



# Human cytomegalovirus multiple-strain infections and viral population diversity in haematopoietic stem cell transplant recipients analysed by high-throughput sequencing

A. Dhingra<sup>1,2</sup> · J. Götting<sup>1,2</sup> · P. R. Varanasi<sup>2,3,6</sup> · L. Steinbrueck<sup>1,2</sup> · S. Camiolo<sup>4</sup> · J. Zischke<sup>1,2</sup> · A. Heim<sup>1,2</sup> · T. F. Schulz<sup>1,2</sup> · E. M. Weissinger<sup>2,3</sup> · P. C. Kay-Fedorov<sup>1,2</sup> · A. J. Davison<sup>4</sup> · N. M. Suárez<sup>4</sup> · T. Ganzenmueller<sup>1,2,5</sup>

Received: 22 June 2021 / Accepted: 21 September 2021 / Published online: 6 October 2021

© The Author(s) 2021

## Abstract

Human cytomegalovirus (HCMV) is an important opportunistic pathogen in allogeneic haematopoietic stem cell transplant (HSCT) recipients. High-throughput sequencing of target-enriched libraries was performed to characterise the diversity of HCMV strains present in this high-risk group. Forty-four HCMV-DNA-positive plasma specimens (median viral input load 321 IU per library) collected at defined time points from 23 HSCT recipients within 80 days of transplantation were sequenced. The genotype distribution for 12 hypervariable HCMV genes and the number of HCMV strains present (i.e. single- vs. multiple-strain infection) were determined for 29 samples from 16 recipients. Multiple-strain infection was observed in seven of these 16 recipients, and five of these seven recipients had the donor (D)/recipient (R) HCMV-serostatus combination D + R +. A very broad range of genotypes was detected, with an intrahost composition that was generally stable over time. Multiple-strain infection was not associated with particular virological or clinical features, such as altered levels or duration of antigenaemia, development of acute graft-versus-host disease or increased mortality. In conclusion, despite relatively low viral plasma loads, a high frequency of multiple-strain HCMV infection and a high strain complexity were demonstrated in systematically collected clinical samples from this cohort early after HSCT. However, robust evaluation of the pathogenic role of intrahost viral diversity and multiple-strain infection will require studies enrolling larger numbers of recipients.

**Keywords** Haematopoietic stem cell transplantation · Human cytomegalovirus · High-throughput sequencing · Sequence diversity · Genotyping · Multiple-strain infection

## Introduction

Despite excellent screening strategies and antiviral treatment regimens applied prophylactically or pre-emptively, human cytomegalovirus (HCMV) remains a significant pathogen following allogeneic haematopoietic stem cell transplantation (HSCT) [1–4]. Primary infection, reinfection and

---

Edited by: Sebastian Voigt.

---

A. Dhingra and J. Götting have contributed equally.

---

✉ T. Ganzenmueller  
Tina.Ganzenmueller@med.uni-tuebingen.de

<sup>1</sup> Hannover Medical School, Institute of Virology, Hannover, Germany

<sup>2</sup> German Center for Infection Research (DZIF), Site Hannover-Braunschweig, Hannover, Germany

<sup>3</sup> Department of Haematology, Haemostasis and Oncology, Hannover Medical School, Hannover, Germany

<sup>4</sup> MRC-University of Glasgow Centre for Virus Research, Glasgow, UK

<sup>5</sup> Institute for Medical Virology and Epidemiology, University Hospital Tuebingen, Elfriede-Aulhorn-Str. 6, 72076 Tuebingen, Germany

<sup>6</sup> Present Address: National Centre for Biological Sciences, Bangalore, India

reactivation can result in subclinical viraemia or systemic disease and end-organ manifestations. Moreover, HCMV is associated with an increased risk of acute graft-versus-host disease (aGvHD) as well as opportunistic infections by other pathogens [5–7]. One well-established risk factor for HCMV infection and related complications after HSCT is the HCMV immunoglobulin G serostatus of the donor (D) and recipient (R), with the highest risk for recipients with the D – R + combination because of delayed HCMV-specific immune reconstitution [8, 9].

HCMV (species *Human betaherpesvirus 5*) has a double-stranded DNA genome of 236 kbp that contains at least 170 open reading frames (ORFs) encoding functional proteins [10]. The diversity of HCMV strains is particularly apparent in a subset of hypervariable genes, each of which presents several stable genotypes, and interstrain recombination during HCMV evolution has generated huge numbers of strains worldwide [11–16]. In addition, reactivation of multiple latent HCMV strains as well as exogenous reinfection with one or more strains can contribute to viral diversity within individuals [17–19]. Prior to the advent of high-throughput sequencing (HTS), many genotyping studies based on polymerase chain reaction (PCR) and Sanger sequencing of single hypervariable genes were aimed at linking clinical outcome parameters to aspects of viral diversity, but were compromised by limitations in understanding strain complexity within individuals [20–24].

In recent years, several studies have used HTS to analyse diversity across the whole HCMV genome [11, 12, 25]. Target capture approaches, which enrich the viral sequences in a sample by hybridisation to bait libraries, have provided further advances, making it possible to generate HCMV sequence data directly from a variety of HCMV-DNA-positive diagnostic specimens. This relatively unbiased approach avoids unnecessary PCR amplification steps or the need to isolate strains in cell culture [14, 19, 26, 27]. In a previous study, we observed that HCMV-DNAemia was linked to different patterns of viral population dynamics in a retrospectively assembled collection of clinical specimens from HCMV-positive immunocompromised recipients [28]. However, to diminish potential bias due to selection of cases (e.g. by preferentially choosing recipients with high viral loads) and thus to analyse the extent of HCMV genome diversity and its potential impact on clinical outcome more robustly, investigations on a prospective and homogeneous cohort are desirable.

Here, we present HTS-generated data on HCMV populations from plasma specimens collected at defined time points from a clinically well-characterised HSCT cohort, focusing on viral intrahost diversity and strain complexity.

## Materials and methods

### Study design and sample collection

Plasma specimens ( $n=253$ ) were obtained prospectively from a cohort of HSCT recipients ( $n=42$ ) who had been transplanted at Hannover Medical School and had the following serostatus combinations: D – R + ( $n=14$ ), D + R + ( $n=22$ ) and D – R – ( $n=6$ ). Three to seven samples were collected from each recipient within defined time periods (7–10, 11–17, 18–24, 25–31, 32–38, 47–53 and 77–83 days post-transplantation) and stored at  $-20\text{ }^{\circ}\text{C}$ . All specimens with detectable HCMV-DNAemia were sequenced. They were designated by the patient number and day of sampling. To simplify the nomenclature, specimens taken during the sampling periods were designated as originating from day 8, 14, 21, 28, 35, 50 or 80, respectively, e.g. 1001–35 signifying plasma obtained in the period around day 35 from patient 1001. Statistical analyses were conducted to investigate correlations between viral sequence features (i.e. presence of multiple strains) and the following clinical features, which were retrieved from the medical records: presence of aGvHD, T cell depletion, subsequent survival and number and duration of HCMV antigenaemia episodes. This study was approved by the institutional review board of Hannover Medical School (no. 2527-2014 and no. 2906-2016).

### Clinical procedures and definitions

Regular (i.e. at least weekly) HCMV screening of recipients after transplantation was based on HCMV pp65 antigenaemia testing [29] according to the standard procedures being applied in the HSCT unit of Hannover Medical School at that time. Plasma samples utilised in the present study were collected independently of samples used for antigenaemia monitoring. HCMV reactivation or infection was defined as  $\geq 5$  pp65 antigen-positive cells per  $4 \times 10^5$  leukocytes or  $\geq 2$  pp65 antigen-positive cells per  $4 \times 10^5$  leukocytes in two or more consecutive samples, and was treated pre-emptively with antiviral drugs at the physicians' discretion. For statistical purposes, the end of an antigenaemia episode was defined as the point at which at least one negative HCMV pp65 antigenaemia result had been obtained; HCMV pp65 antigenaemia values were considered until day 100 for this study. Diagnosis of aGvHD was based on previously established criteria [30]. T cell depletion in vivo was assumed if antithymocyte globulin (ATG) had been part of the conditioning regimen. Overall survival was followed up as long as recipients attended outpatient clinic visits.

## Statistics

Pearson's Chi-square-test or Fisher's exact test was applied to determine the statistical significance of differences between groups (i.e. multiple-strain vs. single-strain infection), using SPSS statistics version 25 (IBM). The Mann–Whitney test was performed to compare median values (e.g. viral loads or antigenaemia levels) among groups, using GraphPad Prism version 7.03 (GraphPad Software). Significance was indicated by *p* values of < 0.05.

