Abundance and diversity of 7α-dehydroxylation *bai* genes. (A) The sum of reads that aligned to a *bai* gene ortholog (normalized to the sum of bacterial reads per sample) in a sample were statistically significantly greater in controls vs. GvHD cases at 30 days post-transplant. Pre-transplant, the (normalized) sums of reads aligned to a *bai* gene ortholog appeared greater in GvHD cases vs. controls, although this was not statistically significant. There was no statistical difference between sum of reads aligned to a *bai* gene at 60 days post-transplant based on GvHD status. **Bai* reads were normalized to the sum of bacterial reads per sample, reported in fragments per million (FPM). (B) Metagenomic reads aligned to statistically fewer *bai* gene orthologs (normalized to the sum of bacterial reads per sample) in GvHD cases vs. controls at day 30 post-transplant. 60 days post-transplant, GvHD cases looked to have fewer *bai* gene orthologs hit vs. controls, although this was not statistically significant. Pre-transplant, GvHD cases and controls had statistically the same number of *bsh* genes hit by the metagenomic reads. **Number of *bai* genes were normalized to the sum of bacterial reads per sample, reported in genes per million (GPM). A read cutoff of 5 was implemented to consider any genes a true hit. The bounds of the boxes, the lines within the boxes, the whiskers, and the consideration of outlying values are the same as those defined in the Figure 6 legend. Mann-Whitney U test was performed to generate p-values.

Timepoint	GvHD status	Number of samples with bai gene ortholog present	Total number of samples	Prevalence (%)	p-value	Adjusted p-value	Odds ratio	Odds ratio 95% CI
Pre-transplant	Case	38	45	84.44	0.786	0.786	0.8	0.23 - 2.70
	Control	39	48	81.25				
Day 30	Case	18	44	40.91	0.004	0.011	4.112	1.50 - 12.01
	Control	29	39	74.36				
Day 60	Case	24	39	61.54	0.066	0.099	2.953	0.85 - 12.09
	Control	**24**	**29**	**82.76**				

Table 13. Prevalence of 7α-dehydroxylation *bai* genes. At least one *bai* gene ortholog was detected in a greater portion of GvHD case samples vs. no-GvHD samples at days 30 and 60 post-transplant. Pre-transplant, at least one *bai* gene ortholog was detected in the same portion of GvHD case and control samples. A read cutoff of 5 was implemented to consider any genes a true hit. Fisher's Exact test was performed to generate p-values. P-values are FDR-adjusted.

3.2.9 Diet and bile acids

As described in Chapter 2, participants in the GvHD study completed dietary questionnaires in addition to providing samples for research purposes. These questionnaires were analyzed for nutrient composition by the Fred Hutchinson Cancer Center Nutrition Assessment shared resource using the University of Minnesota's Nutrition Data System for Research. This enabled correlation analyses between diet/nutrition data and metabolomic data from the same participant(s), longitudinally post-HCT. This section will focus on the links between diet and bile acid metabolomic data.

I generated five principle components (PCs) that maximized variance within each dataset, bile acid metabolomics and dietary/nutrition data. Through Spearman correlation analysis, four statistically significant associations were observed across three dietary PCs and three bile acid PCs (Table 14). The correlation (rho) values' magnitudes ranged from 0.29 to 0.32, although the strongest correlation, 0.32, represented an anti- or negative correlation between dietary PC1 and bile acid PC2. This suggested that further, detailed investigation into the relationship between dietary variables and bile acids may yield interesting and significant findings.

Diet PCs	BA PCs	Rho estimate	p_value	p_adj
PC1	PC2	-0.32	0.002	0.03
PC1	PC3	0.29	0.004	0.03
PC3	PC1	0.29	0.004	0.03
PC4	PC3	0.29	0.004	0.03

Table 14. PCA analysis of bile acids and dietary component data. Rho values were calculated using Spearman correlation. P-values were adjusted for false discovery rate (FDR). PC = principal component, BA = bile acid

Across all 169 diet variables and 32 bile acids, there were 7,605 correlations calculated. Out of the 7,605 correlations, 3,858 (51%) were statistically significant ($p < 0.05$) and 2,208 (29%) were moderately strong to strong (rho > 0.4. or < -0.4) correlations or anti-correlations. Just 2,072 (27%) were both statistically significant and moderately strong to strong (rho > 0.4. or < -0.4) correlations or anti-correlations. Interestingly, 725 out of 2,072 (35%) were correlated but most (1,347 or 65%) were anti-correlated. One such surprising yet significant anti-correlation was between total concentration of glycine-conjugated bile acids (n = 6) and dietary glycine, wherein greater levels of total glycine-conjugated bile acids links with lower levels of dietary glycine (p_{adj} = 6.85E-7, p = 9.43E-9, rho = -0.54; Table 15). This does not support the hypothesis that more ingested glycine resulted in more fecal levels of glycine-conjugated bile acids. A secondary hypothesis was that there was less circulating or free glycine due to glycine getting bound to bile acids; however, the glycine quantities were generated solely from nutrition analysis of dietary components, not from measured levels in stool. Thus, the glycine composition of diet would not directly relate to or interact with free glycine or bile acids. The mechanism(s) underlying this relationship is unclear. Further support for this link is the significant (p_{adj} = 6.46E-8, p = 6.06E-11, rho = -0.60) anti-correlation observed between total concentration of all conjugated bile acids (n = 12, including glycine-conjugated bile acids) in stool and dietary glycine (Table 15). However, the underlying mechanism(s) remain unclear.

Bile acid(s)	Dietary component	p-value	Adjusted p-value	Rho (spearman)
Glycine-conjugated bile acids	Glycine	9.43E-9	6.85E-7	-0.54
Conjugated bile acids	Glycine	6.06E-11	6.46E-8	-0.60
Endogenous secondary bile acids	Cholesterol	1.28E-5	9.45E-5	0.43
Lithocholic acid	Total saturated fatty acids	3.31E-6	3.68E-5	0.45

Table 15. Correlations between individual bile acids and dietary components. Rho values were calculated using Spearman correlation. P-values were adjusted for false discovery rate (FDR).

Total levels of endogenous secondary bile acids (n = 11) significantly correlated with dietary cholesterol (p_{adj} = 9.45E-5, p = 1.28E-5, rho = 0.43), serving as a proof of concept for these correlations (Table 15). Increased dietary cholesterol would push an increase in intestinal bile acids, including the bile acid metabolic end products, secondary bile acids produced endogenously by the gut microbiota. This finding is supported by the correlation between lithocholic acid, an endogenous secondary bile acid, and total dietary saturated fatty acids (p_{adj} = 3.68E-5, p = 3.31E-6, rho = 0.45; Table 15). An increase in saturated fatty acids would increase cholesterol, pushing bile acid transport to the intestine and further production of a microbial bile acid metabolic product, lithocholic acid.

3.3 Discussion

3.3.1 Deficiency in endogenous secondary bile acids associates with GvHD

Despite human leukocyte antigen matching to reduce GvHD post-HCT, many HCT recipients still develop GvHD.[215,216] Why there is such heterogeneity in GvHD development/recovery after HCT occurs and the environmental factors and mechanisms that contribute have yet to be determined. It is known that the gut microbiome is associated with GvHD.[217–219] There are many potential mechanisms for how the gut microbiota may influence GvHD, including through impacts on immune cells, epithelial cells, and the gut barrier; bile acids are known to affect all of these. In this study, we tested stool samples as a proxy for intestinal bile acid dynamics. Stool is not a perfect reflection of metabolite and bile acid concentrations within the intestine, although it may be the best, feasible option for human studies (for greater detail, refer to section 3.3.7: Limitations of the Study). We found a reduction in summed endogenous secondary bile acid concentrations in stool from GvHD cases post-HCT (Fig. 8D).

All HCT recipients are given an exogenous secondary bile acid, UDCA (ursodiol) to prevent liver complications after HCT. Furthermore, six out of eleven endogenous secondary bile acid concentrations measured by the targeted panel in our study were decreased in GvHD cases post-HCT compared to controls (Table 10). Each of these endogenous secondary bile acids have a similar structure to lithocholic acid. Four of these endogenous secondary bile acids are also less prevalent in GvHD vs. no-GvHD controls (Table 11). Allolithocholic acid is an additional endogenous secondary bile acid less prevalent in GvHD (Table 8). Some studies posit that UDCA administration increases LCA levels in blood, yet all participants in the current study received UDCA and we observed differences in LCA levels in stool.[220]

3.3.2 Changes in the gut microbiota influence bile acid metabolism

As expected, all patients lose some microbial diversity due to the antibiotics administered to every patient (Fig. 15). This is reflected by dynamic changes in each patient's microbial community function, such as in the *bai* gene operon that impacts microbial bile acid metabolism and concentrations (Fig. 18). However, this dynamic/variance is more pronounced in the GvHD cases vs. no GvHD controls (Fig. 7F, 15, 18), leading us to ask: why are endogenous secondary bile acid concentrations decreased in patients with GvHD compared to controls? This may be due to a reduction of microbial *bai* genes and *bai* gene diversity in stool from HCT-recipients with GvHD, reflecting loss of 7-alpha hydroxylase activity needed to produce secondary bile acids (Fig. 18). Metagenomic analysis is a more useful indicator of whether *bai* or *bsh* genes are present in stool compared to using microbial taxonomy. We noted that absence of *bai* genes in stool samples is associated with GvHD (Table 13). In contrast, *bsh* genes were detected in every stool sample in our study. This suggests the *bai*-mediated step (not the *bsh*-mediated step) may be rate-limiting in the production of secondary bile acids in these transplant recipients. Additional support for this observation includes lack of any significant difference in the summed levels of *bsh*-modified bile acids (primary and conjugated) comparing cases to controls (Fig. 8B-C). *Bai* gene products are exclusively responsible for the 7α-dehydroxylation step that produces secondary bile acids. Host genetic and other factors do not play a role in secondary bile acid production. The HCT recipients with GvHD may not have sufficient *bai*-possessing microbes. This is reflected by: (A) lower microbial diversity, as observed in our 16S rRNA gene sequencing analysis (Fig. 15), and (B) reductions in both abundance and diversity of *bai* genes that encode enzymes that produce endogenous secondary bile acids, as reflected in our metagenomic analysis (Fig. 18). For example, *Clostridium scindens* encodes *bai* genes and is associated with levels of individual endogenous

secondary bile acids (Table 12). This example supports *bai*-possessing microbes' potential involvement in the observed endogenous secondary bile acid and *bai* gene depletion. *Bsh* was found in every stool sample from patients with and without GvHD, but there was greater *bsh* gene diversity in controls 60 days post-HCT (Fig. 17). This observation suggests that having a diverse set of *bsh* enzymes may be more optimal for converting conjugated bile acids into primary bile acids, and then on to secondary bile acids.

3.3.3 Bile acid levels change with GvHD

A strength of this time analysis study is inclusion of both pre- and post-HCT samples to compare microbiome dynamics and bile acid concentrations over time, but the cadence of sample collection does not distinguish among endogenous secondary bile acids as simply markers of GvHD, or within the pathway to GvHD, or both. Pre-transplant bile acid concentrations did not predict GvHD (Table 8), and by day 30 post-transplant, 30 of 45 patients had developed GvHD in this study. The distribution of bile acids at 30 days post-HCT is described in Fig. 11 and Table 9, where endogenous secondary bile acid levels are lower in the 30 patients with GvHD vs. (a) the 15 patients without GvHD yet and (b) the no GvHD controls. It is possible that changes in endogenous secondary bile acids precede GvHD onset by days or weeks, and the monthly sampling frequency would not capture that change. Alternatively, changes in secondary bile acids may be concordant with onset of GvHD and thus serve as a marker but not a predictor, while still contributing to the maintenance of GvHD. One way to determine if secondary bile acids contribute to the induction or maintenance of GvHD would be to administer these bile acids or similar bile acid receptor agonists in human or animal studies to assess the impact on GvHD.

3.3.4 Bile acids may impact GvHD via several pathways

How might endogenous secondary bile acids contribute to GvHD? Bile acids can signal through receptors on IECs and immune cells.[221,222] This signaling could modulate a range of pleiomorphic host responses, like bile acid production/release, metabolic activity, and immunity.[120–122,223–225] Of particular interest is immune cell activation because GvHD is mediated by T cells. Bile acids 3-oxolithocholic acid and isoallolithocholic acid inhibit T-helper type 17 cells and promote T regulatory cell differentiation in mice.[120–122] Bile acids can also bind cell membrane receptor TGR5 (LCA > DCA > CDCA), present on macrophages and dendritic cells, and nuclear membrane receptor Farnesoid X receptor (FXR; CDCA > DCA > LCA > CA > UDCA), present on T cells, natural killer (NK) cells, NKT cells, and myeloid cells such as neutrophils, macrophages and dendritic cells.[117,221,222,226–228] These connections are highly

relevant because Th17 cells are a main driver of GvHD, whereas Treg cells limit T cell activation and thus GvHD, providing a plausible, direct pathway for bile acids to modulate GvHD.[229–233] Vitamin D receptor (VDR) is another receptor of interest that binds LCA but not other bile acids.[149] VDR is present on IECs and T cells, both of which are crucial for GvHD as IEC-mediated antigen presentation specifically drives the priming and activation of donor T cells that initiate GvHD.[150,234] VDR activity, like that which can be triggered by LCA binding and activation, can promote tight junction protein gene expression to preserve intestinal barrier integrity, abrogate pro-inflammatory IL-12, NFκB, and TNFα gene expression, and impact differentiation of T cells.[151–154] Vitamin D deficiency has been extensively studied in the context of GvHD and linked to poor overall survival post-HCT in humans.[145,146] Case reports describe resolution of chronic GvHD symptoms and reduced relapse rates post-HCT in human recipients supplemented with Vitamin D.[148] Since TGR5 is also present in macrophages and dendritic cells, bile acids can influence these cell subsets that present antigens and indirectly contribute to GvHD by promoting T cell activation.[227,235–239] Additional immunomodulatory effects of bile acids include downregulation of proinflammatory genes and activation of the inflammasome, where the immune cell type involved (e.g. T cells vs. macrophages) and context determine the exact impact on gastrointestinal immune regulation.[240] An alternative pathway draws on bile acids' amphipathic properties permitting their breakdown of dietary fats and their antimicrobial effects. Endogenous secondary bile acids may solubilize bacterial phospholipid outer membranes and reduce the risk of intestinal pathogen growth that would contribute to inflammation and exacerbate GvHD. LCA, an endogenous secondary bile acid, is the most hydrophobic bile acid (LCA > DCA > CDCA > CA > UDCA). LCA exhibits antibacterial activity against GvHD-associated bacteria like *Enterococcus, Staphylococcus, Escherichia,* and others.[241,242] Thus, LCA's observed deficiency (Fig. 12) could permit proliferation of such bacteria to exacerbate GvHD. However, one challenge with using lithocholic acid as an agonist is that high concentrations in the gut can be pro-inflammatory.[243]

3.3.5 Endogenous secondary bile acids and GvHD: reproducible?

Key findings in our study were concordant with findings from another human study by Lindner et al. published in 2024.[116] This includes the deficiency of: LCA (Fig. 13), endogenous secondary bile acids, and the sum of *bai* operon genes in GvHD cases vs. controls. In contrast, summed bile acid levels were reported as lower in GvHD case vs. controls samples in the Lindner study, but not the current study. This difference may reflect different study designs. In Lindner et al., key timepoints of interest were chosen based on individual diagnosis of GvHD (e.g. pre-GvHD onset, peri-GvHD onset). Our study used absolute time to establish timepoints

of interest. Thus, the time between sample collection/analysis and GvHD differs in the two studies. The former retrospective approach using individual diagnosis of GvHD may better capture the biology of GvHD in an individual patient. Ten additional findings from our study were also concordant with Lindner et al. data.[116] This demonstrates our study's findings are not isolated to a single population or center. Although findings were largely concordant, we recognize that each cohort has a unique exposome, including antibiotics, and this may explain differences between these studies.

3.3.6 Endogenous secondary bile acid levels are independent of steroid therapy

It is possible that steroid therapy may impact the findings described in this study by promoting bile acid recycling that in turn, decreases intestinal bile acid levels.[244] Bile acids are primarily reabsorbed through small intestine epithelial cells, but IEC damage and sloughing is characteristic of the conditioning that transplant recipients undergo before transplant. In addition, damage induced by GvHD at the mucosa may reduce ASBT transporters and thus impair bile acid reabsorption, increasing bile acid levels detected in stool. When we excluded samples that were collected from patients during active prednisone/methylprednisolone therapy, the significant deficiency in endogenous bile acids such as lithocholic acid during GvHD was still observed (Fig. 14). This supports that steroid therapy does not bias the fecal bile acid concentrations measured in this study. To the best of our knowledge, this represents a novel contribution to the field, as previous studies characterizing bile acid dynamics in stool post-HCT do not address the effect of steroid therapy on bile acid concentrations. However, this is consistent with reports from the current literature, which support that bile acid levels in stool are unaffected by prednisone therapy.[245]

3.3.7 Limitations of the study

There are some limitations to this study. Antibiotics may affect the gut microbiota and bile acid metabolism. All patients in this study received prophylactic, empiric, or directed antibiotic therapy so there was no unexposed group. There are several factors that may change the gut microbiota and therefore bile acids, like antibiotics and diet; a future analysis will focus on the relationships of specific antibiotics to changes in the gut microbiota and bile acid levels in stool. Microbial RNA was not sequenced so *bsh* or *bai* gene expression was uncharacterized. Participants' stool samples were collected via swab. This facilitated sample collection but prevented reliable measurement of stool weight. Thus, the input stool quantity for LC-MS, 16S rRNA gene sequencing, and metagenomic sequencing was non-normalized. However, the same swabs were used for all participants to make collection consistent. The swabs typically

load 50-100 µL of stool volume. Stool sampling cadence is also limiting. Since the sampling was done approximately monthly, it is unclear how endogenous secondary bile acid levels changed between timepoints. Higher frequency sampling (daily, weekly) is necessary to characterize endogenous secondary bile acids dynamic at greater resolution and determine whether changes happen before or after GvHD.

The nature of the samples was also limiting. Ideally, intestinal contents from each section of the GI tract (e.g. duodenum, jejunum, ileum, cecum, ascending colon, transverse colon, descending colon, sigmoid colon, rectum) of each participant could be directly sampled to capture the precise changes in each bile acid concentration at each region of the GI tract. However, while this can be performed after a mouse experiment, it is not feasible to do so in humans (it would require forensic autopsy[105,246]). This type of sampling would be too invasive and challenging to perform. Thus, stool samples were chosen as a proxy for the dynamic intestinal bile acid concentrations, although this is not a perfect reflection since only approximately 5% of the bile acid pool is excreted in stool. Blood or serum samples are also used to measure bile acid concentrations[247], but the measured bile acid levels could remain constant even if there is very high turnover, processing, and storage of bile acids in the liver. Additionally, the goal was to measure bile acid levels as close as possible to the true intestinal concentrations. A blood or serum sample may more closely resemble bile acid influx to the liver rather than into the intestines. Bile acids can also be stored in the liver or gall bladder such that a decrease in one compartment (e.g. serum) may not necessarily indicate an increase in another (e.g. colon). This discussion is a glimpse into the recurring challenges associated with sampling and measuring bile acid flux (transportation, storage, reabsorption, excretion). Lastly, the GvHD and pre-HCT conditioning damage in the gut likely impacts host bile acid receptors, downstream signaling, and reabsorption rate, and therefore the measured levels in stool as well. Given this complexity, it is noteworthy that statistically significant signal was observed in this study's samples which were collected approximately monthly.

3.3.8 Summary of findings and future work

In summary, we propose that GvHD may be impacted by alterations in gut microbial bile acid metabolism. We showed that bile acid concentrations differ in HCT recipients with vs. without GvHD (Fig. 6). Our study provides evidence that secondary bile acids, especially lithocholic acid, are reduced during GvHD (Fig. 8D, 11, Table 10). The deficiency may result from a loss of the gut microbes that possess *bai* genes and produce secondary bile acids, which can influence gut immunity in the setting of GvHD. These findings are summarized in a visual abstract (Fig. 19). If bile acids contribute to the pathogenesis of GvHD, then lithocholic acid and

the cell receptors that bind it (VDR, TGR5) could represent a novel therapeutic axis to explore for the prevention and treatment of GvHD.

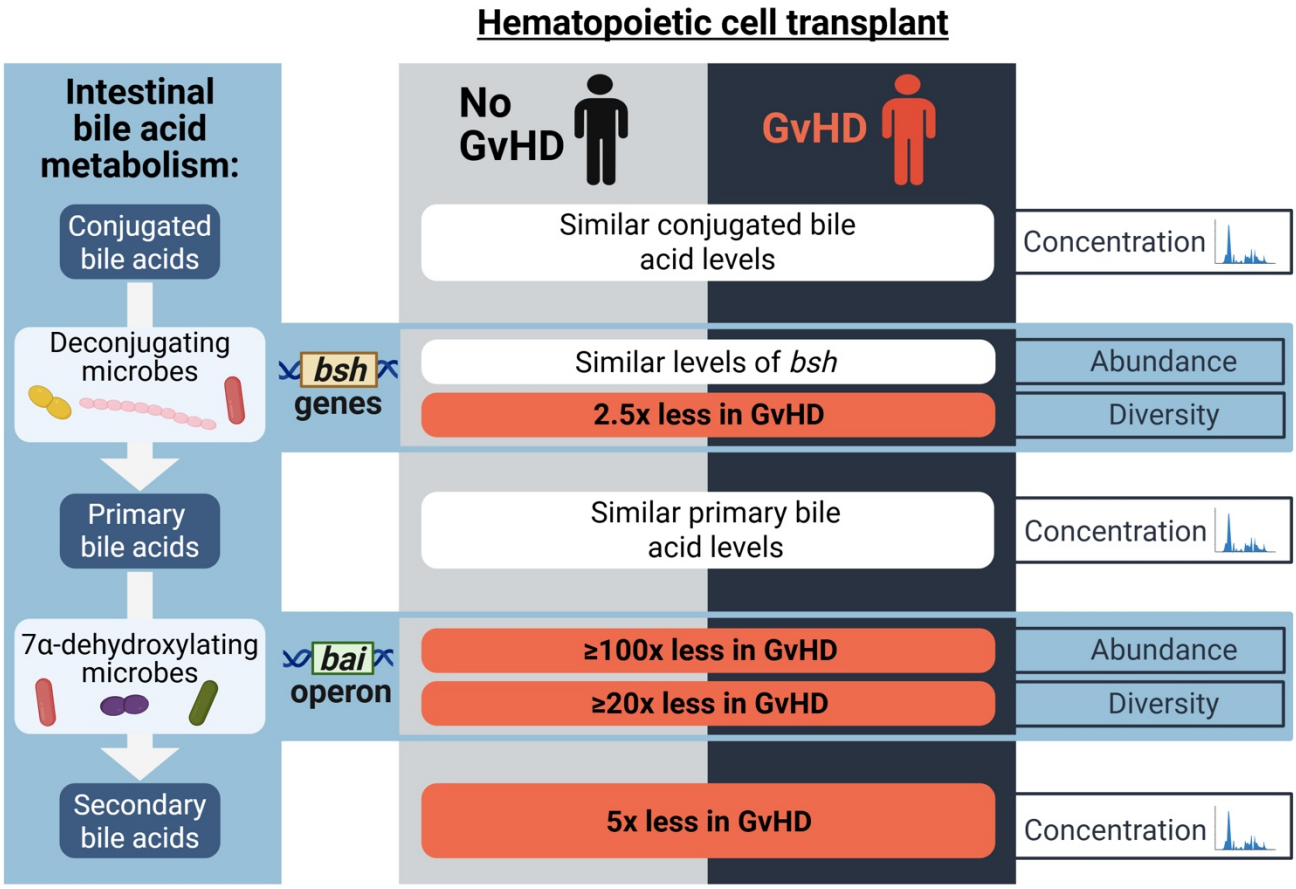


Figure 19. Acute gastrointestinal GvHD is associated with reductions of secondary bile acids, *bsh* genes, and *bai* genes after allo-HCT. This visual abstract summarizes the findings of our study. Created in BioRender. Proll, S. (2026) https://BioRender.com/5z04gp7.

