**FMT procedure**

Patient sample collection protocols were approved by the Memorial Sloan Kettering Cancer Center Institutional Review and Privacy Board (ClinicalTrials.gov identifier: NCT02269150). Briefly, patients’ stool was collected when they first entered the clinic. We chose autologous, as opposed to feces from a heterologous donor, because of potential safety concerns. The stool was tested for the presence of potential intestinal pathogens including *C. difficile* and frozen (−80°C). If a patient was randomized to receive treatment after engraftment of neutrophils, the thawed sample was re-administered via an enema. Due to the strenuous nature of this procedure, it was deemed unethical to administer a mock enema to control patients. Subjects whose pre–allo-HSCT feces demonstrated low microbial diversity (IS index < 2.0) or tested positive for the presence of an intestinal pathogen, for example, C. difficile, were excluded from randomization. After successful hematopoietic cell engraftment (three consecutive blood neutrophil counts ≥500 per mm3), subjects underwent testing of a fecal specimen collected after engraftment to determine the presence of the Bacteroidetes phylum via quantitative PCR. Individuals with low abundance of Bacteroidetes (<0.1% total 16S) were eligible to proceed to randomization and treatment. Eligible subjects were 1:1 randomized to undergo auto-FMT with the subject’s stored pre–allo-HSCT feces versus no fecal transplantation. Randomization was stratified by cord blood source versus non–cord blood source. Subjects could be randomized within a 28-day window after engraftment. Subjects who were critically ill or required prolonged microbiota-perturbing antibiotics through the designated 28-day period were excluded from randomization.

**Dynamic systems analyses**

*Covariates included in interval data*

For a given daily interval, a medication was considered present for at least part of the interval when it was administered on either endpoint. Administration events were obtained from parsing the institutional task data base which contains drug and treatment administrations performed on patients. For the microbiota and blood stream infection data, data from the end day were considered for the interval. Homeostatic feedback calculations used the geometric mean of the white blood cell counts between the two endpoints. We included only those covariates that were present during at least 10 intervals.

Our data comprises >1.6M recorded administrations of 806 different medications with durations provided. All patients analyzed in stage 2 had available medication records, but ~10% of the patients without microbiome data had missing medication records and/or incomplete metadata. In case of missing continuous metadata, missing values were filled with means.

The stem cell graft source is a major determinant of engraftment times and can affect recovery dynamics37,38, and we therefore included intercept terms for unmodified peripheral blood stem cell grafts (PBSC), bone marrow (BM), T-cell depleted graft (*ex‑vivo*) by CD34+selection (TCD) and cord blood (cord) in stage 1.

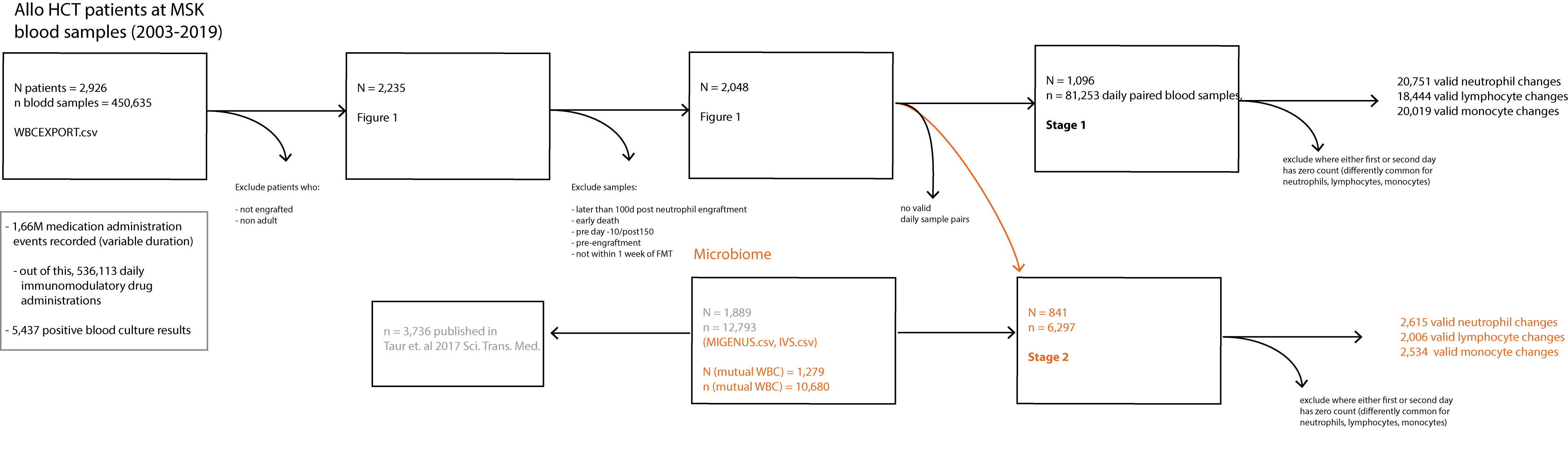
Patients received a variety of conditioning regimens comprised of various doses of chemotherapy and, in some cases, irradiation. There are dozens of conditioning regimes for HCT1. The standard approach in the allo-HCT field for observational studies is to categorize them by the Bacigalupo classification, which uses three graded categories from most to least intense1: Myeloablative (“ABLATIVE”), Reduced Intensity (“REDUCE”), and Nonmyeloablative (“NONABL”). We included the conditioning intensity as indicator variables.

*Data exclusion for white blood cell dynamic (stage 1 and 2)*

Our dynamics analyses focus on the daily changes in white blood cell counts from one day to another during recovery of the circulatory immune cell system. To analyze the kinetics of this reconstitution, we excluded data when (see also flow chart below):

* On the day of HCT, patients were younger than 18 years
* Patients did not engraft
* Patients had a second transplant within 100 days
* Samples were taken outside the window of 100 days starting from neutrophil engraftment
* In case there were multiple blood samples per patient and day, the one closest to noon was chosen
* A patient died within the first 3 months after HCT
* A sample was taken within 1 week of FMT

This data exclusion is encoded as two separate columns in the tidy data table WBC.csv, by a Boolean indicator column named “include”, and the column “exclude\_reason” of string type.



*Shotgun sequencing data processing*

We removed normal optical duplicates in paired FASTQ files using the clumpify.sh tool from the BBMap package (BBMap – Bushnell B. – <https://www.sourceforge.net/projects/bbmap/>), producing a pair of deduped read files. Using the bbduk.sh script in the BBMap package, we trimmmed the right and left side of a read in a pair to Q10 using the Phred algorithm. A pair of reads was dropped if any one of them has a length shorter than 51 nucleotides after trimming. We trim 3’-end adapters using a kmer of length 31, and a shorter kmer of 9 at the other end of the read. One mismatch was allowed in this process, and we allowed adapter trimming based on pair overlap detection (which does not require known adapter sequences) using the ‘tbo’ parameter. We used the ‘tpe’ parameter to trim the pair of reads to the same length.

Removal of human contamination was done using Kneaddata with paired-end reads, employing BMTagger. The BMTagger database was built with human genome assembly GRCh38. After decontamination, the paired-end reads were concatenated to a single FASTQ file as the input for functional profiling with the Humann2 pipeline (main methods).