

1 Cell-free DNA Tissues-of-Origin Profiling to Predict Graft versus Host 2 Disease and Detect Infection after Hematopoietic Cell Transplantation

3
4 **Authors:** Alexandre Pellan Cheng¹, Matthew Pellan Cheng^{2,3}, Joan Sesing Lenz¹, Kaiwen Chen^{2,3}, Philip
5 Burnham⁴, Kaitlyn Marie Timblin^{2,3}, José Luis Orejas^{2,3}, Emily Silverman^{2,3}, Francisco M. Marty^{2,3}, Jerome
6 Ritz^{2,5}, Iwijn De Vlaminck^{1,*}

7 8 **Affiliations:**

9 1 Meinig School of Biomedical Engineering, Cornell University, Ithaca, NY, USA

10 2 Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA

11 3 Division of Infectious Disease, Brigham and Women's Hospital, Boston, MA, USA

12 4 Department of Bioengineering, University of Pennsylvania, Philadelphia, PA, USA

13 5 Department of Medicine, Harvard Medical School, Boston, MA, USA

14 * Corresponding author (vlaminck@cornell.edu)

15
16
17
18 **ABSTRACT:** Allogeneic hematopoietic cell transplantation (HCT) provides effective treatment for
19 hematologic malignancies and immune disorders. Monitoring for immune complications and infection is a
20 critical component of post-HCT therapy, however, current diagnostic options are limited. Here, we propose
21 a blood test that employs genome-wide profiling of methylation marks comprised within circulating cell-
22 free DNA to trace the tissues-of-origin of cell-free DNA (cfDNA), to quantify tissue-specific injury, and to
23 screen for microbial pathogens after allogeneic HCT. We applied this assay to 106 plasma samples collected
24 from 18 HCT recipients at predetermined time points before and after allogeneic HCT. We observe marked
25 dynamics in the composition and abundance of cfDNA from different tissues in response to conditioning
26 chemotherapy and HCT. We find that the abundance of solid-organ derived cfDNA in the blood at one-
27 month after HCT is an early predictor of acute graft-versus-host disease, a frequent immune-related
28 complication of HCT that occurs when donor immune cells attack the patient's own tissues (area under the
29 curve, 0.9, p-value = 0.012). Metagenomic profiling of cfDNA was possible from the same assay and
30 revealed the frequent occurrence of viral and bacterial infection in this patient population. This proof-of-
31 principle study shows that cfDNA has the potential to improve the care of allogeneic HCT recipients by
32 enabling earlier detection and better prediction of immune complications and infection after HCT.

35 INTRODUCTION

36
37 More than 30,000 patients undergo allogeneic hematopoietic cell transplants (HCT) worldwide
38 each year for treatment of a variety of malignant and nonmalignant hematologic diseases¹⁻³. However,
39 immune related complications occur frequently after HCT. Up to 50% of patients experience graft-versus-
40 host disease (GVHD) in the first year after transplantation. GVHD occurs when donor immune cells attack
41 the patient's own tissues^{2,4-6}. Early and accurate diagnosis of GVHD is critical to inform treatment decisions
42 and to prevent serious long-term complications, including organ failure and death. Unfortunately, there are
43 few, noninvasive diagnostic options that reliably identify patients very early after onset of GVHD
44 symptoms. In current clinical practice, diagnosis of GVHD relies almost entirely on clinical criteria and
45 often requires confirmation with invasive procedures, such as a biopsy of the gastrointestinal tract, skin, or
46 liver⁷.

47
48 Small fragments of cell-free DNA (cfDNA) circulate in blood. In the absence of disease, cfDNA
49 originates primarily from apoptosis of cells of the hematopoietic lineage⁸. During disease, a significant
50 proportion of cfDNA can be derived from affected tissues⁹⁻¹⁵. In solid-organ transplantation (SOT), we and
51 others have shown that transplant donor derived cfDNA in the blood is a quantitative noninvasive marker
52 of solid organ transplant injury^{12,13,16,17}. Here, we sought to investigate the utility of circulating cfDNA as a
53 minimally invasive analyte to detect and quantify tissue injury due to GVHD after HCT. To quantify the
54 contributions of any tissue to the mixture of cfDNA in plasma, and thereby the degree of injury to any
55 vascularized tissue in the setting of HCT, we implemented shallow, whole genome bisulfite sequencing
56 (WGBS) to profile 5-methylcytosine (5mC) marks of cfDNA. These marks are cell, tissue and organ type
57 specific¹⁸, and can inform the tissues-of-origin of cell-free DNA^{8,11,19}.

58
59 We collected serial blood samples from a prospective cohort of allogeneic HCT recipients at
60 predetermined time points before and after HCT. We analyzed a total of 106 plasma samples from 18 HCT
61 recipients with and without GVHD in the first 3 months post HCT. We observed rich dynamics in the tissue-
62 origin of cfDNA in response to pre-transplant conditioning chemotherapy and following HCT. The tissue-
63 origin of cfDNA after transplantation was patient-specific and a function of the manifestation of GVHD.
64 For example, the absolute abundance of organ derived cfDNA at one month after transplantation was
65 significantly elevated for patients diagnosed with GVHD within the first 6 months after HCT, indicating
66 that cfDNA tissues-of-origin profiling can be used to predict GVHD and quantify its severity.

67
68 Allogeneic HCT recipients receive therapeutic immunosuppression to manage the risk of GVHD
69 and are consequently at risk of opportunistic infections. Because a wide range of microorganisms can cause
70 disease after HCT, there is a critical need for tools that can broadly inform microbial pathogens. To address
71 this, we investigated the possibility to identify infectious agents via whole-genome bisulfite sequencing of
72 plasma cfDNA, using methods that we previously developed to screen for urinary tract infection¹⁵. We
73 observe significant increases in viral and bacterial derived cfDNA post HCT. cfDNA from anelloviruses,
74 cytomegalovirus (CMV), herpesvirus 6, Epstein-Barr virus, and polyomavirus was frequently observed in
75 this patient population.

76
77 Together, this study provides a proof of principle that cell-free DNA profiling can be used to
78 simultaneously monitor immune and infectious related complications after allogeneic HCT.

79
80
81
82
83
84
85

86 RESULTS

87

88 We performed a prospective cohort study to evaluate the utility of cfDNA to predict and monitor
89 complications after allogeneic HCT. For this study, we selected 18 adults that underwent allogeneic HCT
90 and assayed a total of 106 serial plasma samples collected at six predetermined time points, including before
91 conditioning chemotherapy, on the day of but before hematopoietic cell infusion, after neutrophil
92 engraftment (>500 neutrophils per microliter), and at one, two, and three months post HCT (**Fig. 1a**). The
93 test cohort included patients with both malignant (n=14) and non-malignant blood disorders (n=4)
94 (**supplementary table 1**). In total, nine patients developed acute GVHD (GVHD+) and nine did not
95 (GVHD-), two developed a bloodstream infection and two developed BK virus disease (see Methods and
96 SI).

97

98 We isolated cfDNA from plasma (0.5mL-1.9mL per sample) and implemented whole-genome
99 bisulfite sequencing to profile cytosine methylation marks comprised within cfDNA (**Fig. 1b**). We
100 implemented a single-stranded DNA (ssDNA) library preparation to obtain sequence information after
101 bisulfite conversion^{20,21}. This ssDNA library preparation avoids degradation of adapter-bound molecules
102 which is common for WGBS library preparations that rely on ligation of methylated adapters before
103 bisulfite conversion and avoids amplification biases inherent to WGBS library preparations that implement
104 random priming²². We obtained 41 ± 15 million paired-end reads per sample, corresponding to 0.9 ± 0.3
105 fold per-base human genome coverage (**Fig. 1c**) and achieved a high bisulfite conversion efficiency ($99.4\% \pm 0.5\%$,
106 **Fig. 1d**). We used paired-end read mapping to characterize the length of bisulfite treated cfDNA
107 at single-nucleotide resolution and to investigate potential degradation of cfDNA due to bisulfite treatment.
108 This analysis revealed a fragmentation profile similar to the fragmentation profile previously reported for
109 plasma cfDNA that was not subjected to bisulfite treatment²³. The mode of fragments longer than 100bp
110 was $165 \text{ bp} \pm 7 \text{ bp}$ (**Fig. 1e**), and Fourier analysis revealed a 10.4 bp periodicity in the fragment length
111 profile (**Fig. 1e inset**). A second peak at 60-90 bp in the fragment length profile is characteristic of single-
112 stranded library preparation methods and was reported previously^{21,24}. Overall, we do not find evidence of
113 significant cfDNA fragmentation due to bisulfite treatment, in line with a recent report²⁵.

