Epidemiology of human cytomegalovirus strains through comparison of methodological approaches to explore gN variants

Sara Pignatelli, Paola Dal Monte

Department of Haematology, Oncology and Laboratory Medicine, Clinical Unit of Microbiology,
St. Orsola Malpighi General Hospital, University of Bologna, Bologna, Italy

--- SUMMARY ---

Genomic variation among human cytomegalovirus (HCMV) wild-type strains is a well-documented phenomenon probably implicated in HCMV-induced immunopathogenesis. Extensive genetic polymorphism has been detected for the envelope glycoprotein N (gN) and HCMV clinical isolates have been clustered into seven distinct gN variants (gN-1, gN-2, gN-3a, gN-3b, gN-4a, gN-4b, gN-4c). Several studies from different research groups worldwide have addressed this topic using different methodological approaches (PCR-RFLP, PCR-Cloning, PCR-Sequencing) and sometimes yielding apparently conflicting results. This paper analyses the epidemiology of HCMV strains through analysis of gN variants, criticizing the methodological approaches and study populations by comparison of published reports.

KEY WORDS: Human cytomegalovirus, Glycoprotein N, gN genotypes, PCR primers

INTRODUCTION

Human cytomegalovirus (HCMV) is a pathogen Beta-herpesvirus able to establish lifelong persistent infections, which normally remain asymptomatic. However, severe diseases may develop resulting in a wide spectrum of clinical manifestations, especially in immunologically impaired individuals (e.g. AIDS patients and transplant recipients) or in congenitally-infected newborns (Ho, 1990). The mechanisms that determine the type, duration and severity of clinical manifestations are not completely understood. Host factors (e.g. development of a potent cellular and/or humoral response), together with viral determinants (e.g. viral load and the infecting HCMV strain) may play an important role.

Genetic variability among HCMV strains is currently one of the most widely investigated viral factors implicated in immunopathogenesis and able to influence the clinical manifestations of cytomegalovirus infections (Pignatelli *et al.*, 2004).

Although several studies suggest that different strains of cytomegalovirus are 95% homologous, polymorphic sequences have been detected in coding and non coding regions of the virus genome. In most of the variable regions identified nucleotide changes are strongly clustered determining the existence of dominant genomic variants, usually defined as "genotypes".

HCMV ORF UL73 is one of the most polymorphic genes among HCMV clinical strains. It encodes the immunogen envelope glycoprotein N (gN), a gC-II component implicated in virus attachment to the host cell and spread. The UL73 *locus* showed a polymorphic region (5' sequence

---

Corresponding author
Sara Pignatelli
Department of Haematology
Oncology and Laboratory Medicine
Clinical Unit of Microbiology
St. Orsola Malpighi General Hospital
University of Bologna, Bologna, Italy
Via Massarenti, 9 - 40138 Bologna, Italy
E-mail: sara.pignatelli@unibo.it

TABLE 1 - *Methodological approaches to discriminate gN genotypes. The method used for DNA extraction, PCR primers and conditions, gN cluster assignment method are reported.*

