Title: Cell-free RNA Liquid Biopsy to Monitor Hematopoietic Stem Cell Transplantation

Authors: Conor Loy[1,2], Matthew P. Cheng[3,4], Isabel H. Gonzalez-Bocco[3,4], Joan Lenz[1], Emma Belcher[1], Andrew Bliss[1], Daniel Eweis-LaBolle[2], Tinyi Chu[1], Jerome Ritz[3,5], Iwijn De Vlaminck[1]*

Affiliations:

[1]Meinig School of Biomedical Engineering, Cornell University; Ithaca, 14850, USA.

[2]Department of Molecular Biology and Genetics, Cornell University; Ithaca, 14853, USA.

[3]Department of Medical Oncology, Dana-Farber Cancer Institute; Boston, 02215, USA

[4]Division of Infectious Disease, Brigham and Women's Hospital; Boston, 02215, USA

[5]Department of Medicine, Harvard Medical School, Boston, 02115, USA

*Corresponding author. Email: vlaminck@cornell.edu

Abstract: Hematopoietic Stem Cell Transplantation (HSCT) is one of the oldest and most successful immunotherapies. Yet, despite long-standing success in the use of HSCT for the treatment of blood cancers and severe immune disorders, monitoring post-transplant complications remains a challenge due to a lack of informative diagnostic tests. Here, we investigate the utility of cell-free RNA (cfRNA) in plasma as a liquid biopsy to monitor allogeneic HSCT recipients during and after treatment. We assayed longitudinal samples from 92 HSCT recipients by cfRNA sequencing and show that cfRNA provides insight into treatment and recovery trajectories, immune dynamics in response to transplantation, infection, and solid-tissue injury associated with Graft-Versus-Host Disease. Collectively, our results provide support for the use of plasma cfRNA profiling to monitor complications of HSCT.

Main Text:

Hematopoietic stem cell transplantation (HSCT) is one of the first successful immunotherapies (*1*). In allogeneic HSCT, the recipient's hematopoietic and immune systems are rapidly replaced by donor cells. For patients with leukemia, the engraftment of donor immune cells contributes to a graft-versus-leukemia effect (*2*, *3*). However, the success of HSCT is limited by the frequent occurrence of complications, including graft-versus-host disease (GVHD), relapse, and infection (*4–6*). Most HSCT recipients experience at least one complication and complications can often co-occur, such as GVHD and infection (*7*). Complications can manifest with non-specific symptoms, such as rash or elevated liver enzymes, and progress rapidly, making early prediction and diagnosis critical. The current standard of care involves a battery of diagnostic tests including invasive tissue biopsies (*8–10*). There is a clear need for noninvasive molecular tests to monitor HSCT recipients and predict complications before they occur.

Cell-free DNA (cfDNA) and cell-free RNA (cfRNA) are released in plasma as a byproduct of active excretion and cell death, and therefore reflect cell turnover, immune activity, and solid tissue damage (*11*). Recent studies have shown that methylation profiling of cfDNA can be used to identify complications of allogeneic HSCT (*12–14*). Here, we performed plasma cfRNA profiling on longitudinal samples from allogeneic HSCT recipients and samples from healthy stem cell donors using RNA-sequencing. We found that cfRNA profiles from patients which are complication-free for 9 months after transplantation are similar, whereas cfRNA profiles of patients with recoveries complicated by frequent post allogeneic HSCT complications are all unique and very different from those of complication-free patients. We developed and leveraged multiple analysis techniques, including cell-type of origin deconvolution, distance-to-normal quantification, immune cell turnover estimation, and metagenomic classification, to show that cfRNA reflects immune reconstitution, ablative therapy treatment toxicity, immune activation and solid tissue damage related to GVHD, and differences in metagenomic dynamics likely in response to immunosuppression and inflammation. This study introduces new modalities to analyze cfRNA and demonstrates cfRNA as a highly versatile and informative analyte to monitor the clinical course of allogeneic HSCT recipients.

Results

Clinical cohort

To test the utility of cfRNA to monitor HSCT, we collected and analyzed 382 samples from 92 subjects who underwent allogeneic HSCT at the Dana-Farber Cancer Institute for a variety of conditions, including leukemia, lymphoma, and aplastic anemia (**Table S1**). Samples were collected longitudinally at predetermined time points before and after HSCT (prior to conditioning therapy, PR, at the day of infusion prior to the transplant, D0, at hematopoietic engraftment, E, and at 1 Month, 1M; 2 Month, 2M; 3 Month, 3M; and 6 Month, 6M after transplantation, **Fig. 1A** and **B**). In addition, we collected samples (n=16) at events of interest, such as BK virus reactivation (**Fig. 1B**, Methods), and we collected 18 samples from healthy stem cell donors to use as a reference. Complications were common within the first 9 months of HSCT in this cohort of recipients, including acute and chronic GVHD (aGVHD 53%, cGVHD 32%), BK virus reactivation (BKVR, 53%), cancer relapse (35%), and graft failure (5%). Recipients often had more than one complication, and the timing of complications differed between individuals and within complication categories (**Fig. 1C** and **D**). We grouped patients into two broad categories: those who experienced at least one major complication after engraftment but within 9 months of

transplant (aGVHD, cGVHD, BKV reactivation, CMV reactivation, EBV reactivation, serious infection, veno occlusive disease, hemolytic uremic syndrome, underlying disease relapse, or graft failure, COMP, n=82) and those who did not (noCOMP, n=10) (Methods).

Temporal dynamics of cfRNA in response to HSCT

We first sought to determine if cfRNA can measure a 'distance to normal'. To this end, we calculated the average correlation of each plasma cfRNA profile to those of the healthy stem cell donors (**Fig. 1E**, Methods). We found that cfRNA profiles at the engraftment time point are the least similar to normal profiles, followed by a time-dependent increase in correlation to the normal profiles. The cfRNA profiles from patients with a complication-free course follow a similar trajectory and become highly correlated to the plasma cfRNA profiles of the donors at the six-month time point. We attribute the remaining differences between patients without complications at the six-month time point and the healthy stem cell donors to the effects of immunosuppression and immune reconstitution (**Fig. S1A and B**). Strikingly, the cfRNA profiles of patients with complications are all unique in their own way and very different from those of complication-free patients and the healthy stem cell donors, which can be thought of as an analogy to the Anna Karenina principle in statistics(*15*) (**Fig. 1E** , **Fig. S1C**).

To gain a deeper understanding of the dynamics of plasma cfRNA profiles as a function of treatment stage, we quantified the cell types that contribute cfRNA to the mixture in plasma. To this end, we used BayesPrism(*16*) to computationally deconvolve cfRNA using the Tabula Sapiens single-cell atlas(*17*) as a reference. This analysis revealed rich dynamic changes in the cell-types-of-origin (CTO) of cfRNA in response to conditioning and stem cell infusion (**Fig. 1F**). For example, the proportion of hematopoietic-lineage-cell derived cfRNA was much lower at the day of infusion compared to the engraftment time point, likely reflecting the effects of pre-transplant conditioning and subsequent recovery of donor hematopoiesis.

We next compared the cfRNA CTO with whole blood cell counts obtained from paired samples (**Methods**). We observed similar trends between the cell count and CTO for lymphocytes and leukocytes, but not for neutrophils or monocytes (**Fig. 1G**). Given that blood cell counts measure live cells and cfRNA is a surrogate for dead cells, we suggest that this observed discrepancy can be explained by changes in cellular turnover: while the number of live cells may remain relatively unchanged, the lifespan and cell replacement rate may vary greatly for specific cell types. We calculated the cell turnover rate for neutrophils, monocytes, leukocytes, and lymphocytes by dividing the product of the total cfRNA concentration and the cfRNA fraction derived from each cell type by the cell counts for that cell type (Methods). This provided an estimate of cell turnover that is comparable between samples and can be trended over time. Using this measure, we observe a striking increase in the turnover rate of neutrophils and monocytes at the engraftment time point. This increase in turnover is expected given that neutrophils and monocytes are among the first cells to engraft. The turnover rate of neutrophils followed a bimodal distribution at the subsequent timepoints (**Fig. 1G)**, which we show below separates patients with an inflammatory immune response and those without. Similarly, the monocyte turnover rate displayed a bimodal distribution after the one-month timepoint. Based on these observations, we suggest cell turnover rates measured by cfRNA may provide insights into immune dynamics beyond what can be learned from blood cell counts alone.

cfRNA reflects bone marrow ablation and recolonization

We next examined the dynamic changes of cfRNA profiles in response to bone marrow ablation, stem cell infusion, and engraftment. We performed differential abundance analysis

comparing consecutive timepoints (PR, D0, and E, DESeq2, Methods). Each comparison yielded thousands of significantly differentially abundant genes (BH adjusted p-value < 0.05; PR vs D0 n=3019, D0 vs E n=11992, **Fig. 2A**, **File S1**). There was substantial overlap in the number of differentially abundant genes in both comparisons, and these overlapping genes can be organized into two groups based on whether they are more or less abundant at D0 as compared to the PR and E timepoints. Pathway enrichment analysis indicated that the genes with increased abundance at D0 are related to collagen degradation and cytokine storm signaling. In contrast, the genes with decreased abundance at D0 are related to macrophage IL-12 signaling and production, neutrophil extracellular net signaling, and NK cell signaling (Qiagen IPA, BH adjusted p-value < 0.05, **File S2**). These results indicate that genes elevated at D0 are related to toxicity from treatment and those less abundant at D0 are related to hematopoietic-lineage-cell expansion. In addition to changes at the gene level, we observed significant changes in the relative proportions of cfRNA derived from immune and solid tissue cells (Unpaired Wilcoxon test, BH adjusted p-value<0.05; PR vs D0 n=5, D0 vs E n=40, **Fig. 2B**, **File S3,** Methods). The proportion of cfRNA from immune cells was lowest at the day of infusion and highest at engraftment, reflecting the expected impact of the conditioning therapy. Conversely, the proportion of solid tissue derived cfRNA was highest at the day of infusion and lowest at engraftment, likely due to the toxicity of the HSCT conditioning regimens. These results show that cfRNA reflects the impact of conditioning therapy and infusion on immune and solid tissue cells.

Characterizing and predicting GVHD

GVHD is one of the most common and most severe complications of allogeneic HSCT(*6*). GVHD occurs when engrafted T cells attack the host's organs, which can lead to tissue damage and in severe cases, death. GVHD is further classified into acute GVHD (aGVHD), which typically occurs within the initial 100 days post-transplantation, and chronic GVHD (cGVHD), which develops later(*18*). We analyzed how cfRNA profiles of patients with and without GVHD differ by quantifying the distance to plasma cfRNA of stem cell donors, the degree of solid-tissue specific cfRNA, immune cell turnover, and by implementing RNA biomarker analysis (**Fig. 3A-E**, Methods).

We first compared plasma cfRNA profiles for patients with aGVHD against those without complications by measuring the distance of CTO profiles to those of the stem cell donors (**Fig. 3A,** Methods). We further stratified the aGVHD patient samples based on the timing relative to aGVHD diagnosis. To quantify the distance to normal, we developed a computational method to calculate distance, called cfRNAdist, that accounts for the variation of each cell type and co-variation between them (Methods). Using this metric, we found that cfRNA profiles of aGVHD patients were similar to 'normal' prior to transplant but distinctly farther from 'normal' at all post-transplant time points. This includes a marked difference at engraftment prior to aGVHD diagnosis as well as later time points after diagnosis. This observation indicates that cfRNA likely reflects both early signatures of aGVHD and organ damage caused by aGVHD.