There are several research avenues that should be explored to advance this study's findings towards interventions that could improve outcomes for allo-HCT patients. For instance, even microbial strains within the same genus or species can variably metabolize BAs. These metabolic rates should be well-characterized to inform the strains best suited for therapies. This includes both bsh and bai enzymes since poor diversity of the genes encoding both enzymes was linked with GvHD (Fig. 17-18). Similarly, strategies to enhance *bsh* diversity may be valuable; one such strategy is to administer bacteria with diversity of *bsh*'s represented. Another treatment option could involve prescribing primary bile acids (e.g. chenodeoxycholic acid) as a precursor in combination with the diverse *bsh*-possessing microbes to promote endogenous

conversion to secondary bile acids. Deeper dives into these next steps and more future directions are discussed in greater detail in Ch 6: Future directions, Ch 3: Bile acids and GvHD.

Chapter 4: Short chain fatty acids (SCFAs) and GvHD

4.1 Introduction

4.1.1 Overview of field: SCFAs and protection against vs. priming for GvHD

SCFAs are the end products of microbial fermentation of dietary nondigestible carbohydrates (e.g. fiber, resistant starches) whereas branched chain fatty acids (BCFAs) are the end products of microbial fermentation of proteins. SCFAs have broad physiological impacts; from their microbially-mediated metabolism/production in the colon, they can signal and/or interact with the variety of cells in the intestinal barrier or cross the intestinal barrier and enter the bloodstream for transport throughout the body to a range of organs.[104,106] SCFAs can affect the brain, liver, pancreas, adipose tissue, peripheral nerves and kidneys via receptors such as FFAR2/GPR43, FFAR3/GPR41, GPR109A and OLFR78.[248] It is unclear if BCFAs have distinct or overlapping effects vs. SCFAs, and for the purpose of this thesis, both are referred to as SCFAs. Several microbial taxa are found in the human gastrointestinal tract and capable of producing SCFAs, including but not limited to genera like *Faecalibacterium*[249], *Roseburia*[250], *Clostridium*[251], *Anaerococcus*[252], *Mediterraneibacter*[253], *Fusobacterium*[254], *Anaerostipes*[255], *Bacteroides*[256], and *Blautia*[257]. The total SCFA pool is approximately 130 mmol/kg in the cecum, 80 mmol/kg in the descending colon, and 13 mmol/kg in the ileum.[134] In the colon, the molar ratio of acetate to propionate to butyrate is 60:20:20.[135,258,259] Depending on diet and geographical location, the levels of acetate detected in stool samples range from 40-100mM, for propionate 15-30mM, and for butyrate 10-25mM.[134,135]

4.1.1.1 Tight junctions

There are several proposed mechanisms for how SCFAs exert beneficial and/or protective effects in the intestinal tract. SCFAs have been shown to trigger an influx of intracellular Ca^{2+} that activates 5' adenosine monophosphate-activated protein kinase (AMPK), inhibits myosin light chain kinase (MLCK), phosphorylates protein kinase C beta 2 (PKCB2) to ultimately promote tight junction assembly. SCFAs have also been observed to upregulate expression of claudin-1, a tight junction protein, via the transcription factor specificity protein 1 (SP1). Each of these proposed mechanisms culminates in the strengthening of tight junctions, which regulate paracellular permeability of the intestinal barrier. With more tight junction proteins, the space between cells decreases, enhancing the integrity of the intestinal barrier. An

intact intestinal barrier is not leaky, meaning systemic inflammation is restricted due to decreased translocation of intestinal inflammatory compounds, metabolites, microbes, etc. into the bloodstream.[136,260,261]

4.1.1.2 Anti-inflammation: Histone deacetylase inhibitors (HDACi), GPR-binding

Select SCFAs can reduce intestinal inflammation via multiple proposed pathways. Butyrate and propionate may act as a histone deacetylase inhibitor (HDACi). HDACi affect transcription by keeping histones acetylated, promoting transcription of anti-inflammatory cytokines.[95,262,263] Another anti-inflammatory pathway for SCFAs is mediated by cell signaling on GPRs. Butyrate activation of GPR109A is protective against colonic inflammation via IL-8 and IL-10 induction.[137] Acetate binds GPR43 to increase the amount of intestinal IgA that promotes intestinal immune tolerance as opposed to inflammation.[138] These proposed pathways may contribute to a more stable gut microbiome that is simultaneously more resilient post-HCT and less likely to develop GvHD.

4.1.1.3 Facilitates cell turnover and transcellular damage

SCFAs like butyrate can also act as an energy source to promote intestinal epithelial cell turnover. Where the beneficial effects on tight junctions reduce paracellular permeability of the intestinal barrier, quicker intestinal epithelial cell regeneration may staunch any leakiness of the gut barrier by reducing transcellular permeability/damage, restricting translocation of metabolites or whole microbes by passing through regions of the intestine where cells died and sloughed off.[49] However, this effect is not uniform throughout the intestine; ISC exposure to butyrate has been observed to reduce proliferation.[45] In a healthy state, only low levels of butyrate reach stem cells that are located deep in intestinal crypts. In a dysbiotic state post-HCT, the mucosa and intestinal epithelium may be so damaged that the intestinal stem cells and crypts are exposed to gut luminal butyrate, impeding recover from tissue injury. This occurs in a FoxO3-dependent manner: butyrate stimulates production of transcription factor FoxO3 which binds to the promoter region of negative cell cycle regulators such that the regulators' expression is increased to ultimately inhibit the cell cycle and stem cell progenesis. Since expression levels mediate this effect, butyrate's HDACi role may also come into play by promoting expression of these cell cycle arrest proteins.[264] Butyrogenic bacteria associate with steroid-refractory GvHD, further supporting the spatial and cell-type-dependent nuance of butyrate effects within the intestine.[45] Overall, there are likely several mechanisms at play that combine for the observed impacts of butyrate.

4.2 Results

4.2.1 Targeted SCFA/BCFA metabolomics

4.2.1.1 Individual SCFAs by GvHD case vs. control per timepoint

The targeted SCFA and BCFA metabolomics panel analyzed the concentration of ten SCFAs: acetic acid, propionic acid, butyric acid, isobutyric acid, 2-methylbutyric acid, isovaleric acid, valeric acid, 3-methyl-valeric acid, 4-methyl-valeric acid, and caproic acid. Only 3-methyl-valeric acid was not detected in any of the stool samples analyzed. The SCFAs include acetic acid, propionic acid, and butyric acid. The remaining seven, isobutyric acid, 2-methylbutyric acid, isovaleric acid, valeric acid, 3-methyl-valeric acid, 4-methyl-valeric acid, and caproic acid, are BCFAs.

Hierarchical clustering of all samples on based concentrations of the ten SCFAs investigated did not reveal any trends that linked with GvHD status. However, significant differences emerged on the individual SCFA basis. Pre-transplant, butyric acid concentrations were significantly greater ($p = 0.02$) in controls, 65,195.4 mM, vs. cases, 14,370.2 mM ($p m= 0.02$; Fig. 20). On day 60 post-transplant, concentrations of valeric acid were significantly greater ($p = 0.04$) in controls, 3,007.3 mM, vs. cases, 1,079.9 mM (Fig. 21). 2-methylbutyric acid levels were significantly greater ($p = 0.02$) in controls, 27,706.0 mM, vs. cases, 14,977.5 mM, at day 60 post-HCT (Fig. 22). Concentrations of isobutyric acid are significantly greater ($p = 0.007$) in controls, 49,953.2 mM, vs. cases, 15,114.1 mM, at day 60 post-HCT (Fig. 23). On day 60 post-HCT, there were statistically significantly greater levels of isovaleric acid in controls, 36,465.7 mM, vs. cases, 17,444.9 mM (Fig. 24). There were no statistically significant differences in concentrations of 4-methyl-valeric, propionic (Fig. 25), acetic or caproic acid between cases vs. controls at any timepoint.

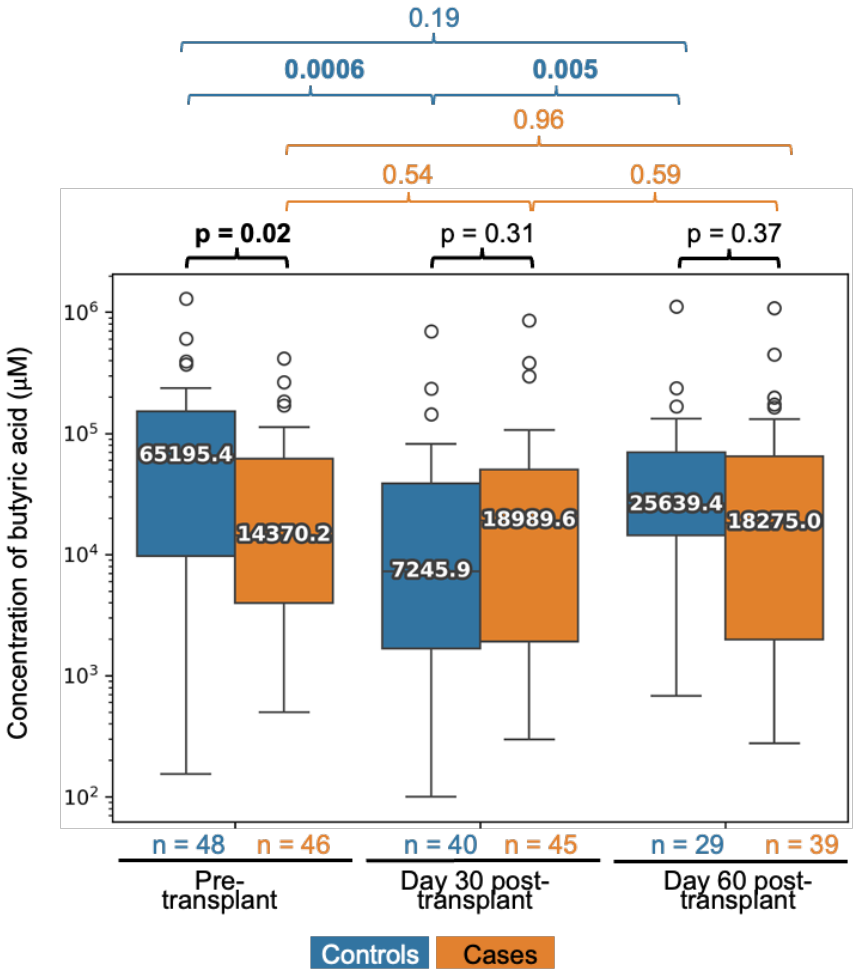


Figure 20. Butyric acid concentrations throughout HCT. Pre-transplant (HCT), there were higher levels of butyric acid in no GvHD controls vs. GvHD cases (p = 0.02). Butyric acid levels in no GvHD controls decreased from pre-HCT to day 30 post-HCT (p = 0.0006). Levels significantly increase from day 30 to day 60 post-HCT (p = 0.005) in no GvHD controls. Concentrations are in millimolar (mM). P-values are generated from the Mann Whitney U test.

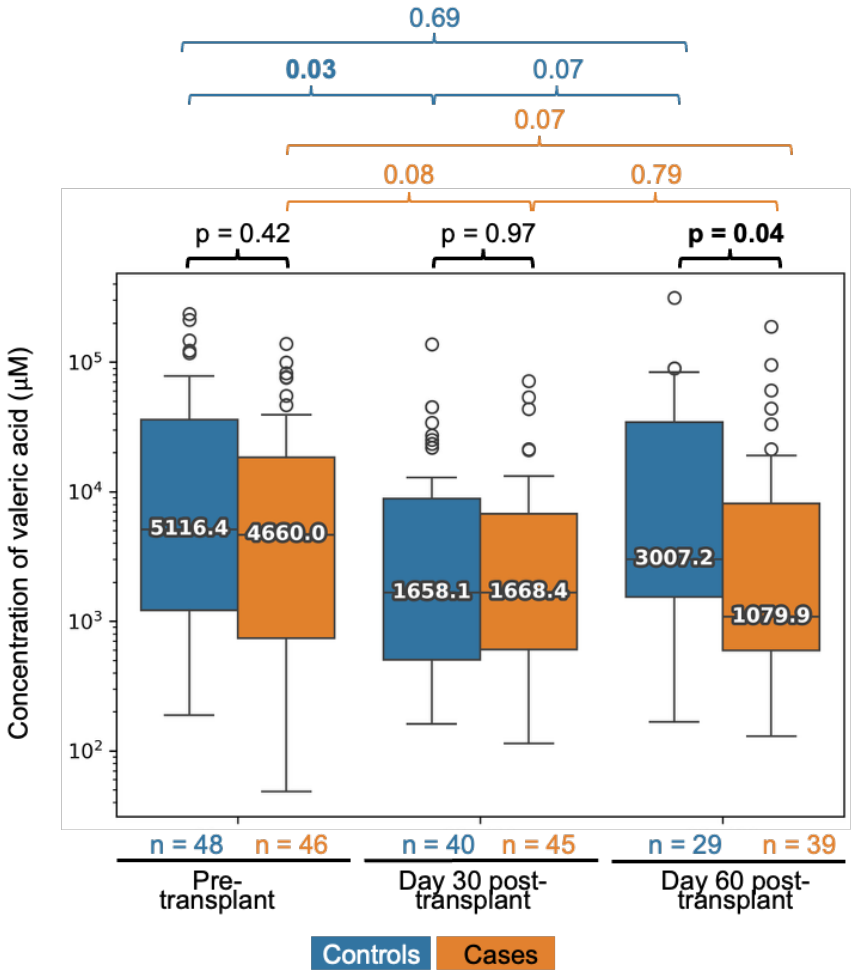


Figure 21. Valeric acid concentrations throughout HCT. At day 60 post-transplant (HCT,) there was greater levels of valeric acid in no GvHD controls vs. GvHD cases (p = 0.04). In no GvHD controls, valeric acid concentrations also decrease significantly from pre-HCT to day 30 post-HCT (p = 0.03). Concentrations are in millimolar (mM). P-values are generated from the Mann Whitney U test.

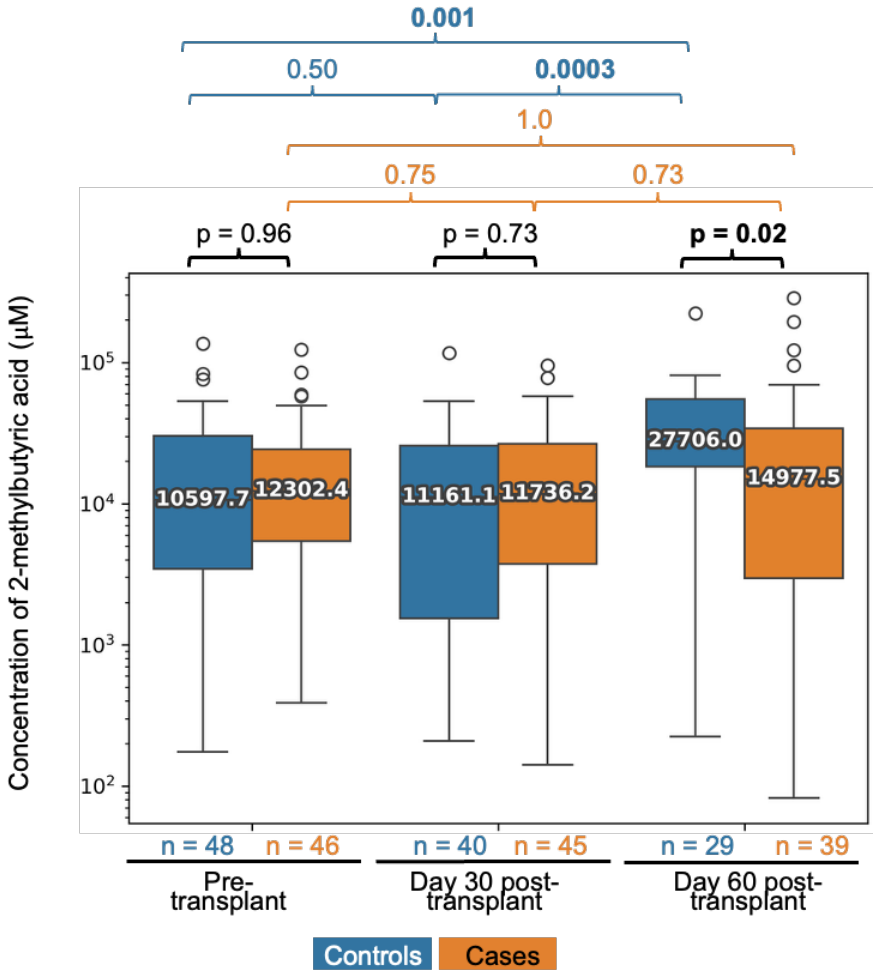


Figure 22. 2-methylbutyric acid concentrations throughout HCT. At day 60 post-transplant (HCT), no GvHD controls had higher concentrations of 2-methylbutyric acid vs. GvHD cases (p = 0.02). The day 60 no GvHD levels were also significantly higher thanday 30 post-HCT (p = 0.0003) and pre-HCT (p = 0.001). Concentrations are in millimolar (mM). P-values are generated from the Mann Whitney U test.

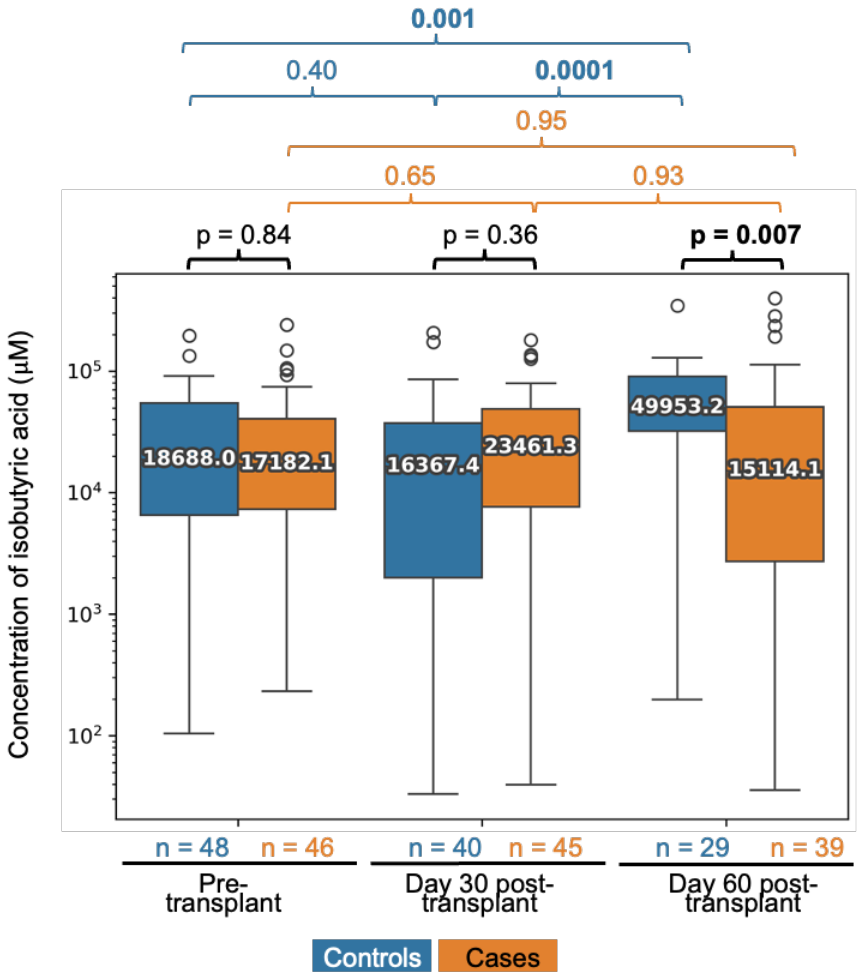


Figure 23. Isobutyric acid concentrations throughout HCT. At day 60 post-HCT, no GvHD controls had higher levels of isobutyric acid vs. GvHD cases (p = 0.007). The day 60 post-HCT levels were also significantly greater than the no GvHD controls at day 30 post-HCT (p = 0.0001) and pre-HCT (p = 0.001). Concentrations are in millimolar (mM). P-values are generated from the Mann Whitney U test.

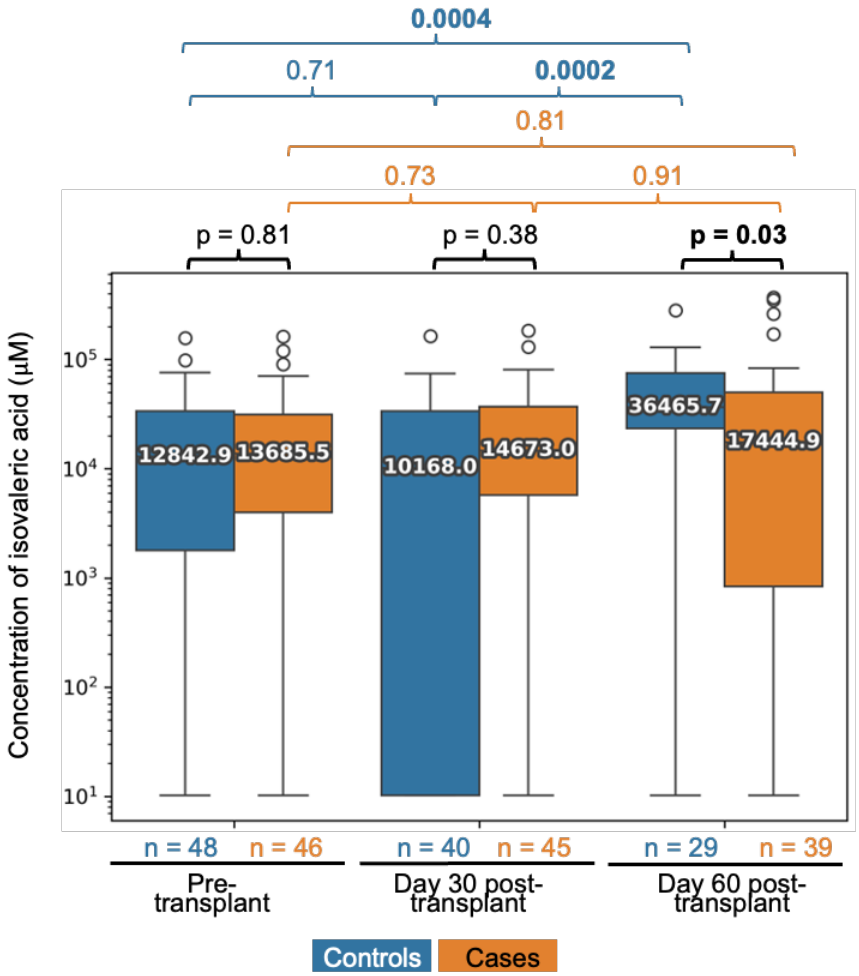


Figure 24. Isovaleric acid concentrations throughout HCT. On day 60 post-HCT, no GvHD controls had higher concentrations of isovaleric acid vs. GvHD cases (p = 0.03). The day 60 post-HCT level was also significantly higher than day 30 post-HCT (p = 0.0002) and pre-HCT (p = 0.0004). Concentrations are in millimolar (mM). P-values are generated from the Mann Whitney U test.