Method	DNA extraction	Primers	PCR conditions	Reference
Nested PCR Sequencing	QIAmp DNA Minikit (QIAGEN, HIlden, Germany)	gN-F1 5'-gacagtaccagttgagagtcg-3' gN-R1 5'-ggaytatctagactcgctgc-3' gN-up 5'-tggtgtgatggagtggaac-3' gN-lw 5'-tagcctttggtggtggttgc-3'	- initial denaturation 5' 96°C; - 40 cycles; - 45" 94°C - 45" 55°C -2' 72°C - final elongation 10' at 72°C	Yan *et al.*, 2008
PCR-RFLP*	Total Nucleic Acid Isolation Kit (Roche Diagnostics, Basel, Switzerland)	UL73-105725F 5'-ggcggtggtgtgatggag-3 UL73-106122R 5'-ttctggaagcagcaatgtcg-3'	- hot start 5' 96°C; - 35 cycles; - 1' 96°C -1' 55°C - 1' 72°C; - final elongation 10' at 72°C	Chen *et al.*, 2008
PCR-Cloning plus sequencing [TOPO TA pCR 2.1 (Invitrogen Inc., Carlsbad, CA-USA]	QIAGEN Blood Spin Column (Qiagen Inc., Chatsworth, CA-USA)	gN-Fw 5'-ggcggtggtgtgatggagtg-3' gN-Rev 5'-aatagcctttggtggtggttgc-3'	- 2' initial denaturation 96°C; - 7 cycles 30" 96°C - 30" 65°C -40" 72°C; - decrease by 1°C of the annealing temperature; - 28 cycles with annealing 58°C; - 7' final extension at 72°C	Novak *et al.*, 2008
PCR-RFLP [SacI, SalI, BsaXI, MmeI]	QIAGEN Blood Spin Column (Qiagen Inc., Chatsworth, CA-USA)	gN-Fw 5'-ggcggtggtgtgatggagtg-3' gN-Rev 5'-aatagcctttggtggtggttgc-3'	same as above	Novak *et al.*, 2008
Nested PCR RFLP [SacI, SalI, ScaI]	Isoquick Nucleic Acid Extraction Kit (ORCA Research Inc., WA, USA)	forward73 (external) 5'-tcccagaatcgctgaaaa-3' L73is (external) 5'-cacccacgtatgtaaaccttac-3' gN-up 5'-tggtgtgatggagtggaac-3' gN-lw 5'-tagcctttggtggtggttgc-3'	- hot start 2' 96°C; - 35 cycles; - 30" 94°C - 30" 55°C -40" 72°C; - final elongation 5' at 72°C	Pignatelli *et al.*, 2006
PCR Sequencing or RFLP [RsaI]	QIAmp DNA Minikit (QIAGEN, HIlden, Germany)	gN1 5'-ggtcggtcaacatcgtaa-3' gN2 5'-ctcgctgctgtttctagatg-3'	- 45 cycles - 30" 94°C - 30" 58°C -40" 72°C	Puchhammer-Stöckl *et al.*, 2006
Touchdown PCR Sequencing	Isoquick Nucleic Acid Extraction Kit (ORCA Research Inc., WA, USA)	gN-up 5'-tggtgtgatggagtggaac-3' gN-lw 5'-tagcctttggtggtggttgc-3'	- hot start 12' 96°C; - 10 cycles 1' 94°C - 1' 67°C -1' 72°C, decreasing the annealing temperature 1°C per cycle; - 30 cycles, 1' 96°C – 1' 57°C -1' 72°C; - final elongation 10' at 72°C	Rossini *et al.*, 2005
PCR Sequencing	QIAmp Kit (QIAGEN, HIlden, Germany)	gN-up 5'-tggtgtgatggagtggaac-3' gN-lw 5'-tagcctttggtggtggttgc-3	see Pignatelli *et al.*, 2003a	Beyari *et al.*, 2005

TABLE 1 - *Methodological approaches to discriminate gN genotypes. The method used for DNA extraction, PCR primers and conditions, gN cluster assignment method are reported.*

Method	DNA extraction	Primers	PCR conditions	Reference
PCR Sequencing*	Isoquick Nucleic Acid Extraction Kit (ORCA Research Inc., WA, USA)	gN-up 5'-tggtgtgatggagtggaac-3' gN-lw 5'-tagcctttggtggtggttgc-3'	see Pignatelli *et al.*, 2003a	Pignatelli *et al.*, 2003c
PCR Sequencing*	Isoquick Nucleic Acid Extraction Kit (ORCA Research Inc., WA, USA)	gN-up 5'-tggtgtgatggagtggaac-3' gN-lw 5'-tagcctttggtggtggttgc-3	see Pignatelli *et al.*, 2003a	Pignatelli *et al.*, 2003b
PCR Sequencing	Isoquick Nucleic Acid Extraction Kit (ORCA Research Inc., WA, USA)	gN-up 5'-tggtgtgatggagtggaac-3' gN-lw 5'-tagcctttggtggtggttgc-3	- hot start 5' 96°C; - 35 cycles; - 1' 96°C -1' 55°C - 1' 72°C; - final elongation 10' at 72°C	Pignatelli *et al.*, 2003a
PCR Sequencing*	Phenol-chlorophorm extraction	forward-U73 5'-ttcggtcggtcaacatcgtaag-3' Reverse-L73 5'-cacccactatgtaaaccttac-3'	- 35 cycles; - 1' 96°C – 1' 50°C - 1' 72°C; - final elongation 10' at 72°C	Pignatelli *et al.*, 2001