Next, we explored whether cfRNA profiling can be used to pinpoint organ-specific damage due to acute and chronic GVHD, focusing on skin, GI, and liver involvement. We stratified patients into four groups for each organ: GVHD+ with organ involvement, GVHD+ without organ involvement, GVHD-, and healthy stem cell donors. We then compared the maximum fraction of relevant cell types for each organ across the sampling time course, which allowed us to control for time of diagnosis (**Fig. 3B**, **File S4**, Methods). Doing so, we observed a significantly higher burden of intrahepatic cholangiocyte-derived cfRNA for patients with liver GVHD, and salivary gland and intestinal stem cell-derived cfRNA for patients with GI GVHD. Notably, we did not observe

tissue-specific cfRNA for patients with skin GVHD. These results indicate that cfRNA reflects organ-specific damage in patients with acute and chronic GVHD.

We next asked whether cfRNA also captures the immune dynamics associated with GVHD prior to diagnosis. To this end, we quantified the turnover rate of lymphocytes, monocytes, neutrophils, and total leukocytes using the approach described above (**Fig. 3C**). To account for the impact of GVHD treatment on immune dynamics, we only used samples collected before diagnosis. We observed elevated immune cell turnover associated with both aGVHD and cGVHD, in particular: elevated leukocyte turnover at Engraftment in aGVHD patients, elevated lymphocyte turnover at the 3M timepoint in patients with late onset aGVHD, elevated monocyte turnover at the 3M timepoint for patients with late onset aGVHD and cGVHD, and elevated leukocyte and neutrophil turnover at the 6M time point for cGVHD patients (**Fig. 3D**, **Fig. S2A**). These findings highlight the immune activity preceding diagnosis and manifestation of GVHD symptoms.

Last, we tested the potential role of cfRNA in the early prediction of aGVHD. We compared cfRNA profiles of samples collected at the engraftment time point from patients who developed aGVHD and those who did not. Differential abundance analysis identified 118 genes significantly differentially abundant between the two groups (DESeq2, BH adjusted p-value < 0.05) (**Fig. 3E**, **File S5**). Pathway analysis revealed enrichment of interferon-alpha/beta and gamma signaling, pathways known to be associated with aGVHD initiation (**Fig. 3F**, **File S6**). Further, the top genes significantly elevated in the aGVHD group have previously been implicated in T cell immune receptor regulation (CD300E), inflammation (CLC), interferon activity (GBP1, IFI35) and cell migration (PRRG4) (**Fig. 3G**). While a larger sample size is needed to test this further, the observed differences and related enriched pathways indicate that cfRNA has the potential to enable prediction of aGVHD as early as at the engraftment timepoint.

cfRNA Metagenomics

Allogeneic HSCT recipients are highly immunocompromised and consequently susceptible to infection(*4*). To test whether cfRNA from viruses and bacteria can be identified in plasma, we mapped all sequences that did not align to the human genome to a microbial reference database (**Methods**). We first characterized the plasma cfRNA virome of the healthy stem cell donors to establish a healthy baseline. We found that four non-mammalian viral families accounted for 98% of all viral reads in these samples (*Siphoviridae*, *Tospoviridae*, *Baculoviridae*, and *Alphafexiviridae*, **Fig. 4A**). These viruses are not known to infect humans; therefore, we suspected that these reads are due to contamination or incorrect classification of human reads. To test this, we correlated the abundance of those 4 viruses to the total RNA concentration and human alignment rates within samples (Methods). We observed strong negative correlation between 3 of the 4 viruses and total RNA concentration (*Tospoviridae*, *Baculoviridae*, and *Alphafexiviridae,* **Fig. S3A**). Such negative correlation is indicative of contamination: samples with lower RNA concentration are proportionally more susceptible to environmental contamination such as from reagents(*19*, *20*). The signal for Siphoviridae was strongly and positively correlated to human alignment rate, indicating misclassification of human reads (**Fig. S3B**). We therefore excluded *Siphoviridae*, *Tospoviridae*, *Baculoviridae*, *Alphafexiviridae*, and all other non-mammalian viral families from our analyses (Methods). After removal of this background, we were left with viral cfRNA from *Flaviviridae*, *Retroviridae*, *Annelloviridae*, and other mammalian viruses that are known to be present in healthy populations, giving us confidence that these reads are bona fide. We next compared the virome for COMP and noCOMP patient samples across the sampling time course and observed significantly higher viral counts at the preconditioning and day zero timepoints for all patient samples regardless of complication status (**Fig. 4B**). However, for

timepoints after stem cell infusion, we observed significantly higher viral cfRNA in plasma of COMP patients relative to noCOMP patients and the healthy stem cell donors.

We analyzed the longitudinal dynamics of the most abundant viral families in samples from patients with complications (*Flaviviridae*, *Anelloviridae*, *Herpesviridae*, and *Retroviridae*, **Fig. 4C**). The burden of *Flaviviridae* cfRNA was strikingly elevated throughout the post-transplant course for patients with complications compared to patients without complications (**Fig. 4C**). In addition, we observed reactivation of *Anelloviridae* in response to the immunosuppressive therapy, for both patients with and without complications (**Fig. 4C**). Similar blooms of *Anelloviridae* have previously been reported for solid-organ transplant patients through the lens of cfDNA(*12*). Last, we observed significantly higher levels of Herpesviridae and *Retroviridae* cfRNA at preconditioning and day of infusion relative to stem cell donors. For patients without complications, the burden of cfRNA from these viruses decreased post-transplantation, whereas for patients with complications, the burden of cfRNA from these viruses remained elevated (**Fig. 4C**). We further analyzed the burden of cfRNA from five viral genera known to affect HSCT recipients: *Betacoronavirus* (coronaviruses), *Betapolyomavirus* (BKV), *Cytomegalovirus* (CMV), *Lymphocryptovirus* (EBV), and *Mastadenovirus* (adenoviruses). We observed overall low viral burden, but higher detection rates in COMP patient samples (**Fig. S3C and D**).

Last, we quantified the bacterial component of the cfRNA microbiome. We observed elevated burden of bacterial cfRNA at preconditioning and the day of infusion relative to stem cell donors (**Fig. 4D**). After infusion we observed significantly higher bacterial cfRNA in COMP patient samples compared to noCOMP patient samples and healthy stem cell donors, while there was with no difference in bacterial cfRNA for noCOMP patient samples and healthy stem cell donors. The majority (60%) of the reads matched to a prokaryotic rRNA operon reference (**Fig. 4E**, Methods). These results show that cfRNA may be a relevant analyte for measuring metagenomic dynamics and that the burden of viral and bacterial cfRNA may be used as an indicator of complications and immunosuppression in HSCT.

Discussion

In this manuscript, we have explored the utility of cfRNA in the setting of allogeneic HSCT. We found that cfRNA reflects bone marrow ablation by conditioning therapy and re-colonization of hematopoietic-lineage-cells following transfusion of donor stem cells. Furthermore, we report cfRNA signatures of HSCT complications, including immune activation and tissue damage related to GVHD. The results presented here provide a proof-of-concept for cfRNA as a versatile analyte in the setting of HSCT and motivate future investigation of the use of cfRNA in immunotherapy more broadly.

Several recent studies highlight applications of RNA liquid biopsy in inflammatory and infectious disease, including pediatric inflammatory diseases, tuberculosis, COVID-19, sepsis, and non-alcoholic fatty liver disease.(*21–26*) There has been only one study on the use of cfRNA in HSCT, and it focused on the immediate post-transfusion period, showing that cfRNA reflects bone marrow reconstitution following allogeneic and autologous HSCT(*27*). This study did not investigate cfRNA dynamics in the context of HSCT complications over a long-time course, as we do here. We aimed to measure the dynamics of cfRNA throughout the allogeneic HSCT treatment course. This approach introduced multiple novel challenges, as HSCT recipients can experience diverse complications that often co-occur at any point during the treatment course(*7*). Unlike previous cfRNA research, which primarily focused on diagnosing a specific condition, the present study demanded a more comprehensive and longitudinal approach to capture the intricate interplay of immune reconstitution, solid organ damage, and microbial dynamics. To deal with challenges of highly heterogeneous sample groups, we used quantification of 'distance-to-normal'. This

measurement revealed that patients free of complications follow highly similar trajectories approaching normal, while patients who did experience complications follow unique, individualized trajectories. These observations follow the Anna Karenina principle in statistics and provide a theoretical framework for addressing similarly heterogeneous patient groups(*15*).

Blood cell counts are ubiquitously used in medicine to gauge immune system health and activity. Yet, blood cell counts alone offer an incomplete picture of immune activity. Increased immune cell death coupled with increased replacement may not lead to a net change in blood cell counts. Here, we suggest quantification of immune cell turnover, as was previously implemented for cfDNA, as a means to track reconstitution of hematopoietic-lineage-cells and measure immune activity prior to GVHD diagnosis(*28*). We show that the turnover of neutrophils after infusion changes dramatically whereas the cell count data does not. Furthermore, we show that turnover levels provide a leading indicator of GVHD, for example increased monocyte and leukocyte turnover at engraftment in patients that develop aGVHD. Further work is needed to compare immune turnover measurements derived from cfDNA and cfRNA, as well as to explore both measurements for other inflammatory diseases.

GVHD arises when the transplanted immune system recognizes the recipient's body as foreign, triggering an immune response that can result in extensive solid tissue damage and mortality(*29*). Beyond differences in 'distance to normal' and immune cell turnover, we observed elevated cfRNA derived from tissues that matched with patient specific GVHD organ involvement. Furthermore, RNA biomarker analysis revealed that cfRNA reflects activation of the interferon gamma pathway, a pivotal pathway in GVHD initiation, at the engraftment timepoint. Recent work using serum protein markers has also shown detection of aGVHD onset 14 days post- transfusion, adding further support for biomarker utility shortly after stem cell infusion(*30*). These results further highlight the potential of cfRNA to predict, diagnose, and characterize GVHD using a single test. Untargeted cfRNA sequencing can provide many layers of information, potentially detecting signal missed by specific biomarker tests such as the aGVHD biomarker assay and algorithm developed by the MAGIC consortium(*31*).

Patients undergoing HSCT are severely immunocompromised, leaving them prone to opportunistic infection(*29*). Here, we show that cfRNA can be used to track viruses and bacteria over the course of HSCT, and we show that the circulating RNA microbiome reflects both inflammation and immunosuppression. Strikingly, we found that patients who experience complications exhibit significantly higher cfRNA burden from viruses in the *Flaviviridae* family. *Flaviviridae* family viruses, including pegivirus, have been shown to have an immunomodulatory effect by reducing host immune activation and are present in serum extracellular vesicles(*32*). Recent studies have found that pegiviruses are present after HSCT at levels higher than in healthy individuals(*33*, *34*). Here, we show that pegivirus load after HSCT is associated with complications, which could possibly be due to immune dysregulation. This opens new avenues for exploration, such as using cfRNA as a multi-faceted metric of inflammation and immune activity, which can be paired with current modalities such as measurements of circulating C-reactive protein (CRP). More research is needed to elucidate the link between *Flaviviridae* viral load and patient outcomes. Another significant advantage of cfRNA is that we could effectively detect DNA virus activity, particularly from the *Anelloviridae* and *Herpesviridae* families, highlighting a key advantage over DNA based methods which exclusively detect DNA viruses. Lastly, our findings reveal discernible differences in prokaryotic read counts between patients with and without complications. We hypothesize that this difference may be associated with a loss of the integrity of the gut-vascular barrier(*35*). However, further investigation is needed to validate these findings.