## DNA extraction and viral load quantification

DNA extraction and quantitative HCMV real-time PCR were carried out as described previously to identify HCMV-DNA-positive plasma samples and quantify viral loads [31]. Calibration was carried out according to the HCMV WHO standard [32], with a lower limit of linear quantification of 1000 IU/ml. HCMV-positive samples with loads below this limit were assigned the extrapolated viral load value calculated by the PCR cycler (ABI 7500) software. The input number of HCMV genomes used to prepare sequencing libraries was calculated in IU (approximately equivalent to genome copies).

## High-throughput sequencing (HTS)

Library preparation and sequencing were performed as described previously with a few modifications [14, 26, 28, 33]. Briefly, DNA was sheared, sequencing libraries were prepared using a library preparation kit (KAPA Biosystems) and PCR pre-amplification (6–14 cycles) was conducted using adapter-specific primers. Up to 750 ng amplified DNA was target-enriched using custom HCMV-specific RNA baits (SureSelect XT kit, Agilent). HCMV-enriched libraries were indexed, amplified (17–20 cycles), multiplexed and sequenced on a MiSeq (Illumina) using 600v3 reagent kits to generate 2 × 300 nt paired-end reads.

## Assembly and analysis of HCMV genome sequences

De novo assembly of reads and generation of HCMV consensus genomes was performed as described previously [28], utilising a pipeline, including fastp (<https://github.com/OpenGene/fastp>) for adapter trimming, bowtie2 (<http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>) for human read filtering, SPAdes (<https://cab.spbu.ru/software/spades/>) for contig assembly and CLC Genomics Workbench (CLC version 10.0.1) for contig scaffolding to the reference HCMV strain Merlin. The resulting draft genomes were trimmed of the inverted repeats at the HCMV genome termini (TR<sub>L</sub> and TR<sub>S</sub>) and manually annotated from the Merlin strain reference. Comparison and visualisation of complete sequences

were performed using BioEdit (<https://bioedit.software.informer.com/7.2/>) version 7.2.0.

## Genotyping and strain enumeration using motif read matching

To distinguish single-strain from multiple-strain infections and to monitor the composition of HCMV populations in individuals longitudinally, we used the previously established approach of motif read matching [14]. This exploits 12 HCMV genes (RL5A, RL6, RL12, RL13, UL1, UL9, UL11, UL73, UL74, UL120, UL146 and UL139), each of which exhibits a high level of nucleotide diversity and falls into several genotypes that may be monitored by searching for conserved motifs specific to each genotype. The numbers of reads containing these motifs (20–24 nucleotides (nt)) were determined for each of the 12 hypervariable genes by screening the data sets as described previously ([https://centre-for-virus-research.github.io/VATK/HCMV\\_pipeline](https://centre-for-virus-research.github.io/VATK/HCMV_pipeline) [14]). The number of HCMV strains represented in each data set was then calculated as the maximum number of genotypes detected for ≥ 2 genes, with a requirement for ≥ 10 reads to contain the relevant motifs and for these to represent ≥ 2 % of the reads identified for all genotypes of the relevant gene. Genotype abundance was visualised in R (<https://www.r-project.org>) using ggplot2 3.3.0 (<https://ggplot2.tidyverse.org>).

## Variant analyses

Analysis of intrahost HCMV genome variation was performed as described previously [28]. Duplicate reads from identical HCMV-DNA fragments were removed from each data set (deduplicated) using Picard v2.3.0. Briefly, variants (single nucleotide polymorphisms (SNPs)) were called after mapping the deduplicated reads from each sample to the consensus sequence derived from the same sample or, in case of data from longitudinal samples, to the consensus sequence derived from the initial sample. Valid variants were defined by the following criteria in deduplicated data sets: read depth ≥ 25, average base call quality ≥ 20, forward/reverse read balance 0.3–0.5 and variant frequency ≥ 2 %.

## Data deposition

Data sets were purged of human reads and deposited in the European Nucleotide Archive under accession numbers ERX4241108–ERX4241138. HCMV genome sequences from recipients infected by single strains were deposited in GenBank under accession numbers MT649468–MT649474.

## Results

### HTS and data set quality

All plasma samples with detectable HCMV loads ( $n=44$ ), obtained from 23 HSCT recipients, were sequenced (Table S1). The median viral input load was 321 IU per library (range 33–155,000). Data sets meeting the quality criteria for genotyping and strain enumeration were obtained from 24 samples (median input 863 IU per library, range 135–155,000) from 14 HSCT recipients. De novo assembly of consensus sequences was achieved for these data sets, with a median average coverage depth of 836 reads/nt (range 103–5925) for all reads and 58 reads/nt (range 11–4624) after deduplication (Table S1). An additional five data sets (median input viral load 221 IU per library) did not meet the quality criteria or yield consensus sequences by de novo assembly, but contained sufficient data for partial genotyping and were retained for further analysis, bringing the number of samples analysed to 29 and the number of recipients represented to 16 (D – R +,  $n=7$ , and D + R +,  $n=9$ ). The clinical and virological characteristics of the HSCT recipients included in the analysis are shown in Table 1. The remaining 15 data sets were excluded because of failure to produce a sequencing library or very poor data quality. These were obtained from samples with viral loads below the limit of quantification of the HCMV-PCR, with a median input of 131 IU per library (range 33–337).

### Genotyping and strain enumeration

Genotyping was carried out by counting reads containing genotype-specific motifs in 12 hypervariable HCMV genes as described above. In total, 7/16 recipients (D + R + ( $n=5$ ) and D – R + ( $n=2$ )) had multiple-strain infections involving two or three HCMV strains (Table 1). Genotyping data were available for a subset of the 12 genes for the five samples that did not meet the quality criteria: three genes for samples 1011–80 and 1032–50 (samples from other time points from these recipients yielded data for the full range of hypervariable genes), 11 genes for 1048–50 and 1074–35 and seven genes for 1074–50. A very wide range of genotypes was detected in the 12 genes analysed (Table 2). The composition of genotypes was generally stable among longitudinal samples from the same recipient during the observed period, although the relative abundance of genotypes in multiple-strain infection varied (Fig. 1), and in one case (recipient 1012) there was a switch of the dominant population over time. Detailed read numbers for each genotype-specific motif are presented in Table S1.

### Variant analysis

Variant analysis was conducted for data sets having a mean coverage depth of  $\geq 25$  deduplicated reads/nt in order to detect minor populations. This analysis (Figure S1) confirmed the distinction made between recipients with single- or multiple-strain infection by genotyping, and also revealed the population dynamics within recipient 1012 (Fig. 2). For this patient, multiple populations were already detectable at the first time point (day 21), with 3174 intrahost variants present at a median frequency of 32 %. At the last time point (day 50), the median frequency of intrahost variants called against the consensus sequence from the initial time point increased to 89 % (Fig. 2a). DNAemia levels increased with the proportion of the variant population (Fig. 2b).

### Association of HCMV diversity with clinical features

Statistical analyses of the clinical and viral diversity features of the 16 recipients with sufficient HTS data indicated that multiple-strain infection was associated with the absence of DNAemia at day 80 post-transplantation ( $p=0.034$ ) (Table 3). However, multiple-strain infection was not associated with antigenaemia duration or level, peak viral load, serostatus of the donor or recipient, in vivo T cell depletion, development of aGvHD, or fatal outcome (Table 3). Statistical tests for associations between the specific genotype of any of the 12 hypervariable genes with clinical features were not conducted because of the large number of genotypes detected and the relatively small cohort size.

## Discussion

Target enrichment protocols combined with HTS have been utilised successfully to investigate the strain complexity, population dynamics and antiviral resistance of HCMV directly in human subjects [14, 26–28, 34]. These studies have generally involved heterogeneous retrospective cohorts. In contrast, the present study focused on a cohort of HSCT recipients from whom plasma samples were collected prospectively at defined time points during a period of 80 days after transplantation. However, probably due to efficient clinical management and pre-emptive therapy, viral loads were generally low in the recipients identified as HCMV positive. Nevertheless, data sets worthy of analysis could be obtained from 29 samples originating from 16 recipients.