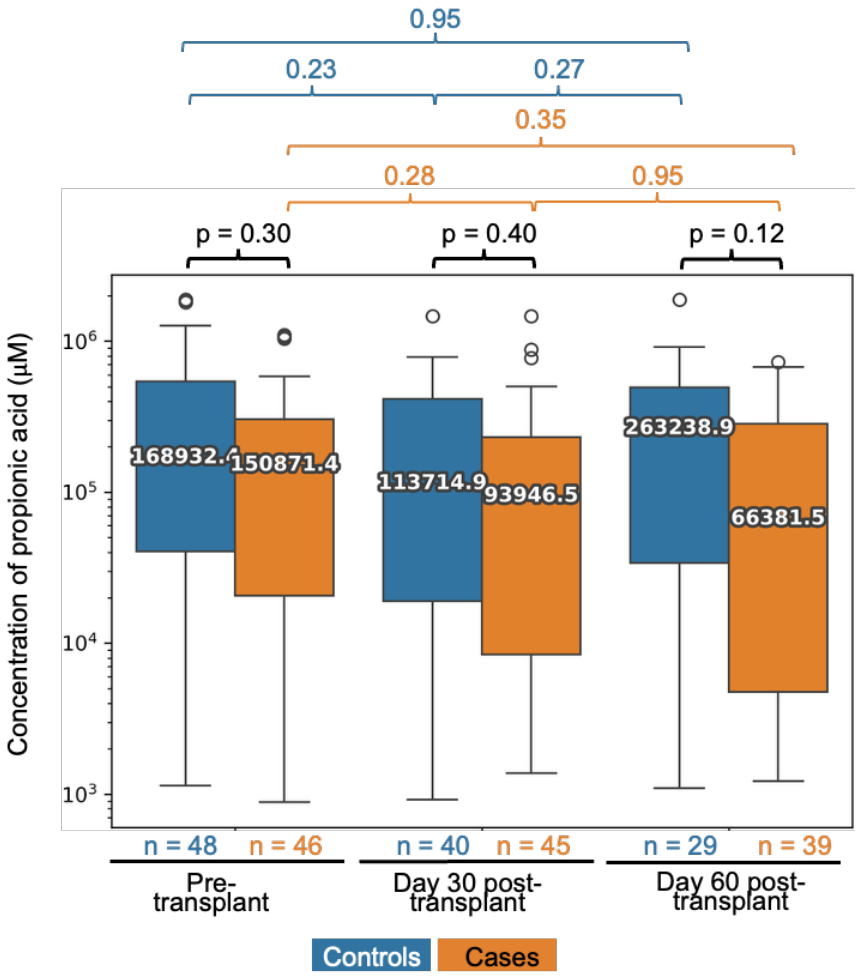


Figure 25. Propionic acid concentrations throughout HCT are similar (p-values > 0.05). Concentrations are in micromolar (μM). P-values are generated from the Mann Whitney U test.

4.2.1.2 Individual SCFAs by timepoint (within case or control group)

It was insightful to compare the delta or change within the cases or control groups longitudinally. Over time, the control concentrations of every SCFA, except propionic acid, statistically significantly changed at least once between timepoints (Figs. 20-25). This was consistent with the expected impacts from HCT. If the change from pre-HCT to day 30 post-HCT was statistically significant, then SCFA concentrations decreased by day 30 post-HCT due to the intensive nature of HCT. If the change from day 30 to day 60 post-HCT was statistically significant, SCFA concentrations increased by day 60 post-HCT reflecting rebound and recovery from HCT. Corresponding changes in appetite and diet could impact these fluctuations since diet varies considerably with impact/recovery from HCT. Overall, the control group concentrations were more dynamic in nature vs. the GvHD case group, supporting the notion

that these patients' SCFAs are responsive to changes in health and diet post-HCT. There were no statistically significant changes in the GvHD case groups' SCFA/BCFA concentrations between timepoints (Figs. 20-25). Notably, levels remained the same between days 30 and 60 post-HCT. Concentrations may stagnate over time since recovery from HCT is impeded by GvHD. It is currently unclear if stagnant SCFA concentrations during GvHD may reflect consistent poor health and diet due to the GvHD symptoms or if this may indicate a persistent gut microbiota state that cannot alter SCFA production in response to environmental changes (e.g. diet).

4.2.1.3 GvHD cases' individual SCFAs on day 30 post-HCT

Since GvHD is diagnosed based on clinical criteria, there can be considerable variability in timing of GvHD onset. When the day 30 post-HCT samples were collected, some patients had already been diagnosed with GvHD whereas some GvHD case participants would not be diagnosed until day 30+. Thus, the day 30 post-HCT samples can be further subdivided into representing pre- or post-GvHD diagnosis. In other words, we can normalize the day 30 post-HCT samples to each individual participant's GvHD status. Even with this normalization, no statistically significant differences in any SCFA/BCFA concentration were observed between pre-GvHD samples, post-GvHD samples, and no GvHD control samples.

4.2.1.4 SCFA correlations

Concentrations of three SCFAs were highly correlated. Isobutyric acid/2-methylbutyric acid, isovaleric acid/2-methylbutyric acid, and isovaleric acid/isobutyric acid had correlation coefficients ranging from 0.92-0.95. The next highest correlation coefficient was 0.75 for propionic and valeric acid, and the remaining correlation coefficients were less than 0.67. The underlying biological reason for these correlations remains unclear.

4.2.2 Diet and SCFAs

I ran individual comparisons to identify nine correlations that were statistically significant upon p-value adjustment. All nine of these correlations involved the SCFA, valeric acid, indicating this SCFA is closely tied to diet. The following significantly correlated with valeric acid: gluten, caffeine, betaine, total grains, refined grains and four saturated fatty acids (Table 16). Valeric acid is produced via a multi-step reaction with ethanol and propionate as reactants. Thus, it is also derived from microbial fermentation of protein and/or carbohydrate, since this produces propionate.[265,266] Some of these observed associations, namely total grains, refined grains, gluten and betaine, may associate with valeric acid because they are found in foods that are rich in carbohydrates and proteins that are only partially digestible by human enzymes. Gluten itself is an example of a protein that microbes can ferment into SCFAs like valeric

acid.[267] The foods/sources of gluten are also often sources of grains and betaine (e.g. wheat), hence why these dietary components also associate with valeric acid.

SCFA	Dietary component	p-value	Adjusted p-value	Rho (spearman)
Valeric acid	Gluten	9.63E-5	0.04	0.38
	Caffeine	7.93E-5	0.04	0.39
	Betaine	5.82E-5	0.04	0.39
	Total grains	4.66E-5	0.04	0.40
	sfa120	0.0002	0.04	0.37
	sfa40	0.0002	0.04	0.37
	Refined grains	0.0003	0.05	0.36
	sfa80	0.0003	0.05	0.36
	sfa100	0.0003	0.05	0.36

Table 16. Correlations between valeric acid and individual dietary components. Since there were significant associations between dietary and SCFA principal components, individual correlations were calculated. Valeric acid was a top hit, with the largest quantity of strong and statistically significant associations to individual dietary components. Spearman correlation was used to generate the rho values and p-values were adjusted for false discovery rate (FDR).

4.2.2.1 Fiber

I helped an undergraduate summer intern examine differences in diet between allo-HCT recipients with vs. without GvHD in both the pre- and post-HCT periods. I was most interested in fiber intake since SCFAs are a byproduct of microbial fiber fermentation. I found that HCT-recipients without GvHD had significantly more fiber in their diet vs. recipients with GvHD pre-HCT and at days 30 and 60 post-HCT (Fig. 26). Dietary fiber was also quantified weekly for nine weeks and similar differences were observed; allo-HCT recipients without GvHD consumed statistically significantly more dietary fiber vs. recipients with GvHD for six of the weeks measured: week -2 (where the week of transplant is week 0), one, three, four, five, and six. During weeks two, seven, and eight there were similar levels of dietary fiber across all participants regardless of GvHD status (Fig. 26).

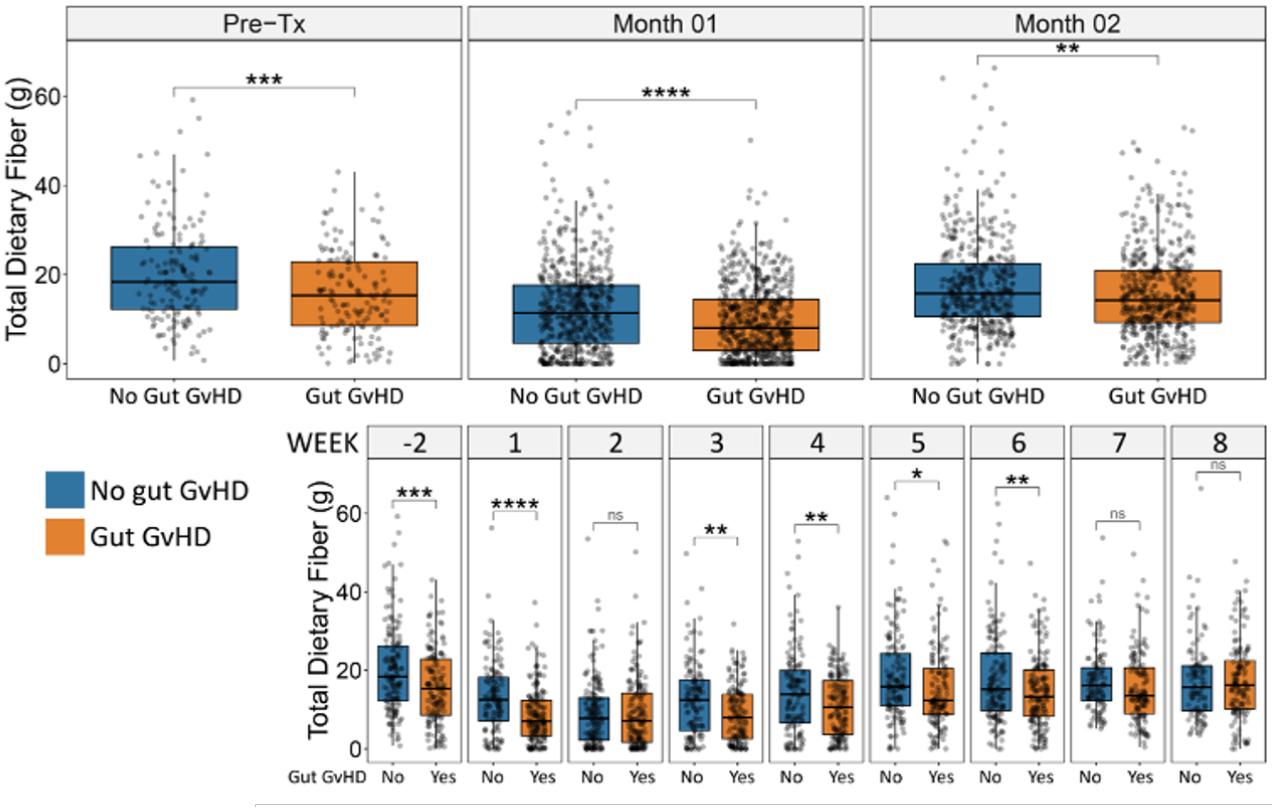


Figure 26. GvHD is associated with fiber deficiency throughout HCT. More fiber was ingested by no gut GvHD controls pre-HCT, day 30 post-HCT/Month 01, and day 60 post-HCT/Month 02. When ingested fiber is parsed out weekly, the same trend is observed (more fiber in no gut GvHD controls) most weeks (week -2, 1, 3, 4, 5 and 6). On weeks 2, 7 and 8, there was no significant difference in ingested fiber between GvHD cases vs. no GvHD controls. Statistical significance was assessed by Mann Whitney U test. *: $p \leq 0.01$, **: $p = 0.001$, ***: $p = 0.0001$, ****: $p < 0.0001$. The 169 patients included in this figure/analysis included all participants in the larger (> 300 participant; Fig. 2) GvHD study for which dietary data and GvHD status were available.

4.3 Discussion

The use of fecal levels of SCFAs to infer systemic or intraintestinal levels is imperfect. It is generally accepted that SCFA measurement in stool does not reflect SCFA production or absorption, but rather a net product of these processes, at best.[268] It has been estimated that just 5-10% of the SCFA pool is excreted in stool.[269,270] Some studies posit that at least the molar ratio of acetate, propionate and butyrate are consistent throughout the large intestine and in fecal samples.[270,271] Additionally, the overall levels of SCFAs in stool may not differ in magnitude of SCFA levels in the colon.[271] Discrepancies between colonic and fecal SCFA levels may be due to intestinal absorption or loss of fecal SCFAs due to their volatility. This study's targeted SCFA LC-MS metabolomic analysis used isotope-labeled derivatization to help stabilize the

volatile SCFAs, facilitating their detection and quantification. Despite these flaws, stool concentrations of SCFAs are a common output measured to assay intestinal, specifically colonic, SCFA concentrations.[269,272]

4.3.1 GvHD and fiber

Fiber, GvHD and the gut microbiota are closely interrelated. Significantly more fiber was consumed in participants without GvHD pre-transplant and at days 30 and 60 post-transplant (Fig. 26). This is concordant with significantly more butyric acid in participants without GvHD vs. with GvHD pre-HCT (Fig. 20). However, the high fiber intake of participants without GvHD at these three timepoints is discordant with statistically similar levels of propionic acid across all participants since these SCFAs are also products of microbial fermentation of fiber (Fig. 25). Differences in the gut microbiota may contribute to these discordant findings, such as differences in producers of butyrate vs. propionate. At week two, the nonsignificant difference in dietary fiber consumption between GvHD and no GvHD allo-HCT recipients may be at least partially explained by all participants experiencing similar symptoms of mucositis and corresponding discomfort, limiting the amount of food/calories that can be tolerated by all recipients (Fig. 26).

Impacts/symptoms of GvHD modulate diet, and dietary fiber promotes a microbial community capable of fermenting fiber, resulting in production of beneficial SCFAs like acetic acid. Preliminary data in the lab points to SCFA-producing microbes, which appear to link with amount of dietary fiber ingested by HCT recipients. Although acetic acid itself is a SCFA which may exert beneficial effects on intestinal health, it is proposed that acetic acid-producers (e.g. *Bifidobacterium adolescentis)* and butyrogenic bacteria cross-feed such that acetic acid promotes microbes such as *Faecalibacterium prausnitzii* which enhances butyrate production.[249,273] In this study, this implies that participants with high dietary fiber and greater levels of acetic acid-producing microbial species could promote growth of SCFA-producers (e.g. *Blautia, Roseburia*) in the gut microbiome, resulting in enhanced SCFA levels.

4.3.2 Pre-transplant butyrate links with protection against GvHD after allo-HCT

The finding that low butyric acid levels pre-transplant associate with GvHD (Fig. 20) is consistent with current literature supporting a protective effect of butyric acid in the gastrointestinal tract. This finding is striking because it suggests butyric acid concentrations pre-HCT could predict GvHD status post-HCT. Further study is required to verify this finding. If replicated in a larger study, then the pre-HCT median (or another level, such as the 25% or 75% quartiles) of the GvHD case groups' levels could be implemented as a cutoff for consideration

as ‘low’ butyric acid pre-HCT. HCT recipients with low butyric acid levels could trigger a specific therapeutic or closer monitoring to attempt GvHD prevention. For instance, butyric acid could be administered pre-HCT to toggle butyric acid’s protective effect, or preemptive steroid therapy to head off GvHD, or Treg-enriched/T-cell depleted grafts considered to reduce the chance of GvHD. Low pre-HCT butyric acid levels’ link with GvHD also suggests that pre-HCT butyric acid levels may be a mechanism for modulating GvHD status post-HCT. Butyric acid concentrations in humans pre-HCT should be examined in an even larger group of HCT recipients to see if the finding is replicated. Experiments incorporating murine models of HCT and GvHD should also be performed to determine if increasing butyric acid levels pre-HCT protects against GvHD and promotes better outcomes post-HCT. If so, studies should focus on the best avenues for therapeutic delivery of butyric acid to the gastrointestinal tract. One possible candidate is tributyrin since it is a butyrate precursor, can reach the colon, and is well-studied.[274] However, as outlined in several publications, multiple GvHD biomarkers may best stratify HCT recipients.[53–56] In the recommended future studies, stool butyric acid’s predictive ability should be considered in addition to other evidence-supported biomarkers, such as fecal α1 antitrypsin or serum ST2 and REG3α.[53–56] Consideration should also be given to the FDA-recommended steps to ensure a biomarker can be used reliably in the clinic.[57]

4.3.3 Propionate does not link with GvHD after allo-HCT

Propionate was not significantly enriched in GvHD cases or no GvHD controls at any point throughout HCT (Fig. 25). This was unexpected because pre-clinical studies supported that propionate had a measurable protective effect against chronic and acute GvHD as well as other diseases due to its production and/or activity within the gastrointestinal tract.[47,94,275] This included human data characterizing a link between low circulating propionate levels in serum samples from HCT recipients with chronic GvHD.[47] Additionally, I hypothesized that even if there are critical mechanistic differences between chronic vs. acute GvHD, there was *in vivo* evidence that propionate could bind receptor G-protein coupled receptor 43 (GPR43) specifically on non-hematopoietic cells to trigger NLRP3 inflammasome activation (via ERK phosphorylation) to ultimately mitigate acute GvHD in a murine model.[94] Even though further research found that propionate can modulate the gut microbiome to reduce an additional disease, vascular calcification[275], I could not replicate these findings. There was no observed link between stool levels of propionate and acute GvHD in our case-control human cohort. One reason may be due to a relatively smaller sample size compared to other human GvHD case-control studies.[44,67,71,77]

Chapter 5: Untargeted/global metabolomics and GvHD

5.1 Introduction

5.1.1 Research value of untargeted metabolomics

The terms ‘untargeted metabolomics’ and ‘global metabolomics’ are used interchangeably within this thesis. This approach takes advantage of every molecule or metabolite’s unique mass and charge upon ionization and is described as a ratio of mass to charge, called a feature. The mass and charge are calculated by analyzing complex samples by liquid chromatography and then mass spectrometry. Metabolites can become positively or negatively charged when ionized, so two modes are used to capture positively ionized compounds and negatively ionized compounds. Both ionization modes are needed for global LC-MS to capture molecules that may exclusively positively or negatively charge when ionized.

A weakness of untargeted metabolomics is that it is only semi-quantitative; this type of metabolomics produces relative concentrations (as opposed to absolute concentrations). Thus, concentrations of a given feature cannot be compared across studies, limiting what can be interpreted from a single dataset. Normalization can help relate global metabolomics data back to the environment or context from which it was generated/derived. However, it is still challenging to directly compare global metabolomic data generated under different conditions, or with targeted metabolomic data. The processing of each data type is too distinct to rationally compare targeted vs. untargeted data, even if both are generated from the same sample(s). Despite these considerations, global metabolomics uniquely permits the discovery of previously unknown compounds that may associate with a disease like GvHD. This includes features that have not been identified or annotated before or that have not been previously associated with a disease state. This ability to study the metabolites in a sample in detail without any prior knowledge of the samples’ metabolites is a powerful discovery tool. For instance, such an untargeted approach facilitated the characterization of microbially-conjugated bile acids (MCBAs).[276] It was previously unknown that MCBAs existed, and the ability of the bile salt hydrolase enzymes to produce MCBAs was unknown.

5.1.2 Microbially-conjugated bile acids (MCBAs)

In early 2024, two publications highlighted an emerging research direction into the conjugation of bile acids with a variety of compounds beyond glycine and taurine, mediated by bile salt hydrolase enzymes.[126,127] It was previously not appreciated that bile salt hydrolases could reconjugate bile acids as well as deconjugate them. (Although as early as 2020 the novel bile acid amidates were characterized without full characterization of the pathway that produced

them.[277]) Further, this reconjugation could be performed on any amino acid, not just the taurine and glycine conjugation that was previously characterized.[126] This led to the discovery that a wide variety of compounds could be conjugated to bile acids, spanning amino acids, polyamines, short-chain fatty acids, at least one neurotransmitter, and more.[127] In the condition cystic fibrosis, diverse bile acid amidates (MCBAs) were noted to be differentially abundant based on disease status.[127,277] Since we observed differences in secondary bile acids based on GvHD status, we were interested in investigating if MCBAs were also associated with GvHD status.

5.2 Results

5.2.1 Summary description of the data

The global metabolomics data yielded 3,963 features in the positive ionization dataset and 3,123 features in the negative ionization set (total: 7,086; Table 17). Over 92% of the stool samples (226 out of 246 samples) called at least 2,000 features in the negative ionization data and at least 2,500 features in the positive ionization data. Only 2-3% of the features called by each ionization MS matched to an entry in the MassBank of North America (MoNA) database. I visualized the spread of the GvHD case vs. control data at each timepoint with a volcano plot of log-base 2 transformed fold change on the X axis vs. log-base 10 transformed p-value on the Y axis. A total of 1,100 features were statistically significantly differentially abundant between GvHD cases and controls across the positive and negative ionization data (Table 17; without any fold change filter applied). In total, 13.2% (n = 524) of the positive ionization data's features were statistically significantly different between GvHD cases vs. controls at a timepoint throughout HCT. For the negative ionization dataset, 18.4% of features (n = 576) were statistically significantly different. Overall, 15.5% of the 7,086 features (n = 1,100) were significantly different between GvHD cases vs. controls. The number of statistically significantly differentially abundant features increased from pre-HCT to day 30 post-HCT to day 60 post-HCT. However, several of these 1,100 targets were eliminated for further study by also applying a fold change filter: greater than two (enriched in cases) or less than 0.5 (enriched in controls). Using both filters, pre-HCT, 11 features were statistically significantly enriched in the positive ionization data and 16 in the negative ionization data (Fig. 27A-B). This figure more than tripled in the positive ionization data to 52 features by day 30 post-HCT, and more than quadrupled to 90 features in the negative ionization data (Fig. 28A-B). By day 60 post-HCT, 248 and 275 features were significantly enriched in the positive and negative ionization data, respectively (Fig. 29A-B). Overall, this sums up to 311 significantly enriched features in the positive data and

381 significantly enriched in the negative data (total: 692). The precise breakdown of significantly different features ($p < 0.05$) without enrichment ($2 \geq FC \geq 0.5$) and vice versa, overall and broken down by timepoint are summarized in Table 17.