In summary, this study introduces analysis techniques that may be broadly useful and highlights the promise of cfRNA to monitor HSCT. Beyond HSCT, this research holds

implications for broader applications in immunotherapy, leveraging HSCT as a foundational example of cellular therapy.

Materials and Methods

Ethics Statement

The Dana-Farber/Harvard Cancer Center's Office of Human Research Studies gave ethical approval for this work. All patients provided written informed consent.

Patient Cohort

Adult patients undergoing allogeneic HCT at Dana-Farber Cancer Institute were prospectively enrolled on a rolling basis from August 2018 to August 2019. Samples were collected longitudinally at predetermined time points: prior to conditioning therapy (Pre-Conditioning, PR), at the day of infusion prior to stem cell infusion (Day of Infusion, D0), at the time of myeloid engraftment (Engraftment, E) and at monthly intervals after the day of infusion (1 Month, 1M; 2 Month, 2M; 3 Month, 3M; 6 Month, 6M). This cohort was initially established to monitor BKV after transplant and additional sampling was done at the onset of BK-related symptoms, disease, or reactivation. In the case of two timepoints overlapping, the sample was preferentially labeled as engraftment or month ½/3/6 (in that order). Samples from healthy stem cell transplant donors were used as a healthy reference control.

Engraftment

Neutrophil engraftment was considered when blood samples contained an absolute neutrophil count greater or equal than 500 cell per microliter blood on two separate measurements. Median day of engraftment was 19 days after transplant.

Patient Categorization

Patients were categorized into three groups: healthy stem cell donors, early complication-free (noCOMP), and early complication-present (COMP). Healthy stem cell donors were individuals who provided stem cell products for patients in this study. Complication-present patients were HSCT recipients that were diagnosed with any of the following complications within 9 months of infusion: relapse, graft failure, acute GVHD (aGVHD), chronic GVHD (cGVHD), and BK viral relapse (BKVR). Complication-free patients were HCT recipients that were not diagnosed with a complication within 9 months of infusion. Medical records of each patient in the complication-free cohort were reviewed to confirm that none of these patients experienced CMV or EBV reactivation, veno occlusive disease (VOD), hemolytic uremic syndrome (HUS) or serious infections after engraftment. If one of those complications were experienced, that patient was categorized as COMP.

Complication Diagnosis Criteria

BK Viral Reactivation was identified using a commercial real-time PCR assay on urine and blood samples (Eurofins Viracor, LLC, Lenexa, KS). Relapse, graft failure, and GVHD: Medical records, including medical oncology notes, clinical laboratory tests and pathology reports were retrospectively reviewed. Data was captured in a clinical management system. Patients were monitored for relapse, graft failure, GVHD, infections and other complications during and after hematopoietic stem cell transplantation. Monitoring involved clinical histories, physical examinations, and regular laboratory assessments to track patients' progress and identify any

potential complications. Supplementary biopsies were conducted to confirm diagnoses of GVHD or relapse. Data collection was conducted retrospectively.

Sample Collection

Blood samples were collected through standard venipuncture in ethylenediaminetetraacetic acid tubes (Becton Dickinson, reference No. 366643). Plasma was extracted through blood centrifugation (2,000 rpm for 10 min using a Beckman Coulter Allegra 6R centrifuge) and stored in 0.5- to 2-mL aliquots at −80 °C. Plasma samples were shipped from the Dana-Farber Cancer Institute to Cornell University on dry ice.

Sample processing and sequencing

Samples were processed using methods described previously(*23*), with the following modifications: sequencing was done on an Illumina NextSeq or NovaSeq and reads were trimmed to 61 base pairs.

Sample quality filtering

Quality filtering was done as previously described(*23*), with the following modifications: Samples were removed from analysis if either the intron to exon ratio was greater than 3, if a sample had less than 10,000,000 reads, if the alignment rate was less than 10%, if the duplication rate was greater than 75%, or if the 5'-3' read alignment ratio bias was greater than 3 z-scores from the mean.

Correlation to healthy stem cell donors

Average Pearson correlation was calculated between each sample and each healthy stem cell donor sample using the top 500 most variable genes after variance stabilizing transformation in the Engraftment and 1, 2, 3, and 6 month follow up timepoint. For healthy stem cell donor samples, correlation to self was excluded from the average calculation. Variance stabilizing transformation was performed using the DESeq2 v*arianceStabilizingTransformation* function.

Cellular turnover calculation

Cell counts were obtained from blood samples collected at the same time as the plasma for cfRNA analysis. Blood samples were collected in EDTA containing tubes and subsequently analyzed at the clinical laboratories of Brigham and Women's Hospital or Dana-Farber Cancer Institute using automated image-based cell counters. Automated cell counters provided WBC differential counts as well as complete blood cell counts. Cell turnover was calculated using deconvolution results, cell counts, and cfRNA concentration. Quantification of cfRNA was done using library qubit values scaled by the number of PCR cycles and volume of input plasma.

$$cfRNA\ Quantification = \frac{Library\ Qubit}{Plasma\ Volume * 2^P}$$

$$P = Number\ of\ PCR\ cycles$$

Turnover for each cell type was calculated by taking the product of the deconvolution result and cfRNA concentration and dividing that by the number of cells.

$$Cell\ Type\ Turnover = \frac{(cfRNA\ CTO) * (cfRNA\ Quantification)}{Blood\ Cell\ Counts}$$

For lymphocytes, the sum of deconvolution results for B cells, T cells, plasma cells, NK cells, and innate lymphoid cells was used. For leukocytes, the sum of deconvolution results for all lymphocytes, along with neutrophils, basophils, mast cells, macrophages, mature conventional dendritic cells, and monocytes was used.

Ablation and recolonization analysis

Differentially abundant RNAs were calculated between PR, D0, and E time point samples using a negative binomial distribution implemented by DESeq2(*36*) and p-values were adjusted using the Benjamini-Hochberg method taking into account patient ID. RNAs significant in both PR vs D0 and D0 vs E comparisons (adjusted p-value < 0.05) were grouped by either being more or less abundant in D0 samples relative to PR and E. Adjusted p-values and log2 fold changes were averaged and used as input for Ingenuity Pathway analysis.

Differential cell type of origin analysis was performed using unpaired Wilcoxon tests and p-values were adjusted using the Benjamini-Hochberg method. Cell types that were significantly differentially abundant (adjusted p-value < 0.05) either in the PR vs D0 or D0 vs E comparison were retained.

Distance to healthy stem cell donors

We introduce a method to measure distance of cell type of origin profiles in cfRNA that accurately reflects the unique characteristics of cell type profiles by considering two critical sources of variation: the inherent variability among different cell types and the correlation patterns between them. This approach ensures that the metric: 1) down-weights the difference for cell types with higher variability and 2) avoids overestimating differences caused by cell types that co-vary within the control population. The process involves several steps: 1) Computing the mean and covariance matrix for cell type fractions from healthy stem cell donors. Given the typical scenario where the number of cell types exceeds the sample size of controls, we apply the nearPD function from the R Matrix package to project the covariance matrix to the nearest positive-definite matrix. 2) Computing the square root of the Mahalanobis distance between the cell type fraction of the query sample and the mean of the healthy stem cell donors under the inverse of the projected covariance matrix:

$$\mathrm{d}(\mathrm{x}, [y_1, \cdots, y_n]) = \sqrt{(\mathrm{x} - \mu)^T \Sigma^{-1} (x - \mu)}$$

where $x \in \Re^k$ denotes the cell type fraction of the query sample across k cell types, $y_1, \ldots, y_n \in \Re^k$ denote the cell type fraction of control samples, μ, and $\sum$ are the mean and projected covariance matrix of samples y. The implementation is based on the Mahalanobis function provided by the R matrixcalc package. The cfRNAdist R package is provided at https://github.com/Team-De-Vlaminck/cfRNAdist.git.

GVHD diagnosis and organ involvement

Acute and chronic GVHD were diagnosed and graded using NIH consensus criteria(*37*). We also categorized patients with acute and chronic GVHD based on involvement of skin, liver, and GI tract. Patients were classified into three groups for each organ: GVHD+ with organ involvement, GVHD+ without organ involvement, GVHD-, and compared with healthy stem cell donors. GVHD+ patients had either acute and/or chronic GVHD. Cell type of origin results for relevant cell types were compared using unpaired Wilcoxon tests. For liver involvement,

hepatocyte and intrahepatic cholangiocyte-derived cfRNA were compared. For skin involvement, melanocyte and keratinocyte-derived cfRNA were compared. For GI involvement, salivary gland, intestinal stem cell, acinary cell of salivary gland, intestinal epithelial cell, intestinal tuft cell, goblet cell, and duodenum glandular cell derived cfRNA were compared.

Metagenomic analysis

Reads unaligned to the human genome were taxonomically assigned using Kraken2 and the standard database with default parameters(*38*). Species abundances were calculated using Braken and formatted using a custom Python script(*39*). Counts per million (CPM) reads were calculated from Braken outputs by dividing the new estimated counts by the number of sequenced reads and multiplying the value by one million. Log10 scaled CPM values were correlated to human alignment rate and cfRNA quantification estimate. The following non-mammalian viral families were removed from analysis: *Siphoviridae*, *Baculoviridae*, *Tospoviridae*, *Alphaflexiviridae*, *Virgaviridae*, *Bromoviridae*, *Myoviridae*, *Solemoviridae*, *Genomoviridae*, *Endornaviridae*, *Tombusviridae.*

We randomly selected 1000 reads from each sample that Kraken2 assigned as prokaryotic in origin to determine the origin of prokaryotic reads. Next, we aligned to a prokaryotic rRNA operon database using nucleotide BLAST+ (num_alignments=1, qcov_hsp_perc=60, perc_identity=80)(*40*). Lastly, we calculated the fraction of reads that aligned successfully.

References

1. N. Cieri, K. Maurer, C. J. Wu, 60 Years Young: The Evolving Role of Allogeneic Hematopoietic Stem Cell Transplantation in Cancer Immunotherapy. *Cancer Res.* **81**, 4373–4384 (2021).
2. P. L. Weiden, N. Flournoy, E. D. Thomas, R. Prentice, A. Fefer, C. D. Buckner, R. Storb, Antileukemic Effect of Graft-versus-Host Disease in Human Recipients of Allogeneic-Marrow Grafts. *N. Engl. J. Med.* **300**, 1068–1073 (1979).
3. D. W. Barnes, M. J. Corp, J. F. Loutit, F. E. Neal, Treatment of murine leukaemia with X rays and homologous bone marrow; preliminary communication. *Br. Med. J.* **2**, 626–627 (1956).
4. C. Annaloro, F. Serpenti, G. Saporiti, G. Galassi, F. Cavallaro, F. Grifoni, M. Goldaniga, L. Baldini, F. Onida, Viral Infections in HSCT: Detection, Monitoring, Clinical Management, and Immunologic Implications. *Front. Immunol.* **11**, 569381 (2021).
5. F. Kreidieh, I. Abou Dalle, N. Moukalled, J. El-Cheikh, E. Brissot, M. Mohty, A. Bazarbachi, Relapse after allogeneic hematopoietic stem cell transplantation in acute myeloid leukemia: an overview of prevention and treatment. *Int. J. Hematol.* **116**, 330–340 (2022).
6. C. McDonald-Hyman, L. A. Turka, B. R. Blazar, Advances and challenges in immunotherapy for solid organ and hematopoietic stem cell transplantation. *Sci. Transl. Med.* **7** (2015).
7. D. A. Jacobsohn, G. B. Vogelsang, Acute graft versus host disease. *Orphanet J. Rare Dis.* **2**, 35 (2007).

