Major article

Increased human cytomegalovirus multiple-strain infections and genome diversity in breast milk from HIV-positive women

Nicolás M. Suárez[1#], Kunda G. Musonda[2,3#], Eric Escriva[2,4], Margaret Njenga[2*], Anthony Agbueze[2,4], Andrew J. Davison[1] and Ursula A. Gompels[2*]

[1]MRC-University of Glasgow Centre for Virus Research, Glasgow, United Kingdom; [2]Pathogen Molecular Biology Department, London School of Hygiene and Tropical Medicine, London, United Kingdom; [3]Virology Laboratory, University Teaching Hospital, Lusaka, Zambia; [4]Birkbeck College, University of London, London, United Kingdom

[#]These authors contributed equally

[*]Present affiliation: Public Health England, Porton Down, Wiltshire, United Kingdom

Running title: HCMV genome diversity in milk

Word counts: Abstract: 200; Text: 3498

Figures/Tables: 7

References: 50

NOTES

Acknowledgements. We thank the participants and clinical staff for facilitating the BFPH study and also Dr L Kasonka, University Teaching Hospital, Lusaka, Zambia and Professor S Filteau, London School of Hygiene and Tropical Medicine, for enabling the follow up analyses. We also thank Dr S. Camiolo, MRC-University of Glasgow Centre for Virus Research, for assisting with data deposition and Drs Taane Clark and Jody Phelan, LSHTM, for facilitating UNIX cluster access and Perl support.

Financial support. This work was supported by the Commonwealth Scholarship Commission, a Bloomsbury Studentship Award and the Medical Research Council (programme grant MC_UU_12014/3).

Potential conflicts of interest. The authors report no conflicts of interest in this study.

Correspondence. U.A. Gompels, PhD, Pathogen Molecular Biology Department, London School of Hygiene and Tropical Medicine, University of London, Keppel Street, WC1E 7HT, United Kingdom; ursula.gompels@lshtm.ac.uk

ABSTRACT

Background. In developed countries, human cytomegalovirus, HCMV, is a major pathogen in congenitally infected people and immunocompromised individuals, where multiple-strain infections link with disease severity. The situation is less understood in developing countries. In Zambia, breast milk, a key route for transmission, carries higher HCMV loads in HIV-positive than HIV-negative women. We investigated whether strain diversity was also higher.

Methods. Strain diversity in breast milk obtained 4-16 weeks post-partum from 15 HIV-positive and 7 HIV-negative women was analysed by high-throughput sequencing, comparison to 100+ reference genomes and genotyping of hypervariable genes.

Results. Multiple-strain infections were detected in 100% HIV-positive and 43% HIV-negative women, showing raised strain-diversity burden with HIV (p=0.005) and up to 6 strains, present in serial samples. There were 95 genotypes in 12 hypervariable genes, combined in 30 strains; these were conserved within individuals, but gave potential for billions of recombinants. Genetic linkage was maintained strongly for adjacent genes UL73/UL74 (encoding entry/exit glycoproteins gN/gO), and RL12/RL13 /UL1 (encoding immunomodulatory glycoproteins), but not for other nonadjacent genes.

Conclusions. Breast milk is infected with multiple-strains of HCMV in HIV-infected women in Zambia. The complexity provides capacity for generating large numbers of recombinant strains and a major source for transmitting diversity.

Keywords. Human cytomegalovirus; high-throughput sequencing; breast milk; target enrichment; virus genomics.

INTRODUCTION

Human cytomegalovirus (HCMV) is a major coinfection in HIV-positive people, in whom as in other immunocompromised individuals such as transplant recipients, it contributes to morbidity and mortality. HCMV is also the most frequent congenital infection, where it causes hearing loss and adverse neurodevelopment, including microcephaly and neonatal morbidity. Postnatal infection, generally by milk in breastfeeding populations, is usually asymptomatic. However it has been linked to morbidity, especially in preterm or underweight infants, and in recent population studies, to adverse developmental effects, especially with HIV exposure in developing countries [1-4]. The most severe HCMV infections, whether due to primary infection or reactivations from latency, can result in end-organ disease, including retinitis, pneumonitis, hepatitis or enterocolitis [5]. Most studies of the diversity, transmission and epidemiology of HCMV have focused on developed countries. Less is known regarding HCMV in developing countries, including those with a high burden of endemic HIV.

HCMV (species *Human betaherpesvirus 5*) has a double stranded DNA genome of 236kbp containing at least 170 protein-coding genes [6]. Like other human herpesviruses, HCMV exhibits overall low diversity among strains, except for a number of genes with levels of hypervariability but existing as distinct stable genotypes. These genes encode vulnerable to immune selection, including the virus entry complex, other glycoproteins and secreted proteins. The recombinant nature of HCMV strains was first identified in serological surveys and then in genomic studies, and is a key consideration for vaccine development [7-16]. Understanding the roles of HCMV diversity is at an early stage [17-19], and is limited by the fact that most analyses have focused on only few hypervariable genes characterised by

polymerase chain reaction (PCR)-based genotyping [7, 12, 20], which is relatively insensitive to multiple-strain infections resulting from *de novo* infection or reactivation from latency. High-throughput sequencing studies have facilitated broader analyses [10, 11, 14, 16, 21], but most have utilised strains isolated in cell-culture, which are prone to strain loss and mutations in surviving strains or have depended on direct sequencing of PCR amplicons generated from clinical specimens [11, 14, 16, 22]. Recent studies of whole-genome diversity have avoided these limitations by using target enrichment to enable direct sequencing of HCMV strains present in clinical specimens from subjects with various health problems (mainly congenital or transplantation-associated) in developed countries [21, 23]. We have applied this approach to examine HCMV diversity in breast milk, from women in Zambia, an HIV endemic population in Africa in which we have previously shown the negative developmental effects of early infant infection with HCMV, particularly with HIV exposure [1, 3].

METHODS

Patients and Specimens

Anonymised breast milk samples were collected with informed consent as a substudy of the Breast Feeding and Postpartum Health (BFPH) study conducted at the University Teaching Hospital (UTH), Lusaka, Zambia, as approved by the ethical committees of UTH and the London School of Hygiene and Tropical Medicine. This sub-study included 29 HCMV-positive breast milk samples available from 22 women (15 HIV-positive and 7 HIV-negative) at 4-16 weeks post-partum for which HCMV loads were determined [3].

Nucleic Acid Extraction and Viral Load Quantitation

DNA was extracted from 200 µl breast-milk using the QIAamp® DNA Mini kit (QIAGEN, Manchester, UK), as described [3]. HCMV load quantification used a HCMV gB TaqMan assay run on the Applied Biosystems® 7500 Fast Real-Time PCR System (Applied Biosystems, Foster City, CA, USA) as described [3].