114

115 Temporal dynamics in response to conditioning therapy and HCT

116

117 To quantify the relative proportion of cfDNA derived from different vascularized tissues and
118 hematologic cell types we analyzed cfDNA methylation profiles against a reference set of methylation
119 profiles of pure cell and tissue types (138 reference tissues, see Methods, **Fig. 1f** and **supplementary**
120 **dataset 1**)²⁶⁻³⁰. We computed the absolute concentration of tissue-specific cfDNA by multiplying the
121 proportion of tissue-specific cfDNA with the concentration of total host-derived cfDNA (Methods). Figures
122 1g,h and 2 summarize these measurements for all patients and time points and reveal rich dynamics in
123 tissue-origin of cfDNA in response to both conditioning chemotherapy and HCT (**Fig. 1g-h, Fig. 2**). The
124 most striking features seen in the data include *i*) a decrease in blood-cell specific cfDNA in response to
125 conditioning therapy performed to deplete the patient's own immune cells, as expected (**Fig 1g, Fig 2a**), *ii*)
126 an increase in total cfDNA concentration at engraftment (**Fig. 1h, Fig. 2b**), *iii*) a decrease in total cfDNA
127 concentration after 60 days for most patients (**Fig. 1h**), and *iv*) an association between tissue-specific
128 cfDNA and the incidence of GVHD (see statistical analysis below).

129

130 We next examined these features in more detail to explore the utility of these measurements to
131 monitor immune related complications of HCT (**Fig. 2**). Prior to conditioning, neutrophils, erythrocyte
132 progenitors and monocytes were the major contributors of cfDNA in plasma (23.0%, 12.3% and 11.7%,
133 respectively, average cfDNA concentration $272 \pm 305 \text{ ng/mL}$ plasma). A variety of HCT conditioning
134 regimens have been developed with varying degrees of organ toxicity and myelosuppression. The majority
135 of patients in our cohort received reduced intensity conditioning therapy (RIC, n=17), whereas a single
136 patient received myeloablative conditioning therapy (id number 005). Comparison of cfDNA tissues-of-

137 origin in plasma before and after conditioning showed a significant drop in blood-derived cfDNA as
138 expected from the function of the conditioning therapy (mean proportion of hematopoietic cell cfDNA
139 decreased from $79\% \pm 10\%$ to $59\% \pm 20\%$, p -value=0.0014, **Fig. 2a**). The proportion of blood-derived
140 cfDNA increased to $84\% \pm 10\%$ at engraftment (p -value = 5.6×10^{-5} , **Fig. 2a**). The most notable effect of
141 stem cell infusion and engraftment was a significant increase in the absolute concentration of cfDNA (mean
142 human-derived cfDNA concentration from 256 ng/mL on day of transplant to 2149 ng/mL at engraftment
143 [p -value = 0.013], **Fig. 2b**).

144 145 **Performance analysis**

146
147 We next evaluated the performance of a cfDNA tissue-of-origin measurement to predict GVHD
148 (**Fig. 2c**). We defined GVHD here as the clinical manifestation of any stage of the disease within the first
149 6 months post HCT (GVHD+, see Methods). We excluded samples collected after GVHD diagnosis, as
150 these patients received additional GVHD treatment. We found that the concentration of solid-organ specific
151 cfDNA was significantly elevated for patients in the GVHD+ group at engraftment, month 1, 2 and 3 (p -
152 values of 0.17, 0.012, 0.0070, 0.020, respectively), but not at the two pre-transplant time points (p = 0.66
153 prior to conditioning, and p = 0.80 prior to hematopoietic cell infusion). Receiver operating characteristic
154 analysis of the performance of cfDNA as a predictive marker of GVHD yielded an area under the curve
155 (AUC) of 0.7, 0.9, 0.9 and 0.9 at engraftment and months 1, 2, and 3, respectively. These results support
156 the notion that cfDNA predicts GVHD occurrence as early as one month after HCT (mean solid organ
157 cfDNA of 995 and 50 ng/mL plasma for GVHD+ and GVHD-, respectively; AUC = 0.9, p -value < 0.012,
158 **Fig. 2c**).

159
160 To evaluate the ability of this assay to pinpoint the site of incidence of GVHD, we quantified the
161 burden of skin-derived cfDNA in the blood of GVHD negative individuals ($n=9$) and individuals who
162 developed cutaneous GVHD ($n=8$). We found that plasma samples from individuals with GVHD had a
163 higher burden of skin-derived cfDNA when compared to samples from individuals who did not develop
164 cutaneous GVHD (mean skin cfDNA of 20.6 ng/mL plasma and 3.2ng/mL plasma, respectively, p -value =
165 0.015 for samples collected post-transplant and pre-diagnosis, **supplementary fig. 1**). The number of
166 samples from patients diagnosed with hepatic and gastrointestinal GVHD was insufficient to test the
167 performance of the assay to pinpoint GVHD related injury to the liver or gut ($n = 1$ and $n = 3$, respectively).

168
169 We next studied the response to GVHD treatment for three similar patients for which samples and
170 cfDNA tissues-of-origin analyses were available after GVHD diagnosis (male patients with RIC
171 chemotherapy and similar GVHD diagnosis timepoints). These patients were diagnosed with GVHD
172 between days 28 and 39 post HCT and two plasma samples after diagnosis were available for each patient.
173 The first patient was diagnosed with mild GVHD (cutaneous stage 1, overall grade I; resolved day 98), and
174 the tissue-of-origins of cfDNA followed a similar pattern observed for GVHD negative patients (**Fig. 3a,b**).
175 The second patient was diagnosed with moderate GVHD (cutaneous stage 3, overall grade II; resolved day
176 137). cfDNA tissue-of-origin profiling identified an increase in solid-organ derived cfDNA after diagnosis
177 (36.5 ng/mL at diagnosis, 199.4 ng/mL and 254.1 ng/mL at months 2 and 3, respectively; **Fig. 3c**). The
178 third patient was diagnosed with severe GVHD (cutaneous stage 4, overall grade IV; unresolved; mortality
179 day 91; **Fig. 3d**). cfDNA tissue-of-origin profiling for samples after diagnosis of this patient revealed an
180 increase in solid-organ derived cfDNA in the blood of this patient despite increasingly potent GVHD
181 treatment (233.8 ng/mL at month 1 and 1217.7 ng/mL at month 2; tacrolimus at month 1, and tacrolimus,
182 sirolimus, ruxolitinib, and glucocorticoids at month 2). These three examples illustrate the potential utility
183 of cfDNA tissue-of-origin profiling to monitor GVHD treatment response and outcome.

184 185 186 **Plasma infectome after HCT**

187

188 It is not only human host cells that shed their DNA into the blood; cfDNA from viruses and bacteria
189 can be detected in the circulation, providing a means to screen for infection via metagenomic cfDNA
190 sequencing^{14,21,31,32}. This may be a particularly powerful approach in the context of HCT, given the high
191 incidence of infectious complications, and the broad range of microorganisms that can cause disease in
192 HCT³³. To test this concept, we mined all data from all patients for microbial derived sequences. In a
193 previous study, we found close agreement between the abundance of organisms measured by shotgun
194 sequencing of untreated and bisulfite-treated cfDNA, despite the reduction in sequence complexity inherent
195 to bisulfite conversion, confirming the possibility to perform metagenomic cfDNA sequencing by WGBS¹⁵.
196 To identify microbial-derived cfDNA after WGBS, we first identified and removed host related sequences
197 and we then aligned the remaining unmapped reads to a set of microbial reference genomes ($0.9 \pm 0.4\%$ of
198 total reads, Materials and Methods). We implemented a background correction algorithm to remove
199 contributions due to alignment noise and environmental contamination³⁴.