8. B. Balakrishnan, U. P. Kulkarni, A. A. Pai, R. S. S. Illangeswaran, E. Mohanan, V. Mathews, B. George, P. Balasubramanian, Biomarkers for early complications post hematopoietic cell transplantation: Insights and challenges. *Front. Immunol.* **14**, 1100306 (2023).

9. S. Jodele, S. M. Davies, A. Lane, J. Khoury, C. Dandoy, J. Goebel, K. Myers, M. Grimley, J. Bleesing, J. El-Bietar, G. Wallace, R. S. Chima, Z. Paff, B. L. Laskin, Diagnostic and risk criteria for HSCT-associated thrombotic microangiopathy: a study in children and young adults. *Blood* **124**, 645–653 (2014).

10. T. Pandey, S. Maximin, P. Bhargava, Imaging of complications from hematopoietic stem cell transplant. *Indian J. Radiol. Imaging* **24**, 327–338 (2014).

11. C. Loy, L. Ahmann, I. De Vlaminck, W. Gu, Liquid Biopsy Based on Cell-Free DNA and RNA. *Annu. Rev. Biomed. Eng.* **26**, annurev-bioeng-110222-111259 (2024).

12. A. P. Cheng, M. P. Cheng, C. J. Loy, J. S. Lenz, K. Chen, S. Smalling, P. Burnham, K. M. Timblin, J. L. Orejas, E. Silverman, P. Polak, F. M. Marty, J. Ritz, I. De Vlaminck, Cell-free DNA profiling informs all major complications of hematopoietic cell transplantation. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2113476118 (2022).

13. B. Avni, D. Neiman, E. Shaked, O. Gal-Rosenberg, S. Grisariu, M. Kuzli, I. Avni, A. Fracchia, P. Stepensky, T. Zuckerman, A. Lev-Sagie, I. Fox-Fisher, S. Piyanzin, J. Moss, S. J. Salpeter, B. Glaser, R. Shemer, Y. Dor, Chronic graft-versus-host disease detected by tissue-specific cell-free DNA methylation biomarkers. *J. Clin. Invest.* **134**, e163541 (2024).

14. Y. Pang, T. E. Andargie, M. K. Jang, H. Kong, W. Park, T. Hill, N. Redekar, Y.-P. Fu, D. A. Parth, N. G. Holtzman, S. Z. Pavletic, S. Agbor-Enoh, Chronic graft-versus-host disease is characterized by high levels and distinctive tissue-of-origin patterns of cell-free DNA. *iScience* **26**, 108160 (2023).

15. W. W. B. Goh, L. Wong, Dealing with Confounders in Omics Analysis. *Trends Biotechnol.* **36**, 488–498 (2018).

16. T. Chu, Z. Wang, D. Pe'er, C. G. Danko, Cell type and gene expression deconvolution with BayesPrism enables Bayesian integrative analysis across bulk and single-cell RNA sequencing in oncology. *Nat. Cancer* **3**, 505–517 (2022).

17. Tabula Sapiens Consortium*, R. C. Jones, J. Karkanias, M. A. Krasnow, A. O. Pisco, S. R. Quake, J. Salzman, N. Yosef, B. Bulthaup, P. Brown, W. Harper, M. Hemenez, R. Ponnusamy, A. Salehi, B. A. Sanagavarapu, E. Spallino, K. A. Aaron, W. Concepcion, J. M. Gardner, B. Kelly, N. Neidlinger, Z. Wang, S. Crasta, S. Kolluru, M. Morri, A. O. Pisco, S. Y. Tan, K. J. Travaglini, C. Xu, M. Alcántara-Hernández, N. Almanzar, J. Antony, B. Beyersdorf, D. Burhan, K. Calcuttawala, M. M. Carter, C. K. F. Chan, C. A. Chang, S. Chang, A. Colville, S. Crasta, R. N. Culver, I. Cvijović, G. D'Amato, C. Ezran, F. X. Galdos, A. Gillich, W. R. Goodyer, Y. Hang, A. Hayashi, S. Houshdaran, X. Huang, J. C. Irwin, S. Jang, J. V. Juanico, A. M. Kershner, S. Kim, B. Kiss, S. Kolluru, W. Kong, M. E. Kumar, A. H. Kuo, R. Leylek, B. Li, G. B. Loeb, W.-J. Lu, S. Mantri, M. Markovic, P. L. McAlpine, A. de Morree, M. Morri, K. Mrouj, S. Mukherjee, T. Muser, P. Neuhöfer, T. D. Nguyen, K. Perez, R. Phansalkar, A. O. Pisco, N. Puluca, Z. Qi, P. Rao, H. Raquer-McKay, N. Schaum, B. Scott, B. Seddighzadeh, J. Segal, S. Sen, S. Sikandar, S. P. Spencer, L. C. Steffes, V. R. Subramaniam, A. Swarup, M. Swift, K. J. Travaglini, W. Van Treuren, E. Trimm, S. Veizades, S. Vijayakumar, K. C. Vo, S. K. Vorperian, W. Wang, H. N. W.

Weinstein, J. Winkler, T. T. H. Wu, J. Xie, A. R. Yung, Y. Zhang, A. M. Detweiler, H. Mekonen, N. F. Neff, R. V. Sit, M. Tan, J. Yan, G. R. Bean, V. Charu, E. Forgó, B. A. Martin, M. G. Ozawa, O. Silva, S. Y. Tan, A. Toland, V. N. P. Vemuri, S. Afik, K. Awayan, O. B. Botvinnik, A. Byrne, M. Chen, R. Dehghannasiri, A. M. Detweiler, A. Gayoso, A. A. Granados, Q. Li, G. Mahmoudabadi, A. McGeever, A. de Morree, J. E. Olivieri, M. Park, A. O. Pisco, N. Ravikumar, J. Salzman, G. Stanley, M. Swift, M. Tan, W. Tan, A. J. Tarashansky, R. Vanheusden, S. K. Vorperian, P. Wang, S. Wang, G. Xing, C. Xu, N. Yosef, M. Alcántara-Hernández, J. Antony, C. K. F. Chan, C. A. Chang, A. Colville, S. Crasta, R. Culver, L. Dethlefsen, C. Ezran, A. Gillich, Y. Hang, P.-Y. Ho, J. C. Irwin, S. Jang, A. M. Kershner, W. Kong, M. E. Kumar, A. H. Kuo, R. Leylek, S. Liu, G. B. Loeb, W.-J. Lu, J. S. Maltzman, R. J. Metzger, A. de Morree, P. Neuhöfer, K. Perez, R. Phansalkar, Z. Qi, P. Rao, H. Raquer-McKay, K. Sasagawa, B. Scott, R. Sinha, H. Song, S. P. Spencer, A. Swarup, M. Swift, K. J. Travaglini, E. Trimm, S. Veizades, S. Vijayakumar, B. Wang, W. Wang, J. Winkler, J. Xie, A. R. Yung, S. E. Artandi, P. A. Beachy, M. F. Clarke, L. C. Giudice, F. W. Huang, K. C. Huang, J. Idoyaga, S. K. Kim, M. Krasnow, C. S. Kuo, P. Nguyen, S. R. Quake, T. A. Rando, K. Red-Horse, J. Reiter, D. A. Relman, J. L. Sonnenburg, B. Wang, A. Wu, S. M. Wu, T. Wyss-Coray, The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. *Science* **376**, eabl4896 (2022).

18. M. H. Jagasia, H. T. Greinix, M. Arora, K. M. Williams, D. Wolff, E. W. Cowen, J. Palmer, D. Weisdorf, N. S. Treister, G.-S. Cheng, H. Kerr, P. Stratton, R. F. Duarte, G. B. McDonald, Y. Inamoto, A. Vigorito, S. Arai, M. B. Datiles, D. Jacobsohn, T. Heller, C. L. Kitko, S. A. Mitchell, P. J. Martin, H. Shulman, R. S. Wu, C. S. Cutler, G. B. Vogelsang, S. J. Lee, S. Z. Pavletic, M. E. D. Flowers, 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).

19. O. Mzava, A. P. Cheng, A. Chang, S. Smalling, L.-A. K. Djomnang, J. S. Lenz, R. Longman, A. Steadman, L. G. Gómez-Escobar, E. J. Schenck, M. Salvatore, M. J. Satlin, M. Suthanthiran, J. R. Lee, C. E. Mason, D. Dadhania, I. De Vlaminck, A metagenomic DNA sequencing assay that is robust against environmental DNA contamination. *Nat. Commun.* **13**, 4197 (2022).

20. P. Burnham, N. Gomez-Lopez, M. Heyang, A. P. Cheng, J. S. Lenz, D. M. Dadhania, J. R. Lee, M. Suthanthiran, R. Romero, I. De Vlaminck, Separating the signal from the noise in metagenomic cell-free DNA sequencing. *Microbiome* **8**, 18 (2020).

21. C. J. Loy, V. Servellita, A. Sotomayor-Gonzalez, A. Bliss, J. Lenz, E. Belcher, W. Suslovic, J. Nguyen, M. E. Williams, M. Oseguera, M. A. Gardiner, Pediatric Emergency Medicine Kawasaki Disease Research Group (PEMKDRG), CHARMS Study Group, J.-H. Choi, H.-M. Hsiao, H. Wang, J. Kim, C. Shimizu, A. Tremoulet, M. Delaney, R. L. DeBiasi, C. A. Rostad, J. C. Burns, C. Y. Chiu, I. D. Vlaminck, Plasma Cell-free RNA Signatures of Inflammatory Syndromes in Children. *MedRxiv Prepr. Serv. Health Sci.*, 2024.03.06.24303645 (2024).

22. A. Chang, C. J. Loy, J. S. Lenz, A. Steadman, A. Andama, N. V. Nhung, C. Yu, W. Worodria, C. M. Denkinger, P. Nahid, A. Cattamanchi, I. De Vlaminck, Circulating Cell-Free RNA in Blood as a Host Response Biomarker for the Detection of Tuberculosis. *MedRxiv Prepr. Serv. Health Sci.*, 2023.01.11.23284433 (2023).

23. C. J. Loy, A. Sotomayor-Gonzalez, V. Servellita, J. Nguyen, J. Lenz, S. Bhattacharya, M. E. Williams, A. P. Cheng, A. Bliss, P. Saldhi, N. Brazer, J. Streithorst, W. Suslovic, C. J. Hsieh, B. Bahar, N. Wood, A. Foresythe, A. Gliwa, K. Bhakta, M. A. Perez, L. Hussaini, E. J. Anderson, A. Chahroudi, M. Delaney, A. J. Butte, R. L. DeBiasi, C. A. Rostad, I. De Vlaminck, C. Y. Chiu, Nucleic acid biomarkers of immune response and cell and tissue damage in children with COVID-19 and MIS-C. *Cell Rep. Med.* **4**, 101034 (2023).

