Major article

Running title: HCMV diversity in clinical samples

(40/40 characters)

Title: Characterization of human cytomegalovirus genome diversity in immunocompromised hosts by whole genomic sequencing directly from clinical specimens

(146/160 characters)

Hage E.[1,5]*, Wilkie G. S.[2]*, Linnenweber-Held S.[3], Dhingra A.[1,5], Suárez N. M.[2], Schmidt J.J.[3], Kay-Fedorov P.[1,5], Mischak-Weissinger E.[4,5], Heim A.[1,5], Schwarz A.[3], Schulz T. F.[1,5], Davison A. J.[2], Ganzenmueller T.[1,5]

[1]Hannover Medical School, Institute of Virology, Hannover, Germany

[2]MRC-University of Glasgow Centre for Virus Research, Glasgow, United Kingdom

[3]Hannover Medical School, Department of Nephrology, Hannover, Germany

[4]Hannover Medical School, Department of Haematology, Haemostasis and Oncology, Hannover, Germany

[5]German Centre for Infection Research, Hannover-Braunschweig site, Germany

*authors contributed equally

word count abstract: 188/200

word count text: 3499/3500

Abstract

Background: Advances in next-generation sequencing (NGS) technologies allow comprehensive studies of genetic diversity over the entire genome of human cytomegalovirus (HCMV), a significant pathogen for immunocompromised individuals.

Methods: NGS was performed on target-enriched sequence libraries prepared directly from a variety of clinical specimens (blood, urine, breast-milk, respiratory samples, biopsies and vitreous humor) obtained longitudinally or from different anatomical compartments from 20 HCMV-infected patients (renal transplant recipients, stem cell transplant recipients and congenitally infected children).

Results: *De novo* assembled HCMV genome sequences were obtained for 57/68 sequenced samples. Analysis of longitudinal or compartmental HCMV diversity revealed various patterns: no major differences were detected among longitudinal, intra-individual blood samples from 9/15 patients and in most of the patients with compartmental samples, whereas a switch of the major HCMV population was observed in six individuals with sequential blood samples and upon compartmental analysis of one patient with HCMV retinitis. Variant analysis revealed additional aspects of minor virus population dynamics and antiviral resistance mutations.

Conclusions: In immunosuppressed patients, HCMV can remain relatively stable or undergo drastic genomic changes that are suggestive of the emergence of minor resident strains or *de novo* infection.

Key words: Next-generation sequencing, immunocompromised, human cytomegalovirus (HCMV), genome diversity, blood, strain switch, evolution

Footnote page

Conflict of interest statement

The authors declare that they do not have a commercial or other association that might pose a conflict of interest.

Funding statement

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], the Volkswagen Foundation, [Communities allied in Infection (COALITION)], and Medical Research Council grant MC_UU_12014/3. E. H. and A. Dhingra were supported by the graduate program "Infection Biology" of the Hannover Biomedical Research School (HBRS).

Meetings

Some of the data have been presented at the IMMEM XI (ESCMID) in March 2016 in Estoril, Portugal (abstract no. PO 67), at the Virus Genomics and Evolution conference in Hinxton 2016 (abstract P-16) and at the German Society of Virology (GfV) Meetings in March 2015, Bochum, Germany (abstract no. OP-69) and March 2016, Münster, Germany (abstract no. P-354), respectively.

Corresponding author contact information

Dr. Tina Ganzenmueller, Institute of Virology, Hannover Medical School, Carl-Neuberg-Straße 1, 30625 Hannover, Germany

Phone: ++49(0)511-532 4328; Fax: ++49(0)511-532 8736

E-mail: Ganzenmueller.Tina@mh-hannover.de

Current affiliations

Current affiliation of Silvia Linnenweber-Held: Hospital Hygiene, Governmental Institute of Public Health of Lower Saxony, Roesebeckstr. 4-6, 30449 Hannover, Germany

Introduction

Human cytomegalovirus (HCMV) contributes significantly to morbidity and mortality in immunocompromised patients, especially after transplantation, as well as in congenitally infected children. Primary HCMV infection or reactivation of latent HCMV can result in acute illness with end-organ manifestations such as retinitis, hepatitis, pneumonitis or enterocolitis [1]. The frequency and severity of HCMV disease in transplant recipients is influenced by risk factors such as donor and recipient HCMV serostatus or antiviral prophylaxis [2, 3].

HCMV (species *Human betaherpesvirus 5*) has a double-stranded DNA genome of approximately 236 kbp containing >170 open reading frames (ORFs) encoding functional proteins [4]. Most of the genome exhibits little variability among strains, although several hypervariable genes exist; in addition, recombination contributes to HCMV variability [5-12]. The picture of HCMV diversity within an individual host is complicated by the fact that prior infections with multiple strains, *de novo* infections and reactivation of latent virus can occur. The role of genome variation in HCMV pathogenesis is not well understood [13-15], although numerous studies have tried to characterize HCMV diversity and its clinical impact by PCR-based genotyping of selected ORFs [16-18].

The high resolution provided by next-generation sequencing (NGS) has made it possible to study diversity across the entire HCMV genome [4, 8, 12, 19]. Several groups have applied NGS approaches to sequencing whole HCMV genomes, albeit mostly employing cell-culture passaged isolates or amplicon sequencing [6, 10, 12, 20-22]. Here, we employed a target enrichment approach to sequence directly from clinical samples, since this offers the advantage that the sequences obtained are more likely to represent accurately the original viral population than those obtained by using previous approaches [23]. We present an analysis based on sequence data obtained from sequential time points or different anatomical compartments from 20 HCMV-infected, immunocompromised patients.

Methods

Patients and specimens

A total of 68 archival, HCMV-positive clinical specimens were obtained from 20 patients for which real-time quantitative PCR (qPCR) for HCMV had been requested: whole blood (n=49), plasma (n=5), urine (n=3), breast milk (n=2), vitreous humor (n=1), tissue biopsies (n=4; from gut or liver) or respiratory material (n=4). Samples originated from four children with (congenital) HCMV disease (including breast milk from two of the children's mothers), eight renal transplant recipients (RTRs) and eight haematopoietic stem cell transplant recipients (SCTRs). For 15/20 patients, blood samples obtained longitudinally at various time points were available (2-8 time points per patient; median 120 days (range 17-383) between first and last samples). Clinical data were retrieved from medical records. This descriptive study was approved by the institutional review board of Hannover Medical School (no. 2527-2014). Table 1 shows an overview of patient demographics, clinical data and sample types.

HCMV monitoring by qPCR (or pp65 antigenemia assay) was performed according to the procedures followed in individual transplant units. Antiviral treatment with ganciclovir, foscarnet or cidofovir and, in a few cases, maribavir, leflunomide or donor lymphocyte infusions were given at the physicians' discretion (pre-emptively or triggered by signs or symptoms of HCMV infection). RTRs usually received antiviral prophylaxis depending on the risk constellation.

Nucleic acid extraction and viral load quantitation

Extraction of DNA from 200 µl clinical material and HCMV qPCR (calibrated to the WHO standard; limit of detection 500 IU/ml) were performed as described previously [24]. The

HCMV genome input concentration was calculated as IU (approximately equivalent to genome equivalents) per sequencing library.

