Referee #1

Comment 1: "*The authors provide a collection of >10,000 fecal microbiota samples longitudinal collected from >1,000 patients, along with some analyses of those specimens. Given the scope of the journal this review focuses on the dataset.*

First, this is a large and rich dataset with excellent potential for re-use. Metagenomes were sequenced on an Illumina MiSeq platform using a paired-end 250 × 250 reads. The dataset includes:

- *Counts of mapped 16S sequences*
- *16S qPCR data for total microbiome estimation.*
- *Metadata about hematopoietic cell transplantation*
- *Timing and route of drug administration*
- *Dates of positive blood cultures*
- *Repeated body temperature measures*
- *Daily measurements of white blood cells, platelets and red blood cells*
- *NCBI SRA accession numbers for the sequencing data*
- *Silva taxonomy tables for OTUs*

*The sequencing data are distributed through NCBI's Sequence Read Archive, and all other data through Figshare. I was able to easily download all these data as CSV files, load them into R, and join them based on sample ID and patient ID. It is a great improvement over the presentation of data and metadata in most publications that are not data-centric. I have only some moderate and minor comments.***"**

Author Response: We thank the referee for taking time to evaluate our data tables and manuscript. We are glad to know that the referee agrees our data is potentially useful.

Comment 2: "*Moderate: "Did the authors provide all of the information needed for others to reuse this dataset, or integrate it with other data?": I selected No here only because the SRR accessions are not publicly visible on SRA, and I didn't see a reviewer link to access them. I imagine these are just still not yet made public, but would want to make sure they are made public with at least minimal metadata on SRA with publication. The data in file tblASVsamples enables linking the sequence data to sampleID and patientID in the figshare metadata.***"**

Author Response: We apologize for the issue where the referee could not access some sequences via SRR accessions. We revised the *tblASVSamples* with accession numbers that we made sure are publicly visible. We also improved the manuscript to explain that some of the samples analyzed here have been submitted to SRA more than once under different projects; the data table provides a correct unique accession number to allow readers to access the raw sequencing data without having to navigate the multiple BioProjects we submitted throughout the years (Lines 61-65, 256-257).

Comment 3: "*The DADA2 taxonomic assignment is great because it doesn't involve OTU picking. But it does use different "Amplicon Sequence Variants" (ASVs) for each sample, so conversion to wide format and naming of taxa is needed for most analysis. It would help some users to also provide a taxa x samples matrix version of the ASV table.***"**

Author Response: We agree with the referee that a taxa-by-sample table of relative abundance may be convenient for readers who will reuse our data. We now provide such tables at different taxonomic levels (ASV, genus, family, order, class, phylum) and describe these tables in the manuscript (Lines 183-184).

Comment 4: "*Minor: The second fig share URL has an error (https://doi/org/10.6084/m9.figshare.12016989 - should be doi.org, not doi/org)***"**

Author Response: Thanks for pointing out the typo. We have merged all four folders into one collection and, therefore, the four figshare URLs (including the incorrect one) has been replaced by a single URL for the entire data collection (https://doi.org/10.6084/m9.figshare.c.5271128.v1).

Comment 5: "*Folder "samples" and Folder "taxonomy" – these are provided as files, not folders.***"**

Author Response: We had created folders for "samples" and "taxonomy" deliberately. Although there is a single file in each folder at the moment, we expect to expand the contents of these folders by cleaning up more data in the future. For example, the taxonomic classification of fungal ITS sequences may be saved to a different file in the "taxonomy" folder. We explained our reasons in the manuscript (Lines 170-172).

Referee #2

Comment 1: "*Nice repository model.***"**

Author Response: We thank the referee for reading and commenting on our manuscript.

Comment 2: "*Some comments: Have you added, in the repo, the episodes of GVHD? How many did you find? Have you identified some risk factors for them?***"**

Author Response: We appreciate the referee's interest in GVHD. We have not added GVHD data as detailed GVHD outcomes are not available for the entire patient cohort without vetting by manual chart review. Regretfully we do not have permission to publish such patient-level granular data on outcomes, even de-identified.

But note that members of our team have previously reported microbiome risk factors for GVHD-related mortality. Those factors include low alpha-diversity, low abundance of Blautia [1] and domination by *Enterococcus* [2] and that chronic GVHD is associated with high circulating concentrations of microbiome-derived short-chain fatty acids [3]. We expanded introduction to discuss these GVHD-microbiome links in the manuscript (Lines: 72-73) which were outside of the main focus of our original paper.

Comment 3: "*About your patients: have you registered other factors that influence gut microbiota composition (e.g. diet, delivery, breastfeeding, PPI use)?***"**

Author Response: We have previously reported drug exposure [4] and host-microbiome interaction [5] as two major drivers of gut microbiota dynamics in our allo-HSTC patients. Both drug administration records and counts of white and red blood cells have been released together with microbiome data in the current data collection.

We do have ongoing projects at our institute to study the associations of these factors such as diet with gut microbiota composition. In the revised manuscript, we acknowledge that factors like diet are not yet available, but we mention that they could be in the future (Lines 469-472).

Reference:

[1] Jenq, Robert R., et al. "Intestinal Blautia is associated with reduced death from graft-versus-host disease." Biology of Blood and Marrow Transplantation 21.8 (2015): 1373-1383.

[2] Stein-Thoeringer, C. K., et al. "Lactose drives Enterococcus expansion to promote graft-versus-host disease." Science 366.6469 (2019): 1143-1149.

[3] Markey, Kate A., et al. "Microbe-derived short chain fatty acids butyrate and propionate are associated with protection from chronic GVHD." Blood Journal (2020): blood-2019003369.

[4] Morjaria, Sejal, et al. "Antibiotic-induced shifts in fecal microbiota density and composition during hematopoietic stem cell transplantation." Infection and immunity 87.9 (2019): e00206-19.

[5] Schluter, Jonas, et al. "The gut microbiota is associated with immune cell dynamics in humans." Nature (2020): 1-5.