Abbreviations: PCR: polymerase chain reaction; RFLP: restriction fragment length polymorphism; y= c + t.
*Clinical isolates were passaged on human embryonic lung fibroblasts.

of the ORF) involving the N-terminal domain of gN and allowing the identification of seven gN genomic variants (gN-1, gN-2, gN-3a, gN-3b, gN-4a, gN-4b, gN-4c) (for review see Pignatelli *et al.*, 2004).

Many research group are engaged in the study of HCMV polymorphic ORFs (e.g. UL55-gB, UL73-gN, UL144-TNF receptor) to establish the epidemiology of cytomegalovirus strains circulating worldwide and establish critical associations between certain genomic variants and virus biological features of clinical relevance.

However, much criticism has emerged from the published studies disclosing discrepancies mainly affected by the population examined and/or the methodological approach to detect gene polymorphisms in clinical HCMV-positive specimens. The aim of this brief review is to elucidate some points on gN genotypes detection in clinical specimens, evaluate mixed infections and the distribution of genotypic frequency critically comparing a number of studies.

Several studies in recent years have documented UL73-gN hypervariability using different approaches (PCR-sequencing; PCR-RFLP; PCR-cloning) to amplify and genotype UL73 from the

HCMV genome. Table 1 summarizes and compares the different methods and sets of primers used to discriminate gN variants. In addition, the relative location of the upper primers designed for PCR reactions are depicted in Figure 1. Briefly, the methodological approaches used in each study may be essentially distinguished as follows:

1) PCR-sequencing: This is the most widespread method adopted, independently from the set of primer designed to amplify UL73 from the HCMV genome. Since 2001, the following sets of primers have been used in standard PCR reactions: forward-U73/Reverse L73 (Pignatelli *et al.*, 2001 and 2003b); gN-up/gN-lw (Pignatelli *et al.*, 2003a, 2003b and 2003c; Beyari *et al.*, 2005); gN1/gN2 (Puchhammer-Stockl *et al.*, 2006).