Next-generation sequencing

DNA was sheared by sonication, and sequencing libraries were prepared (KAPA library preparation kit, KAPA Biosystems, Wilmington, USA) as described [25], with few modifications. Briefly, fragmented DNA was end-repaired, A-tailed and ligated to NEBnext Illumina adaptors (New England Biolabs, Ipswich, USA). After PCR pre-amplification (6-14 cycles), up to 750 ng of DNA was target-enriched for HCMV fragments (SureSelect XT kit, Agilent, CA, USA). RNA probes for the bait library were designed on the basis of 64 complete HCMV genome sequences available to us in April 2014 (details available from A. Davison). HCMV-enriched libraries were indexed using TruGrade oligonucleotides (Integrated DNA Technologies), amplified (KAPA real-time library amplification kit), multiplexed, and sequenced on a MiSeq (Illumina) using reagent kit v3 to generate 2 x 300 base paired-end reads.

Assembly of HCMV genome sequences

Bowtie2 v2.2.6 [26] was used to map reads to the UCSC hg19 human reference genome and these reads were then discarded. The remaining reads were subjected to quality assessment and adaptor trimming using FastQC and Trim Galore v0.4 (http://www.bioinformatics.babraham.ac.uk/projects), retaining reads of quality score ≥20 and length >130 bp. Contigs were assembled *de novo* by using SPAdes v3.6 [27], and ordered against a version of the HCMV reference strain Merlin genome (GenBank AY446894.2) lacking the long and short terminal repeats (TR_L and TR_S) by using Scaffold_builder [28], abacas [29] or CLC Genomics Workbench v8 (Qiagen, https://www.qiagenbioinformatics.com/). Gaps were closed (GapFiller v1-10 [30]), potential

assembly errors corrected by iterative mapping (ICORN2 [31]), and genome annotations were transferred from HCMV strain Merlin (RATT [32]). Subsequent analyses were performed by using the resulting sample-specific consensus sequences ("major genome type"), lacking TR_L and TR_S, having assessed sequence quality by careful inspection of read alignments. The presence of multiple HCMV strains in the original specimen that likely arose by independent infection events (whether by co-infection, serial infection or reactivation) was assessed initially by mapping reads against the genotype-specific sequences of selected hypervariable genes (UL73, UL74, UL139 and UL146) characterised previously from a wide range of independent patients [5, 8, 33] (data not shown), and subsequently by calculating the level of nucleotide diversity by variant analysis. A control experiment was performed to validate this protocol (Figure S1).

Selected complete HCMV genome sequences from patients infected by a single strain were deposited in GenBank (KY123649-KY123653).

Phylogenetic analysis

A multiple sequence alignment of sample-specific consensus sequences was constructed (MAFFT v7 [34]). A neighbor-joining tree was constructed by using MEGA6 [35], applying the Kimura 2-parameter method with 500 bootstrap replicates and complete deletion of gapped sites. Overall mean distance and pairwise distance comparisons (p-distance) were performed (MEGA6) with one representative sample (n=20) from each patient or the whole dataset (n=57) with complete deletion of gapped sites.

Variant analysis

Duplicate reads presumably resulting from two or more PCR-derived copies of the same HCMV-DNA fragment were removed from each dataset by using Picard-v2.3.0

(https://github.com/broadinstitute/picard). Variants (SNPs and indels) were identified by using the low frequency variant detector function in CLC Genomics Workbench.

The dynamics of variants over time were tracked by aligning sequence reads from different time points from a patient against the consensus sequence from the initial time point and performing variant calling. Variants were considered valid under the following conservative criteria: overall read depth ≥50, average basecall quality ≥20, forward/reverse read balance 0.3-0.5 and variant frequency ≥2% (i.e. the relative frequency of a variant at a particular position). Variants detected within homopolymer regions or internal repeats (IR_L/IR_S) were excluded. Variants were plotted using custom R scripts by position in the HCMV genome and frequency at each time point. Variants present at >50% were taken to indicate a switch in the major HCMV genome type compared to the initial time point. In addition, the frequency distribution of variant alleles was visualised in frequency histograms.

Antiviral resistance mutations were identified by using published definitions [36].

Results

Next-generation sequencing of HCMV directly from clinical specimens

To characterize HCMV diversity among samples collected sequentially or from different anatomical compartments, NGS was performed on DNA extracted from a set of 68 different clinical specimens from 20 immunocompromised patients by using target-enrichment. The median HCMV input per library was 1.8 x 10^4 IU (range 1.0 x 10^3 to 9.5 x 10^5) for whole blood and 1.1 x 10^5 IU (range 5.0 x 10^2 to 3.9 x 10^6) for other specimens. The range of percentage of HCMV reads in the enriched libraries was 0.03-88.8%. Consensus HCMV genomes were assembled from 57 sequenced samples, achieving an average coverage of 99.94% (range 98.67-100%). An average read depth was obtained of 17-5,450 (median 359) for whole blood and 113-5,663 (median 2,932) for other specimens. Assembly of consensus

sequences for nine whole blood samples and one breast milk sample was not feasible due to low coverage, probably because of low HCMV input (range 5.0 x 10^2 to 1.8 x 10^4 IU per library). Assembly of data from one blood sample failed due to infection with two strains in approximately equal proportions.

Interstrain diversity and phylogenetic analysis of major HCMV genome types

The average interstrain nucleotide diversity calculated from the consensus sequence of one sample from each patient was 0.022 (i.e. any two HCMV strains differed by an average of 2.193%). Phylogenetic analysis of all sample-specific consensus sequences obtained longitudinally or from different compartments (Figure 1) sorted the patients into two groups: those in whom one genome type was consistently dominant (Child1, Child2, Child3, Child4, RTR2, RTR4, RTR5, RTR9, RTR10, RTR11, SCTR2, SCTR4 and SCTR12) and those in whom the dominant genome type switched during the course of infecton (RTR6 among two compartments, and RTR1, RTR3, SCTR1, SCTR3, SCTR8 and SCTR11 among sequential samples). In the latter group, nucleotide diversity between the initial and final time points was 0.008-0.026, indicating that in some patients the HCMV population after the switch differed from that before the switch to a similar extent to HCMV strains in two different individuals (Table 1).

Genetic diversity among longitudinal samples

The different clustering patterns of consensus sequences obtained from the same patient prompted an investigation of the level of within-host diversity for samples with sufficient read depth. For each patient, the HCMV population diversity was assessed by mapping the read data from each sample against the consensus sequence from the initial time point. Various patterns of within-host diversity were apparent (Table 1).

A switch in the major genome type over time was observed in 6/15 patients. For example, in SCTR1 at day 224 after transplantation, with the ultimately dominant genome type being detectable at low frequency in the initial sample at day 91 (Figure 2). In this sample, a multiple infection was detected by variant analysis of the initial time point data set against its own consensus sequence: 3982 intra-host variants with a median frequency of 7% (range 2-25%) were detected. In the second blood sample (sampled 35 days later) and in the last blood sample, the median variant frequency increased to 36% (range 2-46%) and 83% (range 5-100%), respectively. Similarly, a complete switch of the major genome type between the first and last samples was observed in SCTR11 within 104 days (Figure S2). In SCTR8, 4284 (median frequency 8%, range 2-29%) and 6661 variants (median frequency 52%, range 2-98%) were present at the initial and last time points, respectively. Interestingly, these variants segregated into two clusters, with a frequency range of 25-62% and of 85-95%, respectively (Figure 3). This indicated the emergence of a new dominant genome type at the second time point (which was already detected at the initial time point at low frequency) accompanied by an additional (minor) population perhaps arising due to reactivation. One or more of these additional populations may have been donor-derived in view of the D+/R+ serostatus constellation.