Ionization mode	Timepoint	GvHD status	Significant by FC? (< 0.5 or > 2)	Significant by statistics? (p < 0.05)	Total (count)	% of the ionization mode total	% of the total (7,086)
All	All	All	Y	Y	692	NA	9.77
All	All	All	NA	NA	7086	NA	100.00
Negative	All	All	NA	NA	3123	100.00	44.07
Negative	All	All	Y	Y	381	12.20	5.38
Negative	Day 30 post-HCT	Case	Y	Y	46	1.47	0.65
Negative	Day 30 post-HCT	Control	Y	Y	44	1.41	0.62
Negative	Day 30 post-HCT	All	Y	Y	90	2.88	1.27
Negative	Day 60 post-HCT	Case	Y	Y	220	7.04	3.10
Negative	Day 60 post-HCT	Control	Y	Y	55	1.76	0.78
Negative	Day 60 post-HCT	All	Y	Y	275	8.81	3.88
Negative	Pre-HCT	Case	Y	Y	5	0.16	0.07
Negative	Pre-HCT	Control	Y	Y	11	0.35	0.16
Negative	Pre-HCT	All	Y	Y	16	0.51	0.23
Positive	All	All	NA	NA	3963	100.00	55.93
Positive	All	All	Y	Y	311	7.85	4.39
Positive	Day 30 post-HCT	Case	Y	Y	36	0.91	0.51
Positive	Day 30 post-HCT	Control	Y	Y	16	0.40	0.23
Positive	Day 30 post-HCT	All	Y	Y	52	1.31	0.73
Positive	Day 60 post-HCT	Case	Y	Y	81	2.04	1.14
Positive	Day 60 post-HCT	Control	Y	Y	167	4.21	2.36
Positive	Day 60 post-HCT	All	Y	Y	248	6.26	3.50
Positive	Pre-HCT	Case	Y	Y	7	0.18	0.10
Positive	Pre-HCT	Control	Y	Y	4	0.10	0.06
Positive	Pre-HCT	All	Y	Y	11	0.28	0.16

Table 17. Summary of statistically significantly enriched global features by GvHD status. Thousands of global features were detected and hundred were enriched in GvHD cases or no GVHD controls at each timepoint. FC = fold change, HCT = hematopoietic cell transplant

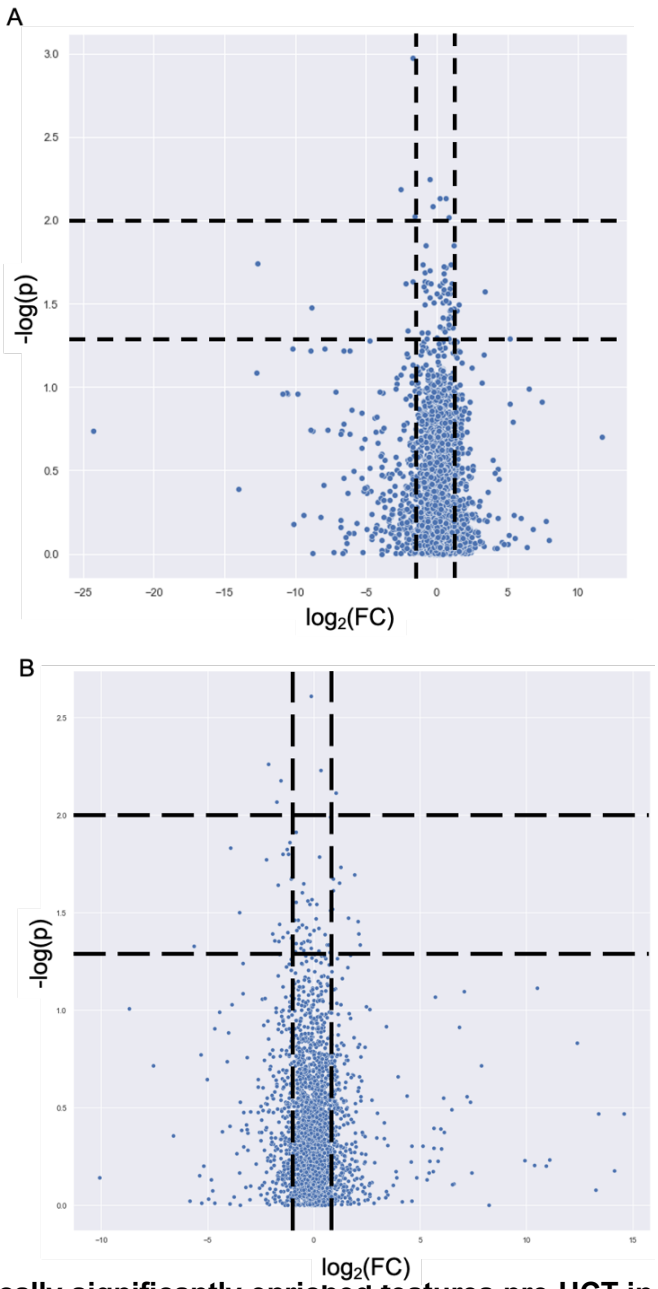


Figure 27. Statistically significantly enriched features pre-HCT in (A) positive and (B) negative ionization data. The inverse of the log transformed p-value is graphed on the y-axis and the log base 2 transformed fold change (FC) is on the x-axis. The FC represents the case median over the control median, therefore, the $\log_2$(FC) values < 0 are enriched in controls vs. FC values > 0 are enriched in cases. FC cutoffs were used to focus on features with a greater magnitude of enrichment. An FC < 0.5 reflects features present at at least twice the concentration in controls vs. cases. An FC > 2 reflects features present at at least twice the concentration in cases vs. controls. The $\log_2(0.5) = -1$ and $\log_2(2) = 1$. These $\log_2$(FC) cutoffs are shown on the by vertical dashed lines. Statistically significantly differentially abundant features are highlighted by horizontal dashed lines at –log(p) = 2 and 1.3, reflecting p = 0.01 and 0.05, respectively. The Mann Whitney U test was used to generate p-values. FC = fold change

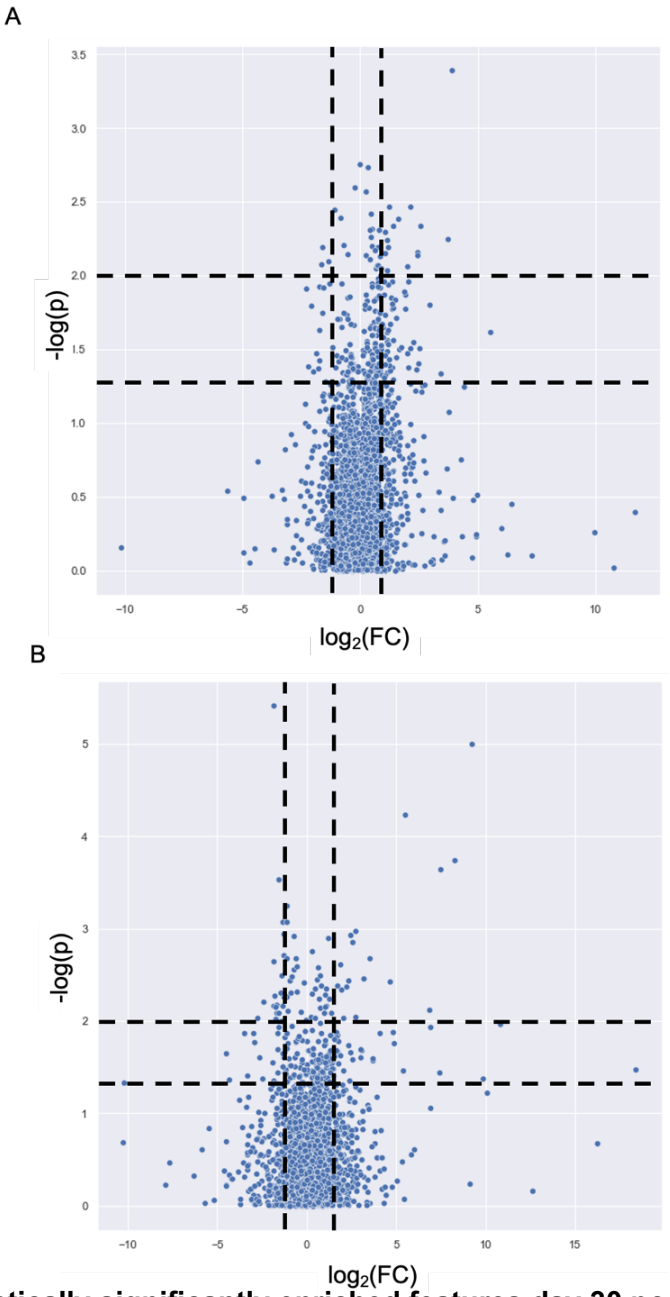


Figure 28. Statistically significantly enriched features day 30 post-HCT in (A) positive and (B) negative ionization data. The inverse of the log transformed p-value is graphed on the y-axis and the log base 2 transformed fold change (FC) is on the x-axis. The FC represents the case median over the control median, therefore, the $\log_2$(FC) values < 0 are enriched in controls vs. FC values > 0 are enriched in cases. FC cutoffs were used to focus on features with a greater magnitude of enrichment. An FC < 0.5 reflects features present at at least twice the concentration in controls vs. cases. An FC > 2 reflects features present at at least twice the concentration in cases vs. controls. The $\log_2(0.5) = -1$ and $\log_2(2) = 1$. These $\log_2$(FC) cutoffs are shown on the by vertical dashed lines. Statistically significantly differentially abundant features are highlighted by horizontal dashed lines at –log(p) = 2 and 1.3, reflecting p = 0.01 and 0.05, respectively. The Mann Whitney U test was used to generate p-values. FC = fold change

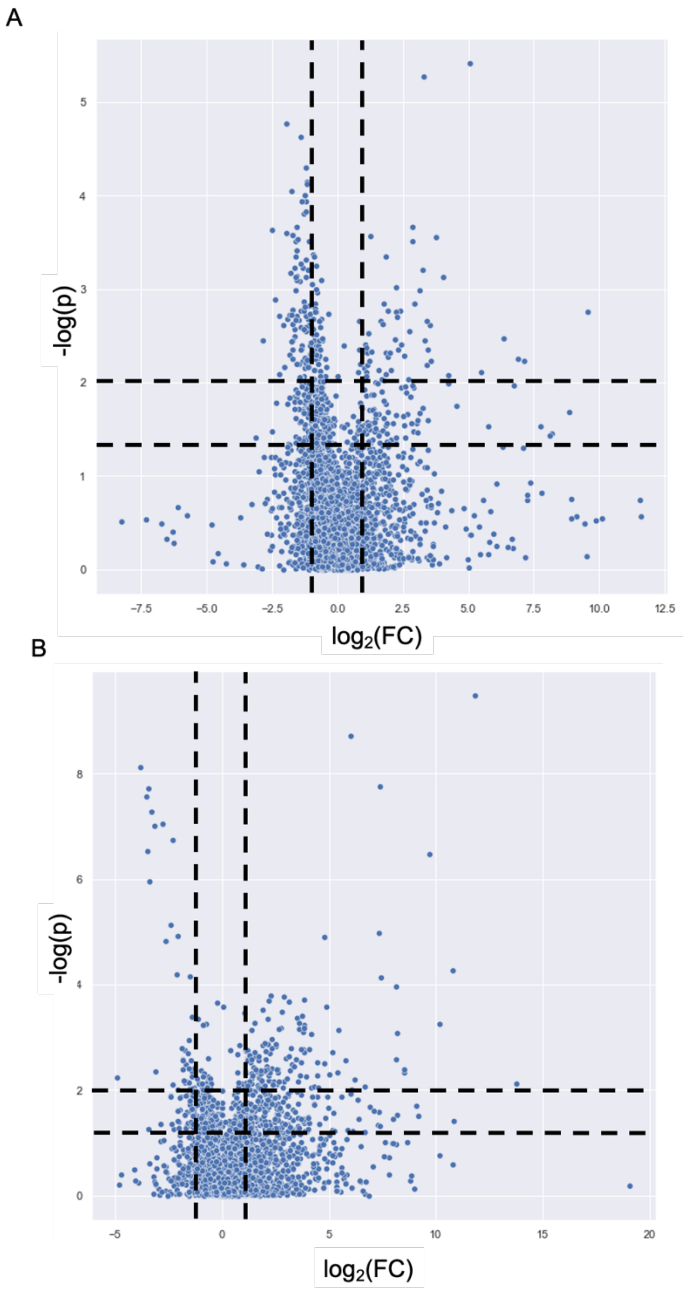


Figure 29. Statistically significantly enriched features day 60 post-HCT in (A) positive and (B) negative ionization data. The inverse of the log transformed p-value is graphed on the y-axis and the log base 2 transformed fold change (FC) is on the x-axis. The FC represents the case median over the control median, therefore, the $\log_2$(FC) values < 0 are enriched in controls vs. FC values > 0 are enriched in cases. FC cutoffs were used to focus on features with a greater magnitude of enrichment. An FC < 0.5 reflects features present at at least twice the concentration in controls vs. cases. An FC > 2 reflects features present at at least twice the concentration in cases vs. controls. The $\log_2(0.5) = -1$ and $\log_2(2) = 1$. These $\log_2$(FC) cutoffs are shown on the by vertical dashed lines. Statistically significantly differentially abundant features are highlighted by horizontal dashed lines at –log(p) = 2 and 1.3, reflecting p = 0.01 and 0.05, respectively. The Mann Whitney U test was used to generate p-values. FC = fold change

Notably, each feature may not necessarily correspond to one single compound; there is redundancy within the features such that several features may correspond to one compound (in different ionization/charge states). Additionally, given the raw number of target features, there is relatively high likelihood for false positives. For instance, some of the 692 features may only appear significantly enriched but are not truly significant. Thus, extra caution should be used when interpreting these data; the findings described in this report are not a final product and require validation in further studies.

5.2.1.1 Zero values within data

5.2.1.1.1 Quantity of nonzero data values

All but one sample each in the negative and positive ionization data hit at least 700 and 1,300 features, respectively. The majority of samples hit 2,000 and 2,500 or greater features per sample, respectively. There was no association between GvHD status and number of features called. Overall, 3,123 unique features were hit in the negative ionization data and 3,963 unique features in the positive ionization data.

5.2.1.1.2 Treatment of zeroes

Global LC-MS produces precise data. For data analysis and interpretation, when the level of a feature is denoted as zero, this was treated as absence of the feature (although true concentration may be below the limit of detection). Thus, it is possible to compare the prevalence of features. I compared the prevalence of zero values reported for each feature in GvHD case vs. control samples, but found no statistically significant differences.

However, to perform the mathematical and statistical analysis for biological interpretation of the untargeted metabolomic data, the zeroes must be addressed because relevant formulas and transformations are incompatible with zero values. There are several options accepted within the field for treatment of such zero values to proceed with critical analysis. The spread/breadth, biological application, and sparsity of the entire dataset are examples of important considerations to weigh when deciding the most accurate way to treat zero values.

5.2.1.1.2.1 Remove NaNs

One option is to remove the zero values from all downstream analyses. This is risky for this project because zero values may represent true absence of a feature, which is informative. For instance, the zero values are necessary to perform prevalence analyses on clinical criteria such as GvHD status or timepoint with respect to HCT.

5.2.1.1.2.2 Impute with one fifth of the minimum value

Another option is to replace/impute the zero values with a relatively small value that permits formulas and transformations of the data, but is so small that the impact of the imputed value on

non-zero values is minimized. For instance, it is common in the field to use one fifth of the minimum detected value within each feature. This makes it such that the imputed value is small relative to each sample or dataset. This method was employed for downstream analyses to balance accurately including the zero values without artificially inflating concentrations or calculations.

5.2.1.2 Normalization strategies

Normalization describes a transformation used to convert datasets with different scales to datasets with a common scale while maintaining the integrity and accuracy of the dataset and what it describes.

5.2.1.2.1 Non-normalized

All data does not need to be normalized, but it can be helpful to better understand results. One example where normalization may not be necessary is if two features that are known to be present at similar concentrations within the sample of interest are being compared. However, it is frequently unknown what compounds correspond to each feature detected, let alone the true physiological concentration range of multiple features in a sample. Thus, normalization can make untargeted metabolomics data more interpretable.

5.2.1.2.2 Quantile normalized

Quantile normalization makes the distribution of different datasets identical. One global metabolomics dataset is generated per sample. Each dataset is arranged in ascending or descending order with the operative component being the ranking of individual data values. The mean of each individual rank across datasets/samples is then calculated. These ranked means replace the corresponding detected values within each rank. The exact feature's concentration (within a sample/dataset) that each individual data value represents must be maintained/preserved throughout the quantile normalization so that the true relative concentration of the feature present at (for example) the highest concentration is still known after transforming the data. This method was performed on the data to facilitate comparisons without losing integrity of the data.

5.2.2 Metaboanalyst functional analysis

MetaboAnalyst (https://www.metaboanalyst.ca/home.xhtml) is a web-based platform for performing a variety of analyses on metabolomic data.[278] It is also available as an R package, although the analysis of the data discussed here was performed on the online platform. Metabolomic data may be evaluated using MetaboAnalyst at nearly any stage. On the earlier end, the user can upload LC-MS spectra for processing, whereas on the later end, a preprocessed metabolite list can be uploaded for a causal analysis pulling in metabolites and

disease outcomes from published metabolite-genome-wide association studies for causal analysis. A preprocessed list of features with relative concentrations may also be uploaded for a pathway analysis to determine which functional pathways may be enriched, drawing on mummichog (series of computational algorithms that predict functional role of metabolomics data[279]) or gene set enrichment analysis annotations. I performed this functional pathway analysis on the negative ionization untargeted metabolomic data and observed the following varied with GvHD status at 60 days post-HCT (in descending order of statistical strength and enrichment): vitamin B metabolism, ascorbate (vitamin C) and aldarate metabolism, and C5-branched dibasic acid metabolism (Fig. 30). It remains unclear how these pathways may be involved in GvHD. Additionally, bile acid biosynthesis was hit but was not statistically significant and only weakly enriched.[280]

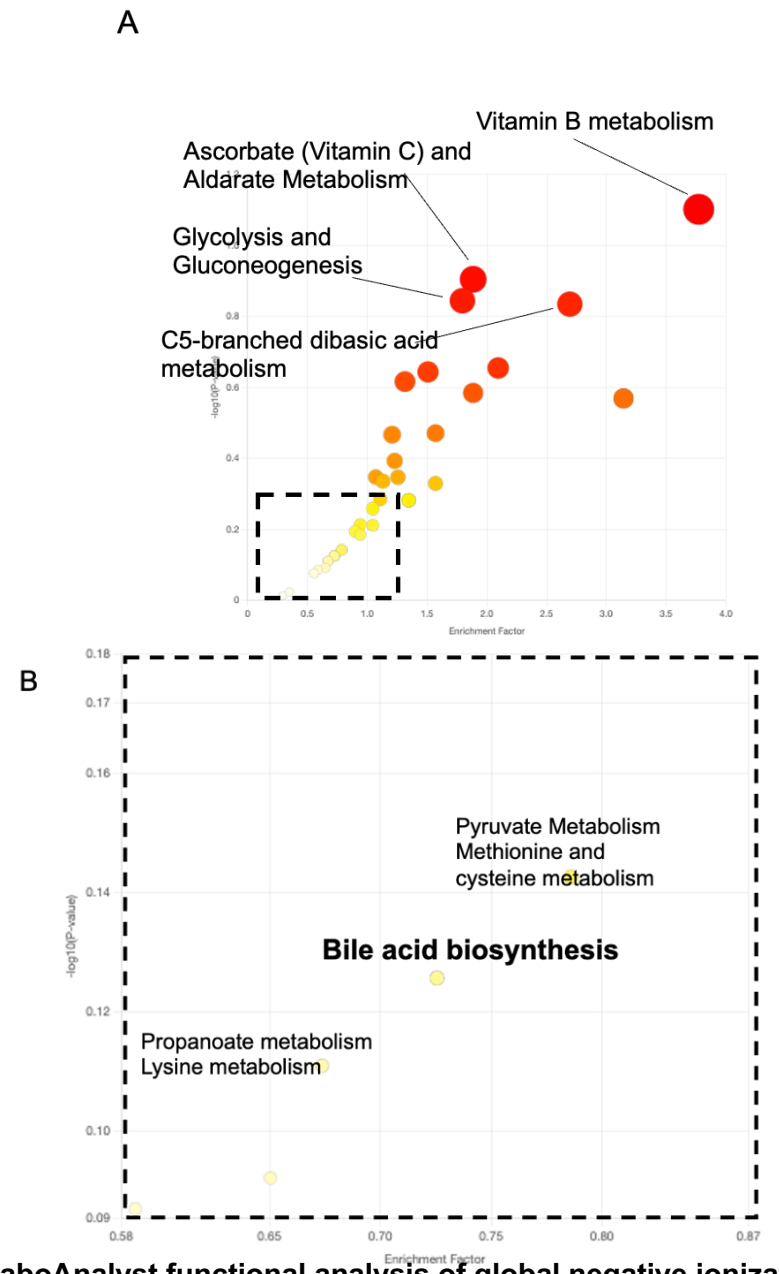


Figure 30. MetaboAnalyst functional analysis of global negative ionization data. The x-axis indicates the level of enrichment via an enrichment factor. The y-axis represents the -10 log transformed p-value. This revealed that the following pathways may be differentially abundant by GvHD cases vs. no GvHD controls, in descending order of enrichment factor and p-value: vitamin B metabolism, ascorbate (vitamin C) and aldarate metabolism, glycolysis and glucogenesis, C5- branched dibasic acid metabolism, pyruvate metabolism, methionine and cysteine metabolism, bile acid biosynthesis and propanoate metabolism/lysine metabolism. The bile acid biosynthesis pathway is concordant with the statistically significant findings observed in the targeted bile acid metabolomics. Panel A shows several of the most statistically significantly enriched pathways, and the dashed square inset is panel B, representing a zoomed-in version of the dashed square inset on panel A. The axes are the same on both panels although the scale varies. Fisher's exact test without adjustment is used as part of the MetaboAnalyst pipeline.

5.2.3 MCBAs and GvHD

5.2.3.1 Process/protocol for targeting MCBAs

After finding interesting features through a combination of statistical analysis and biological contextualization, the next question was: what compound do these features of interest represent? Although the features' chemical names were unknown, the retention time and mass to charge ratio (m/z) was reported in the global metabolomics results. A consultation with Dr. Wentao Zhu yielded a suggested approach for labeling level five feature identifications.[281] According to the metabolomics field, it is best to indicate the confidence of feature identification/annotation through a defined level system. Level five denotes relatively low confidence annotation by matching masses of the feature and suspected metabolite.[281,282] This entailed accessing the public database, MassBank of North America (MoNA), which is composed of spectral records libraries and uploaded by the metabolomics/MoNA-user community. Separate databases are maintained for the positive ionization vs. negative ionization data. Additionally, this approach was developed around having a list of MCBA standards commercially produced by the company BileOmix along with their m/z and neutral mass.