Next-Generation Sequencing

The SureSelect^{XT} v. 1.7 target enrichment system was used with biotinylated cRNA probe baits (Agilent, Stockport, UK) to prepare Illumina sequencing libraries, as described previously [21]. The libraries were sequenced using an Illumina MiSeq with v. 3 chemistry (Illumina, San Diego, CA, USA) to generate paired-end raw FASTQ reads of 300 nucleotide (nt). Raw FASTQ NGS reads were quality checked using FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and trimmed to 100 nt minimum

using Trimmomatic [24] or Trim Galore v 0.4.0 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/).

Variant Detection.

Trimmed reads were optimised for de novo assembly using VelvetOptimiser for reference-independent assembly with Velvet [25] and ABACAS [26], and resulting contigs verified by reference mapping using BWA, SAMtools / BCFtools [27], then GATK [28] for alignment, indexing, mapping, and variant calling. Artemis was used for visualisation [29]. Variants from the consensus were defined as >50% prevalence [21, 30]. Mixed infections with HCMV strains were identified first by mapping reads against the genotype-specific sequences of UL73 and UL74 hypervariable genes in eight genotype specific reference genomes (Supplementary Table 1), followed by ten further hypervariable genes (see below) as defined from patient cohorts with samples from blood, urine or saliva [7-10, 12, 17]. Variant nucleotides were defined as follows: prevalence <50%, overall read depth ≥50, average nucleotide quality ≥30, variant frequency ≥1% for read depths >1000 and >10% for read depths >100, and minimum SNP depth ≥10 [21, 30]. A major genotype switch in serial samples was defined by prevalence exceeding 50% [21].

Complete genome sequences for samples with single or dominant genomes were assembled as described [21], deposited in Genbank under accessions MK290742-MK290744).

Strain quantification using motifs

Short motifs.

FASTQ reads were quality checked using FASTQC and trimmed to 100 bases minimum using Trimmomatic [24]. Variants were verified by cut-offs: overall read depth ≥50, average basecall quality ≥30, variant frequency ≥1% for read depths over 1000 and >10% for read

144 depths over 100, giving minimum SNP depth of 10 as shown [21, 30]. To identify major
145 genotype switches, a cut-off of 50% prevalence was used, as previously validated [21].
146 Genotype-specific nucleotide sequences using short motifs were developed guided by amino
147 acid alignments, phylogenetic analyses and polymorphism plots in DnaSP as demonstrated
148 for hypervariable genes UL73 and UL74 (Supplementary Figure 1, Supplementary Tables 1-
149 3) [7, 10, 12, 18]. 'Short'-motifs of nucleotide sequences of 12-14 bases at three positions
150 were identified and validated against available 163 complete genomes, and over 400 single
151 genes (Genbank Release 211 NCBI, 2015) (Supplementary Table 1). Custom Perl scripts
152 were used to interrogate Genbank FASTA reads and Sample FASTQ reads, allowing
153 maximum mismatches of 1-2bp, to quantify mixed infections by enumerating proportions of
154 individual motifs in relation to read depths and plotted with Microsoft Excel 2013
155 (www.cpan.org).

156 *Long motifs.*

157 The proportions of genotypes of UL73 and UL74 with further ten hypervariable genes
158 (RL5A, RL6, RL12, RL13, UL1, UL9, UL11, UL120, UL146 and UL139) were estimated by
159 counting perfect matches to 'long' motifs (24 nt with one exception at 20 nt) specific to each
160 recognised genotype [7-10, 12, 17]. These motifs were developed from DNA sequence
161 alignments (Suarez et al, submitted accompanying). A genotype detected in >10 reads and
162 >2% of the total number of reads detected for all genotypes was scored as being present, and
163 the number of strains in the sample corresponded to the greatest number of genotypes
164 detected for at least two genes. Pie charts were created for all hyper-variable genotypes using
165 a Perl script filtered using sensitivity cut-offs, then FASTQ reads for each genotype
166 calculated as a percentage and plotted.

167

Statistical Analysis

GraphPad Prism (v. 6.05 GraphPad Software, La Jolla, CA, USA) and Microsoft Excel 2013, and were used for data analysis. Fisher's Exact test was used to analyse contingency tables, Student's t test to analyse sample means and Mann-Whitney U test for medians, with significance determined at p-values<0.05.

RESULTS

Genotypic structure of the UL73-UL74 locus

Our previous analyses using Sanger sequencing at multiple time points post-partum pointed to the presence of multiple-strain HCMV infections in breast-milk [3]. We explored this further by using the differences in sequence between the genotypes of hypervariable genes to estimate the number and proportions of strains present in each sample, taking two approaches. The first focused on adjacent genes UL73 and UL74, as our previous studies had shown that these genes are markedly hypervariable, are almost always genetically linked, and exist as eight genotypes each in breast milk samples. Short motifs capable of identifying individual genotypes were developed, based on one motif in UL73 and three motifs in UL74. The second approach extended the analysis to ten additional hypervariable genes, and involved the use of a single long motif for each gene.

To establish a foundation, the amino acid sequences of UL73 and UL74 were extracted from a large set of GenBank sequences and analysed phylogenetically at the amino acid sequence level (Figure 1). Three recombinant sequences were omitted (highlighted in Supplementary Tables 2-4**)**. This analysis confirmed the existence of eight genotypes for each gene, indicated the high levels of inter-genotypic diversity (25-55% amino acid identity) and low levels of intra-genotypic diversity (<3%) [7, 12], and demonstrated the strong genotypic linkage. It also supported the inference of ancestral recombination giving rise to the gN4c and gO1c linkage as shown in the comparison of the mirrored phylogenetic trees (Figure 1).

Strain identification using sequence motifs

Having established a comprehensive view of UL73 and UL74 genotypes, we developed short sequence motifs capable of identifying them individually. A 14 nt motif was identified in a

197 region near the 5'-end of each UL73 genotype, and 12-13 nt motifs in three regions of each
198 UL74 genotype (Supplementary Table 1). These motifs were used to analyse the datasets
199 derived from the 29 samples. The relative frequencies of genotypes in the datasets were
200 similar to those in the GenBank sequences, which lacked contributions from African samples
201 and from breast milk (Figure 1). Use of short motifs also allowed the proportions of
202 individual genotypes in a multiple-strain infection to be estimated, and showed similar results
203 for samples from the right and left breasts (Figure 2).

204 The analysis was then extended to a further ten hypervariable genes in addition to UL73 and
205 UL74, using a single long sequence motif from each gene (Supplementary Table 5). As with
206 the short motifs, this allowed the number of strains in a sample to be estimated, but reduced
207 the possibility that different strains might by chance have the same UL73 and UL74
208 genotypes. Nonetheless, there was general agreement between the two approaches in
209 estimates of the number of strains present (Table 1 and Supplementary Figure 2, comparing
210 short motifs and long motifs methods in inner and outer circles). In the analysis, the major
211 strain was identified as being present in >50% of the reads, a minor variant in 10-50%, and
212 microvariants in 1-10% at detection limits.