200
201 Using this procedure, we found a significant increase in the burden of cfDNA derived from DNA
202 viruses after HCT (viral cfDNA biomass 1.4×10^{-3} ng/mL and 1.5×10^{-2} ng/mL, day of transplant and at
203 month 3 respectively, p-value = 0.0082 **Fig. 4a**), but not in bacterial cfDNA (bacterial cfDNA biomass
204 9.8×10^{-3} ng/mL and 3.3×10^{-2} ng/mL, day of transplant and at month 3 respectively, p-value = 0.50). We and
205 others have previously reported a link between the abundance in plasma of *Anelloviridae* and the degree of
206 immunosuppression in solid-organ transplantation, and HCT^{30,33}. In line with these observations, the
207 increase in cfDNA derived from DNA viruses was largely due to an increase in the burden of *Anelloviridae*
208 cfDNA in the first months after HCT (**Fig. 4b**)³⁵. *Herpesviridae* and *Polyomaviridae* frequently establish
209 latent infection in adults and may reactivate after allogeneic HCT³⁶. We identified cfDNA from Human
210 *Herpesviridae* and *Polyomaviridae* in 35 of 106 samples from 15 of 18 patients (**Fig. 4c**). In contrast to
211 *Anelloviridae*, we did not observe a consistent increase in the burden of cfDNA from these viruses after
212 HCT (**Fig. 4b**). Our detection of BK polyomavirus is concordant with clinical diagnosis of BK virus disease.
213 Four individuals in our cohort were diagnosed with BK polyomavirus disease, defined here as the
214 combination of clinical BK polyomavirus disease symptoms and a positive blood or urine BK PCR test in
215 the absence of other causes of genitourinary pathology (see Methods). Our metagenomic assay detected
216 BK polyomavirus cfDNA in the plasma of 4 out of 5 patients with a positive BK blood test (Spearman's
217 rho 0.79, p-value $< 2.2 \times 10^{-16}$), but did not for patients with a negative blood test but a positive urine test,
218 (n=15).

219
220 We identified cfDNA from 7 different genera of bacteria (10 species, **supplementary fig. 2**).
221 Interestingly, all of the identified species are well documented intestinal commensal organisms, in
222 agreement with results from a recent study by Armstrong *et al.*³⁷ that suggest a loss of the integrity of the
223 gut vascular barrier associated with GVHD (**supplementary fig. 2**). For a single patient with unresolved
224 stage IV skin GVHD we identified a potential bloodstream infection with *Klebsiella pneumoniae*. Two
225 patients in this cohort developed a clinically diagnosed Streptococcus bloodstream infection within the first
226 6 months post-transplant. We did not detect Streptococcus cfDNA by metagenomic cfDNA sequencing for
227 these two patients, potentially because the infection timepoints were at least six days away from the nearest
228 plasma collection time point and bloodstream infections with Streptococcus species rapidly clear after the
229 initiation of antimicrobial treatment³⁸.

230 231 **DISCUSSION**

232
233 We have described a cfDNA assay with the potential to detect both GVHD related injury and
234 infection after allogeneic HCT. This work was inspired by recent studies that have shown that transplant
235 donor specific cfDNA in the blood of solid-organ transplant recipients is a quantitative marker of solid-
236 organ transplant injury^{12,13}. A variety of commercial assays to quantify donor cfDNA in solid organ
237 transplant recipients are already available^{16,39-41}, and studies are ongoing to test the utility of donor cfDNA
238 to screen for allograft rejection, injury and failure after kidney, liver, heart, pancreas and lung

239 transplantation⁴². We reasoned that cfDNA may also inform injury to vascularized tissues due to GVHD
240 after HCT. To quantify cfDNA derived from any tissue, we implemented bisulfite sequencing of cfDNA,
241 to profile cytosine methylation marks that are comprised within cfDNA and that are cell, tissue and organ
242 type specific. Several other epigenetic marks, including hydroxymethylation⁴³ and histone modifications⁴⁴,
243 can inform the tissues-of-origin of cfDNA, and profiling of these marks may also be useful to monitor
244 GVHD after HCT.

245
246 In recent years, several protein biomarkers have been investigated for the diagnosis of GVHD.
247 Proteomic approaches have yielded candidate markers that are not directly involved in the pathogenesis of
248 GVHD, but that are secreted as a result of end-organ damage^{46,48}. ST2 and REG3 α , which both derive from
249 the gastrointestinal tract, are two such biomarkers with the strongest predictive power. The cfDNA assay
250 presented here may provide inherent advantages over protein biomarker technologies. First, because the
251 concentration of tissue-specific DNA can be directly related to the degree of cellular injury, this assay is
252 easy to interpret, and offers a measure of injury that can in principle be followed over time. Second, the
253 cfDNA assay explored here provides a generalizable approach to measure injury to any tissue, whereas
254 protein injury markers may not be available for all cell and tissue types. Third, this assay is compatible with
255 a variety of quantitative nucleic acid measurement technologies, including digital and quantitative PCR and
256 DNA sequencing. Fourth, this assay does not depend on antibodies, which come with challenges of
257 specificity and reproducibility⁴⁹.

258
259 Whole genome bisulfite sequencing is not only responsive to human host derived cfDNA, but also
260 to microbial cfDNA that may be present in the blood circulation. Several recent studies have demonstrated
261 the value of cfDNA metagenomic sequencing to broadly detect infection in a variety of clinical settings,
262 including urinary tract infection^{14,15}, sepsis⁵⁰ and invasive fungal disease⁵¹. In HCT, sequencing of cell-free
263 DNA has been used to identify pathogens in blood before clinical onset of bloodstream infections⁵². We
264 have previously shown that bisulfite sequencing of cfDNA is compatible with metagenomic analyses,
265 despite the reduction in sequence complexity due to bisulfite conversion of all unmethylated cytosines¹⁵.
266 Here, we investigated the potential to screen for microbial and viral derived cfDNA in plasma of HCT
267 recipients via bisulfite sequencing of cfDNA. BK virus cfDNA was detected using this approach for
268 samples that were BK virus positive in blood, but not for those that were only BK virus positive in urine.
269 In addition, while we note increased herpesvirus genomic abundance, during immunosuppression,
270 differentiating lytic and latent infection and the corresponding host response could be vital in improving
271 patient care. The assay reported here therefore has the potential to simultaneously inform about GVHD,
272 from the tissues-of-origin of host cfDNA, and infection, from metagenomic analysis of microbial cfDNA.
273 Compared to conventional metagenomic sequencing, this assay requires one additional experimental step
274 to bisulfite convert cfDNA, which can be completed within approximately 2 hours and is compatible with
275 multiple existing next-generation sequencing workflows.

276
277 This is a proof-of-principle study with several limitations that need to be addressed in future work.
278 The scope of the current study with 18 patients was not powered to detect any association of cfDNA with
279 acute GVHD involving organs other than skin (liver, gut). Our results suggest that cfDNA tissue-of-origin
280 profiling is predictive of acute GVHD, but larger studies will be needed to extend the current observations
281 to other sites of organ damage, and to assess its utility in detecting and diagnosing chronic GVHD.

282 283 284 **MATERIALS AND METHODS**

285
286 **Study cohort.** We performed a nested case-control study within a prospective cohort of adult patients
287 undergoing allogeneic HCT at Dana-Farber Cancer Institute. Patients were followed for 6 months after
288 HCT. Patients were selected for this study on a rolling basis, and were placed in the GVHD case or control
289 groups based on clinical manifestation of the disease within the first 6 months after HCT. Individuals were

290 excluded from the study if they did not provide blood samples for at least 5 of the 6 studied time points
291 (pre-conditioning, day of transplant, engraftment, months 1, 2 and 3). The study was approved by the Dana-
292 Farber/Harvard Cancer Center's Office of Human Research Studies. All patients provided written informed
293 consent.

294
295 For this study, we used 106 blood samples collected from 18 allogeneic HCT recipients from August 2018-
296 to April -2019. Baseline patient characteristics were recorded. Covariates of interest included HLA
297 matching, donor relatedness and donor-recipient sex mismatch (**supplementary table 1**). Date of onset of
298 GVHD, as well as GVHD prophylaxis and treatment regimens were documented. GVHD was diagnosed
299 clinically and pathologically. GVHD severity was graded according to the Glucksberg criteria⁴³. Other
300 clinical events of interest included the development of bloodstream infections, BK polyomavirus disease,
301 and clinical disease from other DNA viruses.