Beginning with the m/z of the feature of interest from the experimental global metabolomics data, I looked in the appropriate ionization database for an m/z that matched the feature's m/z to the hundredth decimal place. The degree of match varied based on the exact equipment and standards used, but for this experiment, the hundredth decimal place best represented our group's confidence and variability. There were often multiple matching m/z's, which was expected. From any MoNA matches, I pulled the m/z, retention time, compound ID, chemical formula, and listed common adduct formed under LC-MS conditions. I calculated the molecular mass of the listed adduct and performed the specified math of the adduct on the experimental feature of interest to generate the exact or true or neutral mass (M). For example, if the adduct was M+H, then a molecular mass of 1 was added vs. M-H denoted a molecular mass of 1 should be subtracted. I consulted the list of MCBA standards to find any MCBA(s) with a neutral mass that matched a feature of interest's calculated neutral mass. It was expected and acceptable that multiple MCBAs often matched to a calculated feature of interest. For any MCBA standards that matched, I also pulled the m/z, retention time, and ID. For any feature of interest and MCBA standard match, the delta or percent error between each respective m/z was calculated. The formula for percent error in this experiment was dividing the difference between the feature of interest's neutral mass and the MCBA standard's neutral mass by the MCBA standard's neutral mass, and multiplying by 1 million. The final multiplication by a factor of 1

million was necessary because the units of the feature of interest's neutral mass is parts per million (ppm). The absolute value of the percent error or delta had to be less than 100 to accept the match as possible. This represented a 10% error cutoff. Finally, the retention times of the feature of interest and MCBA standard should be similar to increase confidence in a match/annotation.

5.2.3.2 Candidate MCBAs (level five confidence)

Using the above outlined approach, 19 candidate MCBAs were annotated with level five confidence (Table 18). Further studies running the MCBA standards and the experimental sample(s) again are needed to confirm the match or increase the confidence in the annotation (see Chapter 6.4.1: Investigate candidate secondary MCBAs). The possible microbially-conjugated lithocholic acid amidates are the MCBAs of greatest interest because of the observed association between lithocholic acid deficiency and GvHD (see Chapter 3.2 Results). The best way to investigate further is by developing a targeted LC-MS panel using standard microbially conjugated lithocholic acid compounds, available for purchase (for research) from BileOmix.

Candidate MCBA (level five confidence)	Candidate feature	Ionization mode in which feature was captured
Serine-cholic acid	5.52_476.3015m/z	Negative
Alanine-cholic acid	5.64_478.3174m/z	Negative
Glycine-chenodeoxycholic acid 7-sulfate	2.74_510.2533m/z	Negative
Glycine-chenodeoxycholic acid 7-sulfate	5.44_510.2532m/z	Negative
Glutamine-chenodeoxycholic acid	6.24_519.3441m/z	Negative
Taurine-cholic acid 3-sulfate	8.89_594.2489n	Negative
Glycine-cholic acid 3-glucuronide	7.80_624.3391m/z	Negative
5?-cholinic acid-3? 12?-diol-7-one	2.24_407.2794m/z	Positive
Ursocholic acid	2.79_426.3215m/z	Positive
Cysteine-cholic acid	1.39_512.3197n	Positive
Arginine-chenodeoxycholic acid	6.64_549.4010m/z	Positive
Glycine-chenodeoxycholic acid 3-glucuronide	6.62_626.3548m/z	Positive
Glycine-chenodeoxycholic acid 3-glucuronide	7.77_648.3352m/z	Positive
Arginine-lithocholic acid	1.17_550.4969n	Positive
ASN-DCA, PRO-CDCA, PRO-DCA, PRO-UDCA, VAL-OLCA, ILE-LCA, LEU-LCA	1.36_507.3156m/z	Negative
TRP-CDCA, TRP-DCA	4.72_577.3748m/z	Negative
TRP-OLCA	8.57_594.3247m/z	Positive
TRYP-CA	6.06_568.3460n	Positive
TRYPT-CDCA, TRYPT-DCA, TRYPT-UDCA	2.41_552.3006m/z	Negative

Table 18. Candidate MCBAs annotated with level five confidence within global data. Using the approach outlined in Ch 5.2.3.1, 19 features were identified as possible MCBAs across the positive and negative ionization data. Multiple possible MCBAs can correspond to one feature. The confidence level is using the system defined by Schymanski et al. 2014.[255] ASN = aspirigine, PRO = proline, DCA = deoxycholic acid, CDCA = chenodeoxycholic acid, UDCA = ursodeoxycholic acid, VAL = valine, OLCA = 7-oxylithocholic acid, LCA = lithocholic acid, ILE = isoleucine, LEU = leucine, TRP = tryptophan, TRYP = tryptamine and CA = cholic acid.

5.3 Discussion

5.3.1 Comparison/correlation of global feature bile acids to targeted bile acids

Since we also had absolute concentrations from the targeted bile acid metabolomic data, we tried to identify possible bile acid features within the relative concentration global metabolomics data and compare the results to see if the trends were reproduced (Table 19). Using the above outlined approach, five candidate bile acids were identified (with level five confidence). There was an unexpectedly poor correlation between the targeted and untargeted panel bile acid concentrations, with Pearson correlation coefficients ranging from -0.012 though 0.58 (Table 19). Metabolomics experts in the Dan Raftery lab advised that global and targeted metabolomics are not suitable for comparison. These distinct approaches have fundamental differences in how metabolite concentrations are measured, including but not limited to: unique sample preparation methods, separate analyses on different equipment, and discrete quantitation (relative vs. absolute levels).

Candidate conjugated bile acid match (from MoNA; level five confidence)	Pearson R correlation	p-value	Ionization mode in which feature was captured	Feature (from global data)
Cholic acid	0.582900673	1.63E-24	Negative	2.76_407.2804m/z
Taurochenodeoxycholic acid	0.292691229	3.01E-06	Negative	1.67_498.2900m/z
Taurochenodeoxycholic acid	0.290617119	3.56E-06	Negative	2.47_498.2898m/z
Taurocholic acid	0.422891173	4.31E-12	Negative	5.83_514.2845m/z
Glycocholic acid	-0.011687763	0.855	Positive	6.59_466.2335m/z

Table 19. Poor correlation between candidate level five confidence annotated bile acids from global data and bile acids from targeted data. These five features were identified as candidate conjugated bile acids in the global metabolomic data by comparing their masses, indicating a level five confidence in their annotation.[255] These bile acids were detected in the targeted bile acid metabolomic data, although there was unexpectedly poor (Pearson) correlation across the two metabolomics techniques. MoNA = Massbank of North America (global metabolomics public repository)

5.3.2 Broad interpretations and implications of the global data

I thought that a data-driven approach would be the optimal and unbiased approach for interpreting the global metabolomics data with the lens of metabolites relevant to GvHD. For instance, this approach helped make sound decisions on zero-treatment and normalization strategy. However, this did not point towards any particular feature, timepoint, or sample of interest to focus on as there were hundreds of potential features of interest. The number of

significantly different features in GvHD cases vs. controls increased with each timepoint, possibly because pre-HCT, participants are more physiologically similar, but as time goes on and outcomes diverge, more differences can be observed in the gut metabolome (Table 17, Figs. 27-29). Yet, the number of features detected remained similar over time and no GvHD-based patterns emerged by hierarchal clustering. Additionally, I hypothesized that microbial diversity would beget metabolomic diversity, but surprisingly, this was not observed in our dataset (overall or by timepoint). A similar number of features were absent or undetected in GvHD cases and controls. Even though this was unexpected, microbial metabolism is highly diverse and variable by strain, which may explain how a single microbial taxon can produce more metabolites or a greater diversity of metabolites than another without substantially increasing microbial diversity. The MetaboAnalyst functional pathway enrichment analysis pipeline is another tool that could systematically analyze the global metabolomics data. This analysis of our data hit on the bile acid biosynthesis pathway, supporting that the targeted bile acid metabolomic investigation could inform metabolites to target for GvHD (Fig. 30).

5.3.3 Minimal preliminary evidence of MCBAs in stool of allo-HCT recipients

It is unexpected that there was minimal preliminary evidence for MCBAs in the stool samples of allo-HCT recipients (Table 19). A major caveat to this tentative finding is that any annotations were only with a level five confidence and none were confirmed; thus MCBA presence or absence and concentrations remain inconclusive.[281] Based on the observed ubiquity of *bsh* genes in the gut microbial community, we expected that there would be several global features that could be candidate MCBAs using the pipeline/process outlined in Chapter 5.2.3.1. Previous studies detected MCBAs in animal and human stool, and observed differences in MCBAs based on inflammatory bowel disease or cystic fibrosis disease states.[127,277] Especially given the observed difference in endogenous secondary bile acids and lithocholic acid based on GvHD status, the lack of strong MCBA evidence was surprising. Allo-HCT recipients and specifically those that develop GvHD have disrupted gut microbiomes, which could also inhibit MCBA production or detection.[70] Further, allo-HCT recipients experience specific side effects/symptoms like mucositis that can drive diet. A combination of gut microbial dysbiosis and atypical diet may be one reason for poor MCBA evidence. Finally, our investigation was deprioritized due to the weak evidence of MCBAs, so the question of MCBAs in the stool of allo-HCT recipients with vs. without GvHD remains open for future study.

5.3.4 Major challenges remain in identifying metabolic markers or drivers of GvHD

Although at least 692 features were statistically significantly associated with GvHD for at least one timepoint, it was unexpectedly challenging to identify even a single feature. The Raftery lab collaborators affirmed this challenge, suggesting that characterization of one feature would be a considerable lift involving more research hours and resources, such that it would be outside the scope of this research. Even just the 27 features that were significantly enriched pre-HCT presented an insurmountable obstacle for the purpose of this research investigation (Table 17). If a candidate ID was selected, a novel targeted metabolomics approach would have to be created and optimized to confirm the candidate ID. Further, the candidate ID could prove invalid without providing any evidence of the true ID. Chemical adducts are also common in global metabolomics such that an allegedly unknown feature may represent a characterized/known feature with a unique or abnormal adduct. Similarly, surprisingly few features matched to an existing record in the MoNA database and none of those matches/records could be plausibly tied to GvHD. This underscores the true diversity and variability of metabolites in stool, and the undiscovered complexity of metabolites in this research area.

Chapter 6: Future directions

6.1 Overall synthesis/discussion

While it remains unclear why some allo-HCT recipients develop GvHD despite very strong MHC matching between donor and recipient, research is bringing the GvHD field closer to answers and possible solutions. To date, physicians and scientists understand how pre-HCT conditioning and downstream inflammation contribute to GvHD by facilitating activation of donor T cells into Th17 and Th1 cells which incite tissue damage. In the clinic, an individual therapeutic approach is developed for each HCT recipient with GvHD, including combinations of corticosteroids, calcineurin inhibitors, Ruxolitinib, other immunosuppressive medications, and bile acid sequestrants for controlling diarrhea symptoms. We also understand that intestinal barrier integrity, gut microbes and their metabolites, and diet are crucial levers that may modulate GvHD. This study represents a robust, longitudinal investigation into how these factors interact in the context of GvHD.

A 98 participant case-control study design was implemented with collection of 247 stool samples from pre- and post-HCT. We performed 16S rRNA gene sequencing, whole genome metagenomic sequencing, targeted bile acid and SCFA metabolomics, and untargeted metabolomics on these samples to characterize the gut microbiome and identify patterns, markers or modulators that could point towards anti-GvHD interventions. The bile acid

metabolomic panel yielded the strongest observations. Lithocholic acid and endogenous secondary bile acids were deficient at 60 days post-HCT in HCT recipients with GvHD. LCA and the endogenous secondary bile acids are the most promising candidates for further study as markers or modulators that could impact GvHD. Key questions should be answered before introduction of LCA at the clinic: does LCA modulate GvHD, and if so, does the mechanism involve barrier integrity and/or T cell immunomodulation (for greater detail, refer to section 6.2.3: Test hypotheses in an in vitro or murine model system for greater detail)? Is LCA acutely toxic or carcinogenic in humans? Could there be safer alternatives that still protect against GvHD (for greater detail, refer to section 6.2.5: Test if a less toxic conformer of LCA could be administered, like cholic acid 7-sulfate)? If LCA's GvHD-protective mechanism is characterized, variations to toggle this protective effect should be explored. For instance, another compound that activates the same receptor (e.g. Vitamin D and VDR), an LCA-producing microbe (e.g. *C. scindens*), or set of essential microbial genes and enzymes (e.g. enzymes encoded by the *bai* gene operon). Ideally, these questions would be examined in detail through *in vitro* cell culture models or *in vivo* murine models (for greater detail, refer to section 6.2.4: Test LCA as an intervention for GvHD).

From the same participants and sample cohort, although there is ample existing literature supporting health-protective effects of SCFAs like butyrate and propionate, it remains unclear exactly which mechanism(s) SCFAs use to influence outcomes in HCT, such as immune modulation, enhancing tight junction function, HDACi, or facilitating intestinal epithelial cell recovery, all of which are possible pathways. Despite this, butyrate was enriched in no GvHD controls pre-HCT, which is striking regardless of the mechanism. Surprisingly, few of the other SCFAs exhibited patterns that indicated any role in GvHD, including an unexpected lack of association between propionate and GvHD. Similarly, it was difficult to make meaningful links between unidentified global features and GvHD. Although the numbers of statistically significantly enriched features in GvHD cases and controls was striking, there were immense obstacles in selecting which features are most appropriate for further research and in identifying those features. The features that were differentially abundant pre-HCT could be a starting point. The MCBA research is also interesting for advancing basic science by characterizing the capabilities of bile salt hydrolases, but it is less likely to yield a translational finding in the short-term.

There are two highest priority next steps. One involves exploring the potential VDR-mediated protective effect of LCA and endogenous secondary bile acids. Simultaneously, research efforts should be directed towards evaluating the feasibility of administering LCA safely

at the clinic. The current precedent of already administering one secondary bile acid, UDCA, at the clinic makes this direction more feasible and thus high impact. The second involves investigating the pathway and timing of anti-GvHD pre-HCT butyrate. There are several potential pathways at play, but the pre-HCT timing makes this a higher impact future direction. Both of these future directions would benefit from more experimentation with an *in vitro* human tissue co-culture, organoid, or *in vivo* murine model.

6.1.1 Significance and impact

Although the multifactorial nature of GvHD has made it challenging to identify clear causes of and contributors to GvHD, it provides several avenues to leverage as promising treatments that may reduce severity or prevent GvHD. LCA and endogenous secondary bile acid deficiency linked with GvHD, identifying supplementation with LCA as a credible therapy to treat gut GvHD. Strikingly, pre-HCT butyrate deficiency also linked with GvHD, providing another plausible intervention that could prevent GvHD. SCFAs' and bile acids' link with GvHD was notable prior to the research described here, but the significant data on butyrate, LCA and other endogenous secondary bile acids substantially narrows the scope of candidates for further research. Organoid or murine model experiments that probe the mechanism of these associations would advance them closer to clinically meaningful interventions against GvHD.

These findings are relevant for other intestinally-mediated diseases like inflammatory bowel disease, Crohn's disease or ulcerative colitis, increasing their impact. These diseases are all mediated by intestinal inflammation, mucosal barrier damage, acute tissue injury and/or gut microbiome dysbiosis. Butyrate, LCA and other endogenous secondary bile acids could impact each of these, broadening the possible translational applications.

6.2 Bile acids and GvHD (Ch 3)

6.2.1 RNA-sequencing

The microbial genes present in each participant's gastrointestinal tract reveal which bile acid metabolic genes are present. Similarly, concentrations of individual and groups of bile acids depict the activity of the microbial bile acid metabolic enzymes. Even with all this information, it remains unclear which *bsh* or *bai* genes are actively being transcribed and used to modulate bile acids. Are there differences between GvHD cases vs. controls at the transcriptional level? Transcriptional analysis could be performed on stool samples to clarify this. RNA-sequencing is the optimal approach for the *bsh* genes given their poor sequence similarity. Although the *bsh* genes are incompatible with the more precise quantitative real-time polymerase chain reaction (qRT-PCR) because of the low sequence similarity (see Ch 2.6.1), a qRT-PCR approach should

be considered for the *bai* gene operon as the degree of sequence similarity has not been investigated. Thus, there may be sufficient similarity to accommodate a *bai*-wide primer and probe. Although qRT-PCR is more precise, RNAseq may be able to more efficiently assess the entire *bai* operon, which consists of at least nine genes, without having to develop and optimize a new set of primers and probes for each gene. If transcriptional differences are observed between GvHD cases vs. controls, this will strengthen the rationale for performing the additional experiments described in Ch 6.2.3-6.2.5. If there are no transcription-level differences, then a more informed decision can be made to pursue such future directions/studies.

6.2.2 Efficiency/specificity of bsh's

Ch 3.2.9 introduces the idea that gene orthologs of *bsh* within the same taxon may not have similar functions or target specificity/efficiency. Possible interventions for gut GvHD could emerge from characterization of the efficiency and specificity of bsh's. This could permit the administration of bsh's with defined reaction rates that act on target bile acids. For instance, since LCA uniquely binds VDR, a conjugated LCA-targeting bsh could be prescribed to deconjugate conjugated LCA and promote agonism of VDR to in turn enhance intestinal barrier integrity, reduce inflammation and facilitate differentiation of T cells into Treg cells.[149] However, this should be researched in greater depth *in vitro*. Such pre-clinical experiments could begin with adding a defined, initial concentration of conjugated (for example, tauroLCA) or primary (for example, CDCA) bile acids with bacteria that possess bsh('s) of interest (for example, *M. gnavus*[283]). After 10-30 minutes of incubation, the products in solution could be measured with a Bio-Rad Protein assay, as previously described.[284,285] The concentrations of the reactant:product bile acids (for example, LCA:conjugated LCA:CDCA) may also be measured to calculate the ratio of each bile acid product. This could be accomplished by performing LC-MS to quantify the ratio of bile acids.

6.2.3 Test hypotheses in an *in vitro* or murine model system

Before pursuing secondary bile acid or LCA administration as a clinical intervention to reduce GvHD, further pre-clinical studies should be performed. The underlying mechanism(s) behind the observed association between deficiency of endogenous secondary bile acids and LCA with GvHD remain ambiguous. Potential pathways include modulation of T-cell immunity, gastrointestinal barrier integrity, cell signaling or inflammation. Investigation of each mechanism warrants different approaches, markers and/or techniques. An accurate *in vitro* tissue model may be a sufficient system for experimentation that probes each pathway, but such a model remains elusive. Several outstanding problems make an *in vitro* human tissue model for GvHD

challenging. One challenge is that this model would have to co-culture human intestinal epithelial cells and bacterial cells. Although such co-culture systems exist, a major obstacle is that most of the bacteria of interest are obligate or facultative anaerobes whereas the human tissue requires oxygen. At least one group has successfully engineered an oxygen gradient model that should sustain anaerobic microbes and human tissue, but there is significant risk for bacterial and oxygen contamination that would compromise the experiment when handling the models to collect samples and monitor changes.[286] Given these constraints, it is unsurprising that *in vivo* murine GvHD models are the most popular for investigating GvHD mechanisms and treatments. Findings from these models are valuable, but there are significant limitations when translating to humans. Namely, murine immunology, bile acid chemistry and gut microbial communities are distinct from their human counterparts. For example, in humans, bile acids are more commonly conjugated to glycine vs. in mice, bile acids are commonly conjugated to taurine.[287] Humans can also perform sulfation on bile acids where mice cannot.[287] Finally, based on the ratios of primary to secondary and unconjugated to conjugated bile acids, both microbial deconjugation and 7α-dehydroxylation are more robustly performed in humans vs. mice.[287]

Although the currently available systems are imperfect, the connection between VDR and GvHD is the topmost priority hypothesis to study further. This is because the existing link between vitamin D and GvHD, LCA's unique agonism of VDR, and LCA's observed differential abundance by GvHD status.[145,146,148,149,151] Since there is no suitable *in vitro* co-culture model that would evaluate the impact of VDR on T cells and relevant GvHD tissue (e.g. IECs), a murine GvHD model would be the best option.[288,289] Compounds of interest to test include vitamin D as a positive control, and negative controls such as a non-VDR binding comparable compound such as pregnenolone-16α-carbonitrile (PCN, a pregnane X receptor agonist; as used by Makishim et al., 2002) and a non-secondary bile acid like cholic acid. Experimental compounds to test include and LCA, 3-keto-LCA, alloLCA (LCA isomers), tauroLCA (conjugated form of LCA), 12KLCA814, and 5beta-cholanic acid-3beta, 12alpha-diol (similarly structured to LCA and observed to be more prevalent or abundant in no GvHD controls in Ch 3.2). In addition to the compounds of interest, multiple timepoints for intervention could be examined. Each control or experimental compound should be separately studied when administered pre-HCT, immediately post-HCT (3 days post), and later post-HCT (10 days post). Outcomes measured could include overall survival, GvHD severity, white blood cell counts to determine immune strength (via blood samples), ratio of Treg vs. Th17 or Th1 cells to assess immunomodulatory impacts (via flow cytometry on blood, where there are effectors), and inflammatory cytokines (by blood sample) and intestinal length (by collecting the intestinal organ) to assay inflammation.[289]

An even clearer picture could be generated by staining sections of the intestine for villi/crypt damage and apoptotic cells (via hematoxylin and eosin (H&E) staining).[289] If LCA is protective against GvHD, then both LCA, VDR and possibly the LCA isomers would improve overall survival, reduce inflammation, and exhibit longer intestines and less tissue damage vs. negative control mice.