213

214 **Motif analyses show diversity raised with HIV**

215 HIV-positive and negative samples were then compared for HCMV loads and strain
216 diversity. This showed HIV-positive mothers had higher median HCMV loads at week 16 as
217 shown previously [3], together with significantly raised levels of mixed-strain infections
218 shown here (Tables 1 and 2), with twice the average number of strains detected in mixed
219 infections. Up to six strains were detected per sample (Tables 1-3). Given the complexity of
220 the mixed infections (Table 3), only single infections or predominant infections could be

compiled using de novo assembly, primarily from HIV-negative mothers demonstrating the first HCMV genomes from normal donor and from Africa. Genome organisation was similar to HCMV from elsewhere, except for diversity in the 12 hypervariable genes.

Analyses of frequencies of genotypes, their linkages and recombination

Previous genomic analyses have shown evidence for recombination through HCMV while most regions are conserved. Genetic linkage has rarely been observed, but evidence has been shown for adjacent genes UL73/gN and UL74/gO [12, 17, 31], as confirmed above. In addition, the UL11 gene family shows some degenerate linkage through conservation of the IgG binding domain in this gene group [9, 10] and confirmed through linkage disequilibrium analyses [11]. We reasoned that any further genetic linkage in this current cohort could be tested by demonstration of similar frequencies of genotypes across the samples. Individual strains were determined from determining the same relative prevalence of the individual genotypes in each donor. Therefore, a haplotype-virotypes composed of the 12-hypervariable genes genotypes could be constructed for each strain. Both major and minor strain variants in individuals were assembled. Micro variants, at the cut-offs between 1-10% were not included because although detected and indicating burden of mixed-strain infection in an individual, they could not be used to construct a haplotype because at this coverage they would not be statistically equally captured at each loci. In Table 4, major and minor variant strains were tabulated and individual frequencies of each genotype for the 12-hypervariable genes indicated with a prevalence heat map. There were 95/110 possible genotypes detected in the 30 strains identified. The results clearly show that UL73/gN and UL74/gO are most closely linked followed by R12-RL13-UL1, but no evidence for any further genetic linkages in this cohort. Each strain identified within and between individuals represented a unique haplotype-

245 virotype. Most of the genotypes were identified in this cohort, and unexpectedly each strain
246 mixed rare and common haplotypes. Calculating with free recombination this potentially
247 generates 2.5 x 10^{11} different 'haplotype-virotype' strains from multiplying the extant
248 numbers of genotypes per hypervariable gene. However, linkage at UL73/gN-UL74/gO and
249 RL12-RL13-UL1 could reduce these by at least 2 logs. For example, in this cohort alone
250 there were 30 out of a potential for 2 billion strains.

251

252 **Conservation of strain haplotypes within individuals**

253 Having established diversity of strains between individuals we examined within individuals
254 and their maintenance. Donors with samples taken at both week 4 and week 16 or sampled
255 from both left and right breasts were examined further using the same heat map for analysing
256 genotype prevalence. In Table 5, this shows despite the diversity of genotypes shown in
257 unique strains between individuals, that within individuals both single strains and mixtures of
258 major and minor variant strains are maintained both between compartments and
259 longitudinally to 16 weeks post-partum. In donor 243 there is complete conservation of the
260 strains between 4 and 16 weeks post-partum and from samples from either breast. In the other
261 donors, there is little evidence for recombination of major variant with minor variants
262 genotypes. While HIV-positive donors 174 at the RL5A loci in strains in the left breast
263 sample and in donor 259 at UL120 in strains from the right breast sample provide some
264 evidence recombination, these are microvariants at detection limits. Further, in donor 278
265 apparent recombination betwen linked RL12-RL13-UL1 loci in the linked '1' and '9'
266 genotypes at RL13 shown at both 4 and 16 weeks post-partum and from both breast samples,
267 is due to a single SNP. Further analyses in comparative pie charts confirm conservation of
268 strains within individuals but diversity between individuals (Supplementary Figure 3). Micro

variants were detected, but at this coverage, could be reactivations or reinfections postpartum, all the major and minor variants genotypes were detected by 4 weeks post-partum with no evidence for new infections by 16 weeks.

DISCUSSION

Genomic analyses of HCMV directly in clinical specimens is required in order to characterise natural populations and avoid mutations prevalent in tissue culture adapted virus. Target enrichment approaches have been successful [21, 23], but accurate genome assembly can be confounded by mixed-strain infections in clinical samples, particularly in immunosuppressed groups where multiple virus reactivations or new infections were demonstrated [17, 21]. Here we examined breast milk samples as a main source for HCMV transmission, comparing HIV immunosuppressed and uninfected, normal mothers. Previously we showed HIV-positive compared to negative mothers had higher HCMV loads in breast milk, which associated with adverse infant development in Sub Saharan Africa [1, 3]. Despite raised virus burden there is only limited genome analyses in milk or Africa. Here we studied their genome diversity, demonstrating in addition to high viral load, raised levels of mixed-strain infections.

We used modular analyses of 12 hypervariable genes and identified over 95/110(86%) possible genotypes in the 29 samples from the 22 donors, of this cohort alone. There were up to six strains identified in an individual using these analyses consisting of distinct genotypes in the 12-hypervariable genes forming virus haplotype-like ‘virotypes’ giving 30 African strains and distinct from 17 European strains we identified in related studies (Suarez et al, submitted). We demonstrated significantly increased mixed-strain infections in Zambian HIV-positive women in addition to high viral load. This was present from the earliest milk sample, at 4 weeks post partum, and similar in milk samples from both left and right breasts. This indicates these variants were present in the donor prior to virus reactivation in the breast tissue during lactation. Leukocyte infiltrates have been characterised at this time [32] and may be the source of the reactivated virus.

297 Although there were high levels of mixed-strain infections, there was little evidence for de
298 novo infection during the time period analysed to 16 weeks post-partum. Major and minor
299 strains, comprising the genotype mixtures at the 12 loci, were conserved within individuals,
300 but highly diverse between individuals. Intra-genotypic recombination was rare, in agreement
301 with previous analyses on low-passage samples, with decreased opportunity for homologous
302 recombination at these highly variable genes [11, 16]. However, there was an ancestral
303 intragenic recombination event modelled previously in the UL74 C-terminal domain, giving
304 rise to UL74 gO1c, which may drive the positive selection observed [12]. The adjacent
305 UL73/gN and UL74/gO show high gene diversity, to 55% [7, 12, 21], and the strongest
306 linkage in this cohort and Genbank (NCBI), followed only by the gene family RL12-RL13-
307 UL1. All other variable genotypes were unlinked. In addition to restraints on homologous
308 recombination due to their diversity, there may be functional restraints against recombinants,
309 as both UL73/gN and UL74/gO have roles in virus exocytosis, cellular tropism and
310 modulating antibody neutralisation and RL12-RL13-UL1 in immune evasion via functional
311 IgG binding domains [9, 13, 33-36]. In addition, mutations in RL13 may affect UL74
312 influence on HCMV growth [37].