302
303 **Engraftment.** Neutrophil engraftment was considered when blood samples contained an absolute
304 neutrophil count greater or equal than 500 cell per microliter of blood on two separate measurements.

305
306 **BK polyomavirus disease identification.** Patients were identified as BK virus disease positive when they
307 presented BK-related urinary symptoms that correlated with positive BK qPCR test in either urine or blood
308 ($>10^5$ copies/ mL in urine, >0 copies/mL in blood; Viracor BK qPCR test, reference #2500) and did not
309 have evidence of any other cause of genitourinary pathology at the time of symptom onset.

310
311 **Blood sample collection and plasma extraction.** Blood samples were collected through standard
312 venipuncture in EDTA tubes (Becton Dickinson (BD), reference #366643) on admission, before the
313 beginning of the conditioning chemotherapy; on the day of HCT after the completion of the conditioning
314 chemotherapy, at engraftment (usually 14 to 21 days after HCT), and at months 1, 2, and 3 post-HCT.
315 Plasma was extracted through blood centrifugation (2000rpm for 10 minutes using a Beckman Coulter
316 Allegra 6R centrifuge) and stored in 0.5-2mL aliquots at -80°C . Plasma samples were shipped from DFCI
317 to Cornell University on dry ice.

318
319 **Nucleic acid control preparation.** Synthetic oligos were prepared (IDT, **supplementary table 2**), mixed
320 in equal proportions, and diluted at approximately 150 ng/ul. At the time of cfDNA extraction, 8ul of control
321 was added to 1992 μL of 1xPBS and processed as a sample in all downstream experiments.

322
323 **Cell-free DNA extraction.** cfDNA was extracted according to manufacturer recommendations (Qiagen
324 Circulating Nucleic Acid Kit, reference #55114, elution volume 45 μl). Eluted DNA was quantified using a
325 Qubit 3.0 Fluorometer (using 2 μL of eluted DNA). Measured cfDNA concentration was obtained using the
326 following formula:

327
328
$$cfDNA\ concentration = \frac{(Eluted\ cfDNA\ concentration) * (Elution\ volume)}{(Plasma\ volume)}$$

329
330 **Whole-genome bisulfite sequencing.** cfDNA and nucleic acid controls were bisulfite treated according to
331 manufacturer recommendations (Zymo Methylation Lightning Kit, reference #D5030). Sequencing
332 libraries were prepared using a previously described single-stranded library preparation protocol²¹.
333 Libraries were quality-controlled through DNA fragment analysis (Agilent Fragment analyzer) and
334 sequenced on an Illumina NextSeq550 machine using 2x75bp reads. Nucleic acid controls were sequenced
335 a ~1% of the total sequencing lane.

336
337 **Human genome alignment.** Adapter sequences were trimmed using BBTools⁵³. The Bismark⁵⁴ alignment
338 tool was used to align reads to the human genome (version hg19), remove PCR duplicates and calculate
339 methylation densities.

340
341 **Reference tissue methylation profiles and tissue of origin measurement.** Reference tissue methylomes
342 were obtained from publicly available databases²⁶⁻³⁰ (**supplementary dataset 1**). Genomic coordinates
343 from different sources were normalized and converted to a standard 4 column bed file (columns:
344 chromosome, start, end, methylation fraction) using hg19 assembly coordinates. Methylation profiles were
345 grouped by tissue-type and differentially methylated regions were found using Metilene⁵⁵. Tissues and cell-
346 types of origin were determined using quadratic programming as previously described¹⁵.

347
348 **Metagenomic alignment and quantification of microbial cfDNA.** After WGBS, reads were adapter-
349 trimmed using BBTools⁴⁴, and short reads are merged with FLASH⁵⁶. Sequences were aligned to a C-to-T
350 converted genome using Bismark⁵⁴. Unmapped reads were BLASTed⁵⁷ using hs-blastn⁵⁸ to a list of C-to-T
351 converted microbial reference genomes. A relative abundance of all detected organisms was determined
352 using GRAMMy⁵⁹, and relative genomic abundances are measured as previously described³². Microbial
353 cfDNA fraction was calculated by dividing the unique number of reads mapping to microbial species (after
354 adjusting for the length of each microbial genome in the reference set) to the total number of sequenced
355 reads. Human fraction is estimated as 1 - microbial fraction. Microbial species were then filtered for
356 environmental contamination and alignment noise using previously described methods (LBBC, PNAS).
357 For viral species identification (**Fig. 4c**), custom genomic reference was generated from representative
358 genomes for BK polyomavirus (NC_001538.1), cytomegalovirus (NC_006273.1), herpesvirus 6A
359 (NC_001664.4) and 6B (NC_00898.1), human polyomavirus 6 (NC_014406.1) and 7 (NC_014407.1) and
360 ebstein-barr virus (NC_007605). Reads that BLASTed to these species were then re-aligned via Bismark,
361 and a threshold of 1 mapped sequence per 40 million total sequenced reads was used to positively identify
362 a viral species within a plasma sample.

363
364 **cfDNA concentration.** cfDNA concentration of a specific tissue or microbe is calculated as follows:

365
$$\text{Normalized cfDNA concentration} = \frac{(\text{cfDNA concentration}) * (\text{Nucleic acid control input mass})}{\text{Nucleic acid control output mass}}$$

367
$$\text{Tissue specific cfDNA concentration} = (\text{Normalized cfDNA concentration}) * (\text{human read fraction}) * (\text{tissue proportion})$$

371
$$\text{Microbial cfDNA concentration} = (\text{Normalized cfDNA concentration}) * (\text{microbial read fraction})$$

372
373 **Depth of coverage.** The depth of sequencing was measured by summing the depth of coverage for each
374 mapped base pair on the human genome after duplicate removal, and dividing by the total length of the
375 human genome (hg19, without unknown bases).

376
377 **Bisulfite conversion efficiency.** We estimated bisulfite conversion efficiency by quantifying the rate of
378 C[A/T/C] methylation in human-aligned reads (using MethPipe⁶⁰), which are rarely methylated in
379 mammalian genomes.

380
381 **Statistical analysis.** Statistical analysis was performed in R (version 3.5). All tests were performed using
382 a two-sided Wilcoxon test.

383
384 **Data availability.** The sequencing data generated for this study will be available in the database of
385 Genotypes and Phenotypes (dbGaP). All code used to generate figures and analyze primary data is available
386 at www.github.com/alexpcheng/cfDNAme.

389

390 **CONFLICTS OF INTEREST**

391 The authors declare no conflicts of interest.

392

393 **AUTHOR CONTRIBUTIONS**

394 A.P.C., M.P.C., F.M.M., J.R. and I.D.V. designed the study. A.P.C. and J.S.L. performed experiments.

395 M.P.C., F.M.M., J.R., K.C., K.M.T., J.L.O. and E.S. consented patients and acquired clinical data. A.P.C.,

396 M.P.C., P.B., I.D.V., F.M.M. and J.R. analyzed data. A.P.C., M.P.C., I.D.V., F.M.M. and J.R. wrote the

397 manuscript. All authors reviewed and approved the manuscript.