Changes of the dominant HCMV genome type in patients RTR1, RTR3 and SCTR3 were also observed, with nucleotide diversity values between the initial and final time points of 0.011, 0.019 and 0.008, respectively. However, because of low read depth in some samples, variant analysis could not always be interpreted with confidence. Nevertheless, based on the consensus sequence analysis, RTR1 showed an unusal diversity pattern among samples collected from two major reactivation episodes, with major differences in particular within the RL11 gene family (which contains several hypervariable members) and the US6 gene family, these regions being reminiscent of sequences contributed by independent strains (Figure 4). This suggests that the genome type emerging in the second reactivation episode (day 380 after

transplantation) may have resulted from intra-host recombination or from re-infection with, or reactivation of, a closely related strain.

Finally, switches in the major genome type were not detected in the other nine patients with longitudinal samples (e.g. Figure S3), although some contained minor populations indicating multiple strain infections (Figure S4).

HCMV diversity between anatomical compartments

The major HCMV genome types from different anatomical compartments in the same patient were highly similar or identical to each other in blood and urine (blood and bronchoalveolar lavage in one case). These patients included the congenitally infected children (including a corresponding breast milk sample from one mother) and most of the transplant recipients (Table 1). However, similar to the situation in sequential blood samples, minor populations likely corresponding to multiple strain infections were detected in a few patients.

One impressive exception was patient RTR6, who suffered from full-blown HCMV retinitis occurring 8 years after renal transplantation. The major HCMV genome type from vitreous humor demonstrated major differences in comparison with the corresponding blood compartment. Using the consensus genome from blood as a reference, a total of 1000 variants (frequency 2-100% (median 84%)) were detected in the vitreous body compartment (Figure 5A-B). A total of 731 variants were present at >50% (nonsynonymous=145, synonymous=435 and non-coding=151) and located mostly in the region containing genes UL69-UL99 (Figure 5C). Generally, synonymous substitutions were more common than nonsynonymous ones, although UL73 (glycoprotein N) and UL74 (glycoprotein O) displayed high levels of nonsynonymous substitution.

Antiviral resistance mutations

A total of 8/15 longitudinally sampled patients showed mutations associated with HCMV antiviral resistance in at least one sample, in either or both of genes UL54 or UL97. Child4, RTR4, RTR9 and RTR10 did not exhibit resistance mutations initially, but developed UL97 mutations later (Table 1), whereas RTR3 showed a low frequency resistance mutation in UL97 (16%) at the initial time point, which later became dominant. (68%). SCTR11 and SCTR1 developed low-frequency resistance mutations (<40%) in UL54 in blood after a switch of the dominant HCMV population. Interestingly, a dominant UL54 resistance mutation (Glu756Gln) was detected in a gut biopsy from SCTR1 at day 244 (76% frequency), compared to the corresponding blood sample (39% frequency).

In RTR1 (D+/R-), who underwent antiviral prophylaxis, UL97 resistance mutations were already present initially (day 166), but were lost by day 235 when the patient exhibited UL54 resistance mutations (foscarnet treatment had been started at day 186). Following antiviral treatment (two cycles of cidofovir, followed by ganciclovir/valganciclovir) this patient became negative for HCMV-DNAaemia, and, in the second reactivation episode at day 472 (no antiviral treatment at this time), no signs of UL54 or UL97 resistance were detected, suggesting emergence of another HCMV genome type or re-emergence of the parental genome type from the latent repository as a result of selection against the resistance mutants in the absence of drug selection.

Discussion

Due to the low abundance of HCMV-DNA in many clinical specimens, most NGS-based studies analysing the whole HCMV genome have involved sequencing strains isolated in cell culture or amplicons generated by PCR. Here, we sequenced HCMV directly from a variety of clinical specimens using a commercial target enrichment approach that has been applied previously to several pathogens, including herpesviruses and HCMV [12, 37, 38]. Our

carefully designed bait library, which included multiple viral genotypes, proved key to success, as this enabled us to cope with the full range of genome variation and obtain adequate read depths throughout the genome. Also, adaptations of standard library preparation techniques [25] enabled us to sequence complete HCMV genomes even for specimens with relatively low viral loads and high human genomic backgrounds (e.g. whole blood). Current limitations of the method include relatively high costs and a limit of detection to yield complete HCMV genomes (here approximately 5000 IU/library for whole blood). In addition, low sequencing depth due to low viral input may hamper detailed variant analysis. The latter limitations can prove difficult for investigations using whole blood samples from transplant recipients, as high HCMV-DNAaemia is observed only in severe cases or after development of antiviral resistance.

We obtained whole HCMV genomic data from 57 clinical specimens from 20 patients. Levels of interstrain diversity were similar to those estimates proposed previously [10]. Analysis of longitudinal HCMV diversity in blood samples showed a switch in the dominant HCMV population in 6/15 individuals. It is likely that these patients had multiple strain infections, and in most cases the strain that became dominant by the final time point was already detectable at the initial time point by variant analysis. A switch is probably due to a change in the relative levels of replication of strains in the multiple infection, as suggested by a study based on HCMV genotyping in lung transplant recipients [33]. An alternative explanation would be *de novo* infection.

Patients with a longitudinal strain switch had a slightly longer observation time (Mann-Whitney test, p=0.01) but no higher mortality than patients lacking a switch. It is unclear whether the former difference (likely due to the presence of resistance or prolonged, severe disease courses in high risk patients) might have facilitated the detection of a switch or, alternatively, whether a switch is more likely to lead to prolonged, and potentially more severe, episodes of HCMV-DNAemia. The latter hypothesis is supported by previous studies

on clinical outcome in immunocompromised patients, showing that multiple infections can be associated with higher viral loads, delayed clearance and increased HCMV disease rates, graft rejection and co-infections with other herpesviruses [16, 18, 39]. Multiple strain infections are common after transplantation (15-90%), and often found in the D+/R+ and, consistent with data from our cohort, D+/R- or D-/R+ patients, since both donor and recipient may have accumulated multiple strains prior to transplantation [15].

Longitudinal changes in the HCMV population might contribute to pathogenesis, for example via differential properties conferred by polymorphisms in viral glycoproteins, many of which are hypervariable and are known to contribute to immune evasion via antigenic variation [40]. In 4/6 patients showing a marked change of the HCMV population over time, the ultimately dominant genome type encoded variants of glycoprotein and immunomodulatory genes different from the initially dominant type. In contrast, the diversity observed upon a switch of the dominant population in RTR1 and SCTR3 was mostly observed only in genomic regions associated with immunomodulatory functions, which is more consistent with intra-host recombination events. Detection of these different patterns highlights an advantage of sequencing clinical material directly, rather than cell culture isolates, which, during passaging, tend to accumulate variations and mutations in certain glycoprotein and immune evasion genes [6, 8, 41].