313

314 Although variant data could be compiled for all samples, assembly of complete genomes was
315 confounded by the multiple-strain infections. Assembly was possible with single or dominant
316 variants in three donors. To our knowledge this includes the first complete HCMV genome
317 from Africa and a ‘normal’ donor from breast milk, a route of infant virus transmission.
318 There was notable absence of tissue culture adapted, or known antiviral resistant mutations
319 [14, 38]. Previous studies of interstrain diversity did not include UL73 and UL74
320 glycoproteins analyses [11, 16], possibly their diversity initially confounding assembly, and

now corrected on Genbank. Recent studies on transplantation patients used UL73/UL74 for diversity references in deep and whole genome sequencing [17, 21].

The diversity of the strains indicates ancestral modular recombination, but there was little contemporary evidence for this using the current methods and time-span. Further there was no evidence for strain replacement as shown over longer periods in transplantation patients [17, 21]. Micro variant genotypes (<1-10%) present at the latest time point 16 weeks post partum, supports de novo infection, but coverage depth was not sufficient to exclude reactivation from the 4 week sample. The mothers in this cohort all had young children at home who may be sources of HCMV excretion, and all mothers were HCMV positive, therefore opportunities for home and hospital infection. A study in Uganda has shown that mothers can be infected with strains from other children [39].

The varied genotypes in the strains are in virus glycoproteins and immunomodulatory genes, where differences in genotype could provide a growth advantage leading to higher viral loads and pathology. For example, different genotypes of UL74/gO show differences in growth properties [40], and those of UL146/vCXCL1 chemokine have different efficiencies in neutrophil chemotaxis [41]. Similarly variation in the host, for example in immunoglobulin variants GM3/17 affect antibody-dependent cellular cytotoxicity against HCMV via low and higher fc-gamma affinity; of note RL13 binding GM17, would be higher in this cohort and correlates with increased susceptibility [38, 42-45]. Additional variable viral glycoproteins, could participate in antigenic variation. The modular diversity shown here provides a new model for persistent virus infection over previous SNP based antibody-escape mutant analyses for acute infections such as influenza or poliovirus. It is notable that the potential

344 combinatorial diversity between genotypes at the 12 loci, at 10^{12} combinations, if at equal
345 recombinations rates, exceeds some estimates for immune diversity.

346

347 Previous studies on transplantation patients have also shown that multiple infections associate
348 with increased viral loads, as well as pathology such as HCMV disease, graft rejection, other
349 coinfections and may affect delays in virus clearance [18, 20]. In contrast, analyses of HCMV
350 infections in urine from newborns, show primarily single strain genomes, suggesting
351 selection at transmission across the placenta or postnatally by breast-milk or saliva [46]- For
352 example, estimates of saliva transmission indicate only a few virions establishing infection
353 [47]. Infant transmission via breast milk increases with viral load and length of exposure [3,
354 48], and raised mixed-strain infections in HIV-positive women could facilitate their selection.

355

356 We showed high burden of infection via identifying mixed-strains together with raised viral-
357 load and both could modulate HCMV pathology. Previously, we demonstrated that high
358 viral-load and extended breastfeeding linked with increased transmission, and that this
359 associated with poor infant development. Whether these are direct or indirect immune-
360 modulation effects would require further analyses of genome diversity directly from both
361 clinical material as well as normal asymptomatic donors as shown here. Effects on functions
362 of these key hypervariable genes encoding essential glycoproteins and immunomodulatory
363 factors are required both to understand variation and disease, as well as effects on new
364 interventions, such as vaccines.

365

366

367 **Figure Legends.**

368

369 **Figure 1.** HCMV gN and gO phylogenetic analysis show linkage and recombination. This
370 unrooted consensus tree, showing the eight gN and gO genotypes, was inferred from 1000
371 bootstrap replicates to represent the evolutionary history of 163 full-length HCMV gN and
372 gO amino acid sequences from complete HCMV genomes available in GenBank database
373 release 211 (National Center for Biotechnology Information (NCBI), 2015). The evolutionary
374 history was inferred by the Maximum Likelihood method based on the JTT matrix-based
375 model and a discrete Gamma distribution with 5 categories (+G, parameter = 0.8361). All
376 positions with less than 95% site coverage were eliminated, leaving 135 amino acid positions
377 in the final dataset with the branches compressed. All sequences grouped to one of the eight
378 HCMV gN and gO genotypes in the linkage as described [7, 12]. Analyses were conducted in
379 MEGA 6.06 [49]. The distribution was 19(12%) gN1-gO1a, 18(11%) gN2-gO2b, 34(21%)
380 gn3a-gO1b, 14(9%) gN4a-gO3, 37(23%) gN4a-gO3, 11(7%) gN4b-gO4, 10(6%) gN4c-gO1c
381 and 20(12%) gN4d-gO5 out of 163 with the % prevalence indicated. These linked
382 prevalences are not statistically different from those identified in the 30 milk infections
383 shown here or in a follow up survey of 238 complete genomes. The diagonal lines show
384 ancestral recombination event inferred from the phylogeny as described [12] for the origin of
385 selection observed. This shows the recombination required to have mirror phylogenic trees
386 reflecting the linked genotypes.

387

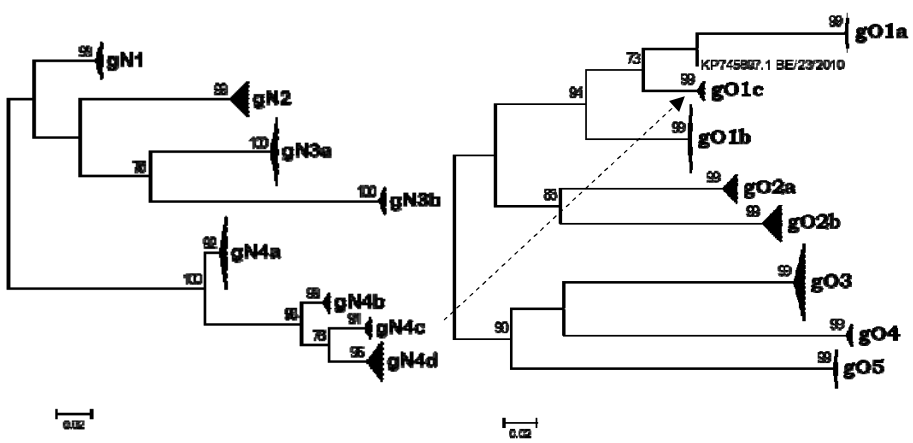
388 **Figure 2.** In vivo duplicates. HCMV gO genotypes in breast milk from four mothers with
389 milk samples collected from both left (L) and right (R) breasts at week 16 postpartum.
390 Sample IDs are indicated in the centre of each doughnut. All four mothers were HIV-positive.