398

399 **ACKNOWLEDGEMENTS**

400 We would like to thank the Cornell Genomics Center for help with sequencing assays, the Cornell

401 Bioinformatics facility for computational assistance, The Pasquarello Tissue bank at the Dana-Farber

402 Cancer Institute for sample processing and cryopreservation and members of the De Vlaminc Lab for

403 helpful discussions. We thank Francoise Vermeulen of the Cornell Statistical Consulting Unit for helpful

404 discussion. This work was supported by R01AI146165 (to I.D.V, J.R, M.P.C. and F.M.), R21AI133331 (to

405 I.D.V.), R21AI124237 (to I.D.V.), DP2AI138242 (to I.D.V.), a National Sciences and Engineering

406 Research Council of Canada fellowship PGS-D3 (to A.P.C).

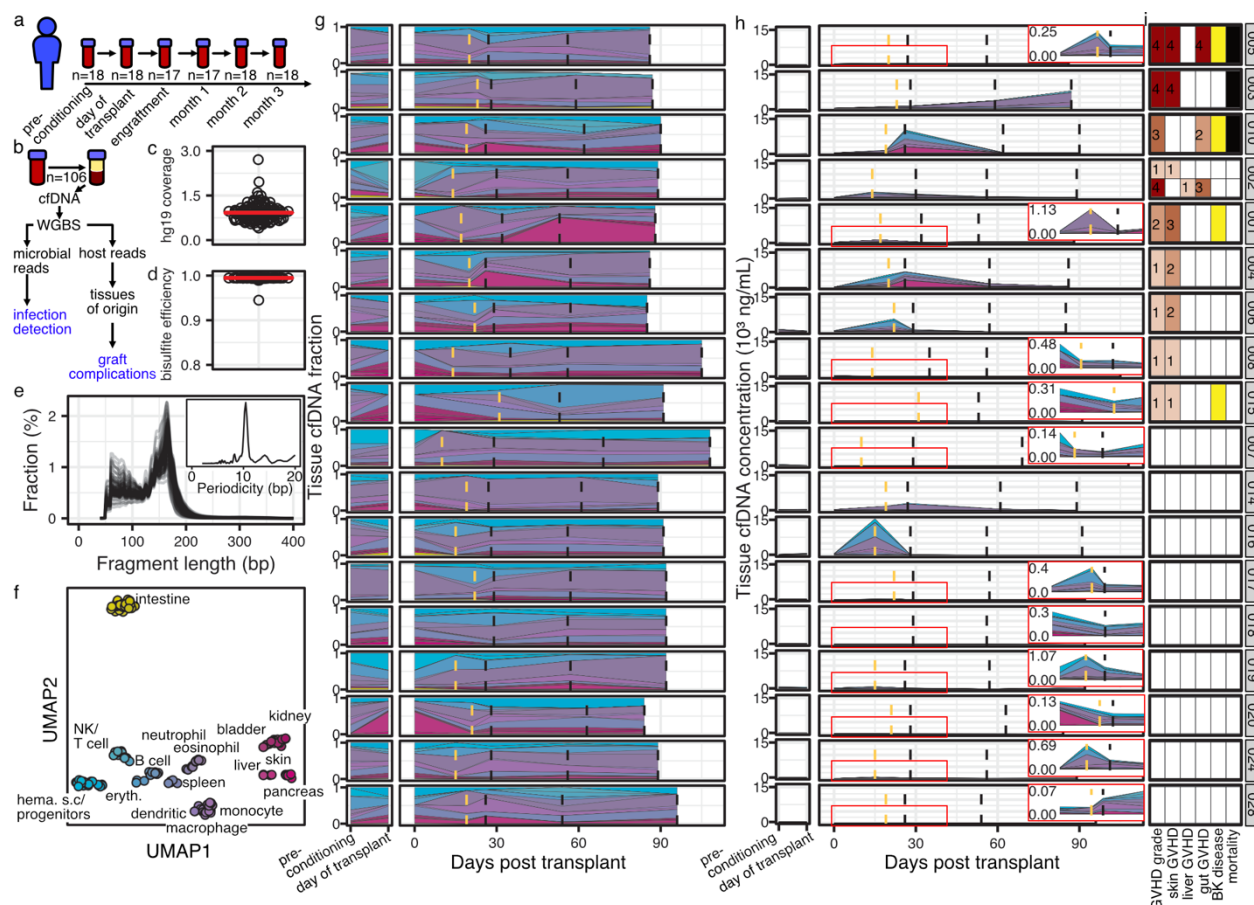
407 **REFERENCES**

- 408
- 409 1. Gratwohl, A. *et al.* Hematopoietic stem cell transplantation: a global perspective. *JAMA* **303**,
- 410 1617–1624 (2010).
- 411 2. McDonald-Hyman, C., Turka, L. A. & Blazar, B. R. Advances and challenges in immunotherapy
- 412 for solid organ and hematopoietic stem cell transplantation. *Sci. Transl. Med.* **7**, 280rv2 (2015).
- 413 3. Niederwieser, D. *et al.* Hematopoietic stem cell transplantation activity worldwide in 2012 and a
- 414 SWOT analysis of the Worldwide Network for Blood and Marrow Transplantation Group including the
- 415 global survey. *Bone Marrow Transplant.* **51**, 778–785 (2016).
- 416 4. Paczesny, S. Discovery and validation of graft-versus-host disease biomarkers. *Blood* **121**, 585–
- 417 594 (2013).
- 418 5. Ferrara, J. L. M. & Deeg, H. J. Graft-versus-Host Disease. *N. Engl. J. Med.* **324**, 667–674 (1991).
- 419 6. Ferrara, J. L. M. & Chaudhry, M. S. GVHD: biology matters. *Blood Adv.* **2**, 3411–3417 (2018).
- 420 7. Vogelsang, G. B., Lee, L. & Bensen-Kennedy, D. M. Pathogenesis and Treatment of Graft-
- 421 Versus-Host Disease After Bone Marrow Transplant. *Annu. Rev. Med.* **54**, 29–52 (2003).
- 422 8. Sun, K. *et al.* Plasma DNA tissue mapping by genome-wide methylation sequencing for
- 423 noninvasive prenatal, cancer, and transplantation assessments. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E5503–
- 424 E5512 (2015).
- 425 9. Mouliere, F. *et al.* Detection of cell-free DNA fragmentation and copy number alterations in
- 426 cerebrospinal fluid from glioma patients. *EMBO Mol. Med.* **10**, e9323 (2018).
- 427 10. Mouliere, F. *et al.* Enhanced detection of circulating tumor DNA by fragment size analysis. *Sci.*
- 428 *Transl. Med.* **10**, (2018).
- 429 11. Lehmann-Werman, R. *et al.* Identification of tissue-specific cell death using methylation patterns
- 430 of circulating DNA. *Proc. Natl. Acad. Sci.* **113**, E1826–E1834 (2016).
- 431 12. Vlaminc, I. D. *et al.* Circulating Cell-Free DNA Enables Noninvasive Diagnosis of Heart
- 432 Transplant Rejection. *Sci. Transl. Med.* **6**, 241ra77-241ra77 (2014).
- 433 13. De Vlaminc, I. *et al.* Noninvasive monitoring of infection and rejection after lung
- 434 transplantation. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 13336–13341 (2015).
- 435 14. Burnham, P. *et al.* Urinary cell-free DNA is a versatile analyte for monitoring infections of the
- 436 urinary tract. *Nat. Commun.* **9**, 1–10 (2018).
- 437 15. Cheng, A. P. *et al.* A cell-free DNA metagenomic sequencing assay that integrates the host injury
- 438 response to infection. *Proc. Natl. Acad. Sci.* **116**, 18738–18744 (2019).
- 439 16. Grskovic, M. *et al.* Validation of a Clinical-Grade Assay to Measure Donor-Derived Cell-Free
- 440 DNA in Solid Organ Transplant Recipients. *J. Mol. Diagn. JMD* **18**, 890–902 (2016).
- 441 17. Sharon, E. *et al.* Quantification of transplant-derived circulating cell-free DNA in absence of a
- 442 donor genotype. *PLOS Comput. Biol.* **13**, e1005629 (2017).
- 443 18. Lokk, K. *et al.* DNA methylome profiling of human tissues identifies global and tissue-specific
- 444 methylation patterns. *Genome Biol.* **15**, 3248 (2014).
- 445 19. Moss, J. *et al.* Comprehensive human cell-type methylation atlas reveals origins of circulating
- 446 cell-free DNA in health and disease. *Nat. Commun.* **9**, 1–12 (2018).
- 447 20. Gansauge, M.-T. & Meyer, M. Single-stranded DNA library preparation for the sequencing of
- 448 ancient or damaged DNA. *Nat. Protoc.* **8**, 737–748 (2013).
- 449 21. Burnham, P. *et al.* Single-stranded DNA library preparation uncovers the origin and diversity of
- 450 ultrashort cell-free DNA in plasma. *Sci. Rep.* **6**, 27859 (2016).