6.2.4 Test LCA as an intervention for GvHD

The largest difference in bile acid concentration between GvHD cases and controls is in LCA, for which there was a 260-fold difference between the median GvHD case vs. control concentrations 60 days post-HCT, with lower concentrations in GvHD. Additionally, lithocholic acid is present in 100% of controls vs. 82% in GvHD cases on day 60 post-HCT. Thus, LCA derivatives and *bai* genes are promising for further studies examining their possible role as drivers or markers of GvHD. The secondary bile acid UDCA is currently administered to all allo-HCT recipients. UDCA was administered therapeutically to humans as early as 1977, although it was not tested in an animal model until 1985.[290,291]. It was also used to treat liver GvHD as early as 1992 despite not testing UDCA's impact on GvHD in an animal model before human use.[292] Cholic acid (Cholbam), chenodeoxycholic acid (chenodiol/Ctexli) and obeticholic acid (Ocaliva) are other examples of FDA-approved bile acid therapies for humans providing a precedent for prescribing bile acids to transplant recipients as prophylaxis against adverse transplant outcomes. The data described in this study are in the early pre-clinical stages, but if there proves to be a relationship with GvHD onset or maintenance, then novel therapeutic interventions may be feasible to mitigate gut GvHD. These interventions could include providing *bai* gene-possessing bacteria like *C. scindens*, to HCT recipients as a live biotherapeutic product. Another therapeutic option would be to directly give endogenous secondary bile acids, like LCA, to HCT recipients post-transplant. This may entail administering a delayed/time-released formulation to target/extend release into the large intestine. Detailed *in vitro* and *in vivo* studies are required to study this concept, since LCA has been observed to have toxic and carcinogenic effects in animal models.[243,293] Less toxic LCA congeners should be considered for GvHD therapy. Well-tolerated alternatives warrant further investigation. Examples include: LCA-sulfate due to its structural similarity to native LCA and cholic acid-7-sulfate's due to its agonism of the same receptor as LCA, TGR5.[294,295]

6.2.5 Test if a less toxic conformer of LCA could be administered, like cholic acid 7-sulfate (CA7S)

LCA can have toxic and/or carcinogenic effects in animal models and therefore may be toxic in humans as well.[243,293] Alternatives LCA-sulfate and cholic acid-7-sulfate (CA7S) could be investigated as non-toxic alternatives.[294,295] However, it would be best to first study the mechanism underlying LCA's beneficial effects against GvHD so that alternative compounds could be explored. For example, if *in vitro* tissue culture or *in vivo* mouse studies (like the experiments outlined in Ch. 6.2.3) support that TGR5 mediates LCA's anti-GvHD effects, then CA7S would be an ideal alternative because it also binds TGR5. Or, if VDR binding appears to be important for LCA and GvHD, then vitamin D itself or another VDR agonist (such as vitamin D metabolic intermediates 25-dihydroxycholesterol, 7-dehydrocholesterol or 1,25-dihydroxycholesterol) may be promising and less toxic alternatives to LCA. Another option to demonstrate LCA's tolerability to humans would be to directly test if LCA has toxic effects. However, this may be challenging since a proposed adverse health effect is carcinogenicity, which would take up to 2 years in a rodent toxicity assay.[296] A short-term assay for mutagenicity like the Ames test could yield important information, but a negative result does not conclusively rule out long-term carcinogenesis. Non-carcinogenic acute toxic effects, like those exerted on specific organs, can be ruled out by testing organ-specific cell lines or animal models with harvest of the organ of interest (e.g. intestine, liver). Further, even the alternatives LCA7S and LCA-sulfate should be evaluated for potential human toxicity.

Another possible pathway for LCA-induced damage could be due to an accumulation of LCA or other therapeutic endogenous secondary bile acids in the esophagus or large intestine as this is not their traditional niche. The primary region for bile acid recirculation is absorption via the small intestine; however, small intestinal tissue injury or damage would lead to bile acid malabsorption and therefore, unusually high concentrations of bile acids in the large intestine. This abundance could lead to off-target acute toxicity mediated by LCA's hydrophobicity and emulsification/breakdown of intestinal cell membranes. Alternatively, the large intestine may still be able to reabsorb the greater levels of bile acids, but such atypical enterohepatic circulation may be less tolerated and also lead to negative health effects.

6.3 SCFAs and GvHD (Ch 4)

6.3.1 Probe mechanism underlying SCFA protection via *in vitro* model

6.3.1.1 Current best model options

An *in vitro* model that replicates *in vivo* human physiology is the best tool to characterize the mechanism behind beneficial effects of SCFAs. Murine models of diseases like GvHD are beneficial for their ability to perform highly controlled *in vivo* experiments. However, critical

differences in murine vs. human immunity, microbial communities and diet restrict the relevance to human biology. One example of a potential viable human model is the gut-on-a-chip model described by Ballerini et al. (2025) containing a diversity of gut epithelial and endothelial cell types, immune cells, and vasculature.[297] It would be best to also incorporate live microbiota that typically inhabit the gut epithelium, as modeled by Bang et al. 2025.[298] This type of co-culture model is challenging to develop because it requires an oxygen gradient be sustained within the model. Native gut microbiota are obligately or facultatively anaerobic whereas the human intestinal epithelial and immune cells require oxygen to respire and survive. The Bang et al. (2025) model can support the oxygen gradient required to grow human with bacterial cells, but their model system uses the immortal cell line Caco-2 cells. These are genetically transformed such that they no longer behave as *in vivo* intestinal cells or primary cell lines. The model described in Kim et al. (2022) checks each of these boxes, maintaining an oxygen gradient required for bacteria and human intestinal cell co-culture with a primary cell line. Such primary cell lines expand to represent the epithelium from which harvested, including a variety of cells types (e.g. Goblet cells, Paneth cells, gut-resident immune cells, etc.) making it the most realistic model.[299]

6.3.1.2 Hypotheses/mechanisms to test

Using a model such as those described above, the mechanism underlying butyric acid's protective effect against GvHD can be probed. There are several proposed pathways, each of which was discussed in detail in Ch 4.1.1. In summary, one potential pathway involved may be strengthening of tight junctions between intestinal cells to improve the integrity of the intestinal barrier. Another pathway is downstream effects that inhibit histone deacetylation, keeping DNA unraveled and primed for quicker transcription of anti-inflammatory genes. SCFAs like butyrate have also been observed to promote cell turnover, which may result in quicker regeneration of an intact intestinal barrier post-injury; although, this conflicts with the observation that butyrogenic bacteria may inhibit ISCs. It is hypothesized that multiple mechanisms may be at play.

6.4 Untargeted/global metabolomics and GvHD (Ch 5)

6.4.1 Follow up on top hits

Untargeted metabolomics produced an abundance of data to interpret; therefore, several research avenues warrant further investigation. As discussed in Ch 5.2.1, there were 27 features that were statistically significantly differentially abundant pre-HCT. Not only is this signal unique (over 350 statistically significant different features at a later timepoint), but the

nature of the timing makes them clinically important. In the GvHD field, it is still unknown why some HCT recipients may develop GvHD whereas others do not and what causes GvHD in these recipients. These 27 features represent opportunities to learn what may cause GvHD and ultimately how to intervene and possibly prevent GvHD. This would entail ordering the candidate compound as a standard, developing a new targeted panel and optimizing the panel to confirm identity of the feature(s) or increase the confidence of the identification. Although, without a match to another feature in the MoNA database or BileOmix standards, it is difficult to land on a single candidate compound. The untargeted data could also be analyzed with the Human Fecal Metabolome Database (www.fecalmetabolome.ca), which is restricted to small molecule metabolites that have been detected in human stool, to try and narrow down the scope to a single candidate compound using a more tailored tool/resource.

In Ch 5.3.2, the downsides of an entirely data-driven and biologically-naïve approach were outlined. The MetaboAnalyst functional pathway enrichment analysis provides a starting point of biologically important pathways and corresponding metabolites to consider. Although the targeted bile acid metabolomic panel capitalized on the bile acid biosynthesis hit from MetaboAnalyst, the strongest hit pathways remain open for investigation. This includes vitamin B metabolism, vitamin C metabolism and aldarate metabolism. Vitamin C has been shown to promote stability of $CD8^+$ regulatory T cells (controlling GvHD while promoting GvL effects). Given the immunomodulatory effects of vitamin C, further research into vitamin C metabolism and GvHD is warranted.[300] The metabolites and compounds involved in these pathways should also be considered as top candidates for 27 pre-HCT significantly different features. Given the statistical strength of the 27 features and the biological emphasis on the vitamin B, C and aldarate metabolism, there may be overlap in these groups. This would take advantage of both a data-driven and biology-informed approach to identify the top priority features.

6.4.2 Investigate candidate secondary MCBAs

The global metabolomics performed on the 247 stool samples identified 19 level five confidence candidate MCBAs. From the targeted metabolomic panel, I observed that a deficiency of lithocholic acid and endogenous secondary bile acids associated with GvHD. There are three level five candidate microbially-conjugated LCA features and three additional level five candidate microbially-conjugated DCA features. Given the targeted metabolomics data supporting a link between GvHD and endogenous secondary bile acids, these six level five candidate MCBAs are of greatest interest to annotate with greater confidence in our samples first. Within the top six candidates, the lithocholic acid amidates should be investigated first because the strongest association observed was with LCA deficiency and GvHD post-HCT.

Higher confidence identification of the six MCBAs of greatest interest and any additional microbially-conjugated secondary bile acids present in these stool samples from allo-HCT recipients (e.g. lithocholic acid conjugates) can be accomplished by ordering the appropriate amidate libraries. This approach was agreed upon after advising, meeting and discussing with the Raftery lab. A company, BileOmix (based at out of Farmington, Connecticut at the University of Connecticut), synthesizes and sells such MCBAs for research use. In theory, purchasing these libraries and then examining our stool samples for the library MCBAs, calculating concentrations by timepoint post-HCT and GvHD status would yield insights into this new class of bile acids in our samples. Upon reaching out to BileOmix, representatives confirmed that they could make the libraries that we were interested in and they would include 21 amidates with, for example, an LCA core compound. However, the Raftery lab advised that we inquire if each individual amidate/MCBA compound is stored separately because the retention time, mass, and charge for each MCBA is likely highly similar due to their having the same core compound (LCA). This would make quantitation of each MCBA challenging because they may run off the LC-MS column simultaneously and appear indistinguishable in analysis. Unfortunately, BileOmix informed us that each library is pooled together in a single vial, making the Raftery lab concerned that each individual MCBA may not look distinct by LC-MS. This challenge is ongoing, but investigation into different types of columns that can tease apart highly structurally similar compounds such as those in the MCBA libraries is an important future direction for determining if MCBAs (particularly, microbially conjugated endogenous secondary bile acids) are present in our stool samples and/or differentially abundant or prevalent by GvHD status or timepoint.

This approach can be replicated and adapted for more bile acids of interest, such as primary or conjugated bile acids. Although the thesis research described did not identify primary or conjugated bile acids of importance for GvHD, it could be insightful to measure their presence and concentrations in stool since there was surprisingly minimal evidence of endogenous secondary MCBAs. Study of the other types of bile acids would more comprehensively address the question of if MCBAs are in stool. This would also require ordering the corresponding bile acid amidate libraries from BileOmix to assay for them in the stool samples. Altogether, this would advance understanding of the true impact of bile acids on GvHD and, more broadly, the gut microbiome beyond the primary/secondary bile acid traditional model.

Conflict of Interest: DNF is a consultant for Seres Therapeutics which is developing an orally administered live biotherapeutic for patients undergoing HCT to prevent bacteremia. None of the companies are associated with the work presented in this study.

Funding: This work was supported by NIH Grant number 1R01HG005966-01 awarded to David N. Fredricks. This work is the result of NIH funding, in whole or in part, and is subject to the NIH Public Access Policy. Through acceptance of this federal funding, the NIH has been given a right to make the work publicly available in PubMed Central.

References

1. Fredricks, D. N. The gut microbiota and graft-versus-host disease. *J Clin Invest* **129**, 1808–1817 (2019).
2. Noth, R. *et al.* Increased intestinal permeability and tight junction disruption by altered expression and localization of occludin in a murine graft versus host disease model. *BMC Gastroenterol* **11**, 109 (2011).
3. Nalle, S. C. *et al.* Graft-versus-host disease propagation depends on increased intestinal epithelial tight junction permeability. *J Clin Invest* **129**, 902–914 (2019).
4. Shono, Y., Docampo, M. D., Peled, J. U., Perobelli, S. M. & Jenq, R. R. Intestinal microbiota-related effects on graft-versus-host disease. *Int J Hematol* **101**, 428–437 (2015).
5. Khoury, H. J. *et al.* Improved survival after acute graft-versus-host disease diagnosis in the modern era. *Haematologica* **102**, 958–966 (2017).
6. Martin, P. J. & Shlomchik, W. D. Overview of Hematopoietic Cell Transplantation Immunology. in *Thomas' Hematopoietic Cell Transplantation* 96–114 (John Wiley & Sons, Ltd, 2015). doi:10.1002/9781118416426.ch9.
7. Yue, X., Zhou, H., Wang, S., Chen, X. & Xiao, H. Gut microbiota, microbiota-derived metabolites, and graft-versus-host disease. *Cancer Medicine* **13**, e6799 (2024).

8. Hill, G. R., Betts, B. C., Tkachev, V., Kean, L. S. & Blazar, B. R. Current Concepts and Advances in Graft-Versus-Host Disease Immunology. *Annual Review of Immunology* **39**, 19–49 (2021).
9. Andersson, B. S., Valdez, B. C. & Jones, R. B. Pharmacologic Basis for High-dose Chemotherapy. in *Thomas' Hematopoietic Cell Transplantation* 211–224 (John Wiley & Sons, Ltd, 2015). doi:10.1002/9781118416426.ch19.
10. Bensinger, W. I. High-dose Preparatory Regimens. in *Thomas' Hematopoietic Cell Transplantation* 223–234 (John Wiley & Sons, Ltd, 2015). doi:10.1002/9781118416426.ch20.
11. Sandmaier, B. M. & Storb, R. Reduced-intensity Allogeneic Transplantation Regimens. in *Thomas' Hematopoietic Cell Transplantation* 232–246 (John Wiley & Sons, Ltd, 2015). doi:10.1002/9781118416426.ch21.
12. Munchel, A. T., Kasamon, Y. L. & Fuchs, E. J. Treatment of hematological malignancies with nonmyeloablative, HLA-haploidentical bone marrow transplantation and high dose, post-transplantation cyclophosphamide. *Best Practice & Research Clinical Haematology* **24**, 359–368 (2011).
13. Shono, Y. *et al.* Increased GVHD-related mortality with broad-spectrum antibiotic use after allogeneic hematopoietic stem cell transplantation in human patients and mice. *Science Translational Medicine* **8**, 339ra71-339ra71 (2016).
14. Glucksberg, H. *et al.* CLINICAL MANIFESTATIONS OF GRAFT-VERSUS-HOST DISEASE IN HUMAN RECIPIENTS OF MARROW FROM HL-A-MATCHED SIBLING DONOR,S. *Transplantation* **18**, 295 (1974).
15. Jagasia, M. H. *et al.* National Institutes of Health Consensus Development Project on Criteria for Clinical Trials in Chronic Graft-versus-Host Disease: I. The 2014 Diagnosis and Staging Working Group Report. *Biol Blood Marrow Transplant* **21**, 389-401.e1 (2015).

16. Przepiorka, D. *et al.* 1994 Consensus Conference on Acute GVHD Grading. *Bone Marrow Transplant* **15**, 825–828 (1995).
17. Singh, R. *et al.* DecodingAcute GVHD: The role of serum biomarkers and imaging in diagnosis and management. *Blood Reviews* 101329 (2025) doi:10.1016/j.blre.2025.101329.
18. Lin, Z. *et al.* Lipid biomarkers of GVHD in allogeneic stem hematopoietic cell transplantation patients. *Front. Immunol.* **16**, (2025).
19. Loeb, K. R., Yeung, C. C. & Shulman, H. M. Pathology of Hematopoietic Cell Transplantation. in *Thomas' Hematopoietic Cell Transplantation* 292–332 (John Wiley & Sons, Ltd, 2015). doi:10.1002/9781118416426.ch26.
20. Cutler, C. & Antin, J. H. Manifestations and Treatment of Acute Graft-versus-Host Disease. in *Thomas' Hematopoietic Cell Transplantation* 1012–1025 (John Wiley & Sons, Ltd, 2015). doi:10.1002/9781118416426.ch83.
21. Commissioner, O. of the. FDA Approves First Drug to Prevent Graft Versus Host Disease. *FDA* https://www.fda.gov/news-events/press-announcements/fda-approves-first-drug-prevent-graft-versus-host-disease (2024).
22. Ye, Y., Hao, W., Malard, F. & Mohty, M. Emerging novel therapies for steroid-refractory acute graft-versus-host disease: recent advances and future directions. *Current Opinion in Immunology* **96**, 102649 (2025).
23. Orange Book: Approved Drug Products with Therapeutic Equivalence Evaluations. https://www.accessdata.fda.gov/scripts/cder/ob/results_product.cfm?Appl_Type=A&Appl_No=205789#39363.
24. Bolaños-Meade, J. *et al.* Post-Transplantation Cyclophosphamide-Based Graft-versus-Host Disease Prophylaxis. *New England Journal of Medicine* **388**, 2338–2348 (2023).
25. Zeiser, R. *et al.* Ruxolitinib for Glucocorticoid-Refractory Acute Graft-versus-Host Disease. *New England Journal of Medicine* **382**, 1800–1810 (2020).

26. Rashidi, A. *et al.* Potential of Fecal Microbiota Transplantation to Prevent Acute GVHD: Analysis from a Phase II Trial. *Clin Cancer Res* **29**, 4920–4929 (2023).
27. Magenau, J. *et al.* A Randomized, Double-Blind, Placebo-Controlled Phase III Trial of Alpha-1-Antitrypsin for Treatment of High Risk Acute Gvhd (BMT CTN 1705). in (Tandem Meetings, 2025).
28. Martini, D. J., Chen, Y.-B. & DeFilipp, Z. Recent FDA Approvals in the Treatment of Graft-Versus-Host Disease. *Oncologist* **27**, 685–693 (2022).
29. PubChem. Ursodiol [USAN:USP]. https://pubchem.ncbi.nlm.nih.gov/substance/134974576.
30. PubChem. Ursodeoxycholic Acid. https://pubchem.ncbi.nlm.nih.gov/compound/31401.
31. Ruutu, T. *et al.* Ursodeoxycholic acid for the prevention of hepatic complications in allogeneic stem cell transplantation. *Blood* **100**, 1977–1983 (2002).
32. Ruutu, T. *et al.* Improved Survival with Ursodeoxycholic Acid Prophylaxis in Allogeneic Stem Cell Transplantation: Long-Term Follow-Up of a Randomized Study. *Biology of Blood and Marrow Transplantation* **20**, 135–138 (2014).
33. Nalle, S. C. & Turner, J. R. Intestinal barrier loss as a critical pathogenic link between inflammatory bowel disease and graft-versus-host disease. *Mucosal Immunol* **8**, 720–730 (2015).
34. Johansson, J.-E. & Ekman, T. Gut toxicity during hemopoietic stem cell transplantation may predict acute graft-versus-host disease severity in patients. *Dig. Dis. Sci.* **52**, 2340–2345 (2007).
35. Hanash, A. M. *et al.* Interleukin-22 Protects Intestinal Stem Cells from Immune-Mediated Tissue Damage and Regulates Sensitivity to Graft versus Host Disease. *Immunity* **37**, 339–350 (2012).
36. Eriguchi, Y. *et al.* Graft-versus-host disease disrupts intestinal microbial ecology by inhibiting Paneth cell production of α-defensins. *Blood* **120**, 223–231 (2012).

37. Weber, D. *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* **12**, e0185265 (2017).
38. Johansson, J.-E. & Ekman, T. Gut Toxicity During Hemopoietic Stem Cell Transplantation May Predict Acute Graft-Versus-Host Disease Severity in Patients. *Dig Dis Sci* **52**, 2340–2345 (2007).
39. Nalle, S. C. *et al.* Recipient NK cell inactivation and intestinal barrier loss are required for MHC-matched graft-versus-host disease. *Science Translational Medicine* **6**, 243ra87-243ra87 (2014).
40. Brown, G. R. *et al.* Tumor necrosis factor inhibitor ameliorates murine intestinal graft-versus-host disease. *Gastroenterology* **116**, 593–601 (1999).
41. Holler, E. *et al.* Modulation of acute graft-versus-host-disease after allogeneic bone marrow transplantation by tumor necrosis factor alpha (TNF alpha) release in the course of pretransplant conditioning: role of conditioning regimens and prophylactic application of a monoclonal antibody neutralizing human TNF alpha (MAK 195F). *Blood* **86**, 890–899 (1995).
42. Noth, R. *et al.* Increased intestinal permeability and tight junction disruption by altered expression and localization of occludin in a murine graft versus host disease model. *BMC Gastroenterol* **11**, 109 (2011).
43. Taur, Y. *et al.* The effects of intestinal tract bacterial diversity on mortality following allogeneic hematopoietic stem cell transplantation. *Blood* **124**, 1174–1182 (2014).
44. Stein-Thoeringer, C. K. *et al.* Lactose drives Enterococcus expansion to promote graft-versus-host disease. *Science* **366**, 1143–1149 (2019).
45. Golob, J. L. *et al.* Butyrogenic bacteria after acute graft-versus-host disease (GVHD) are associated with the development of steroid-refractory GVHD. *Blood Advances* **3**, 2866–2869 (2019).

46. Mathewson, N. *et al.* Unbiased Metabolic Profiling Uncovers a Crucial Role for the Microbial Metabolite Butyrate in Modulating GI Epithelial Cell Damage from Gvhd. *Blood* **124**, 536 (2014).
47. Markey, K. A. *et al.* The microbe-derived short-chain fatty acids butyrate and propionate are associated with protection from chronic GVHD. *Blood* **136**, 130–136 (2020).
48. Meedt, E. *et al.* Prolonged Suppression of Butyrate-Producing Bacteria Is Associated With Acute Gastrointestinal Graft-vs-Host Disease and Transplantation-Related Mortality After Allogeneic Stem Cell Transplantation. *Clin Infect Dis* **74**, 614–621 (2022).
49. Donohoe, D. R. *et al.* The Microbiome and Butyrate Regulate Energy Metabolism and Autophagy in the Mammalian Colon. *Cell Metabolism* **13**, 517–526 (2011).
50. O'Meara, A. *et al.* Fecal calprotectin and α1-antitrypsin dynamics in gastrointestinal GvHD. *Bone Marrow Transplant* **50**, 1105–1109 (2015).
51. Strygler, B., Nicar, M. J., Santangelo, W. C., Porter, J. L. & Fordtran, J. S. *α*1-antitrypsin excretion in stool in normal subjects and in patients with gastrointestinal disorders. *Gastroenterology* **99**, 1380–1387 (1990).
52. Pathirana, W. G. W., Chubb, S. P., Gillett, M. J. & Vasikaran, S. D. Faecal Calprotectin. *Clin Biochem Rev* **39**, 77–90 (2018).
53. Hartwell, M. J. *et al.* An early-biomarker algorithm predicts lethal graft-versus-host disease and survival. *JCI Insight* **2**, e89798.
54. Major-Monfried, H. *et al.* MAGIC biomarkers predict long-term outcomes for steroid-resistant acute GVHD. *Blood* **131**, 2846–2855 (2018).
55. Etra, A. *et al.* Assessment of systemic and gastrointestinal tissue damage biomarkers for GVHD risk stratification. *Blood Adv* **6**, 3707–3715 (2022).
56. Spyrou, N. *et al.* The utility of biomarkers in acute GVHD prognostication. *Blood Adv* **7**, 5152–5155 (2023).