Infection with multiple HCMV strains was detected frequently, in nearly half (44 %) of the recipients ( $n=5$  D + R + and  $n=2$  D – R +). PCR-based genotyping studies have previously implied the presence of multiple strains in solid organ and HSCT recipients and bioinformatic analyses of HTS data have shown previously that the presence

**Table 1** Clinical and virological characteristics of HSCT recipients yielding analysed sequence data

Recipient no.	Age (years)	Sex	Diagnosis	Samples (no.)	HCMV serostatus	Multiple-strain infection (no. of strains)	In vivo T cell depletion	Survival (days after HSCT; cause of death)	aGvHD (grade)	DNAemia at day 80	HCMV reactivation episodes (no.) <sup>b</sup>	Antiviral treatment
1001	48	F	MM	1	D + R +	No	No	Alive (2513)	Yes (II)	No	1	Valganciclovir
1002	19	F	NHL	1	DR +	Yes (2)	Yes	Alive (2484)	No	No	1	Valganciclovir
1011 <sup>a</sup>	24	F	AA	1	D + R +	No	Yes	Alive (2254)	Yes (II)	Yes	0 <sup>c</sup>	None
1012	50	M	MDS	4	D + R +	Yes (2)	Yes	Alive (2397)	Yes (II)	No	1	Cidofovir
1020	51	F	CML	1	D + R +	No	No	Deceased (211; pneumonia)	Yes (II)	Yes	1	Valganciclovir
1028	57	F	AML	2	D – R +	Yes (2)	Yes	Deceased (711; relapse)	No	No	1	Valganciclovir
1032 <sup>a</sup>	54	M	AML	2	D + R +	Yes (3)	Yes	Deceased (134; multiorgan failure)	Yes (II)	No	2	Ganciclovir, foscarnet
1033	43	M	AML	2	D – R +	No	Yes	Alive (2164)	Yes	No	2	Valganciclovir
1036	63	F	MM	2	D – R +	No	Yes	Deceased (254; progress of disease)	No	Yes	2	Ganciclovir, cidofovir
1040	62	M	AML	1	D + R +	Yes (2)	Yes	Alive (1849)	No	No	1	Valganciclovir
1041	27	M	MDS	2	D – R +	No	Yes	Alive (1829)	Yes (II/III)	Yes	1	Valganciclovir
1047	63	M	AML	2	D – R +	No	Yes	Alive (1755)	No	Yes	3	(Val)ganciclovir, foscarnet
1048 <sup>a</sup>	53	M	AML	3	D – R +	No	Yes	Deceased (1445; relapse, pneumonia)	No	No	2	Foscarnet, cidofovir
1065	59	F	CLL	1	D + R +	Yes (2)	No	Alive (1740)	Yes (I-II)	No	1	Valganciclovir
1069	54	M	AML	2	D + R +	Yes (2)	Yes	Alive (1621)	No	No	1	Valganciclovir
1074 <sup>a</sup>	58	M	AML	2	D + R +	No	Yes	Alive (1281)	No	No	1	Valganciclovir

AA aplastic anaemia, aGvHD acute graft-versus-host disease, AML acute myeloid leukaemia, CLL chronic lymphatic leukaemia, CML chronic myeloid leukaemia, D donor, MM multiple myeloma, HSCT haematopoietic stem cell transplantation, MDS myelodysplastic syndrome, NHL non-Hodgkin's lymphoma, R recipient

<sup>a</sup>Sequence data from some samples did not meet the quality criteria but were analysed nonetheless

<sup>b</sup>Based on HCMV antigenaemia data collected until day 100

<sup>c</sup>No detectable HCMV antigenaemia

**Table 2** Genotype analysis and strain enumeration

Data set	Strain number <sup>b</sup>	Genotypes <sup>a</sup>											
		RL5A	RL6	RL12	RL13	UL1	UL9	UL11	UL73	UL74	UL120	UL146	UL139
1001–50	1	6	3	8	8	8	2	1	4D	5	3B	7	6
1002–35	2	1	3, 6	4B, 7	7	4, 7	1	1, 6	3B	2A	1B	7	8
1011–50	1	1	1	4A	4A	4	6	6	4C	1C	2A	12	5
1011–80 <sup>c</sup>	1	–	–	4A	–	–	–	–	–	–	–	12	5
1012–21	2	1	2, 3	4B	2, 4B	4	1, 6	3, 6	2, 3A	1B	1A, 2B	8, 12	5
1012–28	2	1	2, 3	2, 4B	2, 4B	2, 4	1, 6	3, 6	2, 3A	1B, 2B	1A, 2B	8, 12	5
1012–35	2	1	2, 3	2, 4B	2, 4B	2, 4	1, 6	3, 6	2, 3A	1B, 2B	1A, 2B	8, 12	5
1012–50	2	1	2, 3	2, 4B	2, 4B	2, 4	1, 6	3, 6	2, 3A	1B, 2B	1A, 2B	8, 12	5
1020–80	1	1	1	8	8	8	2	1	4A	3	1A	7	2
1028–35	2	1, 2	1, 4	6, 7	6	6, 7	1, 8	7	3A, 4A, 4D	1B, 3	2B	3	4, 7
1028–50	2	1, 2	1, 4, 6	6	6	6	1, 8	7	4A	3	2B	3	4
1032–28	3	1	1, 3	4B, 6, 8	4B, 8	4, 8	1, 9	6	1, 4A, 4D	1A, 3, 5	1A, 4B	2, 12, 13	1A, 3
1032–35	3	1	1, 3	4B, 8	4B, 8	4, 6, 8	1, 9	6, 7	1, 3B, 4A, 4D	1A, 3, 5	1A, 4B	2, 12, 13	1A, 3
1032–50 <sup>c</sup>	1	–	–	–	–	4	–	–	1	–	–	–	1A
1033–28	1	3	5	9	–	9	1	3	2	2B	1A	13	8
1033–50	1	3	5	9	–	4, 9	1	3	2	2B	1A	13	8
1036–50	1	1	2	6	6	6	6	7	2	2B	2A	9	4
1036–80	1	1	2	6	6	6	6	7	2	2B	2A	9	4
1040–50	2	1, 2	1, 4	1B, 7	1, 7	1, 7	1, 4	1	3B, 4A	2A, 3	4A, 4B	7, 13	2, 8
1041–35	1	1	2	2	2	2	1	3	–	2B	1A	4	8
1041–80	1	1	2	2	2	2	1	3	–	2B	1A	4	8
1047–21	1	1	3	–	7	7	1	1	3A	1B	2B	1	7
1047–80	1	1	3	–	7	7	1	1	3A	1B	2B	1	7
1048–50 <sup>c</sup>	1	2	2	1A	1	1	4	1	3B, 4B	–	1A	13	7
1065–50	2	1, 2	3, 4	1A, 4B	1, 4B	1, 4	4, 7	1, 3	4A	1A, 3	1A, 4A	7, 12	2, 5
1069–35	2	5	1, 5	1A, 6	1, 6	1, 6	6	1, 7	1, 2	1A	1A, 3B	12	5, 6
1069–50	2	3, 5	1	1A, 6	1, 6	1, 6	4, 6	1, 7	1, 2	1A, 2B	1A, 3B	8, 12	5, 6
1074–35 <sup>c</sup>	1	–	3	–	6	6	7	3	4A	3	2A	1	5
1074–50 <sup>c</sup>	1	–	–	–	6	6	7	3	4A	–	2A	1	–