From the analyses of HCMV population diversity among anatomical compartments, general compartmental differences in congenitally infected children and most of the transplant recipients were not observed. Recent studies on HCMV diversity in urine and blood from congenitally infected children have reported that multiple strain infections or compartmentalization between blood and urine are rarely detected in these patients [21, 42]. One exception in our study was patient RTR6, suffering from HCMV retinitis, who showed a remarkable diversity in the region encoding genes UL69-UL99, with glycoproteins gO (UL74) and gN (UL73) displaying the highest levels of non-synonymous substitutions. These

glycoproteins play important roles in tissue tropism and protect the virus from neutralizing antibodies [40, 43]. The presence of HCMV genotypes encoding different gB glycoproteins in vitreous humor and blood has been reported previously for AIDS patients with retinitis [44]. The mechanism of compartmentalization is not well understood, but could be attributed to selection of a particular strain (due to cellular tropism), bottlenecking of the viral population in the eye or *de novo* infection [45].

Antiviral resistance of HCMV is a severe clinical problem, especially for transplant recipients [46, 47]. In our study, 8/20 patients showed signs of UL97 or UL54 resistance mutations. In SCTR1, we observed the presence of a dominant UL54 resistance mutation in a gut biopsy but not in the corresponding blood sample. Such a compartmentalization of drug resistant HCMV has been reported previously as characterizing cerebrospinal fluid and blood in SCTRs [48, 49]. Although this was not the main aim of our study, NGS facilitates data mining in an unbiased fashion and allows the detection of low frequency events that are usually not detectable by Sanger sequencing. These advances could facilitate earlier detection of known resistance markers in patients at risk or novel mutations that are currently not accessible to routine diagnostics [50].

Our data show that the NGS of HCMV directly from a variety of diagnostic specimens is feasible and informative, providing new insights into viral diversity in longitudinally sampled patients. We found evidence of a switch of the major genome type in nearly half of the transplant recipients analysed, suggesting that recurrent HCMV-DNAaemia after transplantation does not necessarily reflect reappearance of the previously replicating strain, but rather the emergence of additional strains, possibly with variant biological properties, that were already present or acquired subsequently. This phenomenon could influence the development and course of HCMV disease, although it did not correlate with a fatal outcome in our relatively small patient cohort. Further investigations of the global variability patterns observed in clinical samples, as well as of the role of intra-host recombination and functions

of variants of key viral genes, are necessary to elucidate the role of HCMV diversity in pathogenesis.

Funding

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], the Volkswagen Foundation, Communities allied in Infection (COALITION), and Medical Research Council grant MC_UU_12014/3. E. H. and A. Dhingra were supported by the graduate program "Infection Biology" of the Hannover Biomedical Research School (HBRS).

Acknowledgements

The authors thank Sabine Huebner and Jenny Witthuhn for excellent technical assistance and Guillaume Beauclair for advice on the figure layout.

References

1. Griffiths P, Baraniak I, Reeves M. The pathogenesis of human cytomegalovirus. J Pathol **2015**; 235:288-97.
2. Razonable RR, Hayden RT. Clinical utility of viral load in management of cytomegalovirus infection after solid organ transplantation. Clin Microbiol Rev **2013**; 26:703-27.
3. Ariza-Heredia EJ, Nesher L, Chemaly RF. Cytomegalovirus diseases after hematopoietic stem cell transplantation: a mini-review. Cancer Lett **2014**; 342:1-8.
4. Davison AJ, Holton, M., Dolan, A., Dargan, D.J., Gatherer, D. & Hayward, G.S. . Comparative genomics of primate cytomegaloviruses. In Cytomegaloviruses: from Molecular Pathogenesis to Intervention. Vol. 1. Norwich, UK: Caister Academic Press, **2013** (Reddehase MJ, ed.
5. Bradley AJ, Kovacs IJ, Gatherer D, et al. Genotypic analysis of two hypervariable human cytomegalovirus genes. J Med Virol **2008**; 80:1615-23.
6. Cunningham C, Gatherer D, Hilfrich B, et al. Sequences of complete human cytomegalovirus genomes from infected cell cultures and clinical specimens. J Gen Virol **2010**; 91:605-15.
7. Davison AJ, Dolan A, Akter P, et al. The human cytomegalovirus genome revisited: comparison with the chimpanzee cytomegalovirus genome. J Gen Virol **2003**; 84:17-28.
8. Dolan A, Cunningham C, Hector RD, et al. Genetic content of wild-type human cytomegalovirus. J Gen Virol **2004**; 85:1301-12.
9. Wilkinson GW, Davison AJ, Tomasec P, et al. Human cytomegalovirus: taking the strain. Med Microbiol Immunol **2015**; 204:273-84.
10. Sijmons S, Thys K, Mbong Ngwese M, et al. High-throughput analysis of human cytomegalovirus genome diversity highlights the widespread occurrence of gene-disrupting mutations and pervasive recombination. J Virol **2015**.

11. Rasmussen L, Geissler A, Winters M. Inter- and intragenic variations complicate the molecular epidemiology of human cytomegalovirus. J Infect Dis **2003**; 187:809-19.
12. Lassalle F, Depledge DP, Reeves MB, et al. Islands of linkage in an ocean of pervasive recombination reveals two-speed evolution of human cytomegalovirus genomes. Virus Evolution **2016**; 2.
13. Ross SA, Arora N, Novak Z, Fowler KB, Britt WJ, Boppana SB. Cytomegalovirus reinfections in healthy seroimmune women. J Infect Dis **2010**; 201:386-9.
14. Boeckh M, Geballe AP. Cytomegalovirus: pathogen, paradigm, and puzzle. J Clin Invest **2011**; 121:1673-80.
15. Puchhammer-Stöckl E, Görzer I. Human cytomegalovirus: an enormous variety of strains and their possible clinical significance in the human host. Future Virology **2011**; 6:259-71.
16. Coaquette A, Bourgeois A, Dirand C, Varin A, Chen W, Herbein G. Mixed cytomegalovirus glycoprotein B genotypes in immunocompromised patients. Clin Infect Dis **2004**; 39:155-61.
17. Gorzer I, Kerschner H, Jaksch P, et al. Virus load dynamics of individual CMV-genotypes in lung transplant recipients with mixed-genotype infections. J Med Virol **2008**; 80:1405-14.
18. Manuel O, Asberg A, Pang X, et al. Impact of genetic polymorphisms in cytomegalovirus glycoprotein B on outcomes in solid-organ transplant recipients with cytomegalovirus disease. Clin Infect Dis **2009**; 49:1160-6.
19. Sijmons S, Van Ranst M, Maes P. Genomic and functional characteristics of human cytomegalovirus revealed by next-generation sequencing. Viruses **2014**; 6:1049-72.
20. Renzette N, Gibson L, Bhattacharjee B, et al. Rapid intrahost evolution of human cytomegalovirus is shaped by demography and positive selection. PLoS Genet **2013**; 9:e1003735.
21. Renzette N, Bhattacharjee B, Jensen JD, Gibson L, Kowalik TF. Extensive genome-wide variability of human cytomegalovirus in congenitally infected infants. PLoS Pathog **2011**; 7:e1001344.
22. Bradley AJ, Lurain NS, Ghazal P, et al. High-throughput sequence analysis of variants of human cytomegalovirus strains Towne and AD169. J Gen Virol **2009**; 90:2375-80.
23. Houldcroft CJ, Beale MA, Breuer J. Clinical and biological insights from viral genome sequencing. Nat Rev Microbiol **2017**.
24. Henke-Gendo C, Ganzenmueller T, Kluba J, Harste G, Raggub L, Heim A. Improved quantitative PCR protocols for adenovirus and CMV with an internal inhibition control system and automated nucleic acid isolation. J Med Virol **2012**; 84:890-6.
25. Wilkie GS, Davison AJ, Kerr K, et al. First fatality associated with elephant endotheliotropic herpesvirus 5 in an Asian elephant: pathological findings and complete viral genome sequence. Sci Rep **2014**; 4:6299.
26. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods **2012**; 9:357-9.
27. Nurk S, Bankevich A, Antipov D, et al. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. J Comput Biol **2013**; 20:714-37.
28. Silva GG, Dutilh BE, Matthews TD, et al. Combining de novo and reference-guided assembly with scaffold_builder. Source Code Biol Med **2013**; 8:23.
29. Assefa S, Keane TM, Otto TD, Newbold C, Berriman M. ABACAS: algorithm-based automatic contiguation of assembled sequences. Bioinformatics **2009**; 25:1968-9.
30. Boetzer M, Pirovano W. Toward almost closed genomes with GapFiller. Genome Biol **2012**; 13:R56.
31. Otto TD, Sanders M, Berriman M, Newbold C. Iterative Correction of Reference Nucleotides (iCORN) using second generation sequencing technology. Bioinformatics **2010**; 26:1704-7.