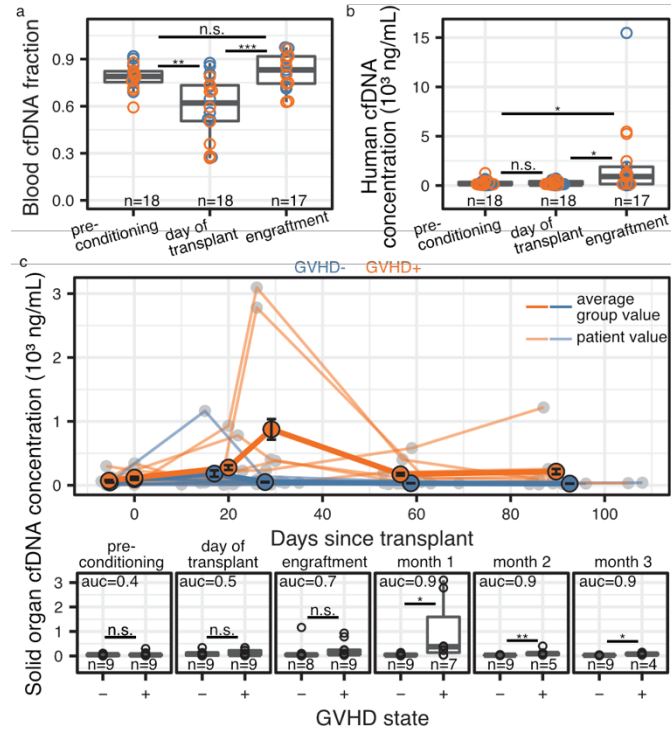
- 451 22. Miura, F., Enomoto, Y., Dairiki, R. & Ito, T. Amplification-free whole-genome bisulfite
452 sequencing by post-bisulfite adaptor tagging. *Nucleic Acids Res.* **40**, e136 (2012).
- 453 23. Jiang, P. & Lo, Y. M. D. The Long and Short of Circulating Cell-Free DNA and the Ins and Outs
454 of Molecular Diagnostics. *Trends Genet.* **32**, 360–371 (2016).
- 455 24. Snyder, M. W., Kircher, M., Hill, A. J., Daza, R. M. & Shendure, J. Cell-free DNA comprises an
456 in vivo nucleosome footprint that informs its tissues-of-origin. *Cell* **164**, 57–68 (2016).
- 457 25. Werner, B. *et al.* Circulating cell-free DNA from plasma undergoes less fragmentation during
458 bisulfite treatment than genomic DNA due to low molecular weight. *PLOS ONE* **14**, e0224338 (2019).
- 459 26. ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project.
460 *Science* **306**, 636–640 (2004).
- 461 27. Bujold, D. *et al.* The International Human Epigenome Consortium Data Portal. *Cell Syst.* **3**, 496-
462 499.e2 (2016).
- 463 28. Albrecht, F., List, M., Bock, C. & Lengauer, T. DeepBlue epigenomic data server: programmatic
464 data retrieval and analysis of epigenome region sets. *Nucleic Acids Res.* **44**, W581–W586 (2016).
- 465 29. Bernstein, B. E. *et al.* The NIH Roadmap Epigenomics Mapping Consortium. *Nat. Biotechnol.*
466 **28**, 1045–1048 (2010).
- 467 30. Fernández, J. M. *et al.* The BLUEPRINT Data Analysis Portal. *Cell Syst.* **3**, 491–495.e5 (2016).
- 468 31. Kowarsky, M. *et al.* Numerous uncharacterized and highly divergent microbes which colonize
469 humans are revealed by circulating cell-free DNA. *Proc. Natl. Acad. Sci.* (2017)
470 doi:10.1073/pnas.1707009114.
- 471 32. De Vlamincq, I. *et al.* Temporal response of the human virome to immunosuppression and
472 antiviral therapy. *Cell* **155**, 1178–1187 (2013).
- 473 33. Tomblyn, M. *et al.* Guidelines for Preventing Infectious Complications among Hematopoietic
474 Cell Transplant Recipients: A Global Perspective. *Biol. Blood Marrow Transplant. J. Am. Soc. Blood*
475 *Marrow Transplant.* **15**, 1143–1238 (2009).
- 476 34. Burnham, P. *et al.* Separating the signal from the noise in metagenomic cell-free DNA
477 sequencing. *bioRxiv* 734756 (2019) doi:10.1101/734756.
- 478 35. Schmitz, J. *et al.* The Value of Torque Teno Virus (TTV) as a Marker for the Degree of
479 Immunosuppression in Adult Patients after Hematopoietic Stem Cell Transplantation (HSCT). *Biol.*
480 *Blood Marrow Transplant.* (2019) doi:10.1016/j.bbmt.2019.11.002.
- 481 36. Marr, K. A. Delayed opportunistic infections in hematopoietic stem cell transplantation patients:
482 a surmountable challenge. *Hematol. Educ. Program Am. Soc. Hematol. Am. Soc. Hematol. Educ.*
483 *Program* **2012**, 265–270 (2012).
- 484 37. Armstrong, A. E. *et al.* Cell-free DNA next-generation sequencing successfully detects infectious
485 pathogens in pediatric oncology and hematopoietic stem cell transplant patients at risk for invasive fungal
486 disease. *Pediatr. Blood Cancer* **66**, e27734 (2019).
- 487 38. Cheng, M. P. *et al.* Blood Culture Results Before and After Antimicrobial Administration in
488 Patients With Severe Manifestations of Sepsis: A Diagnostic Study. *Ann. Intern. Med.* (2019)
489 doi:10.7326/M19-1696.
- 490 39. Sigdel, T. K. *et al.* A rapid noninvasive assay for the detection of renal transplant injury.
491 *Transplantation* **96**, 97–101 (2013).
- 492 40. Beck, J. *et al.* Digital droplet PCR for rapid quantification of donor DNA in the circulation of
493 transplant recipients as a potential universal biomarker of graft injury. *Clin. Chem.* **59**, 1732–1741 (2013).

- 494 41. Gielis, E. M. *et al.* Cell-Free DNA: An Upcoming Biomarker in Transplantation. *Am. J.*
495 *Transplant. Off. J. Am. Soc. Transplant. Am. Soc. Transpl. Surg.* **15**, 2541–2551 (2015).
- 496 42. Knight, S. R., Thorne, A. & Lo Faro, M. L. Donor-specific Cell-free DNA as a Biomarker in
497 Solid Organ Transplantation. A Systematic Review. *Transplantation* **103**, 273–283 (2019).
- 498 43. Song, C.-X. *et al.* 5-Hydroxymethylcytosine signatures in cell-free DNA provide information
499 about tumor types and stages. *Cell Res.* **27**, 1231–1242 (2017).
- 500 44. Sadeh, R. *et al.* ChIP-seq of plasma cell-free nucleosomes identifies cell-of-origin gene
501 expression programs. *bioRxiv* 638643 (2019) doi:10.1101/638643.
- 502 45. Harris, A. C. *et al.* Plasma biomarkers of lower gastrointestinal and liver acute GVHD. *Blood*
503 **119**, 2960 LP – 2963 (2012).
- 504 46. Chen, Y.-B. & Cutler, C. S. Biomarkers for acute GVHD: can we predict the unpredictable? *Bone*
505 *Marrow Transplant.* **48**, 755–760 (2012).
- 506 47. Paczesny, S. Discovery and validation of graft-versus-host disease biomarkers. *Blood* **121**, 585
507 LP – 594 (2013).
- 508 48. Paczesny, S. *et al.* Three Biomarker Panel at Day 7 and 14 Can Predict Development of Grade II-
509 IV Acute Graft-Versus-Host Disease. *Blood* **116**, 675 LP – 675 (2010).
- 510 49. Voskuil, J. L. A. The challenges with the validation of research antibodies. *F1000Research* **6**,
511 (2017).
- 512 50. Blauwkamp, T. A. *et al.* Analytical and clinical validation of a microbial cell-free DNA
513 sequencing test for infectious disease. *Nat. Microbiol.* **4**, 663–674 (2019).
- 514 51. Hong, D. K. *et al.* Liquid biopsy for infectious diseases: sequencing of cell-free plasma to detect
515 pathogen DNA in patients with invasive fungal disease. *Diagn. Microbiol. Infect. Dis.* **92**, 210–213
516 (2018).
- 517 52. Goggin, K. P. *et al.* Evaluation of Plasma Microbial Cell-Free DNA Sequencing to Predict
518 Bloodstream Infection in Pediatric Patients With Relapsed or Refractory Cancer. *JAMA Oncol.* (2019)
519 doi:10.1001/jamaoncol.2019.4120.
- 520 53. Bushnell, B., Rood, J. & Singer, E. BBMerge – Accurate paired shotgun read merging via
521 overlap. *PLOS ONE* **12**, e0185056 (2017).
- 522 54. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq
523 applications. *Bioinformatics* **27**, 1571–1572 (2011).
- 524 55. Jühling, F. *et al.* metilene: fast and sensitive calling of differentially methylated regions from
525 bisulfite sequencing data. *Genome Res.* **26**, 256–262 (2016).
- 526 56. Magoč, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome
527 assemblies. *Bioinformatics* **27**, 2957–2963 (2011).
- 528 57. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search
529 tool. *J. Mol. Biol.* **215**, 403–410 (1990).
- 530 58. Chen, Y., Ye, W., Zhang, Y. & Xu, Y. High speed BLASTN: an accelerated MegaBLAST search
531 tool. *Nucleic Acids Res.* **43**, 7762–7768 (2015).
- 532 59. Xia, L. C., Cram, J. A., Chen, T., Fuhrman, J. A. & Sun, F. Accurate Genome Relative
533 Abundance Estimation Based on Shotgun Metagenomic Reads. *PLoS ONE* **6**, (2011).
- 534 60. Song, Q. *et al.* A Reference Methylome Database and Analysis Pipeline to Facilitate Integrative
535 and Comparative Epigenomics. *PLOS ONE* **8**, e81148 (2013).
- 536