– Indicates insufficient reads meeting the threshold for genotyping that gene

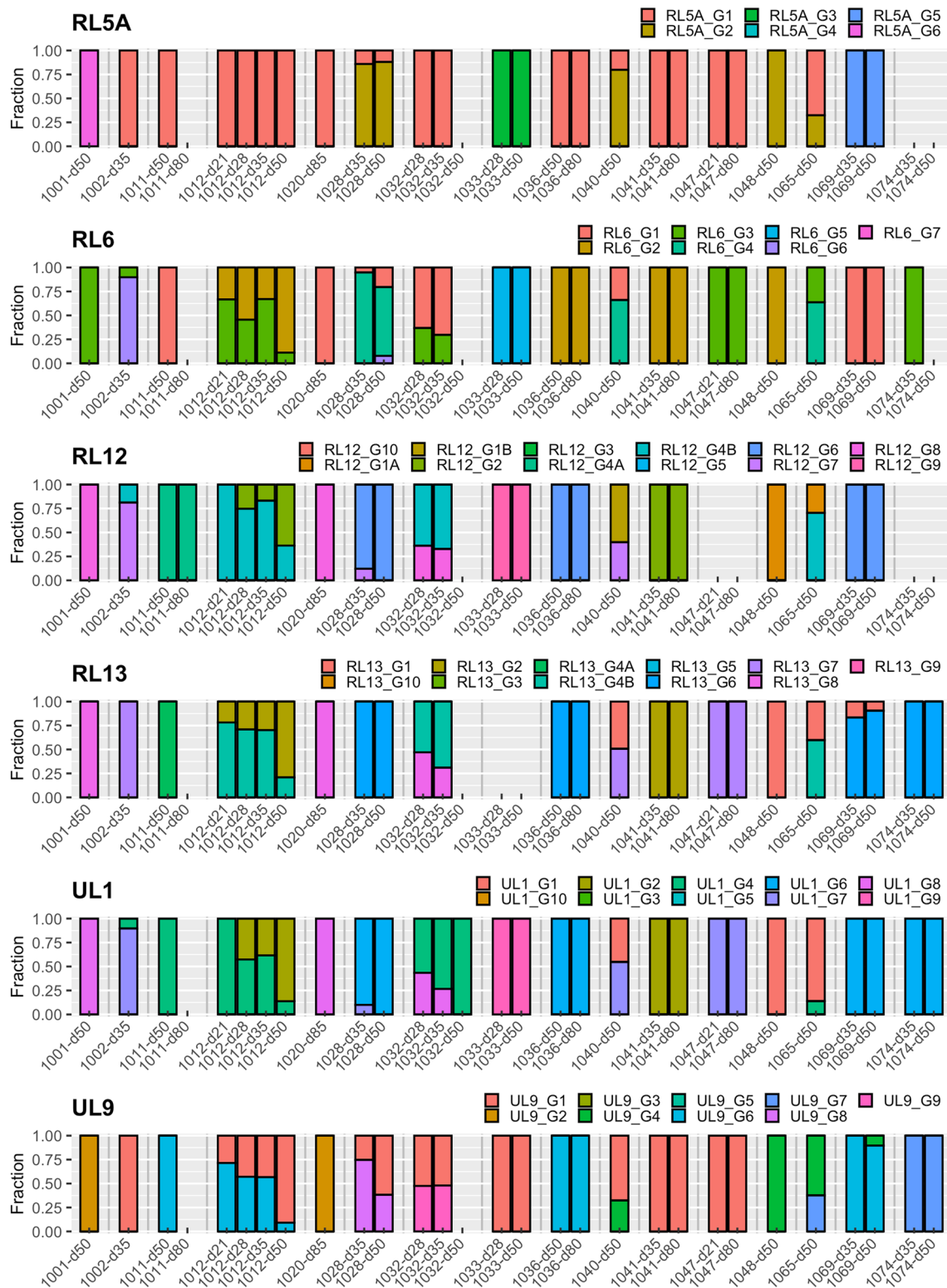
<sup>a</sup>Genotypes assigned to the 29 data sets were determined using motifs for 12 hypervariable HCMV genes

<sup>b</sup>The number of strains detected in a data set was calculated by the maximum number of genotypes detected for at least two genes according to the definitions described in Materials and Methods. Discrepant genotype numbers among different time points in certain genes of a single patient (e.g. UL1 or UL139 in patient 1032) do not necessarily mean that these genotypes emerged de novo or disappeared, but rather were already present in both samples, although in some cases below the cut-off level used to define multiple-strain reactivation. Detailed read numbers for genotype motifs are presented in Table S1

<sup>c</sup>Samples that did not meet the quality criteria for assembly, but yielded genotype data. Interpretations with regard to multiple-strain infections are limited due to relatively low sequencing coverage and should be viewed in the context of other data sets from the same individual

of multiple strains is the major source of HCMV intrahost diversity [14, 17, 19, 35–37]. A study using HTS to genotype genes UL73 and UL74 (which are adjacent and therefore linked genotypically) reported a similar rate of multiple genotypes in 18/29 (62 %) of successfully genotyped specimens from HSCT recipients with active episodes of HCMV infection [38]. The fact that multiple-strain infection may originate from reactivation of multiple strains in the latent repository of a seropositive recipient or from the graft from a seropositive donor may explain the high frequency of such

infection in D + R + recipients. For the two D – R + patients, we speculate that multiple-strain infections may have derived from the latent HCMV repository. As our recent study [37] on HCMV populations in breast milk revealed that approximately half ( $n = 8/15$ ) of immunocompetent HCMV-seropositive women reactivated multiple HCMV strains upon lactation, the presence of multiple, latent HCMV strains in adults and thus also in HSCT recipients may be common. Other rare transmission sources for de novo HCMV re-infection, such as blood transfusions, HCMV-shedding



**Fig. 1** Genotype profile of hypervariable genes. Genotype profile and relative abundance of HCMV genes RL5A–UL139 genotypes in the data sets obtained from 29 plasma samples from 16 HSCT recipients. Each colour represents a different genotype of a hypervariable

gene. Each bar represents a different sample, grouped by subject. The Y-axis indicates the fraction of genotypes in the total sequence reads for that gene in the sample. Underlying data are shown in Table S1. The genotype scoring criteria are described in Materials and Methods