32. Otto TD, Dillon GP, Degrave WS, Berriman M. RATT: Rapid Annotation Transfer Tool. Nucleic Acids Res **2011**; 39:e57.
33. Gorzer I, Guelly C, Trajanoski S, Puchhammer-Stockl E. Deep sequencing reveals highly complex dynamics of human cytomegalovirus genotypes in transplant patients over time. J Virol **2010**; 84:7195-203.
34. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol **2013**; 30:772-80.
35. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Molecular Biology and Evolution **2013**; 30:2725-9.
36. Lurain NS, Chou S. Antiviral drug resistance of human cytomegalovirus. Clin Microbiol Rev **2010**; 23:689-712.
37. Melnikov A, Galinsky K, Rogov P, et al. Hybrid selection for sequencing pathogen genomes from clinical samples. Genome Biology **2011**; 12:R73.
38. Depledge DP, Palser AL, Watson SJ, et al. Specific Capture and Whole-Genome Sequencing of Viruses from Clinical Samples. PLoS ONE **2011**; 6:e27805.
39. Puchhammer-Stockl E, Gorzer I, Zoufaly A, et al. Emergence of multiple cytomegalovirus strains in blood and lung of lung transplant recipients. Transplantation **2006**; 81:187-94.
40. Gardner TJ, Tortorella D. Virion Glycoprotein-Mediated Immune Evasion by Human Cytomegalovirus: a Sticky Virus Makes a Slick Getaway. Microbiol Mol Biol Rev **2016**; 80:663-77.
41. Dargan DJ, Douglas E, Cunningham C, et al. Sequential mutations associated with adaptation of human cytomegalovirus to growth in cell culture. Journal of General Virology **2010**; 91:1535-46.
42. Gorzer I, Trajanoski S, Popow-Kraupp T, Puchhammer-Stockl E. Analysis of human cytomegalovirus strain populations in urine samples of newborns by ultra deep sequencing. J Clin Virol **2015**; 73:101-4.
43. Kropff B, Burkhardt C, Schott J, et al. Glycoprotein N of human cytomegalovirus protects the virus from neutralizing antibodies. PLoS Pathog **2012**; 8:e1002999.
44. Peek R, Verbraak F, Bruinenberg M, Van der Lelij A, Van den Horn G, Kijlstra A. Cytomegalovirus glycoprotein B genotyping in ocular fluids and blood of AIDS patients with cytomegalovirus retinitis. Invest Ophthalmol Vis Sci **1998**; 39:1183-7.
45. Renzette N, Gibson L, Jensen JD, Kowalik TF. Human cytomegalovirus intrahost evolution-a new avenue for understanding and controlling herpesvirus infections. Curr Opin Virol **2014**; 8:109-15.
46. Eid AJ, Razonable RR. New developments in the management of cytomegalovirus infection after solid organ transplantation. Drugs **2010**; 70:965-81.
47. Ljungman P, Hakki M, Boeckh M. Cytomegalovirus in hematopoietic stem cell transplant recipients. Hematology/oncology clinics of North America **2011**; 25:151-69.
48. Hamprecht K, Eckle T, Prix L, Faul C, Einsele H, Jahn G. Ganciclovir-resistant cytomegalovirus disease after allogeneic stem cell transplantation: pitfalls of phenotypic diagnosis by in vitro selection of an UL97 mutant strain. J Infect Dis **2003**; 187:139-43.
49. Frange P, Boutolleau D, Leruez-Ville M, et al. Temporal and spatial compartmentalization of drug-resistant cytomegalovirus (CMV) in a child with CMV meningoencephalitis: implications for sampling in molecular diagnosis. J Clin Microbiol **2013**; 51:4266-9.
50. Chou S. Approach to drug-resistant cytomegalovirus in transplant recipients. Curr Opin Infect Dis **2015**; 28:293-9.

Figure legends

Figure 1: Phylogenetic analysis of HCMV consensus sequences. Sequences from six longitudinally sampled patients (RTR1, RTR3, SCTR1, SCTR3, SCTR8 and SCTR11) and one compartmentally analysed patient (RTR6) that clustered in different parts of the tree are shown in coloured font. Sequences from other patients that clustered together are shown in black font. RTR (triangle): renal transplant recipient; SCTR (circle): stem cell transplant recipient.

Figure 2: Longitudinal and compartmental HCMV sequence diversity in stem cell transplant recipient SCTR1 (D-/R+). The HCMV consensus sequence from day 91 was used as a reference for variant calling. A: Variants are plotted by position on the HCMV genome and frequency at each time point. **B:** The frequency distribution of variants at each time point is illustrated in a histogram. Variants are binned at 1% intervals along the *x*-axis and the number of variants in each bin is shown on the *y*-axis. **C:** HCMV-DNA load (IU/ml; limit of detection 5 x 10^2 IU/ml) in peripheral blood is plotted against the time after transplantation (days) during the follow-up period. Assembled consensus sequences and identified variants from different sampling time points of blood, nasal swab and intestinal biopsy samples are indicated by blue, lilac (# in panel C) and brown (* in panel C), respectively, in panels A-C.

Figure 3: Longitudinal HCMV sequence diversity in stem cell transplant recipient SCTR8 (D+/R+). The consensus sequence from day 55 was used as a reference for variant calling. **A:** Variant alleles are plotted by position on the HCMV genome and frequency at each time point. **B:** The frequency distribution of variants at each time point is plotted as explained in Figure 2. **C:** HCMV-DNA load (IU/ml) in peripheral blood is plotted against the time after transplantation (days) during the follow-up period. Blood samples for which HCMV consensus sequences were assembled are indicated by blue arrows.

Figure 4: Longitudinal analysis of HCMV sequence diversity in renal transplant recipient RTR1 (D+/R-). A: HCMV-DNA load (IU/ml) in peripheral blood is plotted against the time after transplantation (days) during the follow-up period. Blood samples for which HCMV consensus sequences were assembled are indicated by blue arrows. **B**: Nucleotide diversity between consensus sequences from the initial (day 166) and final (day 549) time points was estimated in DnaSP v5.1 by using a sliding window approach with a 500 bp window size and a 100 bp step size, and plotted according to genome position.