In addition, to improve PCR efficiency, in particular when the original sample may host a low amount of HCMV DNA (e.g. peripheral blood monocytes of seropositive healthy donors), two more sensitive PCR reactions have been optimized: 1) a touchdown PCR with primers gN-up/gN-lw (Rossini *et al.*, 2005); 2) a nested PCR, using either the external primers forward73/L73is and as internal primers gN-
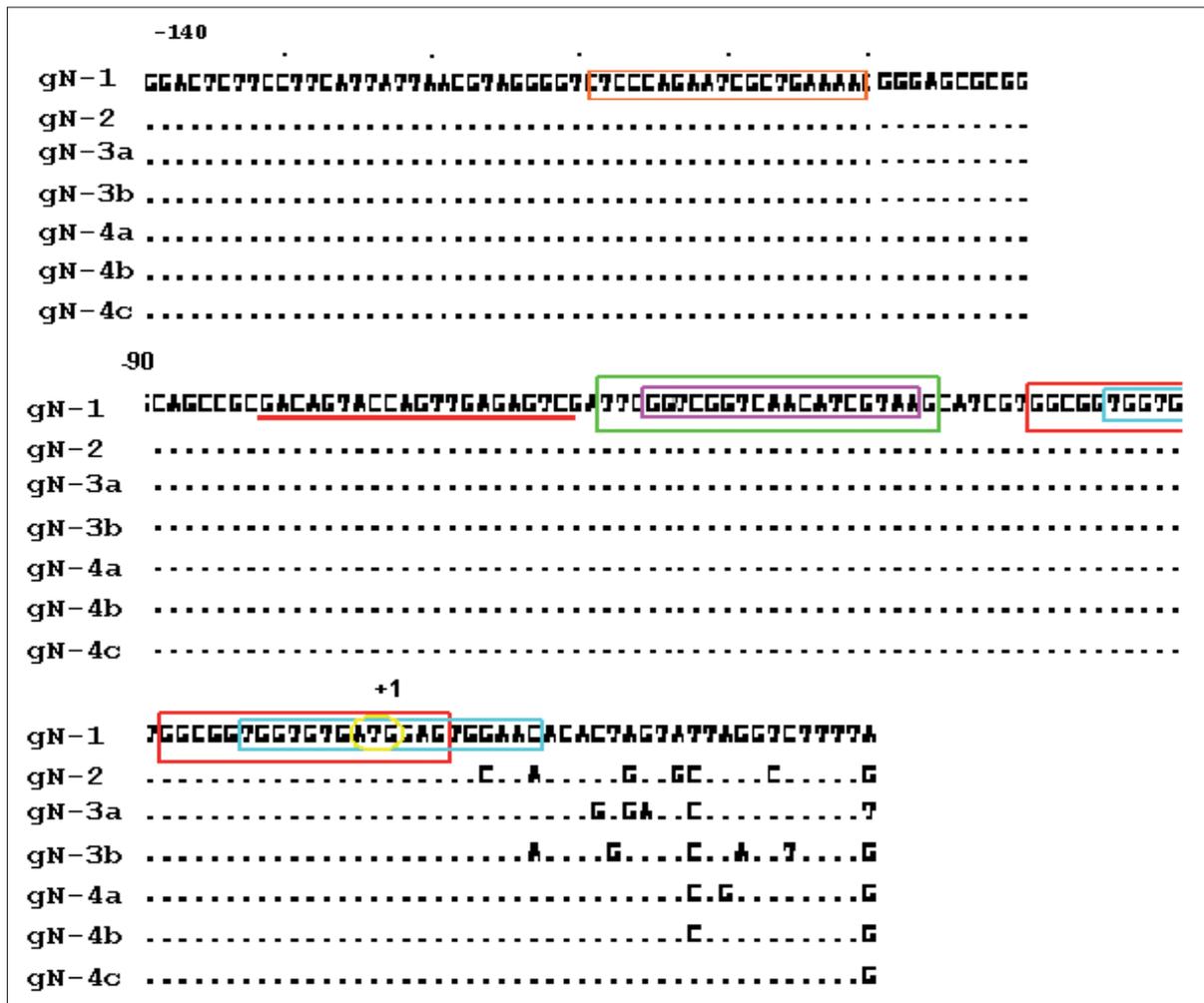
FIGURE 1 - *Alignment of the region surrounding the ATG start codon of the seven UL73(gN) HCMV prototype strains. The AD169/gN-1 was used as reference strain. Dots indicate identity. The upper primers employed in PCR amplification of the gN-encoding gene in each study mentioned in the text are shown in the alignment. Green box: forward-U73 (Pignatelli et al., 2001). Turquoise box: gN-up (Pignatelli et al. 2003a). Pink box: gN-1 (Puchhammer-Stöckl et al., 2006). Orange box: forward73 (Pignatelli et al., 2006). Red box: UL73-105725F (Chen et al., 2008). Underlined in blue: gN-Fw (Novak et al., 2008). Underline red: gN-F1 (Yan et al., 2008). Yellow circle: ATG start codon.*

up/gN-lw (Pignatelli *et al.*, 2006), or a second combination of primers, namely the external gN-F1/gN-R1 plus the internal gN-up/gN-lw, which has been recently reported (Yan *et al.*, 2008). The next genotyping step requires sequencing facilities to obtain the nucleotide sequence of UL73, which will be compared to gN-protoype strains by appropriate software (e.g. Lasergene/Mega Align, Phylip, ClustalW) to assign a wild-type clinical isolate to a gN variant. This approach has some limitations, because results may be strongly affected by the original amount of viral DNA, especially in the presence of mixed infections, which may be underestimated and/or lost during the consecutive PCRs. In addition, the sensitivity and specificity of PCR amplifications optimized in each study may also affect the results, strongly depending on the original specimen and its viral DNA load.

Its strength is that sequencing is the best approach to avoid any mistake in gN genotype assignment and less time-consuming than the approach based on a intervening cloning step (see below for details).