24. Y. Wang, J. Li, L. Zhang, H.-X. Sun, Z. Zhang, J. Xu, Y. Xu, Y. Lin, A. Zhu, Y. Luo, H. Zhou, Y. Wu, S. Lin, Y. Sun, F. Xiao, R. Chen, L. Wen, W. Chen, F. Li, R. Ou, Y. Zhang, T. Kuo, Y. Li, L. Li, J. Sun, K. Sun, Z. Zhuang, H. Lu, Z. Chen, G. Mai, J. Zhuo, P. Qian, J. Chen, H. Yang, J. Wang, X. Xu, N. Zhong, J. Zhao, J. Li, J. Zhao, X. Jin, Plasma cell-free RNA characteristics in COVID-19 patients. *Genome Res.* **32**, 228–241 (2022).

25. K. L. Kalantar, L. Neyton, M. Abdelghany, E. Mick, A. Jauregui, S. Caldera, P. H. Serpa, R. Ghale, J. Albright, A. Sarma, A. Tsitsiklis, A. Leligdowicz, S. A. Christenson, K. Liu, K. N. Kangelaris, C. Hendrickson, P. Sinha, A. Gomez, N. Neff, A. Pisco, S. B. Doernberg, J. L. Derisi, M. A. Matthay, C. S. Calfee, C. R. Langelier, Integrated host-microbe plasma metagenomics for sepsis diagnosis in a prospective cohort of critically ill adults. *Nat. Microbiol.* **7**, 1805–1816 (2022).

26. N. Chalasani, S. Toden, J. J. Sninsky, R. P. Rava, J. V. Braun, S. Gawrieh, J. Zhuang, M. Nerenberg, S. R. Quake, T. Maddala, Noninvasive stratification of nonalcoholic fatty liver disease by whole transcriptome cell-free mRNA characterization. *Am. J. Physiol. Gastrointest. Liver Physiol.* **320**, G439–G449 (2021).

27. A. Ibarra, J. Zhuang, Y. Zhao, N. S. Salathia, V. Huang, A. D. Acosta, J. Aballi, S. Toden, A. P. Karns, I. Purnajo, J. R. Parks, L. Guo, J. Mason, D. Sigal, T. S. Nova, S. R. Quake, M. Nerenberg, Non-invasive characterization of human bone marrow stimulation and reconstitution by cell-free messenger RNA sequencing. *Nat. Commun.* **11**, 400 (2020).

28. I. Fox-Fisher, S. Piyanzin, B. L. Ochana, A. Klochendler, J. Magenheim, A. Peretz, N. Loyfer, J. Moss, D. Cohen, Y. Drori, N. Friedman, M. Mandelboim, M. E. Rothenberg, J. M. Caldwell, M. Rochman, A. Jamshidi, G. Cann, D. Lavi, T. Kaplan, B. Glaser, R. Shemer, Y. Dor, Remote immune processes revealed by immune-derived circulating cell-free DNA. *eLife* **10**, e70520 (2021).

29. S. Ghimire, D. Weber, E. Mavin, X. N. Wang, A. M. Dickinson, E. Holler, Pathophysiology of GvHD and Other HSCT-Related Major Complications. *Front. Immunol.* **8** (2017).

30. N. Spyrou, Y. Akahoshi, S. Kowalyk, G. Morales, R. Beheshti, P. Aguayo-Hiraldo, M. M. Al Malki, F. Ayuk, P. Bader, J. Baez, A. Capellini, H. Choe, Z. DeFilipp, M. Eder, G. Eng, A. Etra, S. Gleich, S. A. Grupp, E. Hexner, M. Hoepting, W. J. Hogan, S. Kasikis, N. Katsivelos, A. Khan, C. L. Kitko, S. Kraus, D. Kwon, P. Merli, J. Portelli, M. Qayed, R. Reshef, T. Schechter, I. Vasova, M. Wölfl, K. Wudhikarn, R. Young, E. Holler, Y.-B. Chen, R. Nakamura, J. E. Levine, J. L. M. Ferrara, A Day 14 Endpoint for Acute GVHD Clinical Trials. *Transplant. Cell. Ther.* **30**, 421–432 (2024).

31. M. J. Hartwell, U. Özbek, E. Holler, A. S. Renteria, H. Major-Monfried, P. Reddy, M. Aziz, W. J. Hogan, F. Ayuk, Y. A. Efebera, E. O. Hexner, U. Bunworasate, M. Qayed, R. Ordemann, M. Wölfl, S. Mielke, A. Pawarode, Y.-B. Chen, S. Devine, A. C. Harris, M. Jagasia, C. L. Kitko, M. R. Litzow, N. Kröger, F. Locatelli, G. Morales, R. Nakamura, R. Reshef, W. Rösler, D. Weber, K. Wudhikarn, G. A. Yanik, J. E. Levine, J. L. M. Ferrara, An

early-biomarker algorithm predicts lethal graft-versus-host disease and survival. *JCI Insight* **2** (2017).

32. E. T. Chivero, J. T. Stapleton, Tropism of human pegivirus (formerly known as GB virus C/hepatitis G virus) and host immunomodulation: insights into a highly successful viral infection. *J. Gen. Virol.* **96**, 1521–1532 (2015).

33. Z. Li, Y. Li, Y. Liang, L. Hu, S. Chen, Prevalence and risk factors of human pegivirus type 1 infection in hematopoietic stem cell transplantation patients. *Int. J. Infect. Dis.* **85**, 111–113 (2019).

34. D.-L. Vu, S. Cordey, F. Simonetta, F. Brito, M. Docquier, L. Turin, C. Van Delden, E. Boely, C. Dantin, A. Pradier, E. Roosnek, Y. Chalandon, E. M. Zdobnov, S. Masouridi-Levrat, L. Kaiser, Human pegivirus persistence in human blood virome after allogeneic haematopoietic stem-cell transplantation. *Clin. Microbiol. Infect.* **25**, 225–232 (2019).

35. J. Mouries, P. Brescia, A. Silvestri, I. Spadoni, M. Sorribas, R. Wiest, E. Mileti, M. Galbiati, P. Invernizzi, L. Adorini, G. Penna, M. Rescigno, Microbiota-driven gut vascular barrier disruption is a prerequisite for non-alcoholic steatohepatitis development. *J. Hepatol.* **71**, 1216–1228 (2019).

36. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

37. on behalf of the EBMT (European Society for Blood and Marrow Transplantation) Transplant Complications Working Party and the "EBMT−NIH (National Institutes of Health)−CIBMTR (Center for International Blood and Marrow Transplant Research) GvHD Task Force," H. M. Schoemans, S. J. Lee, J. L. Ferrara, D. Wolff, J. E. Levine, K. R. Schultz, B. E. Shaw, M. E. Flowers, T. Ruutu, H. Greinix, E. Holler, G. Basak, R. F. Duarte, S. Z. Pavletic, EBMT−NIH−CIBMTR Task Force position statement on standardized terminology & guidance for graft-versus-host disease assessment. *Bone Marrow Transplant.* **53**, 1401–1415 (2018).

38. D. E. Wood, J. Lu, B. Langmead, Improved metagenomic analysis with Kraken 2. *Genome Biol.* **20**, 257 (2019).

39. J. Lu, F. P. Breitwieser, P. Thielen, S. L. Salzberg, Bracken: estimating species abundance in metagenomics data. *PeerJ Comput. Sci.* **3**, e104 (2017).

40. C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T. L. Madden, BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).

Acknowledgments: This work would not have been possible without the vision, guidance, and leadership of the late Dr. Francisco M. Marty, who was instrumental in launching this study. The co-authors are grateful for his friendship, mentorship, and for everything he taught us. We thank the patients and their families for their help to further our understanding of pediatric inflammatory conditions hematopoietic stem cell transplantation.

Funding: This work was funded by National Institutes of Health (NIH) grant R01AI146165 (JR, IDV). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions: Each author's contribution(s) to the paper should be listed [we encourage you to follow the CRediT model]. Each CRediT role should have its own line, and there should not be any punctuation in the initials.

Examples:

Conceived and designed the study: CJL, MPC, JR, IDV

Performed sequencing experiments: AB, JL, and EB

Identified and collected patient samples and clinical metadata: IHGB, MPC and JR.

Provided input for data analysis: MPC, TC, DEL

Analyzed sequencing data: CJL, TC, IDV

Supervised the study: MPC, JR, IDV

Wrote the manuscript and prepared the figures: CJL, IDV

All authors read and edited the manuscript and agreed to its contents.

Competing interests: C.J.L and I.D.V are inventors on submitted patents pertaining to cell-free nucleic acids (US patent applications 63/237,367 and 63/429,733). I.D.V. is a member of the Scientific Advisory Board of Karius Inc., Kanvas Biosciences and GenDX. I.D.V. is listed as an inventor on submitted patents pertaining to cell-free nucleic acids (US patent applications 63/237,367, 63/056,249, 63/015,095, 16/500,929, 41614P-10551-01-US) and receives consulting fees from Eurofins Viracor. JR receives research funding from Kite/Gilead, Novartis and Oncternal Therapeutics and serves on advisory boards for Garuda Therapeutics, LifeVault Bio, Smart Immune and TriArm Therapeutics. MPC reports research support from the Canadian Institutes of Health Research and is supported by the Fonds de Recherche du Québec – Santé; research contracts from Cidara Therapeutics, Scynexis, and Amplyx Pharmaceuticals; honoraria from AstraZeneca, Takeda, Merck, and Pfizer; 3 pending patents (Methods for detecting tissue damage, graft-versus-host disease, and infections using cell-free DNA profiling; Methods for assessing the severity and progression of SARS-CoV-2 infections using cell-free DNA; and rapid identification of antimicrobial resistance and other microbial phenotypes using highly-multiplexed fluorescence in situ hybridization); stock options as a member of the scientific advisory board for GEn1E Lifesciences and Nomic Bio; and equity as co-founder of Kanvas Biosciences.

Data and materials availability: De-identified RNA-seq count matrices have been uploaded to the NCBI (National Center for Biotechnology Information) GEO (Gene Expression Omnibus) database and will be publicly available upon publication. All code has been deposited on GitHub and will be available upon publication.