Figure 5: HCMV compartmentalization in renal transplant recipient RTR6 suffering from HCMV retinitis. The consensus sequence from peripheral blood was used as a reference for variant calling. **A:** Variant alleles are plotted by position on the HCMV genome and frequency in each specimen. **B:** The frequency distribution of variants in each specimen is plotted as explained in Figure 2. Assembled consensus sequences and identified variants from peripheral blood and vitreous humor are indicated by blue and orange, respectively. **C:** Synonymous (dS) and non-synonymous (dN) dominant changes present in vitreous humor (>50% frequency) are depicted according to HCMV gene.

Figure legends (for greyscale in print version)

Figure 1: Phylogenetic analysis of HCMV consensus sequences. Sequences from six longitudinally sampled patients (RTR1, RTR3, SCTR1, SCTR3, SCTR8 and SCTR11) and one compartmentally analysed patient (RTR6) that showed high inter-sample sequence diversity are labelled with a triangle (RTR) or circle (SCTR). RTR=renal transplant recipient; SCTR=stem cell transplant recipient.

Figure 2: Longitudinal and compartmental HCMV sequence diversity in stem cell transplant recipient SCTR1 (D-/R+). The HCMV consensus sequence from day 91 was used as a reference for variant calling. **A:** Variants are plotted by position on the HCMV genome and frequency at each time point. **B:** The frequency distribution of variants at each time point is illustrated in a histogram. Variants are binned at 1% intervals along the *x*-axis and the number of variants in each bin is shown on the *y*-axis. **C:** HCMV-DNA load (IU/ml; limit of detection 5 x 10^2 IU/ml) in peripheral blood is plotted against the time after transplantation (days) during the follow-up period. Assembled consensus sequences and identified variants from different sampling time points of blood (days: 91, 126, 194, 224, 231, and 245) are indicated by arrows with solid lines. The nasal swab (day 130) and intestinal biopsy (day 244) samples are indicated by arrows with dashed lines and labelled with # and *, respectively.

Table 1: Patient demographics, clinical data and overview of sequencing results.

Patient	Gender, age (years)	Underlying disease/HCMV manifestation	D/R HCMV serostatus	Outcome (days after transplantation or birth)	Specimens (no. sequential blood samples)	Time span of samples (days)	Change of major HCMV population detected?	Mean distance between intial and final timepoint or compartments[a]	Multiple HCMV strain infection?	HCMV antiviral resistance mutations detected (% and timepoints)	Treatment
Child-1	M, 0	cCMV disease/microcephaly, dystrophy & hearing loss	NA	Alive (1203)	Blood, urine, mother´s breast milk	0	No	0.0000000000	No	None	GCV, VGCV
Child-2 twin-A	F, 0	cCMV disease/hearing loss	NA	Alive (589)	Blood, urine, mother´s breast milk	1	No	0.0000000000	No	None	GCV, VGCV
Child-3 twin-B	F, 0	cCMV disease/hearing loss	NA	Alive (786)	Blood, urine	1	No	0.0000000000	No	None	GCV, VGCV
Child-4	F, 0	Immunodeficiency/pneumonia	NA	Died (164)	Blood (n=2), BAL	69	No	0.0000044904	No	UL97: Met460Val (35.61% - day 161); Met460Ile (2.73% - day 161); His520Gln (45.46% - day 161); Leu595Ser (2.32% - day 161)	GCV, FCV, leflunomide, CMV-specific IgG
RTR1	F, 68	Renal transplanation/multiple reactivations	D+/R-	Alive (1460)	Blood (n=6)	383	Yes	0.0110013965	Unclear	UL97: Leu595Ser (70% - day 166, 25% - day 235); UL54: Leu545Ser (50% - day 235)	VGCV, GCV, FCV
RTR2	M, 70	Renal transplanation /fever, graft rejection	D+/R-	Alive (1424)	Blood (n=3)	173	No	0.0000000000	No		GCV, VGCV
RTR3	F, 62	Renal transplanation/reactivation under prophylaxis	D+/R-	Alive (1311)	Blood (n=3)	170	Yes	0.0191334492	Yes	UL97: Met460Ile (16% - day 367; 68.06% - day 408)	GCV
RTR4	M, 62	Renal transplanation/enterocolitis, hepatitis	D+/R-	Alive (670)	Blood (n=3)	105	No	0.0000044904	No	UL97: Leu595Phe (63.03% - day 355)	GCV, VGCV
RTR5	M, 77	Renal transplanation/systemic diesease (liver, lung & GI-tract affected)	D+/R-	Alive (497)	Blood, biopsy, BAL	17	No	0.0000000000	No	None	GCV, FCV
RTR6	M, 57	Renal transplanation/HCMV retinitis	D+/R+	Alive (no data)	Vitreous body fluid, blood	1	Yes	0.0032106116	Yes	None	GCV, VGCV
RTR9	F, 52	Renal and pancreas tranplanation/systemic disease, enterocolitis	D+/R-	Alive (1640)	Blood (n=2), biopsy	128	No	0.0000044904	No	UL97: Leu595Ser (100% - day 310)	GCV, VGCV
RTR10	M, 68	Renal transplanation/systemic disease	D+R-	Alive (2193)	Blood (n=2)	120	No	0.0000000000	No	UL97: Leu595Ser (18.10% - day 237); Lys599_Ser604del (60.18% - day 237)	VGCV
SCTR1	F, 52	Allo-HSCT/persistent reactivation with systemic disease	D-/R+	Died (792)	Blood (n=8), nasal swab, biopsy	263	Yes	0.0253750578	Yes	UL54: Thr813Ser (7.46% - day 224, 22.86% - day 245); Ala809Val (4.48% - day 224); Glu756Gln (23.08% - day 224, 29.82% - day 231, 75.94% - day 244[b], 39.29% - day 245); Ala834Pro (6.12% - day 231, 8.57% - day 245)	GCV, FCV, CDV, DLI, maribavir
SCTR2	M, 36	Allo-HSCT/HCMV pneumonia	D-/R+	Alive (491)	Blood (n=2), BAL	17	No	0.0000000000	Yes	None	GCV, FCV, CDV, VGCV
SCTR3	F, 46	Allo-HSCT (twice)/reactivations, hepatitis	D+/R+ & D-/R+	Died (320)	Blood (n=5)	184	Yes	0.0080646972	Yes	None	GCV, FCV
SCTR4	F, 64	Allo-HSCT/subclinical reactivations	D-/R+	Died (254)	Blood (n=2)	64	No	0.0000000000	No	None	GCV, FCV, CDV
SCTR8	F, 36	Allo-HSCT/reactivations	D+/R+	Died (314)	Blood (n=3)	235	Yes	0.0256848931	Yes	None	FCV,CDV,DLI
SCTR9	M, 43	Allo-HSCT (twice)/systemic disease, hepatitis,duodenitis	D+/R-	Died (187)	Blood (n=3), liver biopsy	94	No	0.0000000000	Yes	None	GCV, FCV, CDV
SCTR11	F, 36	Allo-HSCT/enterocolitis	D-/R+	Alive 2153)	Blood (n=3)	104	Yes	0.0252089143	Yes	UL54: Val781Ile (15.05% - day 192)	GCV, FCV, CDV, CMV-specific IgG, leflunomide

SCTR12	F, 0	Congenital immunodeficiency, allo-HSCT/systemic disease with lung involvement	D+/R+	Alive (2132)	Blood (n=2)	84	No	0.0000000000	No	None	GCV, FCV

[a]p-distance of HCMV genome sequences; [b]gut biopsy

RTR=renal transplant recipient; SCTR=stem cell transplant recipient; cCMV=congenital CMV infection; allo-HSCT=allogeneic haematopoietic stem cell transplantation; D=donor; R=recipient;

GCV=ganciclovir; VGCV=valganciclovir; FCV=foscarnet; CDV=cidofovir; DLI= donor lymphocyte infusion

Figure 1.