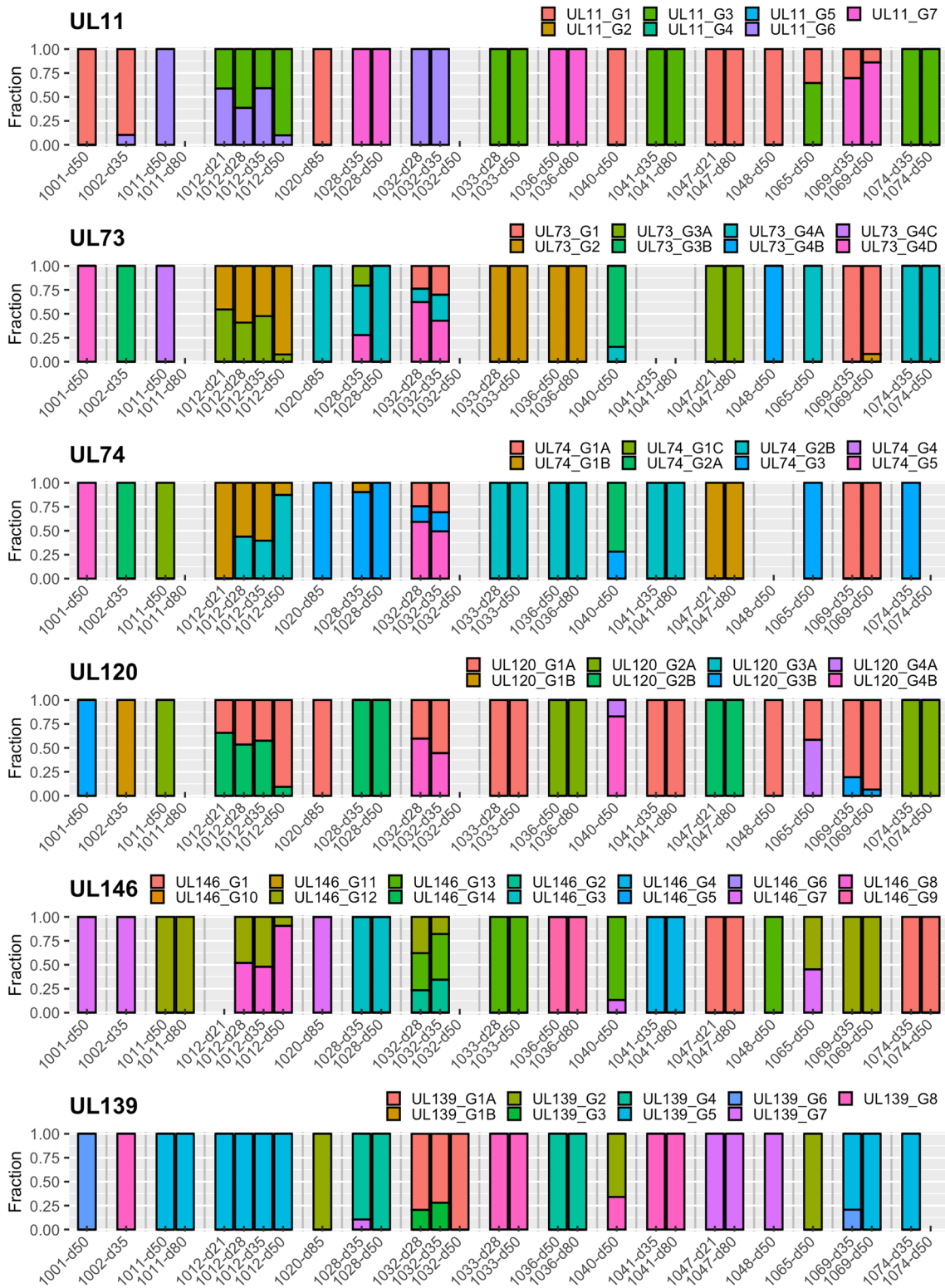


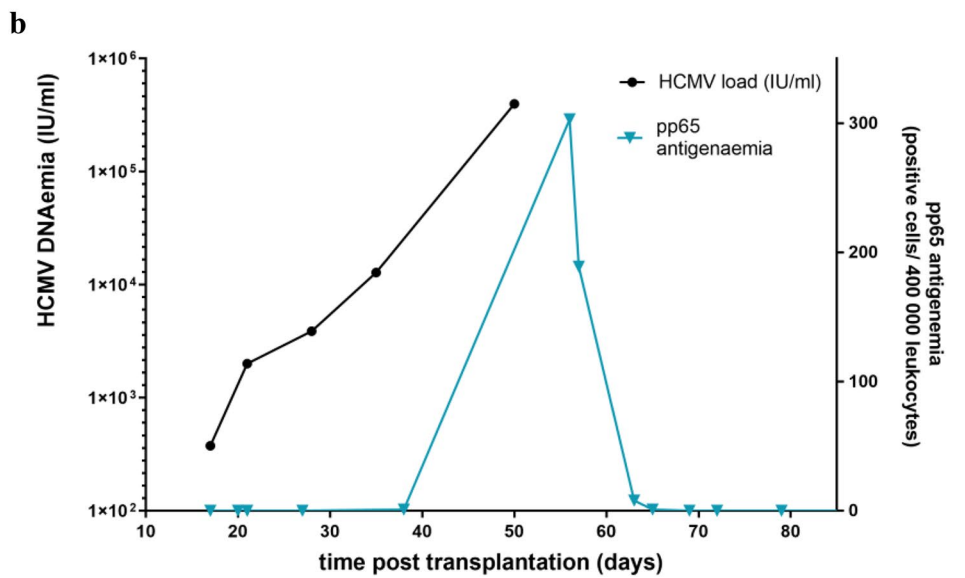
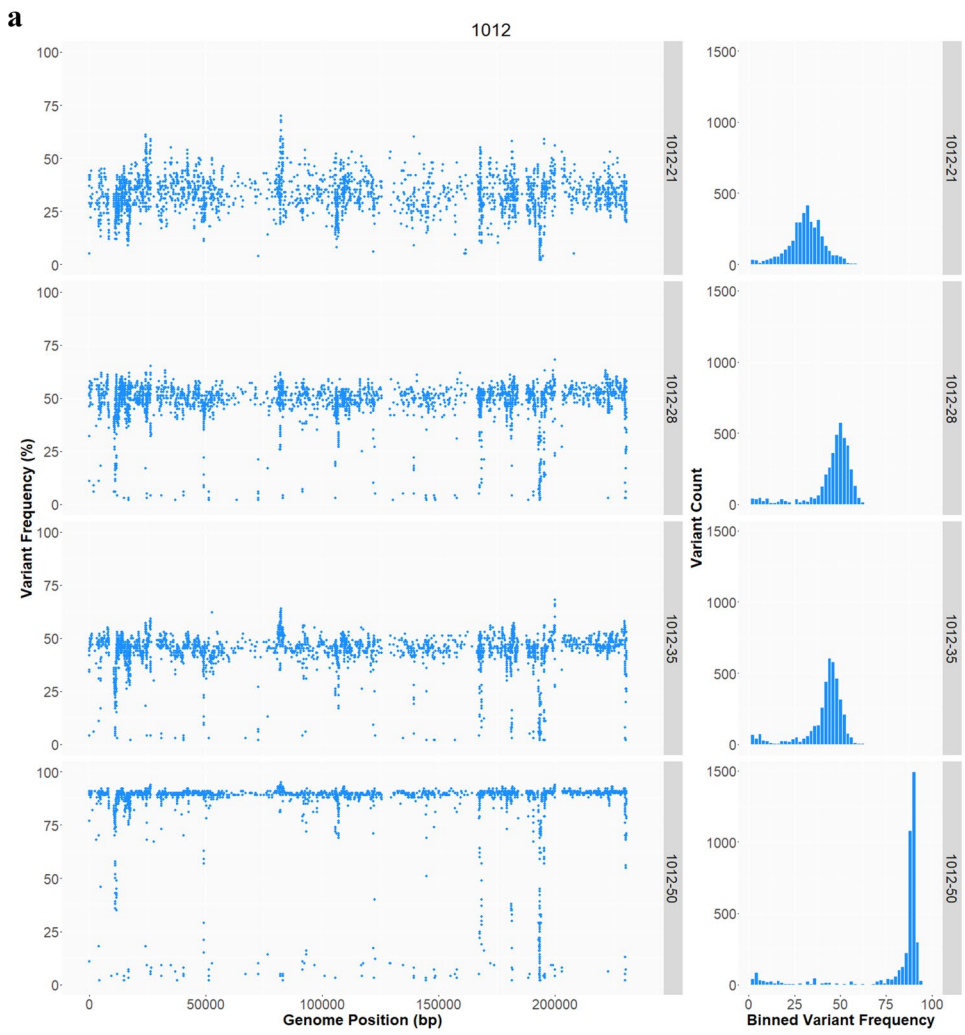
Fig. 1 (continued)

household contacts (e.g. infants) or nosocomial exposure, cannot be ruled out, but seem rather unlikely due the use

of leucocyte-depleted, HCMV-negative blood products and strict hygienic precautions.



**Fig. 2** Longitudinal HCMV sequence diversity in recipient 1012. **a** Variant analysis was conducted in relation to the consensus sequence from day 21. The frequency of variants (Y-axis) was plotted against the position on the genome (X-axis). The histogram on the left shows the variants binned according to their frequency percentage (X-axis) plotted against the absolute number of variants (Y-axis). **b** HCMV DNA load (IU/ml, left Y-axis) and pp65 antigenaemia (positive cells/400,000 leukocytes, right Y-axis) in plasma were plotted against time after transplantation (days, X-axis). Samples from four time points (days 21, 28, 35 and 50 post-transplantation) were sequenced



**Table 3** Comparison of clinical and virological parameters between recipients infected with single or multiple HCMV strains

	Recipients with single-strain infection ( <i>n</i> = 9)	Recipients with multiple-strain infection ( <i>n</i> = 7)	<i>p</i> value
Sex male, <i>n</i> (%)	5 (56)	4 (57)	1
Mean age (range), years	51 (24–63)	51 (19–62)	0.633
Donor/recipient HCMV serostatus, <i>n</i> (%)			0.358
D + R +	4 (44)	5 (71)	
D – R +	5 (56)	2 (29)	
In vivo T cell depletion, <i>n</i> (%)	7 (78)	6 (85)	1
aGvHD, <i>n</i> (%)	5 (56)	3 (42)	1
Median time to onset of aGvHD (range), days after HSCT	35 (10–66)	36 (28–47)	0.786
Fatal outcome, <i>n</i> (%)	3 (33)	2 (29)	1
Median time to fatal outcome (range), days after HSCT	254 (211–1445)	423 (134–711)	0.8
More than one reactivation episode, <i>n</i> (%)	4 (44)	1 (14)	0.308
DNAemia detected at day 80, <i>n</i> (%)	5 (56)	0 (0)	<b>0.034</b>
Median HCMV load (range), IU/ml plasma <sup>a</sup>	1680 (303–5790)	2010 (348–398,000)	0.374
Median viral load peak value (range), IU/ml	2550 (463–5790)	1630 (692–398,000)	0.918
Median duration of individual antigenaemia episodes (range), days after HSCT	15 (0–38)	11 (7–19)	0.202
Median duration of overall HCMV-antigenaemia (range), days	25 (0–49)	14 (9–22)	0.142
Median antigenaemia peak value (range), pp65 positive cells/400 000 leukocytes	33 (0–400)	88 (13–303)	0.837

Bold value indicates statistically significant *p* values of < 0.05

*D* donor, *aGvHD* acute graft-versus-host disease, *HSCT* haematopoietic stem cell transplantation, *R* recipient

<sup>a</sup>For all plasma samples with sequence data of the respective group

In our study, genotyping of hypervariable genes detected a remarkable variety of genotypes that remained stable over the observed time period in longitudinal samples from the same individual, and yet no two recipients had identical combinations of genotypes, consistent with the view that interstrain recombination has generated enormous numbers of strains during HCMV evolution. Similarly, extensive HCMV strain variability among different individuals with high intrahost stability has also been reported for breast milk samples from European and African women in studies that also used genotyping of hypervariable genes to distinguish single-strain from multiple-strain infections [26, 37]. However, despite genotypic stability, the relative abundance of genotypes and the strains they represent may change over time in an individual, as observed in the present study. These findings are in line with our previous study of HCMV genomes in retrospectively collated sample sets of renal transplant and HSCT recipients, and also with other studies focused on a small number of hypervariable genes in lung transplant recipients [28, 39]. The complexity of HCMV populations in multiple-strain infection can fluctuate in a relatively short period of time, as observed in recipient 1012, in whom the dominant population switched within a time period of 29 days. This phenomenon illustrates how fast the strain repertoire can change in infected individuals, and is consistent with the previously reported short doubling time of HCMV [40].