391 Genotypes are colour-coded as follows: gO1a – grey; gO1b – black; gO1c – magenta; gO2a –

392 orange; gO2b – dark red; gO3 – yellow; gO4 – green; gO5 – blue.

393

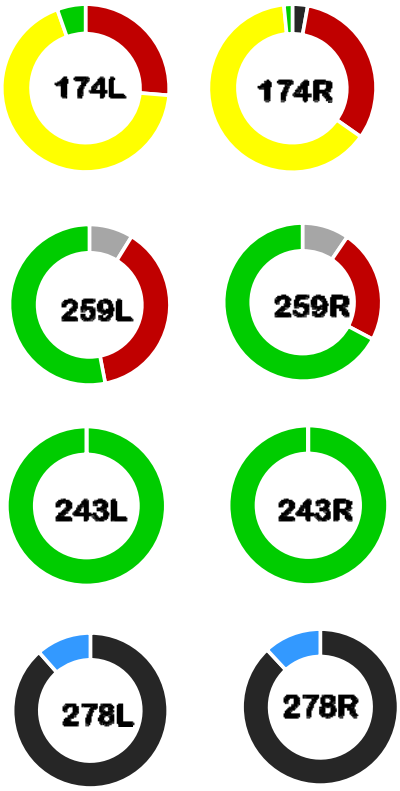
Figure 1



linkage	NCBI	milk
gN1-gO1a	12%	13%
gN2-gO2b	11%	10%
gN3a-gO1b	21%	29%
gN3b-gO2a	9%	3%
gN4a-gO3	23%	16%
gN4b-gO4	7%	13%
gN4c-gO1c	6%	3%
gN4d-gO5	12%	13%

404 Figure 2

405



406

407

ID[a]	HIV Status	Viral load Genomes/ml	Breast sample	W4	W16	Median[b] coverage/nt	Strains Detected[c]
158	Negative	8×10^5	Left		✓	>100	2
166	Negative	3×10^5	Left		✓	50-100	1
193*	Negative	2×10^5	Right		✓	>1000	1
232	Negative	5×10^5	Right	✓		>1000	2
239	Negative	5×10^6	Right	✓		>100	2
263	Negative	5×10^6	Left	✓		>50	1
280	Negative	8×10^6	Right	✓		>1000	1
141	Positive	3×10^5	Right		✓	>1000	3
154	Positive	2×10^6	Left		✓	>1000	3
173	Positive	3×10^5	Right		✓	>1000	6
174	Positive	2×10^6	Left		✓	>100	5
"	"	2×10^6	Right		✓	>100	4
181	Positive	2×10^6	Left	✓		>1000	2
243	Positive	6×10^5	Left		✓	>1000	2
"	"	7×10^7	Right	✓		>1000	2
"	"	8×10^5	Right		✓	>1000	2
248*	Positive	4×10^5	Right	✓		>1000	2
258	Positive	6×10^5	Right	✓		>1000	2
259	Positive	2×10^6	Left		✓	>1000	4
"	"	5×10^6	Right		✓	>1000	3
264	Positive	4×10^5	Left		✓	>1000	2
277	Positive	3×10^5	Right		✓	>1000	3
278	Positive	4×10^6	Left		✓	>1000	2
"	"	3×10^8	Right	✓		>1000	2
"	"	4×10^6	Right		✓	>1000	2
281	Positive	2×10^7	Right	✓		>1000	3
283*	Positive	3×10^5	Right		✓	>1000	2
288	Positive	3×10^7	Right	✓		>1000	4

[a]Median HCMV load at week 16, HIV-negative vs. HIV-positive: 2.5×10^5 vs 1.4×10^6 genomes/ml, z-score=2.02, p=0.04.
[b]Derived consensus sequence; *, complete genome sequence obtained.
[c]Using short motifs for UL73 and UL74 and long motifs for 12 genes.

408
409

Table 2. Higher HCMV mixed infections in HIV+ mothers breast-milk

HIV	Mixed(%)	Single(%)	Total Donors	Avg No. Strains
Negative	3(43)	4(57)	7	1.4
Positive	15(100)	0(0)	15	3
Total	18	4	22	

P=0.005 fishers exact test two tailed

410

411

412

413

Table 3. CMV mixed infections in breast-milk: Genotype analyses identify multiple genome haplotypes defining strains

ID	W	HIV	Strains	RL5A	RL6	RL12-----RL13------UL1			UL9	UL11	UL73----UL74		UL120	UL146	UL139
158	16	-	**2**	4	7	4A	4A/3	4/8	1	1	4A	3	4B	13/11	5
166	16	-	**1**	-	3	9	-	-	9	-	-	1A	4A	-	3
193*	16	-	**1**	1	3	8	8	8	4	1	3A	1B	2B/4B	12	3
232	4	-	**2**	1	1	1A	8	-	6	4	4D	5	4B	14/1	4/7
239	4	-	**2**	6	1	10	-	-	9	-	4B	-	2A	2	3/7
263	16	-	**1**	-	2	10	10	10	8	7	3A	1B	2A	-	4
280	4	-	**1**	1	-	-	6	6	7	-	1	1A	-	-	-
141	16	+	**3**	1/6	3/2	4B/2	2	4/2	2/3	6/5	2:3A /4D	2B:1B	4B/3B	12/2	3/8
154	16	+	**3**	1	-	-	-	-	3	-	-	1B	3A	2/1/8	5/2
173	16	+	**6**	1/2:6	3/2:4	8/6/1B: 3:4B:7	8/6/1/9	8/6 :1 3:2	4:7/ 2:9:6	1/3:5	3A:4A/ 4B:1:2	1B/3:4: 1A/2B: 2A	1A/4B/ 3A:4A	9/3:12 /7	1A/7: 1B/3:6
174R	16	+	**4**	2/ 1	4:2/ 1:3	1B:6 /4A:10	1/6 :10	1/6	4/6/ 7:8	1:4/ 7	4A/2/ 1:3A	3/2B/1 B	2B:3A/ 4B	9/7/ 12	5/7/ 3
174L	16	+	**5**	2	4/2	1B/6	1/6/ 10	1:6 /4	4:6/ 1	1:4/ 7	4A/2/ 1:3A :4B	3/2B/4 :1A	2B/3A 2A	9/7/ 12	5/7/ 2:4
181	16	+	**2**	1	2	4A	6/4A	6	6/9	1	4A	1A/3	2A/4B	8/14	4/7
243R	4	+	**2**	1	6/3	2/6	2/6	2/4	3/7	6/1	4B	4	2A	8	7
243R	16	+	**2**	1	6/3	2	2/1	2/6	3	6/3	4B	4	2A/3A	8	7
243L*	16	+	**2**	1	6	2/10	2/1	2	3	6/1	4B	4	2A/3A	8/9	7
248*	4	+	**2**	1	1	4A	4A/2	4	1	1	3B	2A	4B	-	4/7
258	4	+	**3**	6	2	3	3	3	3	6	3A/4A/ 4B	1B/3	2B/2A	9/2	5
259R	16	+	**3**	1/2	3/4/2	8/1A/6	8:1/6	8/1 /6	3:1/6	1:6/4	4B/2:1	4/2B/1	4B/	9/12/1	8/2
259L	16	+	**4**	1:2	3:4/2	8:1A/6	8:1/6	8/1 /6	3:1/6 /4	1:6/4	4B:2/1	4:2B/1	4B/1A	9:12/1	8:2
264	16	+	**2**	1	2/1	10/8	10/8	10/ 8	8	7/4	4B/3A	4/1B	3A/1A	10	3/5:8
277	16	+	**4**	1	3	4A/6	4A/6/3	4/6 /8	9/6	1/4	4A/1	3/1A/ 1B/5	3A/4A/ 4B	1/9	3/4
278R	4	+	**2**	6/2	3/4	9/1A	1/9	9/1	9/1	6/1	3A/4D	1B/5	4B/3A	8/9	2
278R	16	+	**2**	6/2	3/4	9/1A	1/9	9/1	9/1	6/1	3A/4D	1B/5	4B/3A	8/9	2
278L	16	+	**2**	6/2	3/4	9/1A	1/9	9/1	9/1	6/1	3A/4D	1B/5	4B/3A	8/9	2
281	4	+	**3**	6/1	3	1A/6 :10	1/6 :10	1/6 :10	4/7	1/3:7	4D/3A	5/1B	4B/3B	1	5
283R*	16	+	**2**	-	4	4B	4B	4	2	5	3A/3B	1B/2A	3A	1	5
288	4	+	**4**	2/1	4/3	6/7	6/8	6/: 5/7	4/6/9	1/4	4D/1: 4B	5/1A: 4/3	2B/4B	3/1	4/5/3
Genotypes detected/total:96/110				4/6	6/7	10/12	10/11	10/11	8/9	6/7	7/8	7/8	7/8	11/14	9/9