57. Bidgoli, A. *et al.* Current Definitions and Clinical Implications of Biomarkers in Graft-versus-Host Disease. *Transplantation and Cellular Therapy* **28**, 657–666 (2022).
58. Rafei, H. & Jenq, R. R. Microbiome-intestine cross talk during acute graft-versus-host disease. *Blood* **136**, 401–409 (2020).
59. Hill, G. R. & Koyama, M. Cytokines and costimulation in acute graft-versus-host disease. *Blood* **136**, 418–428 (2020).
60. Fegan, C., Poynton, C. H. & Whittaker, J. A. The gut mucosal barrier in bone marrow transplantation. *Bone Marrow Transplant* **5**, 373–377 (1990).
61. Koltun, W. A., Bloomer, M. M., Colony, P. & Kauffman, G. L. Increased intestinal permeability in rats with graft versus host disease. *Gut* **39**, 291–298 (1996).
62. Malard, F., Gasc, C., Plantamura, E. & Doré, J. High gastrointestinal microbial diversity and clinical outcome in graft-versus-host disease patients. *Bone Marrow Transplant* **53**, 1493–1497 (2018).
63. Taur, Y. *et al.* The effects of intestinal tract bacterial diversity on mortality following allogeneic hematopoietic stem cell transplantation. *Blood* **124**, 1174–1182 (2014).
64. Galloway-Peña, J. R. *et al.* Fecal Microbiome, Metabolites, and Stem Cell Transplant Outcomes: A Single-Center Pilot Study. *Open Forum Infect Dis* **6**, ofz173 (2019).
65. Han, L. *et al.* Intestinal Microbiota at Engraftment Influence Acute Graft-Versus-Host Disease via the Treg/Th17 Balance in Allo-HSCT Recipients. *Front. Immunol.* **9**, (2018).
66. Han, L. *et al.* Intestinal Microbiota Can Predict Acute Graft-versus-Host Disease Following Allogeneic Hematopoietic Stem Cell Transplantation. *Biology of Blood and Marrow Transplantation* **25**, 1944–1955 (2019).
67. Peled, J. U. *et al.* Microbiota as Predictor of Mortality in Allogeneic Hematopoietic-Cell Transplantation. *New England Journal of Medicine* **382**, 822–834 (2020).
68. Greco, R. *et al.* Microbiome markers are early predictors of acute GVHD in allogeneic hematopoietic stem cell transplant recipients. *Blood* **137**, 1556–1559 (2021).

69. Song, A. *et al.* Exploration of the relationship between intestinal flora changes and gut acute graft-versus-host disease after hematopoietic stem cell transplantation. *Translational Pediatrics* **10**, 28395–28295 (2021).
70. Jenq, R. R. *et al.* Regulation of intestinal inflammation by microbiota following allogeneic bone marrow transplantation. *Journal of Experimental Medicine* **209**, 903–911 (2012).
71. Jenq, R. R. *et al.* Intestinal Blautia Is Associated with Reduced Death from Graft-versus-Host Disease. *Biology of Blood and Marrow Transplantation* **21**, 1373–1383 (2015).
72. Simms-Waldrip, T. R. *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. *Biology of Blood and Marrow Transplantation* **23**, 820–829 (2017).
73. Ilett, E. E. *et al.* Associations of the gut microbiome and clinical factors with acute GVHD in allogeneic HSCT recipients. *Blood Adv* **4**, 5797–5809 (2020).
74. Weber, D. *et al.* Microbiota Disruption Induced by Early Use of Broad-Spectrum Antibiotics Is an Independent Risk Factor of Outcome after Allogeneic Stem Cell Transplantation. *Biology of Blood and Marrow Transplantation* **23**, 845–852 (2017).
75. Holler, E. *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. *Biology of Blood and Marrow Transplantation* **20**, 640–645 (2014).
76. Payen, M. *et al.* Functional and phylogenetic alterations in gut microbiome are linked to graft-versus-host disease severity. *Blood Adv* **4**, 1824–1832 (2020).
77. Burgos da Silva, M. *et al.* Preservation of the fecal microbiome is associated with reduced severity of graft-versus-host disease. *Blood* **140**, 2385–2397 (2022).
78. Peled, J. U. *et al.* Intestinal Microbiota and Relapse After Hematopoietic-Cell Transplantation. *J Clin Oncol* **35**, 1650–1659 (2017).

79. Golob, J. L. *et al.* Stool Microbiota at Neutrophil Recovery Is Predictive for Severe Acute Graft vs Host Disease After Hematopoietic Cell Transplantation. *Clinical Infectious Diseases* **65**, 1984–1991 (2017).
80. Biagi, E. *et al.* Gut microbiota trajectory in pediatric patients undergoing hematopoietic SCT. *Bone Marrow Transplant* **50**, 992–998 (2015).
81. Staffas, A., Burgos da Silva, M. & van den Brink, M. R. M. The intestinal microbiota in allogeneic hematopoietic cell transplant and graft-versus-host disease. *Blood* **129**, 927–933 (2017).
82. Gavriilaki, E. *et al.* Alteration of Gut Microbiota Composition and Diversity in Acute and/or Chronic Graft-versus-Host Disease Following Hematopoietic Stem Cell Transplantation: A Prospective Cohort Study. *International Journal of Molecular Sciences* **25**, 5789 (2024).
83. Hino, A. *et al.* Prolonged gut microbial alterations in post-transplant survivors of allogeneic haematopoietic stem cell transplantation. *British Journal of Haematology* **201**, 725–737 (2023).
84. Stoma, I. *et al.* Compositional Flux Within the Intestinal Microbiota and Risk for Bloodstream Infection With Gram-negative Bacteria. *Clin Infect Dis* **73**, e4627–e4635 (2021).
85. Weber, D. *et al.* Microbiota Disruption Induced by Early Use of Broad-Spectrum Antibiotics Is an Independent Risk Factor of Outcome after Allogeneic Stem Cell Transplantation. *Biology of Blood and Marrow Transplantation* **23**, 845–852 (2017).
86. Ostojic, S. M. Inadequate Production of H2 by Gut Microbiota and Parkinson Disease. *Trends in Endocrinology & Metabolism* **29**, 286–288 (2018).
87. Ohta, S. Molecular hydrogen as a preventive and therapeutic medical gas: initiation, development and potential of hydrogen medicine. *Pharmacology & Therapeutics* **144**, 1–11 (2014).
88. Ellison, C. A. *et al.* The role of interferon-γ, nitric oxide and lipopolysaccharide in intestinal graft-versus-host disease developing in F1-hybrid mice. *Immunology* **109**, 440–449 (2003).

89. Dayang, E.-Z. *et al.* Identification of LPS-Activated Endothelial Subpopulations With Distinct Inflammatory Phenotypes and Regulatory Signaling Mechanisms. *Front. Immunol.* **10**, (2019).
90. La Mura, V. *et al.* Liver sinusoidal endothelial dysfunction after LPS administration: A role for inducible-nitric oxide synthase. *Journal of Hepatology* **61**, 1321–1327 (2014).
91. Mantel, C. R. *et al.* Enhancing Hematopoietic Stem Cell Transplantation Efficacy by Mitigating Oxygen Shock. *Cell* **161**, 1553–1565 (2015).
92. Woolthuis, C. M. *et al.* Loss of quiescence and impaired function of CD34+/CD38low cells one year following autologous stem cell transplantation. *Haematologica* **98**, 1964–1971 (2013).
93. Wu, K. *et al.* The gut microbial metabolite trimethylamine N-oxide aggravates GVHD by inducing M1 macrophage polarization in mice. *Blood* **136**, 501–515 (2020).
94. Fujiwara, H. *et al.* Microbial metabolite sensor GPR43 controls severity of experimental GVHD. *Nat Commun* **9**, 3674 (2018).
95. Mathewson, N. D. *et al.* Gut microbiome-derived metabolites modulate intestinal epithelial cell damage and mitigate graft-versus-host disease. *Nat Immunol* **17**, 505–513 (2016).
96. Takahashi, D. *et al.* Microbiota-derived butyrate limits the autoimmune response by promoting the differentiation of follicular regulatory T cells. *eBioMedicine* **58**, (2020).
97. Weber, D. *et al.* Low urinary indoxyl sulfate levels early after transplantation reflect a disrupted microbiome and are associated with poor outcome. *Blood* **126**, 1723–1728 (2015).
98. Swimm, A. *et al.* Indoles derived from intestinal microbiota act via type I interferon signaling to limit graft-versus-host disease. *Blood* **132**, 2506–2519 (2018).
99. Li, X. *et al.* Tyrosine supplement ameliorates murine aGVHD by modulation of gut microbiome and metabolome. *eBioMedicine* **61**, (2020).

100. Toubai, T. *et al.* Host NLRP6 exacerbates graft-versus-host disease independent of gut microbial composition. *Nat Microbiol* **4**, 800–812 (2019).

101. Jasperson, L. K. *et al.* Indoleamine 2,3-dioxygenase is a critical regulator of acute graft-versus-host disease lethality. *Blood* **111**, 3257–3265 (2008).

102. Liu, M., Nieuwdorp, M., de Vos, W. M. & Rampanelli, E. Microbial Tryptophan Metabolism Tunes Host Immunity, Metabolism, and Extraintestinal Disorders. *Metabolites* **12**, 834 (2022).

103. Shaw, C., Hess, M. & Weimer, B. C. Microbial-Derived Tryptophan Metabolites and Their Role in Neurological Disease: Anthranilic Acid and Anthranilic Acid Derivatives. *Microorganisms* **11**, 1825 (2023).

104. Flint, H. J., Scott, K. P., Duncan, S. H., Louis, P. & Forano, E. Microbial degradation of complex carbohydrates in the gut. *Gut Microbes* **3**, 289–306 (2012).

105. Chinda, D. *et al.* Spatial distribution of live gut microbiota and bile acid metabolism in various parts of human large intestine. *Sci Rep* **12**, 3593 (2022).

106. Yang, K. *et al.* Distribution of gut microbiota across intestinal segments and their impact on human physiological and pathological processes. *Cell Biosci* **15**, 47 (2025).

107. Bansal, T., Alaniz, R. C., Wood, T. K. & Jayaraman, A. The bacterial signal indole increases epithelial-cell tight-junction resistance and attenuates indicators of inflammation. *Proceedings of the National Academy of Sciences* **107**, 228–233 (2010).

108. Venkatesh, M. *et al.* Symbiotic Bacterial Metabolites Regulate Gastrointestinal Barrier Function via the Xenobiotic Sensor PXR and Toll-like Receptor 4. *Immunity* **41**, 296–310 (2014).

109. Wlodarska, M. *et al.* Indoleacrylic Acid Produced by Commensal Peptostreptococcus Species Suppresses Inflammation. *Cell Host & Microbe* **22**, 25-37.e6 (2017).

110. Hubbard, T. D. *et al.* Adaptation of the human aryl hydrocarbon receptor to sense microbiota-derived indoles. *Sci Rep* **5**, 12689 (2015).

111. Williams, B. B. *et al.* Discovery and Characterization of Gut Microbiota Decarboxylases that Can Produce the Neurotransmitter Tryptamine. *Cell Host & Microbe* **16**, 495–503 (2014).

112. Lamas, B. *et al.* CARD9 impacts colitis by altering gut microbiota metabolism of tryptophan into aryl hydrocarbon receptor ligands. *Nat Med* **22**, 598–605 (2016).

113. Jansson, J. *et al.* Metabolomics Reveals Metabolic Biomarkers of Crohn's Disease. *PLOS ONE* **4**, e6386 (2009).

114. Jasperson, L. K. *et al.* Inducing the tryptophan catabolic pathway, indoleamine 2,3-dioxygenase (IDO), for suppression of graft-versus-host disease (GVHD) lethality. *Blood* **114**, 5062–5070 (2009).

115. Landfried, K. *et al.* Tryptophan catabolism is associated with acute GVHD after human allogeneic stem cell transplantation and indicates activation of indoleamine 2,3-dioxygenase. *Blood* **118**, 6971–6974 (2011).

116. Lindner, S. *et al.* Altered microbial bile acid metabolism exacerbates T cell-driven inflammation during graft-versus-host disease. *Nat Microbiol* **9**, 614–630 (2024).

117. Ding, L., Yang, L., Wang, Z. & Huang, W. Bile acid nuclear receptor FXR and digestive system diseases. *Acta Pharm Sin B* **5**, 135–144 (2015).

118. Wan, Y.-J. Y. & Sheng, L. Regulation of bile acid receptor activity. *Liver Res* **2**, 180–185 (2018).

119. Hanafi, N. I., Mohamed, A. S., Sheikh Abdul Kadir, S. H. & Othman, M. H. D. Overview of Bile Acids Signaling and Perspective on the Signal of Ursodeoxycholic Acid, the Most Hydrophilic Bile Acid, in the Heart. *Biomolecules* **8**, 159 (2018).

120. Campbell, C. *et al.* Bacterial metabolism of bile acids promotes generation of peripheral regulatory T cells. *Nature* **581**, 475–479 (2020).

121. Paik, D. *et al.* Human gut bacteria produce TH17-modulating bile acid metabolites. *Nature* **603**, 907–912 (2022).

122. Hang, S. *et al.* Bile acid metabolites control TH17 and Treg cell differentiation. *Nature* **576**, 143–148 (2019).

123. Islam, K. B. M. S. *et al.* Bile Acid Is a Host Factor That Regulates the Composition of the Cecal Microbiota in Rats. *Gastroenterology* **141**, 1773–1781 (2011).

124. Swann, J. R. *et al.* Systemic gut microbial modulation of bile acid metabolism in host tissue compartments. *PNAS* **108**, 4523–4530 (2011).

125. Song, X. *et al.* Microbial bile acid metabolites modulate gut RORγ+ regulatory T cell homeostasis. *Nature* **577**, 410–415 (2020).

126. Guzior, D. V. *et al.* Bile salt hydrolase acyltransferase activity expands bile acid diversity. *Nature* **626**, 852–858 (2024).

127. Mohanty, I. *et al.* The underappreciated diversity of bile acid modifications. *Cell* **187**, 1801-1818.e20 (2024).

128. Song, Z. *et al.* Taxonomic profiling and populational patterns of bacterial bile salt hydrolase (BSH) genes based on worldwide human gut microbiome. *Microbiome* **7**, 9 (2019).

129. Hylemon, P., Melone, P., Franklund, C., Lund, E. & Björkhem, I. Mechanism of intestinal 7 alpha-dehydroxylation of cholic acid: evidence that allo-deoxycholic acid is an inducible side-product. *Journal of Lipid Research* **32**, 89–96 (1991).

130. Pitt, H. A. & Nakeeb, A. Chapter 8 - Bile secretion and pathophysiology of biliary tract obstruction. in *Blumgart's Surgery of the Liver, Biliary Tract and Pancreas, 2-Volume Set (Sixth Edition)* (ed. Jarnagin, W. R.) 123-132.e1 (Elsevier, Philadelphia, 2017). doi:10.1016/B978-0-323-34062-5.00008-X.

131. Devlin, A. S. & Fischbach, M. A. A biosynthetic pathway for a prominent class of microbiota-derived bile acids. *Nature Chemical Biology* **11**, 685–690 (2015).

132. Mann, G., Mora, S., Madu, G. & Adegoke, O. A. J. Branched-chain Amino Acids: Catabolism in Skeletal Muscle and Implications for Muscle and Whole-body Metabolism. *Frontiers in Physiology* **12**, (2021).

133. Salazar, N., González, S., de los Reyes Gavilan, C. G. & Rios-Covian, D. Branched Short-Chain Fatty Acids as Biological Indicators of Microbiota Health and Links with Anthropometry. in *Biomarkers in Nutrition* (eds Patel, V. B. & Preedy, V. R.) 1–17 (Springer International Publishing, Cham, 2022). doi:10.1007/978-3-030-81304-8_4-1.

134. Topping, D. L. & Clifton, P. M. Short-Chain Fatty Acids and Human Colonic Function: Roles of Resistant Starch and Nonstarch Polysaccharides. *Physiological Reviews* **81**, 1031–1064 (2001).

135. Cummings, J. H., Pomare, E. W., Branch, W. J., Naylor, C. P. & Macfarlane, G. T. Short chain fatty acids in human large intestine, portal, hepatic and venous blood. https://doi.org/10.1136/gut.28.10.1221 (1987) doi:10.1136/gut.28.10.1221.

136. Miao, W. *et al.* Sodium Butyrate Promotes Reassembly of Tight Junctions in Caco-2 Monolayers Involving Inhibition of MLCK/MLC2 Pathway and Phosphorylation of PKCβ2. *International Journal of Molecular Sciences* **17**, 1696 (2016).

137. Singh, N. *et al.* Activation of Gpr109a, Receptor for Niacin and the Commensal Metabolite Butyrate, Suppresses Colonic Inflammation and Carcinogenesis. *Immunity* **40**, 128–139 (2014).

138. Wu, W. *et al.* Microbiota metabolite short-chain fatty acid acetate promotes intestinal IgA response to microbiota which is mediated by GPR43. *Mucosal Immunology* **10**, 946–956 (2017).

139. Makki, K., Deehan, E. C., Walter, J. & Bäckhed, F. The Impact of Dietary Fiber on Gut Microbiota in Host Health and Disease. *Cell Host & Microbe* **23**, 705–715 (2018).

140. McCullough, F., Cheung, J. & Miller, L. J. A Systematic Review Evaluating the Impact of Fibre Supplementation on Gut Health and Other Clinical Outcomes in Adults with

Haematological Malignancies During Haematopoietic Stem Cell Transplantation. *Nutrients* **17**, 2973 (2025).

141. Paredes, J. *et al.* Increased Fiber Intake Results in Better Overall Survival and Lower GI-aGVHD in Allo-HCT Recipients and Pre-Clinical Gvhd Models. *Blood* **144**, 259 (2024).

142. Zou, J. *et al.* Fiber-Mediated Nourishment of Gut Microbiota Protects against Diet-Induced Obesity by Restoring IL-22-Mediated Colonic Health. *Cell Host & Microbe* **23**, 41-53.e4 (2018).

143. Andermann, T. M. *et al.* A Fructo-Oligosaccharide Prebiotic Is Well Tolerated in Adults Undergoing Allogeneic Hematopoietic Stem Cell Transplantation: A Phase I Dose-Escalation Trial. *Transplantation and Cellular Therapy* **27**, 932.e1-932.e11 (2021).

144. Yoshifuji, K. *et al.* Prebiotics protect against acute graft-versus-host disease and preserve the gut microbiota in stem cell transplantation. *Blood Adv* **4**, 4607–4617 (2020).

145. Arain, A. & Matthiesen, C. Vitamin D deficiency and graft-versus-host disease in hematopoietic stem cell transplant population. *Hematology/Oncology and Stem Cell Therapy* **12**, 133–139 (2019).

146. Silva, F. *et al.* Effect of vitamin D treatment in chronic GVHD. *Bone Marrow Transplant* **46**, 1395–1397 (2011).

147. Holick, M. F. Resurrection of vitamin D deficiency and rickets. *J Clin Invest* **116**, 2062–2072 (2006).

148. Yigenoglu, T. N. *et al.* Is there a relationship between vitamin D levels and graft versus host disease? *Transfusion and Apheresis Science* **64**, 104054 (2025).

149. Makishima, M. *et al.* Vitamin D Receptor As an Intestinal Bile Acid Sensor. *Science* **296**, 1313–1316 (2002).

150. Koyama, M. & Hill, G. R. The primacy of gastrointestinal tract antigen-presenting cells in lethal graft-versus-host disease. *Blood* **134**, 2139–2148 (2019).

151. Rosenblatt, J. *et al.* Immunomodulatory effects of vitamin D: implications for GVHD. *Bone Marrow Transplant* **45**, 1463–1468 (2010).

152. Kong, J. *et al.* Novel role of the vitamin D receptor in maintaining the integrity of the intestinal mucosal barrier. *American Journal of Physiology-Gastrointestinal and Liver Physiology* **294**, G208–G216 (2008).

153. Kongsbak, M., Levring, T. B., Geisler, C. & von Essen, M. R. The Vitamin D Receptor and T Cell Function. *Front. Immunol.* **4**, (2013).

154. D'Ambrosio, D. *et al.* Inhibition of IL-12 production by 1,25-dihydroxyvitamin D3. Involvement of NF-kappaB downregulation in transcriptional repression of the p40 gene. *J Clin Invest* **101**, 252–262 (1998).

155. Srinivasan, S. *et al.* Vaginal Bacteria and Proinflammatory Host Immune Mediators as Biomarkers of Human Immunodeficiency Virus Acquisition Risk Among African Women. *The Journal of Infectious Diseases* **230**, 1444–1455 (2024).

156. Srinivasan, S. *et al.* Bacterial Communities in Women with Bacterial Vaginosis: High Resolution Phylogenetic Analyses Reveal Relationships of Microbiota to Clinical Criteria. *PLOS ONE* **7**, e37818 (2012).

157. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**, 581–583 (2016).

158. Kamp, K. J. *et al.* Bile Acids and Microbiome Among Individuals With Irritable Bowel Syndrome and Healthy Volunteers. *Biol Res Nurs* **23**, 65–74 (2021).

159. Ginos, B. N. R. *et al.* Circulating bile acids in healthy adults respond differently to a dietary pattern characterized by whole grains, legumes and fruits and vegetables compared to a diet high in refined grains and added sugars: A randomized, controlled, crossover feeding study. *Metabolism* **83**, 197–204 (2018).

160. Raftery, D. *Mass Spectrometry in Metabolomics: Methods and Protocols*. (2014).

161. Sarafian, M. H. *et al.* Bile Acid Profiling and Quantification in Biofluids Using Ultra-Performance Liquid Chromatography Tandem Mass Spectrometry. *Anal. Chem.* **87**, 9662–9670 (2015).

162. Funabashi, M. *et al.* A metabolic pathway for bile acid dehydroxylation by the gut microbiome. *Nature* **582**, 566–570 (2020).

163. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).

164. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–410 (1990).

165. Python Language Reference. Python Software Foundation.

166. pandas - Python Data Analysis Library. https://pandas.pydata.org/.

167. NumPy. https://numpy.org/.

168. openpyxl - A Python library to read/write Excel 2010 xlsx/xlsm files — openpyxl 3.1.3 documentation. https://openpyxl.readthedocs.io/en/stable/.

169. Josef Perktold *et al.* statsmodels/statsmodels: Release 0.14.2. Zenodo https://doi.org/10.5281/ZENODO.593847 (2024).

170. SciPy. https://scipy.org/.

171. scikit-learn: A set of python modules for machine learning and data mining.

172. Home — scikit-bio 0.7.2 documentation. https://scikit.bio/docs/latest/index.html.

173. skbio.stats.distance.permanova — scikit-bio 0.7.3-dev documentation. https://scikit.bio/docs/dev/generated/skbio.stats.distance.permanova.html.

174. Waskom, M. seaborn: statistical data visualization. *JOSS* **6**, 3021 (2021).

175. Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* **9**, 90–95 (2007).

176. McMurdie, P. J. & Holmes, S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLOS ONE* **8**, e61217 (2013).

177. Barnett, D. J. m, Arts, I. C. w & Penders, J. microViz: an R package for microbiome data visualization and statistics. *Journal of Open Source Software* **6**, 3201 (2021).

178. Make ggplot2 Graphics Interactive. https://davidgohel.github.io/ggiraph/index.html.

179. Wickham, H. The Split-Apply-Combine Strategy for Data Analysis. *Journal of Statistical Software* **40**, 1–29 (2011).

180. CRAN: Package RColorBrewer. https://cran.r-project.org/web/packages/RColorBrewer/index.html.

181. A Grammar of Data Manipulation. https://dplyr.tidyverse.org/.

182. Create Elegant Data Visualisations Using the Grammar of Graphics. https://ggplot2.tidyverse.org/.

183. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.