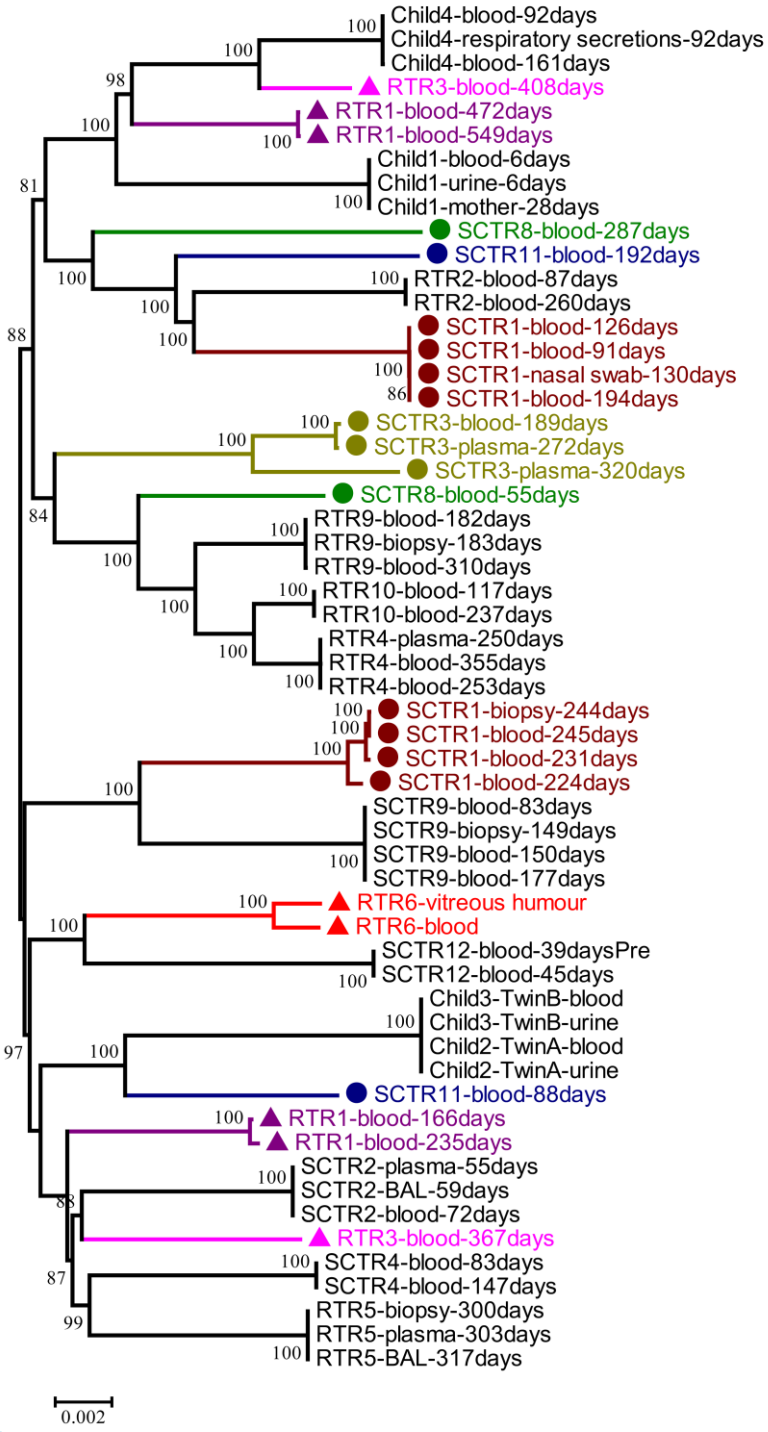
Child4-blood-92days
Child4-respiratory secretions-92days
Child4-blood-161days
RTR3-blood-408days
RTR1-blood-472days
RTR1-blood-549days
Child1-blood-6days
Child1-urine-6days
Child1-mother-28days
SCTR8-blood-287days
SCTR11-blood-192days
RTR2-blood-87days
RTR2-blood-260days
SCTR1-blood-126days
SCTR1-blood-91days
SCTR1-nasal swab-130days
SCTR1-blood-194days
SCTR3-blood-189days
SCTR3-plasma-272days
SCTR3-plasma-320days
SCTR8-blood-55days
RTR9-blood-182days
RTR9-biopsy-183days
RTR9-blood-310days
RTR10-blood-117days
RTR10-blood-237days
RTR4-plasma-250days
RTR4-blood-355days
RTR4-blood-253days
SCTR1-biopsy-244days
SCTR1-blood-245days
SCTR1-blood-231days
SCTR1-blood-224days
SCTR9-blood-83days
SCTR9-biopsy-149days
SCTR9-blood-150days
SCTR9-blood-177days
RTR6-vitreous humour
RTR6-blood
SCTR12-blood-39daysPre
SCTR12-blood-45days
Child3-TwinB-blood
Child3-TwinB-urine
Child2-TwinA-blood
Child2-TwinA-urine
SCTR11-blood-88days
RTR1-blood-166days
RTR1-blood-235days
SCTR2-plasma-55days
SCTR2-BAL-59days
SCTR2-blood-72days
RTR3-blood-367days
SCTR4-blood-83days
SCTR4-blood-147days
RTR5-biopsy-300days
RTR5-plasma-303days
RTR5-BAL-317days
0.002

Figure 2.