*Indicates complete genome compiled; R indicates right tissue; L indicates left tissue; ** coverage vs reference Merlin; prevalence indicated by order, first over 50%,: indicates equivalent within 10%, grey indicates at cut-off of 1-10%, - indicates below cut-off; minimum 10 reads validates a SNP with gene coverage of 100-1000 for cut-off of 1-10%

414

Table 4. HCMV haplotypes of hypervariable genes are diverse between individuals and genotype prevalence shows restricted genetic linkage

				Genotypes hypervariable genes											
ID	Mix	Strain	HIV	RL5A	RL6	**RL12**	**RL13**	**UL1**	UL9	UL11	**UL73**	**UL74**	UL120	UL146	UL139
154	major	z-a	-	**4**	**7**	**4A**	**4A**	**4**	**1**	**1**	**4A**	**3**	**4B**	**13**	**5**
166	single	z-b	-	-	**3**	**9**	-	-	**9**	-	-	**1A**	**4A**	-	**3**
193	single	z-c	-	**1**	**3**	**8**	**8**	**8**	**4**	**1**	**3A**	**1B**	**2B**	**12**	**3**
232	major	z-d	-	**1**	**1**	**1A**	**8**	-	**6**	**4**	**4D**	**5**	**4B**	**14**	**7**
239	major	z-e	-	**6**	**1**	**10**	-	-	**9**	-	**4B**	-	**2A**	**2**	**7**
263	single	z-f	-	-	**2**	**10**	**10**	**10**	**8**	**7**	**3A**	**1B**	**2A**	-	**4**
280	single	z-g	-	**1**	-	-	**6**	**6**	**7**	-	**1**	**1A**	-	-	-
141	major	z-h	+	**1**	**3**	**4B**	**2**	**4**	**2**	**6**	**2**	**2B**	**4B**	**12**	**3**
141	minor	z-i	+	**6**	**2**	**2**	**2**	**2**	**3**	**5**	**3A**	**1B**	**3B**	**2**	**8**
173	major	z-j	+	**1**	**3**	**8**	**8**	**8**	**4**	**1**	**3A**	**1B**	**1A**	**9**	**1A**
173	minor	z-k	+	**2**	**2**	**6**	**6**	**6**	**8**	**3**	**4A**	**3**	**4B**	**3**	**7**
174	major	z-l	+	**2**	**4**	**1B**	**1**	**1**	**4**	**1**	**4A**	**3**	**2B**	**9**	**5**
174	minor	z-m	+	**1**	**2**	**6**	**6**	**6**	**6**	**4**	**2**	**2B**	**3A**	**7**	**7**
181	major	z-n	+	**1**	**2**	**4A**	**6**	**6**	**6**	**1**	**4A**	**3**	**2A**	**8**	**4**
243	major	z-o	+	**1**	**6**	**2**	**2**	**2**	**3**	**6**	**4B**	**4**	**2A**	**8**	**7**
248	major	z-p	+	**1**	**1**	**4A**	**4A**	**4**	**1**	**1**	**3B**	**2A**	**4B**	-	**4**
258	major	z-q	+	**6**	**2**	**3**	**3**	**3**	**3**	**6**	**3A**	**1B**	**2B**	**9**	**2**
259	major	z-r	+	**1**	**3**	**8**	**8**	**8**	**3**	**1**	**4B**	**4**	**4B**	**9**	**8**
259	minor	z-s	+	**2**	**4**	**1A**	**1**	**1**	**1**	**6**	**2**	**2B**	**1A**	**12**	**2**
264	major	z-t	+	**1**	**2**	**10**	**10**	**10**	**8**	**7**	**4B**	**4**	**3A**	**10**	**3**
264	minor	z-u	+	**1**	**1**	**8**	**8**	**8**	**8**	**4**	**3A**	**1B**	**1A**	**10**	**5**
277	major	z-v	+	**1**	**3**	**4A**	**4A**	**4**	**9**	**1**	**4A**	**3**	**3A**	**1**	**3**
277	minor	z-w	+	**1**	**3**	**6**	**6**	**6**	**6**	**4**	**1**	**1A**	**4A**	**9**	**4**
278	major	z-x	+	**6**	**3**	**9**	**1**	**9**	**9**	**6**	**3A**	**1B**	**4B**	**8**	**2**
278	minor	z-y	+	**2**	**4**	**1A**	**9**	**1**	**1**	**1**	**4D**	**5**	**3A**	**9**	**2**
281	major	z-z	+	**6**	**3**	**1A**	**1**	**1**	**4**	**1**	**4D**	**5**	**4B**	**1**	**5**
281	minor	z-a1	+	**1**	**3**	**6**	**6**	**6**	**7**	**3**	**3A**	**1B**	**3B**	**1**	**5**
283	major	z-b1	+	-	**4**	**4B**	**4B**	**4**	**2**	**5**	**3A**	**1B**	**3A**	**1**	**5**
288	major	z-c1	+	**2**	**4**	**6**	**6**	**6**	**4**	**1**	**4D**	**5**	**2B**	**3**	**4**
288	minor	z-d1	+	**1**	**3**	**7**	**8**	**8**	**6**	**4**	**1**	**1A**	**4B**	**1**	**5**
Number of total genotypes				4/6	6/7	10/12	10/11	8/11	8/9	6/7	7/8	7/8	7/8	10/14	7/9
Prevalence key			>35%		>30%	>25%		>20%		>15%		>10%		>5%	>1%