Several studies have reported statistical trends associating the presence of multiple HCMV genotypes with higher peak load and longer duration of DNAemia as well as slower viral clearance after transplantation [22, 36, 38, 41]. Some studies have also reported an association of multiple-strain infection with a worse clinical course, characterised, for example, by an increased number of episodes of HCMV disease, graft loss or opportunistic infection with other pathogens [17, 35]. However, other studies have not found such associations [41–43]. In the present work, multiple-strain infection was not significantly associated with prolonged duration or increased levels of antigenaemia or any other investigated clinical feature during the first 100 days after HSCT. In contrast, all seven recipients with multiple-strain infection had undetectable DNAemia at 80 days post-infection (the last available sampling point) compared with 5/9 recipients with single-strain infection. This statistically slightly significant difference may have been due to the higher occurrence of multiple-strain infection in D + R + recipients, who have an overall decreased risk of severe or recurrent HCMV reactivation [8, 44]. HCMV-specific immune reconstitution occurs faster in recipients with a seropositive donor [9, 45], and recurrent reactivation occurred less often in R + D + recipients than in R + D – recipients in the present study, corroborating data obtained by others.

It is not clear whether and how multiple-strain HCMV infection influences viral load dynamics. The immune system may deal differently with antigenically different strains, leading to altered viral clearance properties in comparison with single-strain infection. In addition, complementation among strains or the contributions of viral factors with differing effects on virulence may enhance viral replication overall [38, 46]. However, even in a prospective study, higher viral loads or prolonged episodes of viraemia may make it easier technically to detect multiple HCMV strains without multiple-strain infection necessarily being associated with an inferior clinical outcome. In the present study, HTS analysis was performed without a preselection bias towards high viral loads, but sequencing failed for many samples with very low viral loads. This outcome led to a limitation of the present study, the relatively small number of samples from which HTS data sets of sufficient quality were generated. Although a substantial number ( $n = 253$ ) of plasma samples were collected, the proportion that contained detectable levels of HCMV DNA was low (17 %), and generally low viral loads reduced the proportion of useable data sets even further (11 %). This challenge is likely to be encountered in future studies of this type, with major limitations set by both the type of clinical specimen and the concentration of HCMV DNA in the specimen. Nevertheless, the relatively low concentration of human DNA in plasma compared to whole blood and the application of an effective target enrichment protocol made it possible to obtain tractable sequence data sets from samples with an input of about  $\geq 400$  IU per library.

In order to investigate the key clinical questions more robustly, including the influence of multiple-strain infection or other viral markers on clinical outcome, it will be necessary to conduct large, prospective studies (e.g. studies in multicentric transplant recipient cohorts, which are currently being established) [47, 48]. In the HSCT setting, donor/recipient HCMV serostatus is an important factor in the risk and course of HCMV reactivation and disease, and this, as well as the incidence of multiple-strain infection, must be considered carefully in future studies. Development of more sensitive HTS protocols capable of sequencing the latent HCMV repository would also support efforts to elucidate HCMV population structure in both donor and recipient. HCMV strain composition differs greatly between individuals, and convincing associations between HCMV strain and clinical outcome have proven elusive in this cohort and in the literature generally. Investigations have also been conducted on the influence of the host side, i.e. of human genetic variants associated with HCMV reactivation and disease after HSCT. A recent study from Casto and colleagues identified a biologically convincing polymorphism that influences intracellular concentrations of immunosuppressive drugs after transplantation and reduces the risk of

HCMV reactivation by 20 % [49]. This study also suggested that other, previously recognised variants mediating HCMV-related phenotypes do not significantly affect the risk of HCMV reactivation or disease after HSCT. These observations indicate that, if clinical associations with virus or host genotype exist, their contributions are unlikely to be strong individually, and that the predictive power of genotyping to identify virus or host biomarkers might be limited in comparison to other factors involved in HCMV pathogenesis.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00430-021-00722-5>.

**Acknowledgements** The authors thank Jenny Witthuhn and Regina Schuster (Institute of Virology, Hannover Medical School) for excellent technical assistance.

**Author contributions** PCK-F, TG, TFS and EMW contributed to the study conception and design. Material preparation and data collection were performed by PCK-F, TG, PRV, EMW and JZ. Experiments and analysis were conducted by SC, AJD, AD, JG, TG, AH, LS and NM. The manuscript draft was written by AD and JG and revised by AJD, TG, TFS and NMS. All the authors commented on previous versions of the manuscript and read and approved the final manuscript. AD and JG contributed equally to this work.

**Funding** Open Access funding enabled and organized by Projekt DEAL. This work was supported by the Deutsche Forschungsgemeinschaft (DFG) Collaborative Research Centre (SFB) 900, core project Z1 [Grant number SFB-900, Z1], the German Center for Infection Research (DZIF) [Grant number TTU IICH 07 801] and the Medical Research Council [grant number MC\_UU\_12014/3]. A. Dhingra and J. Götting were supported by the graduate programme “Infection Biology” of the Hannover Biomedical Research School (HBRS). S. Camiolo was supported by the Wellcome Trust (WT) [Grant number 204870/Z/16/Z].

**Data availability** Data sets were purged of human reads and deposited in the European Nucleotide Archive (accession numbers ERX4241108–ERX4241138). Genome sequences from recipients infected by single strains were deposited in GenBank (MT649468–MT649474).

**Code availability** [https://centre-for-virus-research.github.io/VATK/HCMV\\_pipeline](https://centre-for-virus-research.github.io/VATK/HCMV_pipeline)

## Declarations

**Conflict of interest** The authors declare no conflict of interest or competing interest.

**Ethical approval** Sequencing of blood samples and analysis of clinical data was approved by the institutional review board of Hannover Medical School (no. 2527-2014 and no. 2906-2016).