537 **FIGURES and CAPTIONS**
538

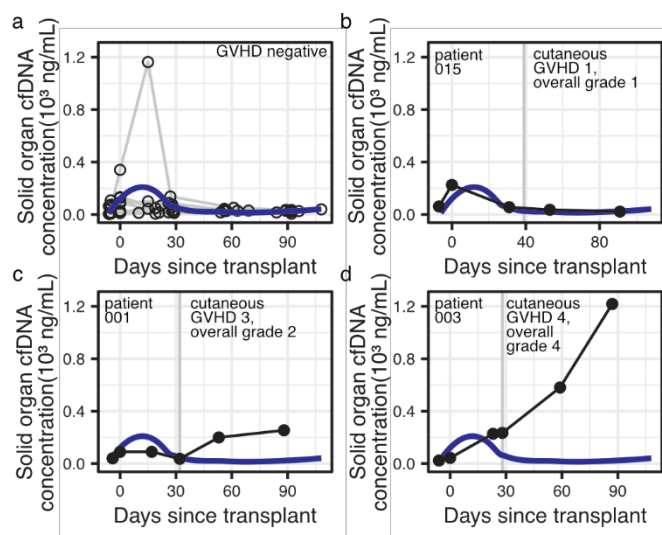


539 **Figure 1.** Study workflow. **a** Blood samples from hematopoietic cell transplant recipients (n=18) were
540 prospectively collected at six time points. **b** WGBS is performed on cfDNA extracted from patient plasma.
541 Sequenced cfDNA is processed through a custom bioinformatics pipeline. **c,d** hg19 sequence coverage (**c**)
542 and bisulfite conversion efficiency (**d**) of sequenced cfDNA (n=106). Red lines indicate the median. **e**
543 Fragment length profiles of 106 cfDNA samples after bisulfite treatment. Inset: Fourier analysis reveals a
544 10.4 bp periodicity in the fragment length profiles of bisulfite treated cfDNA. **f** UMAP dimensional
545 reduction of cell and tissue methylation profiles. Individual tissues are colored by UMAP coordinates using
546 a linear gradient where each of the four corners is either cyan, magenta, yellow or black. **g,h** Tissue cfDNA
547 fraction (**g**) and tissue cfDNA concentration (**h**) of 18 patients with and without aGVHD; tissues are colored
548 by their average color provided by the UMAP coordinate-based color scheme. Vertical dashed lines
549 represent engraftment (yellow) and month 1, 2, and 3 (black) timepoints. Red boxes in **h** are magnified
550 areas of engraftment and month 1 timepoints. **i** Patient GVHD and BK virus disease. GVHD grades and
551 stages are indicated within the individual boxes. Patient identification numbers are listed on the right. Rows
552 in **h,g,i** are sorted by mortality, overall GVHD grade and GVHD stage.
553
554



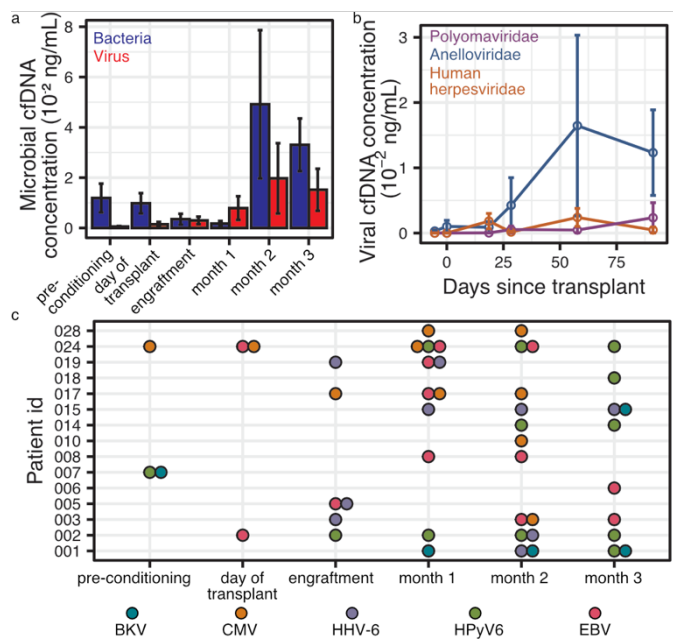
555
556
557
558
559
560
561
562

Figure 2. Host-derived cfDNA dynamics before and after HCT. **a,b** Effect of conditioning and HCT infusion on cfDNA composition (a) and absolute concentration (b). **c** Solid organ derived cfDNA concentration in plasma. Top row: dark lines represent mean solid-organ cfDNA and days post transplant for each patient time point. Error Bars represent standard error of the mean. Bottom row: solid organ cfDNA by time point. Samples are removed from analysis if plasma was collected after aGVHD diagnosis. * p-value < 0.05; ** p-value < 0.01



563
564
565
566
567
568

Figure 3. Solid organ cfDNA concentration dynamics. **a** Solid organ cfDNA concentration in GVHD negative individuals. **b,c,d** Solid organ cfDNA concentration in three patients. Blue line represents loess-smoothed solid organ cfDNA in GVHD negative patients.



569
570
571
572
573
574

Figure 4. Plasma infectome. **a** Microbial cfDNA concentration by time point. **b** Polyomavirus, anellovirus and human herpesvirus abundance in plasma before and after HCT. **c** Human herpesvirus and polyomavirus species detected per patient (n=15) per time point (patients without detectable herpesvirus or human polyomavirus are not shown (n=3)). Error bars represent standard error of the mean.

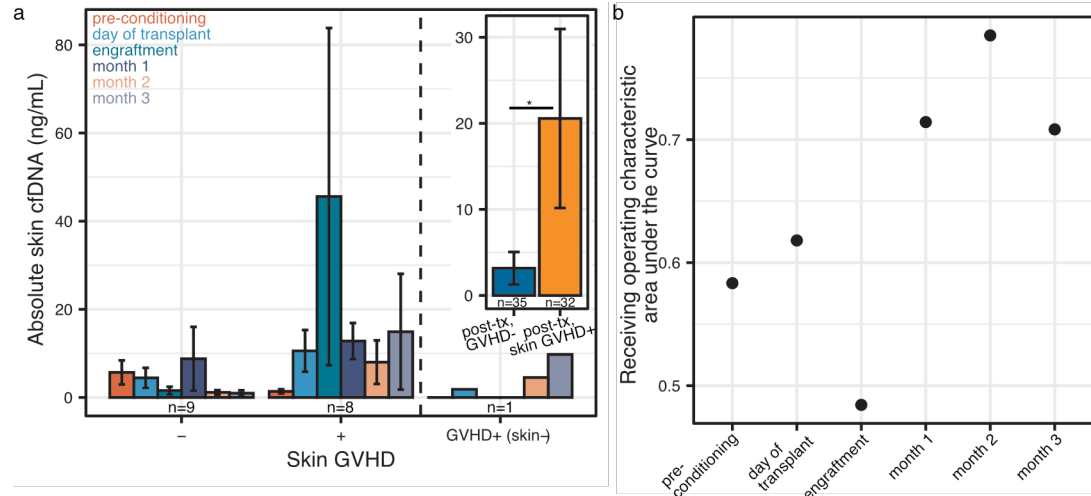
Supplemental Materials for:

Cell-free DNA Tissues-of-Origin Profiling to Predict Graft versus Host Disease and Detect Infection after Hematopoietic Cell Transplantation

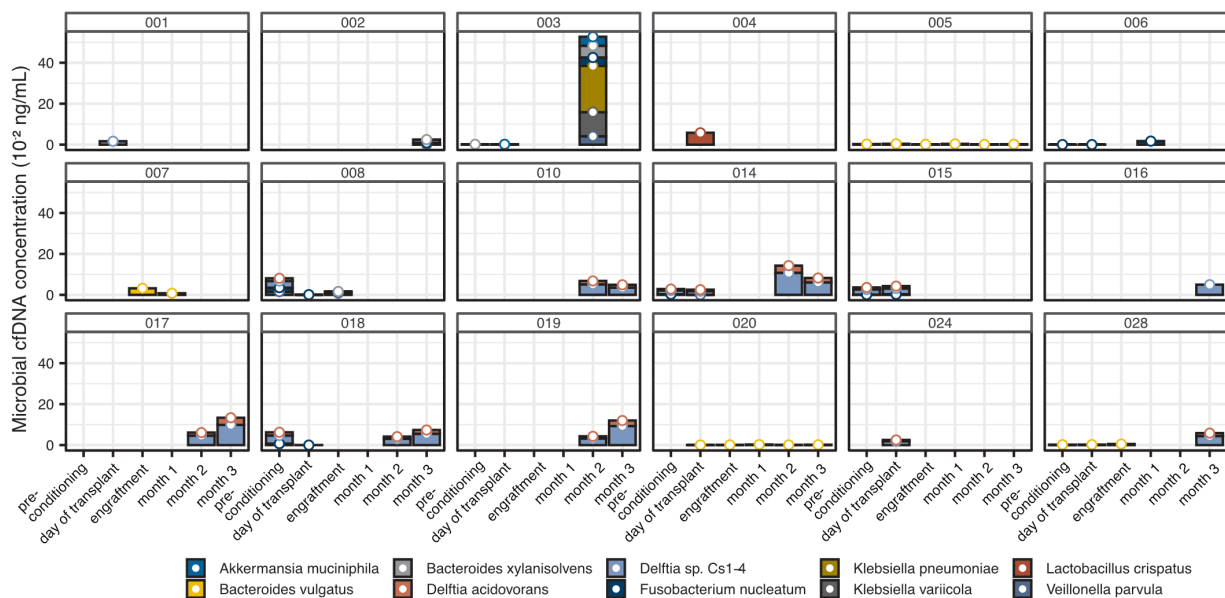
Authors: Alexandre Pellan Cheng¹, Matthew Pellan Cheng^{2,3}, Joan Sesing Lenz¹, Kaiwen Chen^{2,3}, Philip Burnham⁴, Kaitlyn Marie Timblin^{2,3}, José Luis Orejas^{2,3}, Emily Silverman^{2,3}, Francisco M. Marty^{2,3}, Jerome Ritz^{2,5}, Iwijn De Vlamincck^{1,*}

Affiliations:

- 1 Meinig School of Biomedical Engineering, Cornell University, Ithaca, NY, USA
- 2 Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA
- 3 Division of Infectious Disease, Brigham and Women's Hospital, Boston, MA, USA
- 4 Department of Bioengineering, University of Pennsylvania, Philadelphia, PA, USA
- 5 Department of Medicine, Harvard Medical School, Boston, MA, USA
- * Corresponding author ([vlaminck@cornell.edu](mailto:vlamincck@cornell.edu))



Supplementary figure 1. Skin-derived cfDNA in patients with aGVHD. a Skin cfDNA concentration for each individual timepoint. Inset: comparison between GVHD negative and skin GVHD positive individuals (all post-transplant collection timepoints aggregated). **b** Receiving operating characteristic area under the curves for each timepoint between GVHD negative and skin GVHD positive individuals (post-diagnosis samples also considered). * p-value <0.05. Error bars represent standard error of the mean.



Supplementary figure 2. Bacterial species detected in HCT patient plasma through WGBS. Each box represents a single patient at each profiled time point.

Supplementary table 1. Clinical information

Age at enrollment -- median (range) (years)	63 (20-73)	Relation to donor -- no (%)	
Female sex -- no (%)	7 (38%)	Unrelated	15 (83%)
Race / Ethnicity -- no (%)		Related	3 (17%)
Caucasian	17 (94%)	Recipient CMV status -- no (%)	
American Indian/Alaskan Native	1 (6%)	R+	10 (56%)
Reason for hematopoietic cell transplant -- no (%)¹		R-	8 (44%)
Acute myeloid leukemia	4 (22%)	GVHD status -- no (%)²	
Acute lymphocytic leukemia	4 (22%)	Overall grade I	5 (28%)
Mantle cell lymphoma	2 (11%)	Overall grade II	1 (6%)
Myelodysplastic syndrome	2 (11%)	Overall grade III	1 (6%)
T-Cell lymphoma	2 (11%)	Overall grade IV	3 (17%)
Aplastic anemia	1 (6%)	Skin staging I	3 (17%)
Chronic myelomonocytic leukemia	1 (6%)	Skin staging II	2 (11%)
Chronic lymphocytic leukemia	1 (6%)	Skin staging III	1 (6%)
Myelofibrosis	1 (6%)	Skin staging IV	2 (11%)
Paroxysmal nocturnal hemoglobinuria	1 (6%)	Liver staging I	1 (6%)
Source of HCT -- no (%)		Liver staging II	0 (0%)
Peripheral blood	16 (88%)	Liver staging III	0 (0%)
Umbilical cord	1 (6%)	Liver staging IV	0 (0%)
Bone marrow	1 (6%)	Gut staging I	0 (0%)
HLA matching -- no (%)		Gut staging II	1 (6%)
Match	14 (77%)	Gut staging III	1 (6%)
Mismatch	3 (17%)	Gut staging IV	1 (6%)
Haploidentical	1 (6%)	Conditioning regimen -- no (%)	
Conditioning regimen -- no (%)		Reduced intensity	17 (94%)
Busulfan, Fluradabine	10 (56%)	Myeloablative	1 (6%)
Cyclophosphamide, Fludarabine, total body irradiation (TBI)	2 (11%)	Other characteristics -- no (%)	
Fludarabine, Melphalan	2 (11%)	Mortality	3 (17%)
Clyclophoamamide, TBI	1 (6%)	Previous HCT	0 (0%)
Clyclophosphamide, Fludarabine, TBI, Mesna	1 (6%)	Time to GVHD onset (median \pm std)	95 \pm 64 days
Cyclophosphamide, Fludarabine, TBI, anti-thymocyte globulin	1 (6%)	T-cell depletion	0 (0%)
Busulfan, Fluradabine, Melphalan	1 (6%)	GVHD treatment -- no (%)	
GVHD prophylaxis -- no (%)		Glucocorticoids	6 (67%)
Methotrexate, Tacrolimus, Sirolimus	9 (50%)	Jakafi, Glucocorticoids	3 (33%)
Methotrexate, Tacrolimus	5 (28%)		
Mycophenolate mofetil, Tacrolimus, Post-transplant cyclophosphamide	2 (11%)		

Tacrolimus, Sirolimus	1 (6%)		
Mycophenolate mofetil, Tacrolimus	1 (6%)		

1. One individual received an HCT for two blood disorders
2. One individual had two separate incidences of GVHD

Supplementary table 2. Oligonucleotides comprising nucleic acid control.

	Sequence (5'-3')
oligo1	TTTAACGCATAAACATGCGTTTTGGGTAGTGTTTTTTGGAAACACAGATCCGTGCGCACACCTGGTGGAG
oligo2	ATAAACATGCGTTTTGGGTAGTGTTTTTTGGAAACACAGATCCGTGCGCACACCT
oligo3	GCGTTTTGGGTAGTGTTTTTTGGAAACACAGATCCGTGCG
oligo4	GGTAGTGTTTTTTGGAAACACAGAT