415

416

Table 5. Conservation of HCMV haplotypes mixtures in HIV+ individuals and cases of modular recombination in minor variants

						Genotypes hypervariable genes											
ID	Cite	W	Burden	Strain	HIV	RL5A	RL6	RL12	**RL13**	**UL1**	UL9	UL11	**UL73**	**UL74**	UL120	UL146	UL139
174	R	16	major	z-i	+	**2**	**4**	**1B**	**1**	**1**	**4**	**1**	**4A**	**3**	**2B**	**9**	**5**
174	L	16	major	z-i	+	**2**	**4**	**1B**	**1**	**1**	**4**	**1**	**4A**	**3**	**2B**	**9**	**5**
174	R	16	minor	z-j	+	**1**	**2**	**6**	**6**	**6**	**6**	**4**	**2**	**2B**	**3A**	**7**	**7**
174	L	16	minor	z-jA	+	**2**	**2**	**6**	**6**	**6**	**6**	**4**	**2**	**2B**	**3A**	**7**	**7**
243	R	4	major	z-l	+	**1**	**6**	**2**	**2**	**2**	**3**	**6**	**4B**	**4**	**2A**	**8**	**7**
243	R	16	major	z-l	+	**1**	**6**	**2**	**2**	**2**	**3**	**6**	**4B**	**4**	**2A**	**8**	**7**
243	L	16	major	z-l	+	**1**	**6**	**2**	**2**	**2**	**3**	**6**	**4B**	**4**	**2A**	**8**	**7**
259	R	16	major	z-o	+	**1**	**3**	**8**	**8**	**8**	**3**	**1**	**4B**	**4**	**4B**	**9**	**8**
259	L	16	major	z-o	+	**1**	**3**	**8**	**8**	**8**	**3**	**1**	**4B**	**4**	**4B**	**9**	**8**
259	R	16	minor	z-pA	+	**2**	**4**	**1A**	**1**	**1**	**1**	**6**	**2**	**2B**	**4B**	**12**	**2**
259	L	16	minor	z-p	+	**2**	**4**	**1A**	**1**	**1**	**1**	**6**	**2**	**2B**	**1A**	**12**	**2**
278	R	4	major	z-u	+	**6**	**3**	**9**	**1**	**9**	**9**	**6**	**3A**	**1B**	**4B**	**8**	**2**
278	R	16	major	z-u	+	**6**	**3**	**9**	**1**	**9**	**9**	**6**	**3A**	**1B**	**4B**	**8**	**2**
278	L	16	major	z-u	+	**6**	**3**	**9**	**1**	**9**	**9**	**6**	**3A**	**1B**	**4B**	**8**	**2**
278	R	4	minor	z-w	+	**2**	**4**	**1A**	**9**	**1**	**1**	**1**	**4D**	**5**	**3A**	**9**	**2**
278	R	16	minor	z-w	+	**2**	**4**	**1A**	**9**	**1**	**1**	**1**	**4D**	**5**	**3A**	**9**	**2**
278	L	16	minor	z-w	+	**2**	**4**	**1A**	**9**	**1**	**1**	**1**	**4D**	**5**	**3A**	**9**	**2**
Number of total genotypes						3/6	3/6	6/12	5/11	5/11	5/9	3/7	5/8	5/8	5/8	6/14	4/9
Genotype prevalence key						>35%		>30%			>20%		>15%		>10%	>5%	>1%

417

418

419

420 REFERENCES

421

422 1. Gompels UA, Larke N, Sanz-Ramos M, et al. Human cytomegalovirus infant infection
423 adversely affects growth and development in maternally HIV-exposed and unexposed infants
424 in Zambia. Clin Infect Dis 2012;54:434-42
425 2. Josephson CD, Caliendo AM, Easley KA, et al. Blood transfusion and breast milk
426 transmission of cytomegalovirus in very low-birth-weight infants: a prospective cohort study.
427 JAMA Pediatr 2014;168:1054-62
428 3. Musonda KG, Nyonda M, Filteau S, Kasonka L, Monze M and Gompels UA. Increased
429 Cytomegalovirus Secretion and Risks of Infant Infection by Breastfeeding Duration From
430 Maternal Human Immunodeficiency Virus Positive Compared to Negative Mothers in Sub-
431 Saharan Africa. J Pediatric Infect Dis Soc 2016;5:138-46
432 4. Hamprecht K, Goelz R. Postnatal Cytomegalovirus Infection Through Human Milk in
433 Preterm Infants: Transmission, Clinical Presentation, and Prevention. Clin Perinatol
434 2017;44:121-130
435 5. Griffiths P, Baraniak I and Reeves M. The pathogenesis of human cytomegalovirus. J
436 Pathol 2015;235:288-97
437 6. Gatherer D, Seirafian S, Cunningham C, et al. High-resolution human cytomegalovirus
438 transcriptome. Proc Natl Acad Sci U S A 2011;108:19755-60
439 7. Bates M, Monze M, Bima H, et al. High human cytomegalovirus loads and diverse linked
440 variable genotypes in both HIV-1 infected and exposed, but uninfected, children in Africa.
441 Virology 2008;382:28-36
442 8. Bradley AJ, Kovacs IJ, Gatherer D, et al. Genotypic analysis of two hypervariable human
443 cytomegalovirus genes. J Med Virol 2008;80:1615-23

444 9. Davison AJ, Akter P, Cunningham C, et al. Homology between the human
445 cytomegalovirus RL11 gene family and human adenovirus E3 genes. J Gen Virol
446 2003;84:657-63
447 10. Dolan A, Cunningham C, Hector RD, et al. Genetic content of wild-type human
448 cytomegalovirus. J Gen Virol 2004;85:1301-12
449 11. Lassalle F, Depledge D, Reeves MB, et al. Islands of linkage in an ocean of pervasive
450 recombination reveals two-speed evolution of human cytomegalovirus genomes. Virus
451 Evolution, 2016;2
452 12. Mattick C, Dewin D, Polley S, et al. Linkage of human cytomegalovirus glycoprotein gO
453 variant groups identified from worldwide clinical isolates with gN genotypes, implications
454 for disease associations and evidence for N-terminal sites of positive selection. Virology
455 2004;318:582-97
456 13. Paterson DA, Dyer AP, Milne RS, Sevilla-Reyes E and Gompels UA. A role for human
457 cytomegalovirus glycoprotein O (gO) in cell fusion and a new hypervariable locus. Virology
458 2002;293:281-94
459 14. Cunningham C, Gatherer D, Hilfrich B, et al. Sequences of complete human
460 cytomegalovirus genomes from infected cell cultures and clinical specimens. J Gen Virol
461 2010;91:605-15
462 15. Rasmussen L, Geissler A and Winters M. Inter- and intragenic variations complicate the
463 molecular epidemiology of human cytomegalovirus. J Infect Dis 2003;187:809-19
464 16. Sijmons S, Thys K, Mbong Ngwese M, et al. High-throughput analysis of human
465 cytomegalovirus genome diversity highlights the widespread occurrence of gene-disrupting
466 mutations and pervasive recombination. J Virol 2015