**Consent to participate** Informed consent was obtained from all participants.

**Consent for publication** Not applicable.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing,

adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Boeckh M, Ljungman P (2009) How we treat cytomegalovirus in hematopoietic cell transplant recipients. *Blood* 113(23):5711–5719. <https://doi.org/10.1182/blood-2008-10-143560>
- Green ML, Leisenring W, Xie H, Mast TC, Cui Y, Sandmaier BM, Sorror ML, Goyal S, Ozkok S, Yi J, Sahoo F, Kimball LE, Jerome KR, Marks MA, Boeckh M (2016) Cytomegalovirus viral load and mortality after haemopoietic stem cell transplantation in the era of pre-emptive therapy: a retrospective cohort study. *Lancet Haematol* 3(3):e119–127. [https://doi.org/10.1016/S2352-3026\(15\)00289-6](https://doi.org/10.1016/S2352-3026(15)00289-6)
- Saullo JL, Li Y, Messina JA, Thompson J, Dalton T, Giri VK, Reed SD, Miller R, Horwitz ME, Alexander BD, Sung AD (2020) Cytomegalovirus in allogeneic hematopoietic transplantation: impact on costs and clinical outcomes using a preemptive strategy. *Biol Blood Marrow Transplant* 26(3):568–580. <https://doi.org/10.1016/j.bbmt.2019.11.005>
- Hiwarkar P, Gaspar HB, Gilmour K, Jagani M, Chiesa R, Bennett-Rees N, Breuer J, Rao K, Cale C, Goulden N, Davies G, Amrolia P, Veys P, Qasim W (2013) Impact of viral reactivations in the era of pre-emptive antiviral drug therapy following allogeneic haematopoietic SCT in paediatric recipients. *Bone Marrow Transplant* 48(6):803–808. <https://doi.org/10.1038/bmt.2012.221>
- Ljungman P, Boeckh M, Hirsch HH, Josephson F, Lundgren J, Nichols G, Pikis A, Razonable RR, Miller V, Griffiths PD, Disease Definitions Working Group of the Cytomegalovirus Drug Development F (2017) Definitions of cytomegalovirus infection and disease in transplant patients for use in clinical trials. *Clin Infect Dis* 64(1):87–91. <https://doi.org/10.1093/cid/ciw668>
- Ljungman P, Hakki M, Boeckh M (2011) Cytomegalovirus in hematopoietic stem cell transplant recipients. *Hematol Oncol Clin North Am* 25(1):151–169. <https://doi.org/10.1016/j.hoc.2010.11.011>
- Ariza-Heredia EJ, Nesher L, Chemaly RF (2014) Cytomegalovirus diseases after hematopoietic stem cell transplantation: a mini-review. *Cancer Lett* 342(1):1–8. <https://doi.org/10.1016/j.canlet.2013.09.004>
- Ljungman P, Brand R, Hoek J, de la Camara R, Cordonnier C, Einsele H, Styczynski J, Ward KN, Cesaro S, Infectious Diseases Working Party of the European Group for B, Marrow T (2014) Donor cytomegalovirus status influences the outcome of allogeneic stem cell transplant: a study by the European group for blood and marrow transplantation. *Clin Infect Dis* 59(4):473–481. <https://doi.org/10.1093/cid/ciu364>
- Lilleri D, Fornara C, Chiesa A, Caldera D, Alessandrino EP, Gerna G (2008) Human cytomegalovirus-specific CD4+ and CD8+ T-cell reconstitution in adult allogeneic hematopoietic stem cell transplant recipients and immune control of viral infection. *Haematologica* 93(2):248–256. <https://doi.org/10.3324/haematol.11912>
- Davison AJ, Holton M, Dolan A, Dargan DJ, Gatherer D, Hayward GS (2013) Comparative genomics of primate cytomegaloviruses. In: Reddehase MJ (ed) *Cytomegaloviruses: from molecular pathogenesis to intervention*, vol 1. Caister Academic Press, Norwich
- Sijmons S, Thys K, Mbong Ngwese M, Van Damme E, Dvorak J, Van Loock M, Li G, Tachezy R, Busson L, Aerssens J, Van Ranst M, Maes P (2015) High-throughput analysis of human cytomegalovirus genome diversity highlights the widespread occurrence of gene-disrupting mutations and pervasive recombination. *J Virol* 89(15):7673–7695. <https://doi.org/10.1128/JVI.00578-15>
- Lassalle F, Depledge DP, Reeves MB, Brown AC, Christiansen MT, Tutill HJ, Williams RJ, Einer-Jensen K, Holdstock J, Atkinson C, Brown JR, van Loenen FB, Clark DA, Griffiths PD, Verjans G, Schutten M, Milne RSB, Balloux F, Breuer J (2016) Islands of linkage in an ocean of pervasive recombination reveals two-speed evolution of human cytomegalovirus genomes. *Virus Evol.* <https://doi.org/10.1093/ve/vew017>
- Dolan A, Cunningham C, Hector RD, Hassan-Walker AF, Lee L, Addison C, Dargan DJ, McGeoch DJ, Gatherer D, Emery VC, Griffiths PD, Sinzger C, McSharry BP, Wilkinson GWG, Davison AJ (2004) Genetic content of wild-type human cytomegalovirus. *J Gen Virol* 85(Pt 5):1301–1312. <https://doi.org/10.1099/vir.0.79888-0>
- Suarez NM, Wilkie GS, Hage E, Camiolo S, Holton M, Hughes J, Maabar M, Vattipally SB, Dhingra A, Gompels UA, Wilkinson GWG, Baldanti F, Furione M, Lilleri D, Arossa A, Ganzemueller T, Gerna G, Hubacek P, Schulz TF, Wolf D, Zavattoni M, Davison AJ (2019) Human cytomegalovirus genomes sequenced directly from clinical material: variation, multiple-strain infection, recombination, and gene loss. *J Infect Dis* 220(5):781–791. <https://doi.org/10.1093/infdis/jiz208>
- Marti-Carreras J, Maes P (2019) Human cytomegalovirus genomics and transcriptomics through the lens of next-generation sequencing: revision and future challenges. *Virus Genes* 55(2):138–164. <https://doi.org/10.1007/s11262-018-1627-3>
- Rasmussen L, Geissler A, Winters M (2003) Inter- and intragenic variations complicate the molecular epidemiology of human cytomegalovirus. *J Infect Dis* 187(5):809–819. <https://doi.org/10.1086/367900>
- Puchhammer-Stockl E, Gorzer I, Zoufaly A, Jaksch P, Bauer CC, Klepetko W, Popow-Kraupp T (2006) Emergence of multiple cytomegalovirus strains in blood and lung of lung transplant recipients. *Transplantation* 81(2):187–194. <https://doi.org/10.1097/01.tp.0000194858.50812.cb>
- Ljungman P, Griffiths P, Paya C (2002) Definitions of cytomegalovirus infection and disease in transplant recipients. *Clin Infect Dis* 34(8):1094–1097. <https://doi.org/10.1086/339329>
- Cudini J, Roy S, Houldcroft CJ, Bryant JM, Depledge DP, Tutill H, Veys P, Williams R, Worth AJJ, Tamuri AU, Goldstein RA, Breuer J (2019) Human cytomegalovirus haplotype reconstruction reveals high diversity due to superinfection and evidence of within-host recombination. *Proc Natl Acad Sci USA* 116(12):5693–5698. <https://doi.org/10.1073/pnas.1818130116>
- Gorzer I, Kerschner H, Jaksch P, Bauer C, Seebacher G, Klepetko W, Puchhammer-Stockl E (2008) Virus load dynamics of individual CMV-genotypes in lung transplant recipients with mixed-genotype infections. *J Med Virol* 80(8):1405–1414. <https://doi.org/10.1002/jmv.21225>
- Manuel O, Åsberg A, Pang X, Rollag H, Emery VC, Preiksaitis JK, Kumar D, Pescovitz MD, Bignamini AA, Hartmann A, Jardine AG, Humar A (2009) Impact of genetic polymorphisms in cytomegalovirus glycoprotein B on outcomes in solid-organ transplant recipients with cytomegalovirus disease. *Clin Infect Dis* 49(8):1160–1166. <https://doi.org/10.1086/605633>
- Lisboa LF, Tong Y, Kumar D, Pang XL, Asberg A, Hartmann A, Rollag H, Jardine AG, Pescovitz MD, Humar A (2012) Analysis