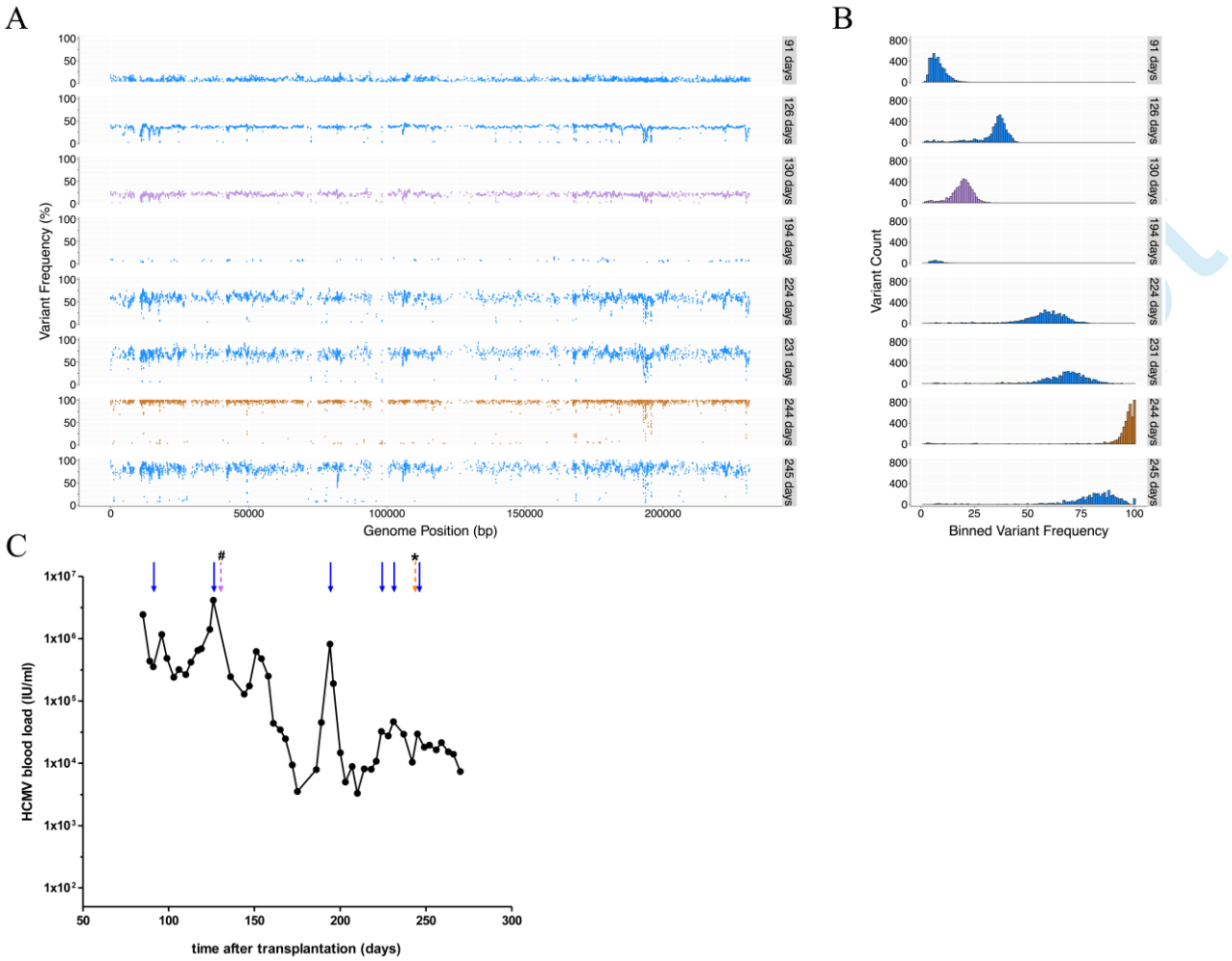

Figure 3.

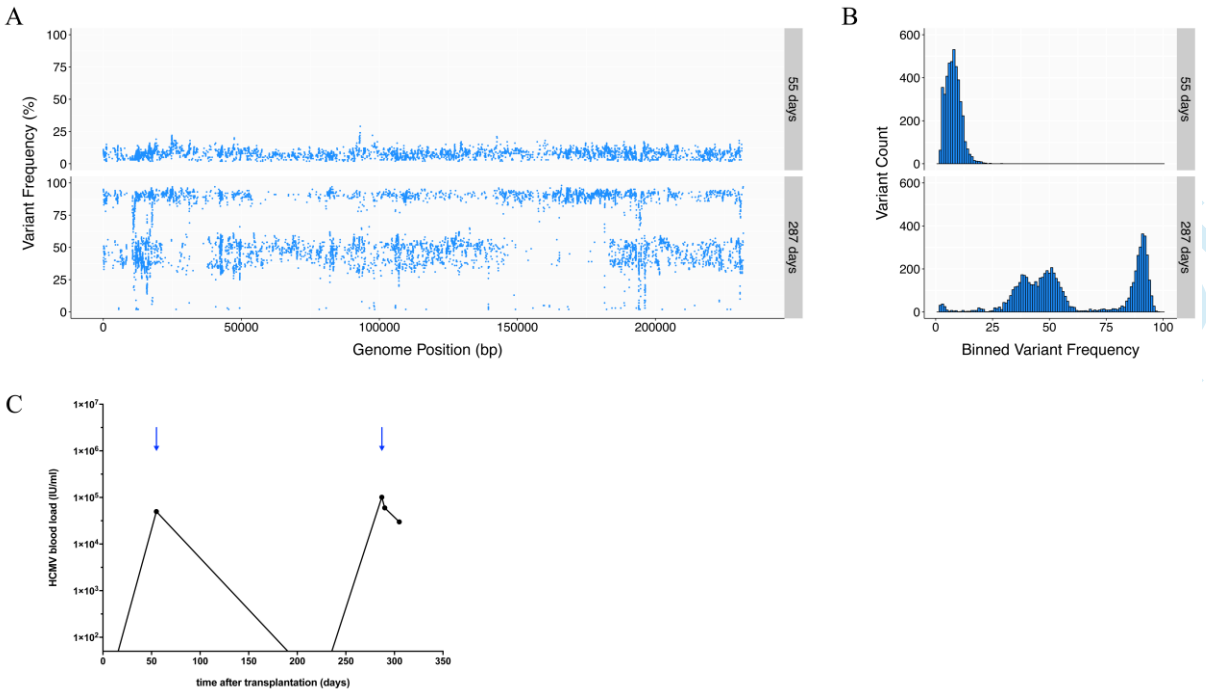

Figure 4.

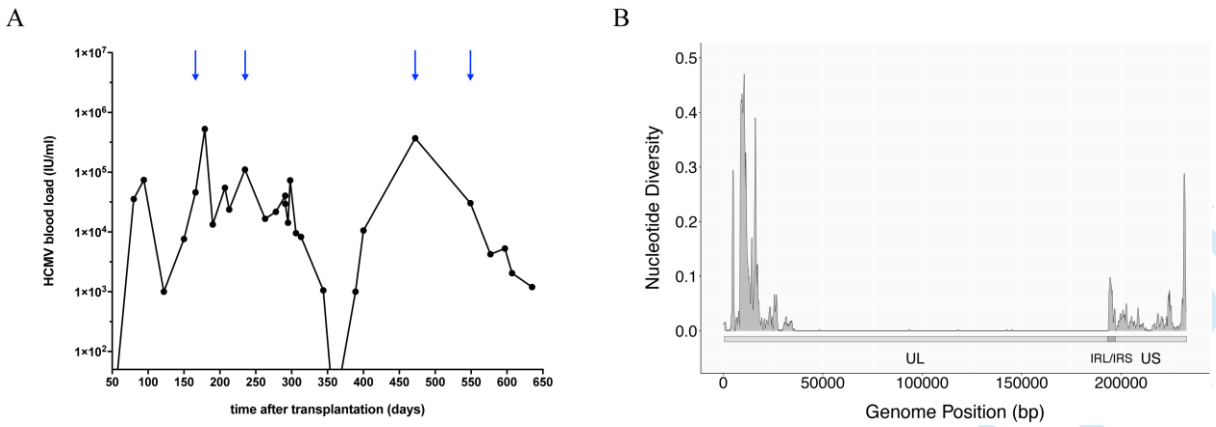

Figure 5.