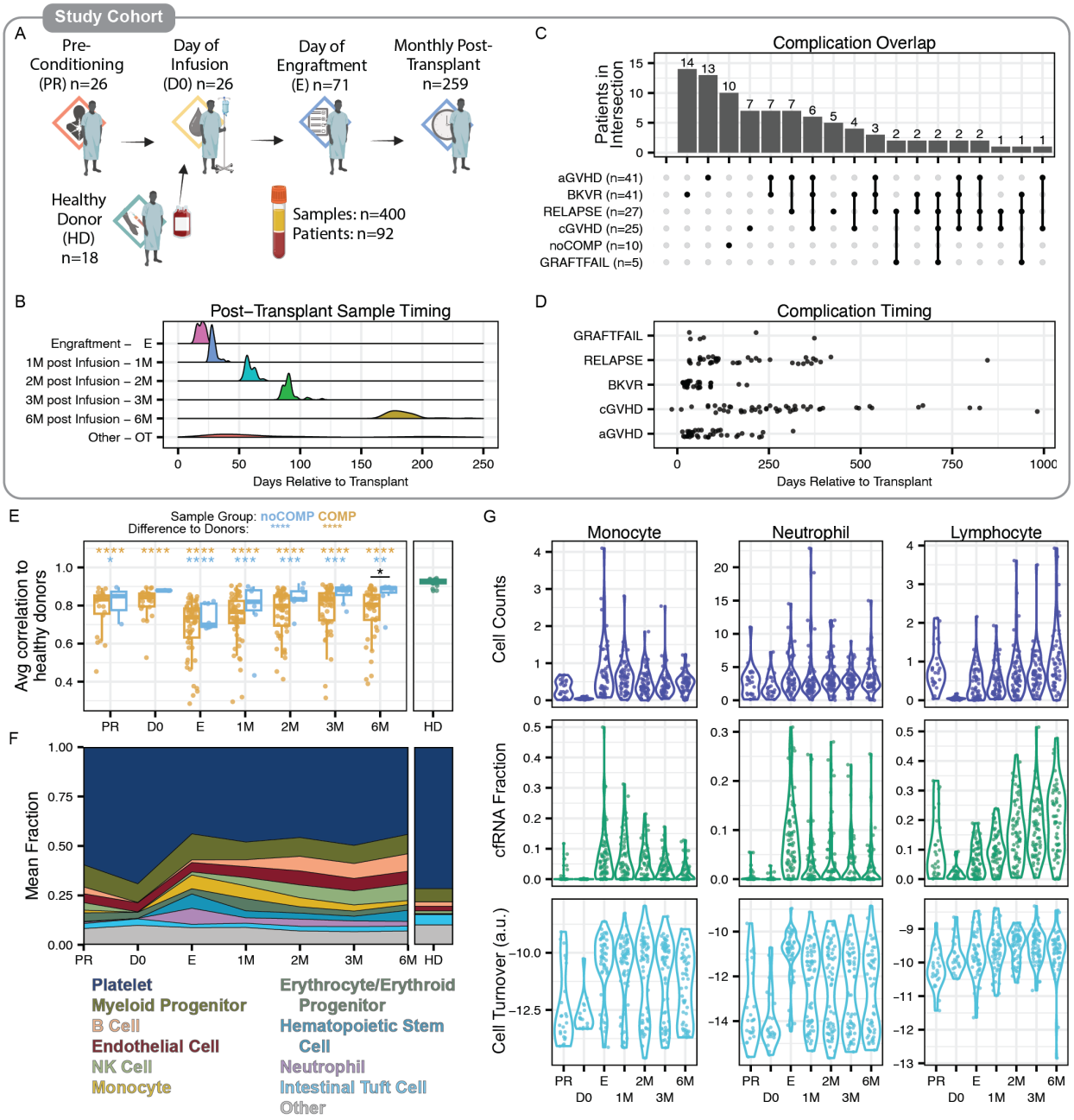


Fig. 1. Sample cohort overview (**A**) Sample collection overview. (**B**) Day relative to transplantation samples were collected for each timepoint. (**C**) Upset plot of complications. (**D**) Timing of complications relative to transplantation. (**E**) Correlation of cfRNA abundance to healthy stem cell donors of samples from noCOMP and COMP patients (Methods). Asterisks indicate statistical significance by Mann-Whitney U test, unless otherwise noted, using Benjamini-Hochberg adjusted p-values as follows: ns, non-significant; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.001$. (**F**) Stacked area plot of mean fraction of top ten most abundant

cell types across the time course (**G**) Paired cell counts, cfRNA cell type of origin fractions, and estimated turnover in arbitrary units for paired samples.

5

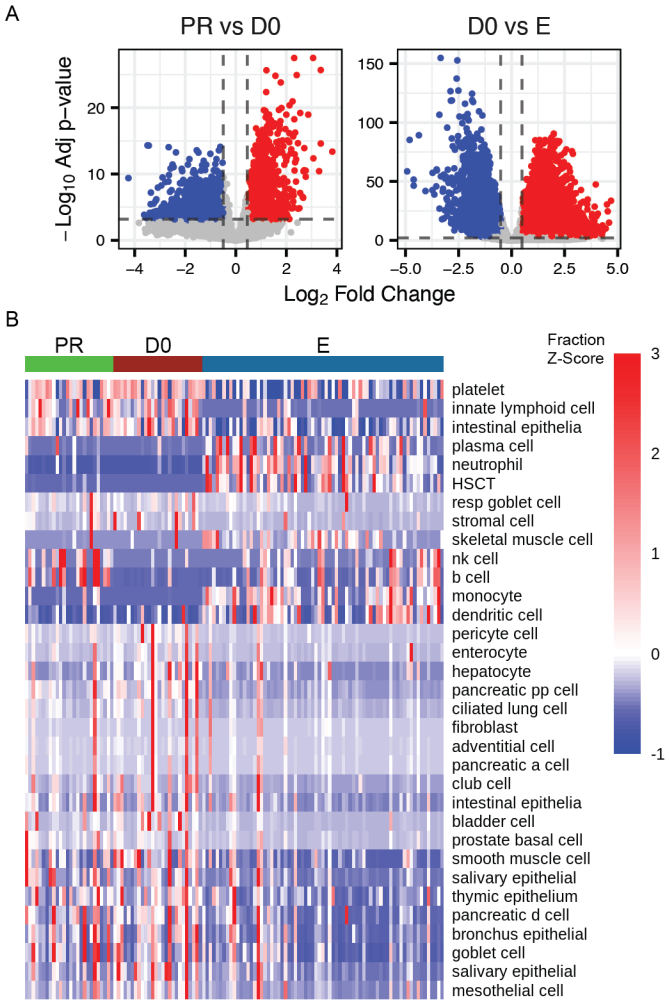


Fig. 2. cfRNA dynamics during treatment (**A**) Volcano plots of the results from the PR vs D0 and D0 vs E comparisons using DESeq2. Points colored blue are down regulated (BH adjusted p-value<0.05, log2FoldChange<-0.05) and points colored red are upregulated (BH adjusted p-value<0.05, log2FoldChange>0.05) (**B**) Cell type of origin estimates of cell types with significantly different contributions between D0vPR or D0vsE.

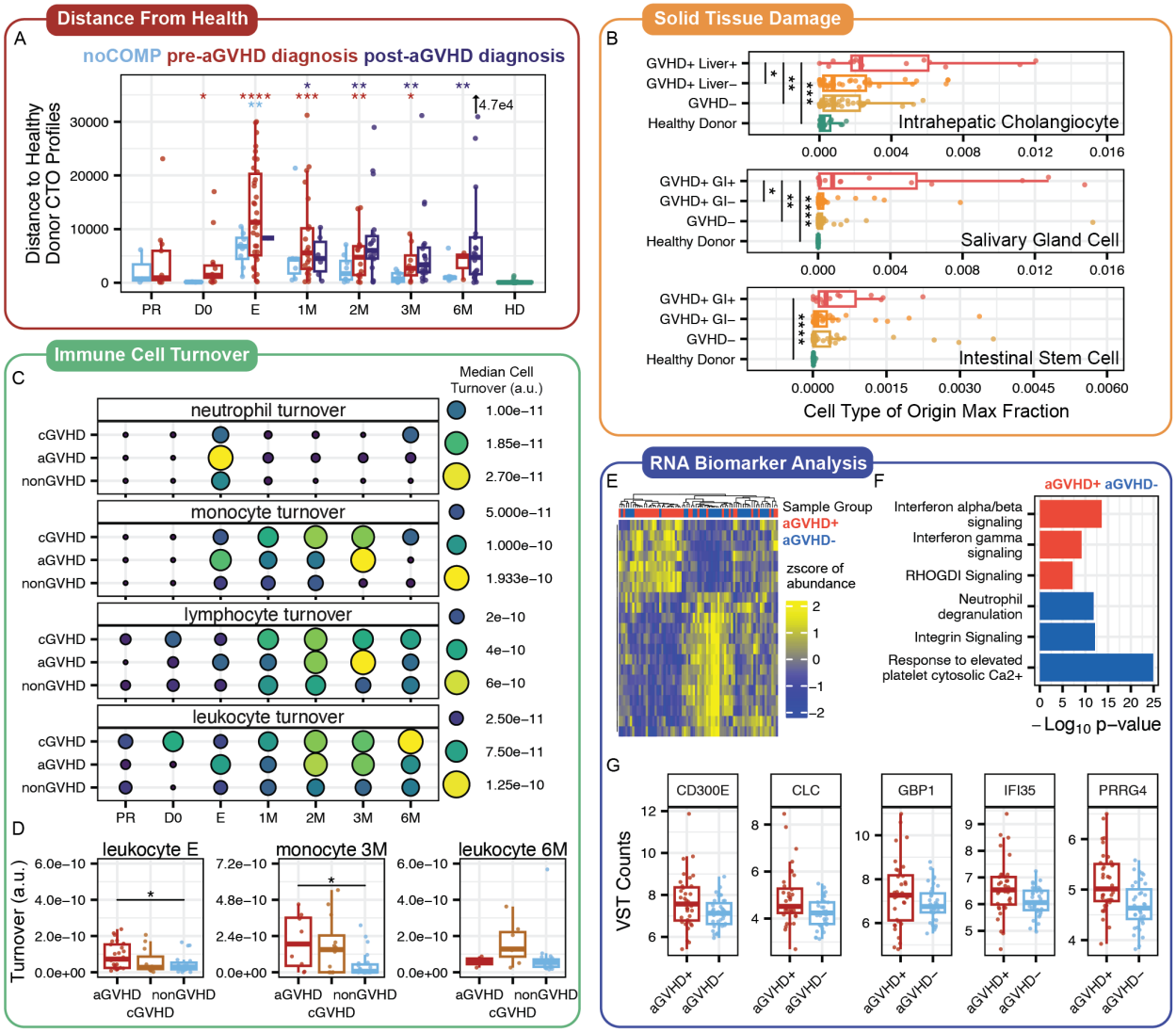


Fig.3. cfRNA in GVHD (**A**) Distance of cell type of origin (CTO) results from patients with aGVHD compared to noCOMP patients. Samples from patients with aGVHD further split by pre- and post- aGVHD diagnosis. Statistical comparisons done using a Student's t-test. (**B**) Maximum cell type of origin measurement for Intrahepatic cholangiocyte, salivary gland cell, and intestinal stem cells for samples split by GVHD status and organ involvement. (**C**) Median immune cell turnover measurements from samples of patients with cGVHD, aGVHD, or non-GVHD all prior to GVHD diagnosis. Dot size and color represent median cell turnover in arbitrary units (a.u.). (**D**) Cell turnover measurements of patients with cGVHD, aGVHD, and non-GVHD all prior to GVHD diagnosis at specific timepoints. (**E**) Variance stabilization transformation values of significantly differentially abundant genes (DAGs) at the engraftment timepoint of samples from patients that develop aGVHD (aGVHD+) or do not develop aGVHD (aGVHD-) (DESeq2, Benjamini-Hochberg adjusted p-value < 0.01, |Log2FoldChange| > 1.5). Samples and genes are clustered based on correlation. (**F**) Enriched pathways of significantly differentially abundant RNAs at the engraftment timepoint of samples from patients that develop aGVHD (aGVHD+) or do not develop aGVHD (aGVHD-). (**G**) Variance stabilization transformation values of the top five differentially abundant genes elevated in patients with GVHD. Asterisks indicate statistical

significance by Mann-Whitney U test unless otherwise noted using Benjamini-Hochberg adjusted p-values as follows: ns, non-significant; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.001$.