- and clinical correlation of genetic variation in cytomegalovirus. *Transpl Infect Dis* 14(2):132–140. <https://doi.org/10.1111/j.1399-3062.2011.00685.x>
23. Dieamant DC, Bonon SH, Peres RM, Costa CR, Albuquerque DM, Miranda EC, Aranha FJ, Oliveira-Duarte G, Fernandes VC, De Souza CA, Costa SC, Vigorito AC (2013) Cytomegalovirus (CMV) genotype in allogeneic hematopoietic stem cell transplantation. *BMC Infect Dis* 13:310. <https://doi.org/10.1186/1471-2334-13-310>
  24. Nijman J, Mandemaker FS, Verboon-Macielek MA, Aitken SC, van Loon AM, de Vries LS, Schuurman R (2014) Genotype distribution, viral load and clinical characteristics of infants with postnatal or congenital cytomegalovirus infection. *PLoS ONE* 9(9):e108018. <https://doi.org/10.1371/journal.pone.0108018>
  25. Renzette N, Bhattacharjee B, Jensen JD, Gibson L, Kowalik TF (2011) Extensive genome-wide variability of human cytomegalovirus in congenitally infected infants. *PLoS Pathog* 7(5):e1001344. <https://doi.org/10.1371/journal.ppat.1001344>
  26. Suarez NM, Musonda KG, Escriva E, Njenga M, Agbueze A, Camiolo S, Davison AJ, Gompels UA (2019) Multiple-strain infections of human cytomegalovirus with high genomic diversity are common in breast milk from human immunodeficiency virus-infected women in zambia. *J Infect Dis* 220(5):792–801. <https://doi.org/10.1093/infdis/jiz209>
  27. Depledge DP, Palser AL, Watson SJ, Lai IY, Gray ER, Grant P, Kanda RK, Leproust E, Kellam P, Breuer J (2011) Specific capture and whole-genome sequencing of viruses from clinical samples. *PLoS ONE* 6(11):e27805. <https://doi.org/10.1371/journal.pone.0027805>
  28. Hage E, Wilkie GS, Linnenweber-Held S, Dhingra A, Suarez NM, Schmidt JJ, Kay-Fedorov PC, Mischak-Weissinger E, Heim A, Schwarz A, Schulz TF, Davison AJ, Ganzenmueller T (2017) Characterization of human cytomegalovirus genome diversity in immunocompromised hosts by whole-genome sequencing directly from clinical specimens. *J Infect Dis* 215(11):1673–1683. <https://doi.org/10.1093/infdis/jix157>
  29. Engelmann I, Welte T, Fuhner T, Simon AR, Mattner F, Hoy L, Schulz TF, Gottlieb J (2009) Detection of Epstein-Barr virus DNA in peripheral blood is associated with the development of bronchiolitis obliterans syndrome after lung transplantation. *J Clin Virol* 45(1):47–53. <https://doi.org/10.1016/j.jcv.2009.02.005>
  30. Przepiora D, Weisdorf D, Martin P, Klingemann HG, Beatty P, Hovs J, Thomas ED (1995) 1994 consensus conference on acute GVHD grading. *Bone Marrow Transplant* 15(6):825–828
  31. Henke-Gendo C, Ganzenmueller T, Kluba J, Harste G, Raggub L, Heim A (2012) Improved quantitative PCR protocols for adenovirus and CMV with an internal inhibition control system and automated nucleic acid isolation. *J Med Virol* 84(6):890–896. <https://doi.org/10.1002/jmv.23285>
  32. Fryer JF, Heath AB, Minor PD, Collaborative Study G (2016) A collaborative study to establish the 1st WHO International Standard for human cytomegalovirus for nucleic acid amplification technology. *Biologicals* 44(4):242–251. <https://doi.org/10.1016/j.biologicals.2016.04.005>
  33. Wilkie GS, Davison AJ, Kerr K, Stidworthy MF, Redrobe S, Steinbach F, Dastjerdi A, Denk D (2014) First fatality associated with elephant endotheliotropic herpesvirus 5 in an Asian elephant: pathological findings and complete viral genome sequence. *Sci Rep* 4:6299. <https://doi.org/10.1038/srep06299>
  34. Houldcroft CJ, Bryant JM, Depledge DP, Margetts BK, Simmonds J, Nicolaou S, Tutill HJ, Williams R, Worth AJ, Marks SD, Veys P, Whittaker E, Breuer J (2016) Detection of low frequency multidrug resistance and novel putative maribavir resistance in immunocompromised pediatric patients with cytomegalovirus. *Front Microbiol* 7:1317. <https://doi.org/10.3389/fmicb.2016.01317>
  35. Coaquette A, Bourgeois A, Dirand C, Varin A, Chen W, Herbein G (2004) Mixed cytomegalovirus glycoprotein B genotypes in immunocompromised patients. *Clin Infect Dis* 39(2):155–161. <https://doi.org/10.1086/421496>
  36. Ciotti M, Cella E, Ritta M, Ciccozzi M, Cavallo R, Perno CF, Costa C (2017) Cytomegalovirus glycoprotein B genotype distribution in italian transplant patients. *Intervirology* 60(4):165–170. <https://doi.org/10.1159/000486593>
  37. Götting J, Lazar K, Suarez NM, Steinbruck L, Rabe T, Goelz R, Schulz TF, Davison AJ, Hamprecht K, Ganzenmueller T (2021) Human cytomegalovirus genome diversity in longitudinally collected breast milk samples. *Front Cell Infect Microbiol* 11(274):664247. <https://doi.org/10.3389/fcimb.2021.664247>
  38. Vinuesa V, Bracho MA, Albert E, Solano C, Torres-Puente M, Gimenez E, Gonzalez-Candela F, Navarro D (2017) The impact of virus population diversity on the dynamics of cytomegalovirus DNAemia in allogeneic stem cell transplant recipients. *J Gen Virol* 98(10):2530–2542. <https://doi.org/10.1099/jgv.0.000916>
  39. Gorzer I, Guelly C, Trajanoski S, Puchhammer-Stockl E (2010) Deep sequencing reveals highly complex dynamics of human cytomegalovirus genotypes in transplant patients over time. *J Virol* 84(14):7195–7203. <https://doi.org/10.1128/JVI.00475-10>
  40. Emery VC, Cope AV, Bowen EF, Gor D, Griffiths PD (1999) The dynamics of human cytomegalovirus replication in vivo. *J Exp Med* 190(2):177–182. <https://doi.org/10.1084/jem.190.2.177>
  41. Manuel O, Asberg A, Pang X, Rollag H, Emery VC, Preiksaitis JK, Kumar D, Pescovitz MD, Bignamini AA, Hartmann A, Jardine AG, Humar A (2009) Impact of genetic polymorphisms in cytomegalovirus glycoprotein B on outcomes in solid-organ transplant recipients with cytomegalovirus disease. *Clin Infect Dis* 49(8):1160–1166. <https://doi.org/10.1086/605633>
  42. Rycel M, Wujcicka W, Zawilinska B, Paradowska E, Suski P, Gaj Z, Wilczynski J, Lesnikowski Z, Nowakowska D (2015) Mixed infections with distinct cytomegalovirus glycoprotein B genotypes in Polish pregnant women, fetuses, and newborns. *Eur J Clin Microbiol Infect Dis* 34(3):585–591. <https://doi.org/10.1007/s10096-014-2266-9>
  43. Pati SK, Pinninti S, Novak Z, Chowdhury N, Patro RK, Fowler K, Ross S, Boppana S, Investigators NCS (2013) Genotypic diversity and mixed infection in newborn disease and hearing loss in congenital cytomegalovirus infection. *Pediatr Infect Dis J* 32(10):1050–1054. <https://doi.org/10.1097/INF.0b013e31829bb0b9>
  44. Serrais S, Dumontier N, Biard L, Schnepf N, Resche-Rigon M, Pefault de Latour R, Scieux C, Robin M, Meunier M, Xhaard A, Sicre de Fontbrune F, Le Goff J, Socie G, Simon F, Mazon MC (2016) Response to antiviral therapy in haematopoietic stem cell transplant recipients with cytomegalovirus (CMV) reactivation according to the donor CMV serological status. *Clin Microbiol Infect* 22(3):289 e281–287. <https://doi.org/10.1016/j.cmi.2015.11.006>
  45. Camargo JF, Komanduri KV (2017) Emerging concepts in cytomegalovirus infection following hematopoietic stem cell transplantation. *Hematol Oncol Stem Cell Ther* 10(4):233–238. <https://doi.org/10.1016/j.hemonc.2017.05.001>
  46. Cicin-Sain L, Podlech J, Messerle M, Reddehase MJ, Koszinowski UH (2005) Frequent coinfection of cells explains functional in vivo complementation between cytomegalovirus variants in the multiply infected host. *J Virol* 79(15):9492–9502. <https://doi.org/10.1128/JVI.79.15.9492-9502.2005>
  47. Karch A, Schindler D, Kuhn-Stein A, Blaser R, Kuhn KA, Sandmann L, Sommerer C, Guba M, Heemann U, Strohacker J, Glockner S, Mikolajczyk R, Busch DH, Schulz TF, for the Transplant Cohort of the German Center for Infection Research C (2021) The transplant cohort of the German center for infection research (DZIF Tx-Cohort): study design and baseline

- characteristics. *Eur J Epidemiol* 36(2):233–241. <https://doi.org/10.1007/s10654-020-00715-3>
48. Koller MT, van Delden C, Muller NJ, Baumann P, Lovis C, Marti HP, Fehr T, Binet I, De Geest S, Bucher HC, Meylan P, Pascual M, Steiger J (2013) Design and methodology of the Swiss Transplant Cohort Study (STCS): a comprehensive prospective nationwide long-term follow-up cohort. *Eur J Epidemiol* 28(4):347–355. <https://doi.org/10.1007/s10654-012-9754-y>
49. Casto A, Seo S, Levine D, Storer BE, Dong X, Hansen JA, Boeckh M, Martin PJ (2021) Genetic variants associated with cytomegalovirus infection after allogeneic hematopoietic cell transplantation. *Blood*. <https://doi.org/10.1182/blood.2021012153>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.