184. Kolde, R. pheatmap: Pretty Heatmaps. (2025).

185. Ooms [aut, J., cre & details, J. M. (Author of libxlsxwriter (see A. and C. files for details)) writexl author. writexl: Export Data Frames to Excel 'xlsx' Format. (2025).

186. Boogaart, K. G. van den, Tolosana-Delgado, R. & Bren, M. compositions: Compositional Data Analysis. (2025).

187. Wickham, H. *et al.* readxl: Read Excel Files. (2025).

188. Wickham, H., Henry, L., Software, P., PBC [cph & fnd. purrr: Functional Programming Tools. (2026).

189. Pedersen, T. L. & RStudio. ggraph: An Implementation of Grammar of Graphics for Graphs and Networks. (2025).

190. Pedersen, T. L. tidygraph: A Tidy API for Graph Manipulation. (2024).

191. Wickham, H. *et al.* tidyr: Tidy Messy Data. (2025).

192. Wickham, H. *et al.* Welcome to the Tidyverse. *Journal of Open Source Software* **4**, 1686 (2019).

193. Li, W. & Li, J. scImpute: Accurate and robust imputation of single-cell RNA sequencing data. (2018).

194. Nalle, S. C. *et al.* Graft-versus-host disease propagation depends on increased intestinal epithelial tight junction permeability. *Journal of Clinical Investigation* **129**, 902–914 (2019).

195. Ridlon, J. M., Harris, S. C., Bhowmik, S., Kang, D.-J. & Hylemon, P. B. Consequences of bile salt biotransformations by intestinal bacteria. *Gut Microbes* **7**, 22–39 (2016).

196. Daniel, S. L. & Ridlon, J. M. Clostridium scindens: history and current outlook for a keystone species in the mammalian gut involved in bile acid and steroid metabolism. *FEMS Microbiology Reviews* **49**, fuaf016 (2025).

197. Hamilton, J. P. *et al.* Human cecal bile acids: concentration and spectrum. *American Journal of Physiology-Gastrointestinal and Liver Physiology* **293**, G256–G263 (2007).

198. Vijayvargiya, P. & Camilleri, M. Current Practice in the Diagnosis of Bile Acid Diarrhea. *Gastroenterology* **156**, 1233–1238 (2019).

199. Lee, J. W. *et al.* Formation of secondary allo-bile acids by novel enzymes from gut Firmicutes. *Gut Microbes* **14**, 2132903 (2022).

200. [Clostridium] scindens G10 DNA, complete genome. (2021).

201. Dysosmobacter welbionis k43-0107-D3 DNA, complete genome. (2024).

202. [Ruminococcus] torques isolate RtorquesLFYP15 genome assembly, contig: 14, whole genome shotgun sequence. (2019).

203. Roseburia intestinalis strain L1 82 chromosome, complete genome. (2022).

204. Faecalibacterium prausnitzii A2-165 genomic scaffold Scfld0, whole genome shotgun sequence. (2015).

205. Gemmiger formicilis strain ATCC 27749 genome assembly, contig: EI45DRAFT_scaffold00013.13, whole genome shotgun sequence. (2017).

206. Jones, B. V., Begley, M., Hill, C., Gahan, C. G. M. & Marchesi, J. R. Functional and comparative metagenomic analysis of bile salt hydrolase activity in the human gut microbiome. *PNAS* **105**, 13580–13585 (2008).

207. Daly, J. W., Keely, S. J. & Gahan, C. G. M. Functional and Phylogenetic Diversity of BSH and PVA Enzymes. *Microorganisms* **9**, 732 (2021).

208. O'Flaherty, S., Briner Crawley, A., Theriot, C. M. & Barrangou, R. The Lactobacillus Bile Salt Hydrolase Repertoire Reveals Niche-Specific Adaptation. *mSphere* **3**, 10.1128/msphere.00140-18 (2018).

209. Rimal, B. *et al.* Bile salt hydrolase catalyses formation of amine-conjugated bile acids. *Nature* **626**, 859–863 (2024).

210. Lambert, J. M., Bongers, R. S., de Vos, W. M. & Kleerebezem, M. Functional Analysis of Four Bile Salt Hydrolase and Penicillin Acylase Family Members in Lactobacillus plantarum WCFS1. *Applied and Environmental Microbiology* **74**, 4719–4726 (2008).

211. Liang, L. *et al.* A Comprehensive Genome Survey Provides Novel Insights into Bile Salt Hydrolase (BSH) in Lactobacillaceae. *Molecules* **23**, 1157 (2018).

212. Franz, C. M., Specht, I., Haberer, P. & Holzapfel, W. H. Bile salt hydrolase activity of Enterococci isolated from food: screening and quantitative determination. *J Food Prot* **64**, 725–729 (2001).

213. Tanaka, H., Hashiba, H., Kok, J. & Mierau, I. Bile Salt Hydrolase of Bifidobacterium longum—Biochemical and Genetic Characterization. *Applied and Environmental Microbiology* **66**, 2502 (2000).

214. Begley, M., Sleator, R. D., Gahan, C. G. M. & Hill, C. Contribution of Three Bile-Associated Loci, bsh, pva, and btlB, to Gastrointestinal Persistence and Bile Tolerance of Listeria monocytogenes. *Infection and Immunity* **73**, 894–904 (2005).

215. Arai, S. *et al.* Increasing Incidence of Chronic Graft-versus-Host Disease in Allogeneic Transplantation: A Report from the Center for International Blood and Marrow Transplant Research. *Biology of Blood and Marrow Transplantation* **21**, 266–274 (2015).

216. Storb, R. *et al.* Allogeneic Hematopoietic Cell Transplantation following Minimal Intensity Conditioning: Predicting Acute Graft-versus-Host Disease and Graft-versus-Tumor Effects. *Biology of Blood and Marrow Transplantation* **19**, 792–798 (2013).

217. Staffas, A., Burgos da Silva, M. & van den Brink, M. R. M. The intestinal microbiota in allogeneic hematopoietic cell transplant and graft-versus-host disease. *Blood* **129**, 927–933 (2017).

218. Andermann, T. M. *et al.* The Microbiome and Hematopoietic Cell Transplantation: Past, Present, and Future. *Biology of Blood and Marrow Transplantation* **24**, 1322–1340 (2018).

219. Shono, Y. & van den Brink, M. R. M. Gut microbiota injury in allogeneic haematopoietic stem cell transplantation. *Nat Rev Cancer* **18**, 283–295 (2018).

220. Sinakos, E. *et al.* Bile acid changes after high-dose ursodeoxycholic acid treatment in primary sclerosing cholangitis: relation to disease progression. *Hepatology* **52**, 197–203 (2010).

221. Wang, H., Chen, J., Hollister, K., Sowers, L. C. & Forman, B. M. Endogenous Bile Acids Are Ligands for the Nuclear Receptor FXR/BAR. *Molecular Cell* **3**, 543–553 (1999).

222. Kawamata, Y. *et al.* A G Protein-coupled Receptor Responsive to Bile Acids *. *Journal of Biological Chemistry* **278**, 9435–9440 (2003).

223. Guo, C. *et al.* Bile Acids Control Inflammation and Metabolic Disorder through Inhibition of NLRP3 Inflammasome. *Immunity* **45**, 944 (2016).

224. Haselow, K. *et al.* Bile acids PKA-dependently induce a switch of the IL-10/IL-12 ratio and reduce proinflammatory capability of human macrophages. *Journal of Leukocyte Biology* **94**, 1253–1264 (2013).

225. Perino, A. *et al.* TGR5 reduces macrophage migration through mTOR-induced C/EBPβ differential translation. *J Clin Invest* **124**, 5424–5436 (2014).

226. Campbell, C. *et al.* FXR mediates T cell-intrinsic responses to reduced feeding during infection. *Proc Natl Acad Sci U S A* **117**, 33446–33454 (2020).

227. Vavassori, P., Mencarelli, A., Renga, B., Distrutti, E. & Fiorucci, S. The Bile Acid Receptor FXR Is a Modulator of Intestinal Innate Immunity1. *The Journal of Immunology* **183**, 6251–6261 (2009).

228. Jones, B. A. *et al.* Farnesoid X receptor prevents neutrophil extracellular traps via reduced sphingosine-1-phosphate in chronic kidney disease. *American Journal of Physiology-Renal Physiology* **325**, F792–F810 (2023).

229. Yi, T. *et al.* Reciprocal differentiation and tissue-specific pathogenesis of Th1, Th2, and Th17 cells in graft-versus-host disease. *Blood* **114**, 3101–3112 (2009).

230. Gartlan, K. H. *et al.* Th17 plasticity and transition toward a pathogenic cytokine signature are regulated by cyclosporine after allogeneic SCT. *Blood Advances* **1**, 341–351 (2017).

231. Bettelli, E. *et al.* Reciprocal developmental pathways for the generation of pathogenic effector TH17 and regulatory T cells. *Nature* **441**, 235–238 (2006).

232. Edinger, M. *et al.* CD4+CD25+ regulatory T cells preserve graft-versus-tumor activity while inhibiting graft-versus-host disease after bone marrow transplantation. *Nat Med* **9**, 1144–1150 (2003).

233. Gagliani, N. *et al.* Th17 cells transdifferentiate into regulatory T cells during resolution of inflammation. *Nature* **523**, 221–225 (2015).

234. Koyama, M. *et al.* MHC Class II Antigen Presentation by the Intestinal Epithelium Initiates Graft-versus-Host Disease and Is Influenced by the Microbiota. *Immunity* **51**, 885-898.e7 (2019).

235. Gadaleta, R. M. *et al.* Farnesoid X receptor activation inhibits inflammation and preserves the intestinal barrier in inflammatory bowel disease. *Gut* **60**, 463–472 (2011).

236. Fiorucci, S., Biagioli, M., Zampella, A. & Distrutti, E. Bile Acids Activated Receptors Regulate Innate Immunity. *Front. Immunol.* **9**, (2018).

237. Penna, G. & Adorini, L. 1α,25-Dihydroxyvitamin D3 Inhibits Differentiation, Maturation, Activation, and Survival of Dendritic Cells Leading to Impaired Alloreactive T Cell Activation. *The Journal of Immunology* **164**, 2405–2411 (2000).

238. Griffin, M. D. *et al.* Dendritic cell modulation by 1α,25 dihydroxyvitamin D3 and its analogs: A vitamin D receptor-dependent pathway that promotes a persistent state of immaturity in vitro and in vivo. *Proceedings of the National Academy of Sciences* **98**, 6800–6805 (2001).

239. Piemonti, L. *et al.* Vitamin D3 Affects Differentiation, Maturation, and Function of Human Monocyte-Derived Dendritic Cells1. *The Journal of Immunology* **164**, 4443–4451 (2000).

240. Fiorucci, S., Zampella, A., Ricci, P., Distrutti, E. & Biagioli, M. Immunomodulatory functions of FXR. *Molecular and Cellular Endocrinology* **551**, 111650 (2022).

241. McKenney, P. T. *et al.* Intestinal Bile Acids Induce a Morphotype Switch in Vancomycin-Resistant *Enterococcus* that Facilitates Intestinal Colonization. *Cell Host & Microbe* **25**, 695-705.e5 (2019).

242. do Nascimento, P. G. G. *et al.* Lithocholic acid and derivatives: Antibacterial activity. *Steroids* **104**, 8–15 (2015).

243. Cai, S.-Y. *et al.* Bile acids initiate cholestatic liver injury by triggering a hepatocyte-specific inflammatory response. *JCI Insight* **2**, (2017).

244. Out, C. *et al.* Prednisolone increases enterohepatic cycling of bile acids by induction of Asbt and promotes reverse cholesterol transport. *Journal of Hepatology* **61**, 351–357 (2014).

245. Esparteiro, D. *et al.* Serum bile acids profiles are altered without change of the gut microbiota composition following a seven-day prednisolone therapy in severe alcoholic hepatitis. *Gut Microbes* **16**, 2382767.

246. Hamilton, J. P. *et al.* Human cecal bile acids: concentration and spectrum. *Am J Physiol Gastrointest Liver Physiol* **293**, G256-263 (2007).

247. Azer, S. A. & Hasanato, R. Use of bile acids as potential markers of liver dysfunction in humans. *Medicine (Baltimore)* **100**, e27464 (2021).

248. Deleu, S., Machiels, K., Raes, J., Verbeke, K. & Vermeire, S. Short chain fatty acids and its producing organisms: An overlooked therapy for IBD? *eBioMedicine* **66**, (2021).

249. Rios-Covian, D., Gueimonde, M., Duncan, S. H., Flint, H. J. & de los Reyes-Gavilan, C. G. Enhanced butyrate formation by cross-feeding between Faecalibacterium prausnitzii and Bifidobacterium adolescentis. *FEMS Microbiol Lett* **362**, fnv176 (2015).

250. Duncan, S. H., Hold, G. L., Barcenilla, A., Stewart, C. S. & Flint, H. J. Roseburia intestinalis sp. nov., a novel saccharolytic, butyrate-producing bacterium from human faeces. *International Journal of Systematic and Evolutionary Microbiology* **52**, 1615–1620 (2002).

251. Zhu, Z., Zhu, L. & Jiang, L. Dynamic regulation of gut Clostridium-derived short-chain fatty acids. *Trends in Biotechnology* **40**, 266–270 (2022).

252. Ezaki, T. *et al.* Proposal of the genera Anaerococcus gen. nov., Peptoniphilus gen. nov. and Gallicola gen. nov. for members of the genus Peptostreptococcus. *International Journal of Systematic and Evolutionary Microbiology* **51**, 1521–1528 (2001).

253. Kim, J.-S. *et al.* Mediterraneibacter butyricigenes sp. nov., a butyrate-producing bacterium isolated from human faeces. *J Microbiol.* **57**, 38–44 (2019).

254. Barker, H. A., Kahn, J. M. & Hedrick, L. Pathway of lysine degradation in Fusobacterium nucleatum. *J Bacteriol* **152**, 201–207 (1982).

255. Schwiertz, A. *et al. Anaerostipes caccae* gen. nov., sp. nov., a New Saccharolytic, Acetate-utilising, Butyrate-producing Bacterium from Human Faeces. *Systematic and Applied Microbiology* **25**, 46–51 (2002).

256. Price, C. E. *et al.* Intestinal Bacteroides modulates inflammation, systemic cytokines, and microbial ecology via propionate in a mouse model of cystic fibrosis. *mBio* **15**, e03144-23 (2024).

257. Holmberg, S. M. *et al.* The gut commensal Blautia maintains colonic mucus function under low-fiber consumption through secretion of short-chain fatty acids. *Nat Commun* **15**, 3502 (2024).

258. den Besten, G. *et al.* The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. *J Lipid Res* **54**, 2325–2340 (2013).

259. Hijova, E. & Chmelarova, A. Short chain fatty acids and colonic health. *Bratisl Lek Listy* **108**, 354–358 (2007).

260. Wang, H.-B., Wang, P.-Y., Wang, X., Wan, Y.-L. & Liu, Y.-C. Butyrate enhances intestinal epithelial barrier function via up-regulation of tight junction protein Claudin-1 transcription. *Dig Dis Sci* **57**, 3126–3135 (2012).

261. Peng, L., Li, Z.-R., Green, R. S., Holzman, I. R. & Lin, J. Butyrate Enhances the Intestinal Barrier by Facilitating Tight Junction Assembly via Activation of AMP-Activated Protein Kinase in Caco-2 Cell Monolayers. *The Journal of Nutrition* **139**, 1619–1625 (2009).

262. Vinolo, M. A. R. *et al.* Suppressive effect of short-chain fatty acids on production of proinflammatory mediators by neutrophils. *The Journal of Nutritional Biochemistry* **22**, 849–855 (2011).

263. Pérez-Reytor, D., Puebla, C., Karahanian, E. & García, K. Use of Short-Chain Fatty Acids for the Recovery of the Intestinal Epithelial Barrier Affected by Bacterial Toxins. *Front. Physiol.* **12**, (2021).

264. Kaiko, G. E. *et al.* The Colonic Crypt Protects Stem Cells from Microbiota-Derived Metabolites. *Cell* **165**, 1708–1720 (2016).

265. Paciolla, C. *et al.* Valeric Acid: A Gut-Derived Metabolite as a Potential Epigenetic Modulator of Neuroinflammation in the Gut–Brain Axis. *Cells* **14**, 1823 (2025).

266. Kim, H., Jeon, B. S. & Sang, B.-I. An Efficient New Process for the Selective Production of Odd-Chain Carboxylic Acids by Simple Carbon Elongation Using Megasphaera hexanoica. *Sci Rep* **9**, 11999 (2019).

267. Caminero, A. *et al.* Differences in gluten metabolism among healthy volunteers, coeliac disease patients and first-degree relatives. *British Journal of Nutrition* **114**, 1157–1167 (2015).

268. Ecklu-Mensah, G. *et al.* Gut microbiota and fecal short chain fatty acids differ with adiposity and country of origin: the METS-microbiome study. *Nat Commun* **14**, 5160 (2023).

269. Wong, J. M. W., de Souza, R., Kendall, C. W. C., Emam, A. & Jenkins, D. J. A. Colonic health: fermentation and short chain fatty acids. *J Clin Gastroenterol* **40**, 235–243 (2006).

270. Canfora, E. E., Jocken, J. W. & Blaak, E. E. Short-chain fatty acids in control of body weight and insulin sensitivity. *Nat Rev Endocrinol* **11**, 577–591 (2015).

271. Cummings, J. H., Pomare, E. W., Branch, W. J., Naylor, C. P. & Macfarlane, G. T. Short chain fatty acids in human large intestine, portal, hepatic and venous blood. *Gut* **28**, 1221–1227 (1987).

272. Sun, Q., Jia, Q., Song, L. & Duan, L. Alterations in fecal short-chain fatty acids in patients with irritable bowel syndrome. *Medicine (Baltimore)* **98**, e14513 (2019).

273. Hosomi, K. *et al.* Oral administration of Blautia wexlerae ameliorates obesity and type 2 diabetes via metabolic remodeling of the gut microbiota. *Nat Commun* **13**, 4477 (2022).

274. Yang, N. *et al.* Tributyrin alleviates gut microbiota dysbiosis to repair intestinal damage in antibiotic-treated mice. *PLoS One* **18**, e0289364 (2023).

275. Yan, J. *et al.* Beneficial effect of the short-chain fatty acid propionate on vascular calcification through intestinal microbiota remodelling. *Microbiome* **10**, 195 (2022).

276. Guzior, D. V. *et al.* Bile salt hydrolase acyltransferase activity expands bile acid diversity. *Nature* 1–7 (2024) doi:10.1038/s41586-024-07017-8.

277. Quinn, R. A. *et al.* Global chemical effects of the microbiome include new bile-acid conjugations. *Nature* **579**, 123–129 (2020).

278. Pang, Z. *et al.* MetaboAnalystR 4.0: a unified LC-MS workflow for global metabolomics. *Nat Commun* **15**, 3675 (2024).

279. Li, S. *et al.* Predicting Network Activity from High Throughput Metabolomics. *PLoS Comput Biol* **9**, e1003123 (2013).

280. MetaboAnalyst. https://www.metaboanalyst.ca/MetaboAnalyst/ModuleView.xhtml.

281. Schymanski, E. L. *et al.* Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. *Environ. Sci. Technol.* **48**, 2097–2098 (2014).

282. Theodoridis, G. *et al.* What's in a name? Metabolite identification: challenges and pitfalls in untargeted metabolomics. *Metabolomics* **22**, 22 (2026).

283. choloylglycine hydrolase. *NCBI* https://www.ncbi.nlm.nih.gov/datasets/gene/57432353/.

284. Tanaka, H., Hashiba, H., Kok, J. & Mierau, I. Bile Salt Hydrolase of Bifidobacterium longum—Biochemical and Genetic Characterization. *Applied and Environmental Microbiology* **66**, 2502–2512 (2000).

285. Foley, M. H. *et al.* Lactobacillus bile salt hydrolase substrate specificity governs bacterial fitness and host colonization. *Proceedings of the National Academy of Sciences* **118**, e2017709118 (2021).

286. Kim, R. *et al.* An in vitro intestinal platform with a self-sustaining oxygen gradient to study the human gut/microbiome interface. *Biofabrication* **12**, 015006 (2019).

287. Zheng, D. *et al.* Comparative profiling of serum, urine, and feces bile acids in humans, rats, and mice. *Commun Biol* **7**, 641 (2024).

288. Heimesaat, M. M. *et al.* MyD88/TLR9 mediated immunopathology and gut microbiota dynamics in a novel murine model of intestinal graft-versus-host disease. *Gut* **59**, 1079–1087 (2010).

289. Riesner, K., Kalupa, M., Shi, Y., Elezkurtaj, S. & Penack, O. A preclinical acute GVHD mouse model based on chemotherapy conditioning and MHC-matched transplantation. *Bone Marrow Transplant* **51**, 410–417 (2016).

290. Nakagawa, S., Makino, I., Ishizaki, T. & Dohi, I. Dissolution of cholesterol gallstones by ursodeoxycholic acid. *Lancet* **2**, 367–369 (1977).

291. Kitani, K., Ohta, M. & Kanai, S. Tauroursodeoxycholate prevents biliary protein excretion induced by other bile salts in the rat. *Am J Physiol* **248**, G407-417 (1985).

292. Fried, R. H. *et al.* Ursodeoxycholic Acid Treatment of Refractory Chronic Graft-versus-Host Disease of the Liver. *Ann Intern Med* **116**, 624–629 (1992).

293. Baek, M. K. *et al.* Lithocholic acid upregulates uPAR and cell invasiveness via MAPK and AP-1 signaling in colon cancer cells. *Cancer Letters* **290**, 123–128 (2010).

294. Ito, E. *et al.* Sulfated bile acid is a host-derived ligand for MAIT cells. *Science Immunology* **9**, eade6924 (2024).

295. Chaudhari, S. N. *et al.* Bariatric surgery reveals a gut-restricted TGR5 agonist with anti-diabetic effects. *Nat Chem Biol* **17**, 20–29 (2021).

296. Carcinogens, N. R. C. (US) C. on C. T. of N. O. Methods for Evaluating Potential Carcinogens and Anticarcinogens. in *Carcinogens and Anticarcinogens in the Human Diet: A Comparison of Naturally Occurring and Synthetic Substances* (National Academies Press (US), 1996).

297. Ballerini, M. *et al.* A gut-on-a-chip incorporating human faecal samples and peristalsis predicts responses to immune checkpoint inhibitors for melanoma. *Nat. Biomed. Eng* **9**, 967–984 (2025).

298. Bang, L. L. *et al.* An anaerobic in vitro flow model for studying interactions at the gastrointestinal host–microbe interface. *npj Biofilms Microbiomes* **11**, 160 (2025).

299. Kim, R. *et al.* An in vitro intestinal platform with a self-sustaining oxygen gradient to study the human gut/microbiome interface. *Biofabrication* **12**, 015006 (2019).

300. Iamsawat, S. *et al.* Vitamin C stabilizes CD8+ iTregs and enhances their therapeutic potential in controlling murine GVHD and leukemia relapse. *Blood Adv* **3**, 4187–4201 (2019).