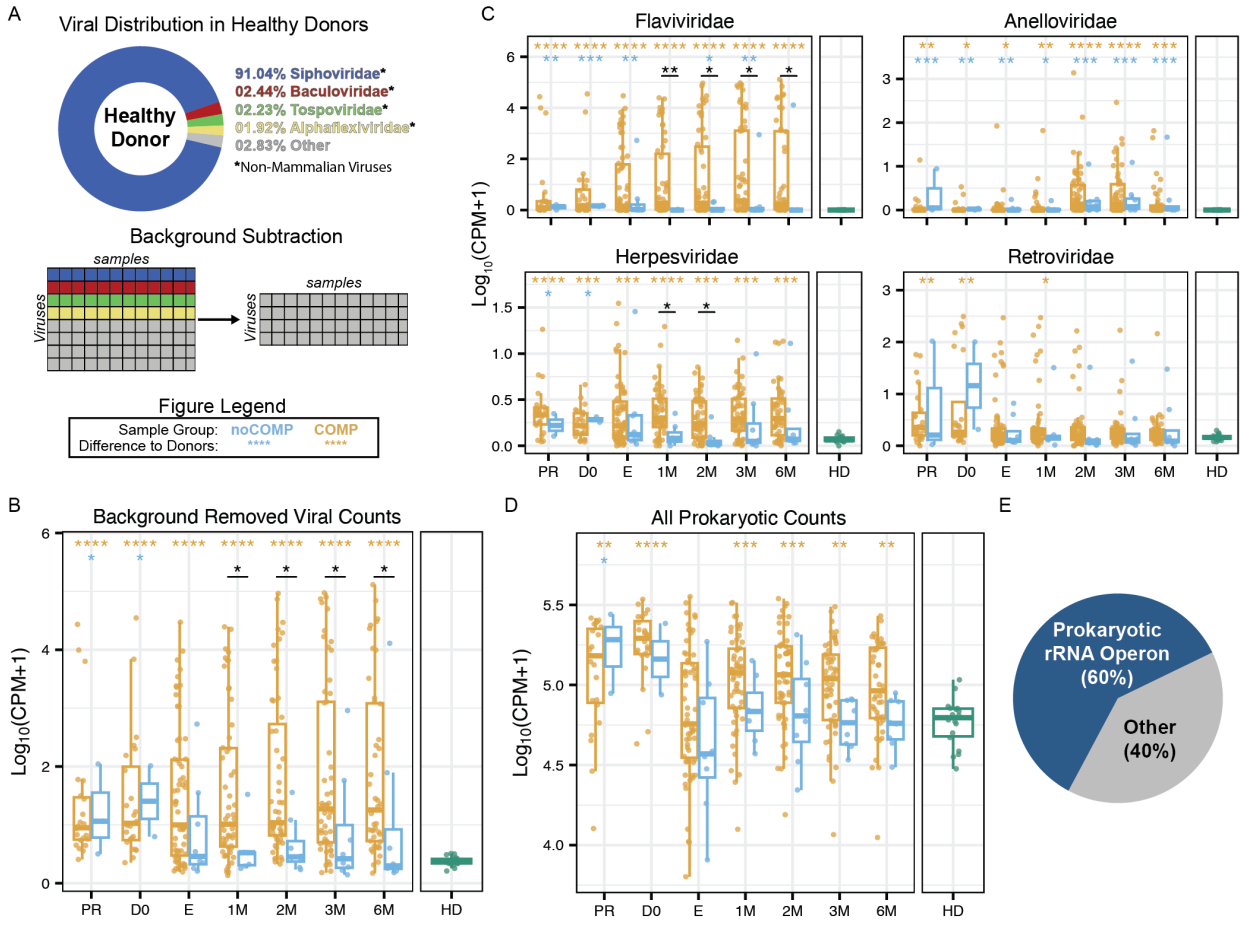
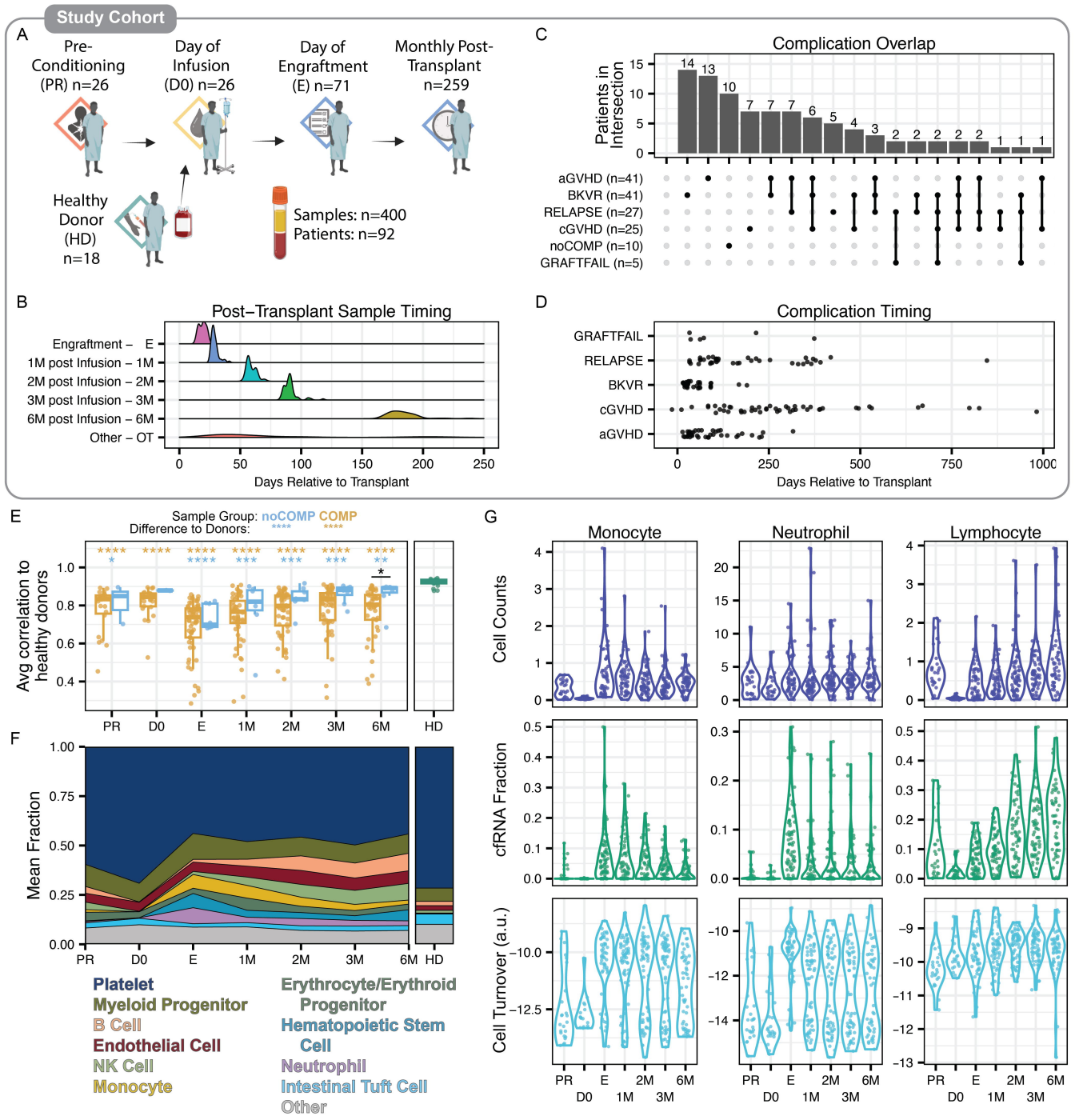
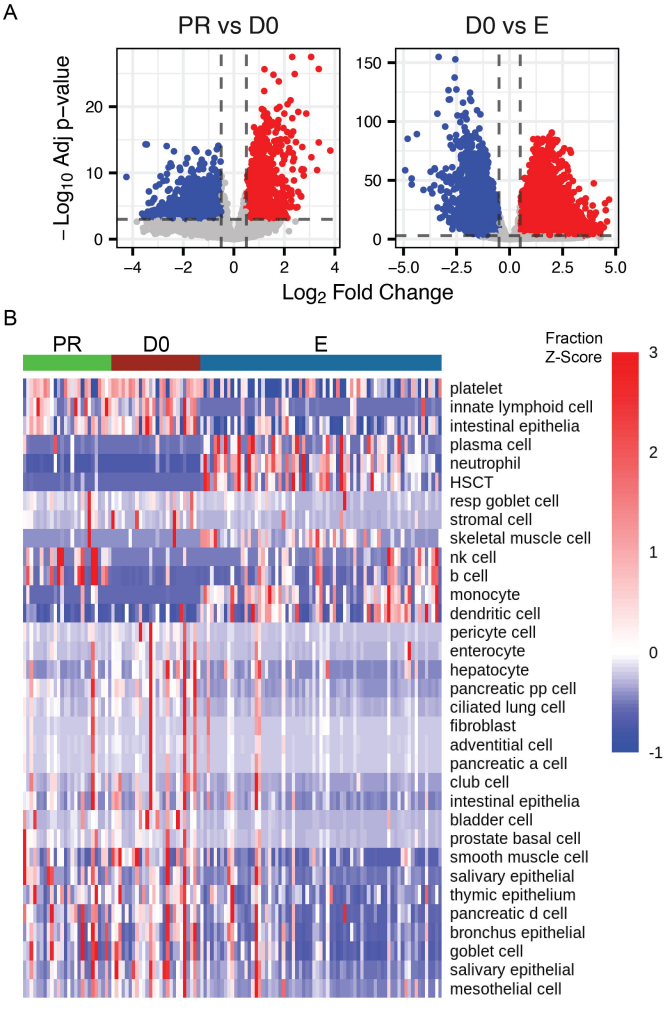
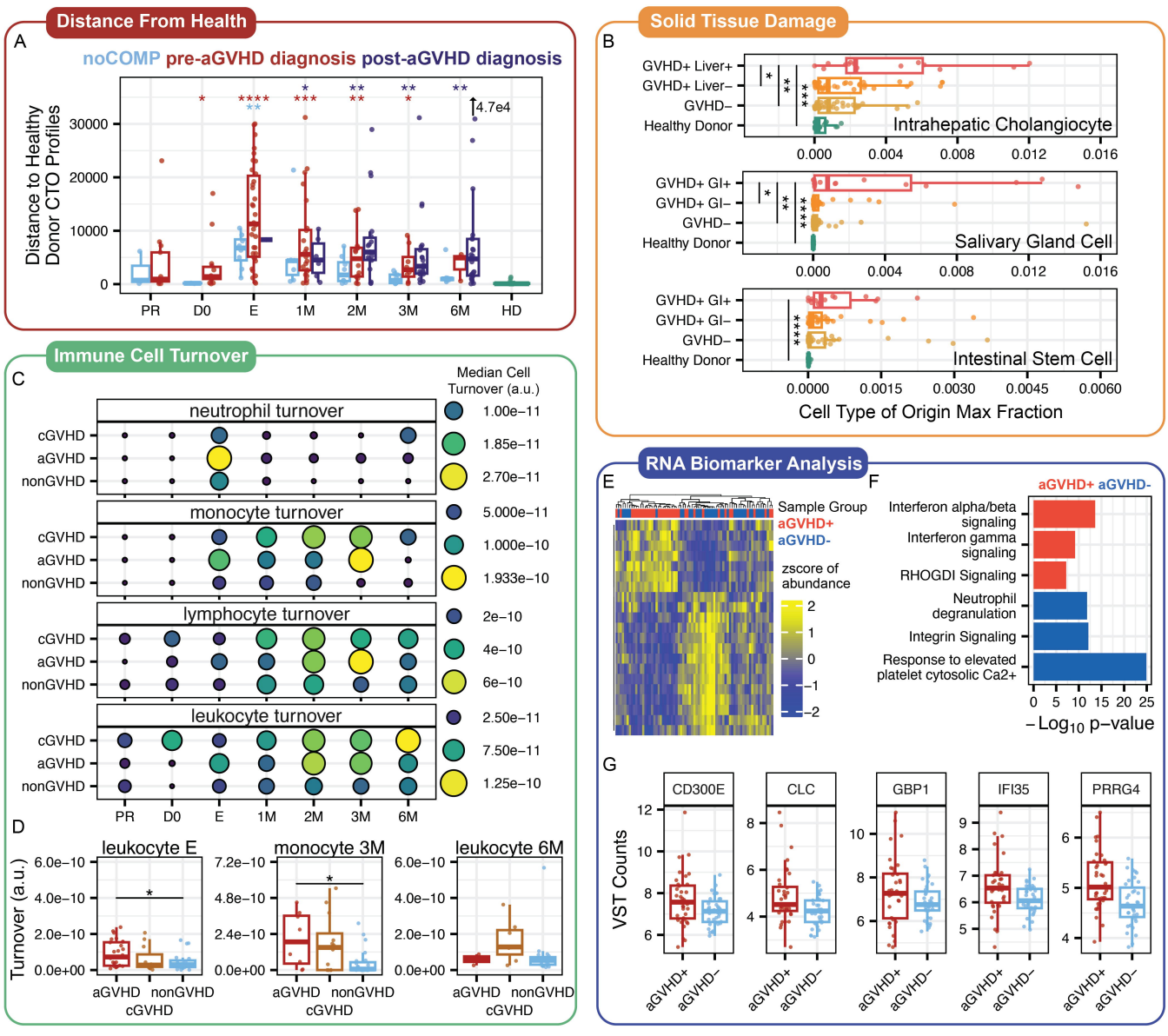
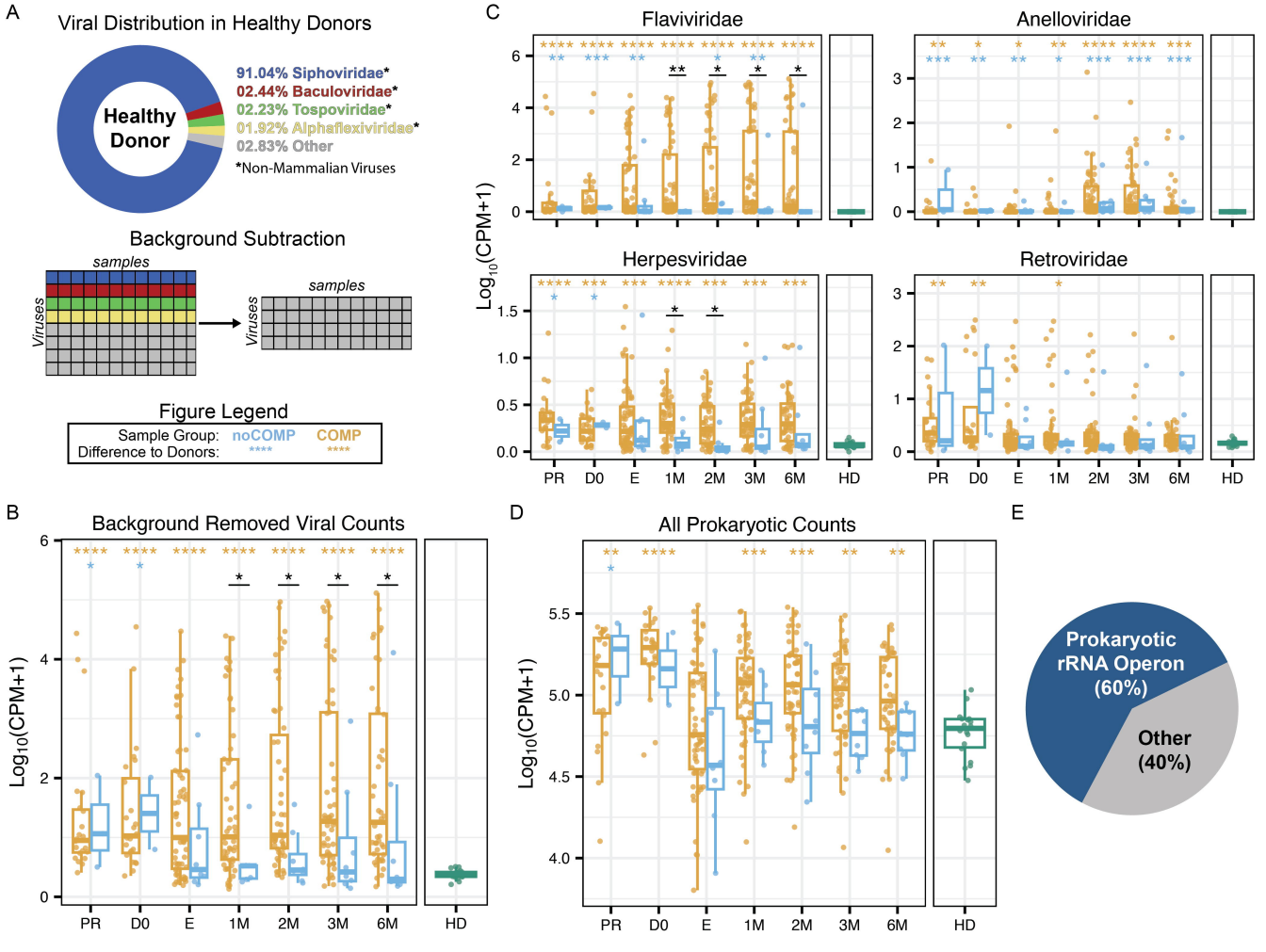