17. Gorzer I, Guelly C, Trajanoski S and Puchhammer-Stockl E. Deep sequencing reveals highly complex dynamics of human cytomegalovirus genotypes in transplant patients over time. J Virol 2010;84:7195-203

18. 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

19. Ross SA, Arora N, Novak Z, Fowler KB, Britt WJ and Boppana SB. Cytomegalovirus reinfections in healthy seroimmune women. J Infect Dis 2010;201:386-9

20. 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

21. Hage E, Wilkie GS, Linnenweber-Held S, et al. Characterization of Human Cytomegalovirus Genome Diversity in Immunocompromised Hosts by Whole-Genome Sequencing Directly From Clinical Specimens. J Infect Dis 2017;215:1673-1683

22. 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

23. Houldcroft CJ, Beale MA and Breuer J. Clinical and biological insights from viral genome sequencing. Nat Rev Microbiol 2017;15:183-192

24. Bolger AM, Lohse M and Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014;30:2114-20

25. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 2008;18:821-9

26. Assefa S, Keane TM, Otto TD, Newbold C and Berriman M. ABACAS: algorithm-based automatic contiguation of assembled sequences. Bioinformatics 2009;25:1968-9

492 27. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and
493 SAMtools. Bioinformatics 2009;25:2078-9
494 28. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce
495 framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297-
496 303
497 29. Rutherford K, Parkhill J, Crook J, et al. Artemis: sequence visualization and annotation.
498 Bioinformatics 2000;16:944-5
499 30. Tweedy JG, Escriva E, Topf M and Gompels UA. Analyses of Tissue Culture Adaptation
500 of Human Herpesvirus-6A by Whole Genome Deep Sequencing Redefines the Reference
501 Sequence and Identifies Virus Entry Complex Changes. Viruses 2017;10
502 31. Yan H, Koyano S, Inami Y, et al. Genetic linkage among human cytomegalovirus
503 glycoprotein N (gN) and gO genes, with evidence for recombination from congenitally and
504 post-natally infected Japanese infants. J Gen Virol 2008;89:2275-9
505 32. Maschmann J, Goelz R, Witzel S, et al. Characterization of human breast milk leukocytes
506 and their potential role in cytomegalovirus transmission to newborns. Neonatology
507 2015;107:213-9
508 33. Jiang XJ, Adler B, Sampaio KL, et al. UL74 of human cytomegalovirus contributes to
509 virus release by promoting secondary envelopment of virions. J Virol 2008;82:2802-12
510 34. Kropff B, Burkhardt C, Schott J, et al. Glycoprotein N of human cytomegalovirus
511 protects the virus from neutralizing antibodies. PLoS Pathog 2012;8:e1002999
512 35. Scrivano L, Sinzger C, Nitschko H, Koszinowski UH and Adler B. HCMV spread and
513 cell tropism are determined by distinct virus populations. PLoS Pathog 2011;7:e1001256
514 36. Wu Y, Prager A, Boos S, et al. Human cytomegalovirus glycoprotein complex gH/gL/gO
515 uses PDGFR-alpha as a key for entry. PLoS Pathog 2017;13:e1006281

516 37. Laib Sampaio K, Stegmann C, Brizic I, Adler B, Stanton RJ and Sinzger C. The
517 contribution of pUL74 to growth of human cytomegalovirus is masked in the presence of
518 RL13 and UL128 expression. J Gen Virol 2016;97:1917-27
519 38. Stanton RJ, Baluchova K, Dargan DJ, et al. Reconstruction of the complete human
520 cytomegalovirus genome in a BAC reveals RL13 to be a potent inhibitor of replication. J Clin
521 Invest 2010;120:3191-208
522 39. Boucoiran I, Mayer BT, Krantz EM, et al. Nonprimary Maternal Cytomegalovirus
523 Infection After Viral Shedding in Infants. Pediatr Infect Dis J 2018;37:627-631
524 40. Kalser J, Adler B, Mach M, Kropff B, Puchhammer-Stockl E and Gorzer I. Differences in
525 Growth Properties among Two Human Cytomegalovirus Glycoprotein O Genotypes. Front
526 Microbiol 2017;8:1609
527 41. Heo J, Dogra P, Masi TJ, et al. Novel Human Cytomegalovirus Viral Chemokines,
528 vCXCL-1s, Display Functional Selectivity for Neutrophil Signaling and Function. J Immunol
529 2015;195:227-36
530 42. Corrales-Aguilar E, Trilling M, Hunold K, et al. Human cytomegalovirus Fcgamma
531 binding proteins gp34 and gp68 antagonize Fcgamma receptors I, II and III. PLoS Pathog
532 2014;10:e1004131
533 43. Cortese M, Calo S, D'Aurizio R, Lilja A, Pacchiani N and Merola M. Recombinant
534 human cytomegalovirus (HCMV) RL13 binds human immunoglobulin G Fc. PLoS One
535 2012;7:e50166
536 44. Di Bona D, Accardi G, Aiello A, et al. Association between gamma marker, human
537 leucocyte antigens and killer immunoglobulin-like receptors and the natural course of human
538 cytomegalovirus infection: a pilot study performed in a Sicilian population. Immunology
539 2018;153:523-531

45. Pandey JP, Namboodiri AM, Mohan S, Nietert PJ and Peterson L. Genetic markers of immunoglobulin G and immunity to cytomegalovirus in patients with breast cancer. Cell Immunol 2017;312:67-70

46. Gorzer I, Trajanoski S, Popow-Kraupp T and Puchhammer-Stockl E. Analysis of human cytomegalovirus strain populations in urine samples of newborns by ultra deep sequencing. J Clin Virol 2015;73:101-104

47. Mayer BT, Krantz EM, Swan D, et al. Transient Oral Human Cytomegalovirus Infections Indicate Inefficient Viral Spread from Very Few Initially Infected Cells. J Virol 2017;91

48. Slyker J, Farquhar C, Atkinson C, et al. Compartmentalized cytomegalovirus replication and transmission in the setting of maternal HIV-1 infection. Clin Infect Dis 2014;58:564-72

49. Tamura K, Stecher G, Peterson D, Filipski A and Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 2013;30:2725-9