Fig. 4. Metagenomic dynamics of cfRNA in HSCT (**A**) Family count distribution of all viral reads assigned to healthy stem cell donors and schematic of background removal. (**B**) Counts per million background removed viral counts in noCOMP, COMP, and healthy patients. (**C**) Counts per million reads assigned to the Flaviviridae, *Anelloviridae*, *Herpesviridae*, and *Retroviridae* viral families in noCOMP, COMP, and healthy patients. (**D**) Counts per million reads assigned prokaryotic in origin. (**E**) Fraction of randomly sampled bacterial reads from each sample that aligned to a prokaryotic rRNA operon reference. Asterisks indicate statistical significance by Mann-Whitney U test unless otherwise noted using Benjamini-Hochberg adjusted p-values as follows: ns, non-significant; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.001$.


Study Cohort
A
Pre-Conditioning (PR) n=26
Day of Infusion (D0) n=26
Day of Engraftment (E) n=71
Monthly Post-Transplant n=259
Healthy Donor (HD) n=18
Samples: n=400
Patients: n=92
B
Post–Transplant Sample Timing
Engraftment – E
1M post Infusion – 1M
2M post Infusion – 2M
3M post Infusion – 3M
6M post Infusion – 6M
Other – OT
Days Relative to Transplant
C
Complication Overlap
Patients in Intersection
aGVHD (n=41)
BKVR (n=41)
RELAPSE (n=27)
cGVHD (n=25)
noCOMP (n=10)
GRAFTFAIL (n=5)
D
Complication Timing
GRAFTFAIL
RELAPSE
BKVR
cGVHD
aGVHD
Days Relative to Transplant
E
Sample Group: noCOMP COMP
Difference to Donors:
Avg correlation to healthy donors
PR D0 E 1M 2M 3M 6M HD
F
Mean Fraction
PR D0 E 1M 2M 3M 6M HD
Platelet
Myeloid Progenitor
B Cell
Endothelial Cell
NK Cell
Monocyte
Erythrocyte/Erythroid Progenitor
Hematopoietic Stem Cell
Neutrophil
Intestinal Tuft Cell
Other
G
Monocyte
Neutrophil
Lymphocyte
Cell Counts
cfRNA Fraction
Cell Turnover (a.u.)
PR D0 E 1M 2M 3M 6M


A
PR vs D0
D0 vs E
– Log10 Adj p–value
Log2 Fold Change
B
PR
D0
E
Fraction
Z-Score
platelet
innate lymphoid cell
intestinal epithelia
plasma cell
neutrophil
HSCT
resp goblet cell
stromal cell
skeletal muscle cell
nk cell
b cell
monocyte
dendritic cell
pericyte cell
enterocyte
hepatocyte
pancreatic pp cell
ciliated lung cell
fibroblast
adventitial cell
pancreatic a cell
club cell
intestinal epithelia
bladder cell
prostate basal cell
smooth muscle cell
salivary epithelial
thymic epithelium
pancreatic d cell
bronchus epithelial
goblet cell
salivary epithelial
mesothelial cell


Distance From Health
A
noCOMP pre-aGVHD diagnosis post-aGVHD diagnosis
Distance to Healthy Donor CTO Profiles
4.7e4
PR D0 E 1M 2M 3M 6M HD
Solid Tissue Damage
B
GVHD+ Liver+
GVHD+ Liver−
GVHD−
Healthy Donor
Intrahepatic Cholangiocyte
GVHD+ GI+
GVHD+ GI−
Salivary Gland Cell
Intestinal Stem Cell
Cell Type of Origin Max Fraction
Immune Cell Turnover
C
Median Cell Turnover (a.u.)
neutrophil turnover
monocyte turnover
lymphocyte turnover
leukocyte turnover
cGVHD
aGVHD
nonGVHD
1.00e−11
1.85e−11
2.70e−11
5.000e−11
1.000e−10
1.933e−10
2e−10
4e−10
6e−10
2.50e−11
7.50e−11
1.25e−10
D
leukocyte E
monocyte 3M
leukocyte 6M
Turnover (a.u.)
RNA Biomarker Analysis
E
Sample Group
aGVHD+
aGVHD-
zscore of abundance
F
aGVHD+ aGVHD-
Interferon alpha/beta signaling
Interferon gamma signaling
RHOGDI Signaling
Neutrophil degranulation
Integrin Signaling
Response to elevated platelet cytosolic Ca2+
−Log10 p−value
G
CD300E
CLC
GBP1
IFI35
PRRG4
VST Counts
aGVHD+
aGVHD−


A
Viral Distribution in Healthy Donors
Healthy Donor
91.04% Siphoviridae*
02.44% Baculoviridae*
02.23% Tospoviridae*
01.92% Alphaflexiviridae*
02.83% Other
*Non-Mammalian Viruses
Background Subtraction
samples
Viruses
samples
Viruses
Figure Legend
Sample Group: noCOMP COMP
Difference to Donors: **** ****
B
Background Removed Viral Counts
Log10(CPM+1)
PR D0 E 1M 2M 3M 6M HD
C
Flaviviridae
Anelloviridae
Herpesviridae
Retroviridae
Log10(CPM+1)
PR D0 E 1M 2M 3M 6M HD
D
All Prokaryotic Counts
Log10(CPM+1)
PR D0 E 1M 2M 3M 6M HD
E
Prokaryotic rRNA Operon (60%)
Other (40%)