2) PCR-RFLP: the joining of UL73 amplification by PCR to restriction fragment length polymorphism (RFLP) analysis avoids the time-consuming and expensive sequencing step, allowing the clustering of UL73 PCR products simply observing differences in cutting patterns by restriction endonucleases, visualized by agarose gel electrophoresis. PCR-RFLP was first proposed by Pignatelli *et al.*, 2003a (PCR primers: gN-up/gN-lw) and was based on the enzymes SacI, SalI and ScaI.

This method has been recently applied also by Chen *et al.*, 2008 (PCR-primers: UL73105725F/ UL73-106122R) with some discrepancies reported and commented in the following paragraphs. Later on, changes in the RFLP analysis were added by Puchhammer-Stockl *et al.*, 2006 (PCR primers: gN1/gN2) by simply cutting the PCR products with the endonuclease RsaI (gN-4b and 4-c are not discriminated) and by Novak *et al.*, 2008 (PCR primers: gN-Fw/gN-Rev), who introduced a quadruple digestion with SalI, SacI, BsaXI and MmeI. Although time and costs are lowered, the RFLP analysis presents some problems depending on the amount of DNA in the samples tested, which may determine an underestimation of mixed infections, and mistakes in assignment of genotypes, due to minute differences in band size or small and single changes at nucleotide level, affecting the restriction pattern.

3) PCR-Cloning plus sequencing: this approach was introduced by Novak *et al.*, 2008 and is based on:

a) PCR-amplification of UL73 from the HCMV genome using the primers gN-Fw and gN-Rev;

b) cloning into TOPO TA cloning vector pCR2.1 (Invitrogen Inc., Carlsbad, CA);

c) screening of colonies for the presence of the gN-insert by direct PCR; d) growth and gN-typing of bacterial colonies by sequencing using the gN-Fw primer.

The cloning-sequencing method is not affected by the relative amount of type-specific DNA in a single sample, thus allowing a better identification of mixed strain infection. However, the sensitivity and specificity of PCR amplification and the number of bacterial colonies that can be screened give serious limitations to this time-consuming approach.

Finally, as reported in Table 1, independently from the method used, some studies (Pignatelli *et al.*, 2001; Pignatelli *et al.*, 2003b and 2003c; Chen *et al.*, 2008) analysed the virus isolates after propagation in cell culture rather than DNA extracted from the original specimen (e.g. urine, saliva, blood, amniotic fluid). This step could select among the multiple strains potentially present in the original clinical sample, depending on their relative amount, thus it should be avoided.

Given this overview of the methodological approaches optimized over the years to genotype HCMV gN, let us comparatively review the results, discrepancies and strength of the main studies.

As clearly shown in Table 2, the studies describing gN variants distribution analysed HCMV strains from distinct patient populations, differing both in terms of number of cases, geographical origin of strains and clinical status of patients (e.g. solid organ transplant recipients, congenital infections or AIDS patients).

This *premise* may obviously explain the main differences in the genotypic frequencies reported in each study, but also differences in the methodological approaches should be taken into account. Drawing from the recently published study by Chen *et al.* (2008), who analysed the distribution of human cytomegalovirus gN genotypes among the Chinese population of Taiwan, gN genotypic frequencies were shown to differ significantly from those reported by Pignatelli *et al.* (2003a), especially those referred to gN-2 and gN-3 variants. The authors suggest that primer design influenced the previously reported results, in that a one (for gN-3) or two (for gN-2) nucleotide mismatch in the upper primer might have led to an underestimation of gN-2 and gN-3 genotypes among Chinese HCMV wild type isolates.

However, other explanations might be considered to justify the reported discrepancies further. In the study by Chen *et al.* HCMV strains were genotyped after inoculation on lung fibroblasts, as also performed in the earliest studies by Pignatelli *et al.* (2001). This preliminary step to increase the infecting strain in cell culture is absolutely unnecessary when the modern PCR techniques are employed to amplify the viral genome. In addition, this step may select the strain(s) present in the original specimen (Sinzger *et al.*, 1999; Novak *et al.*, 2008), thus potentially causing an artifactual result.

Secondly, the PCR conditions used by Pignatelli *et al.* (2003a) and Chen *et al.* (2008) were slightly

TABLE 2 - *Main studies on gN genotypes. The method used for genotyping, number of cases, population examined, and gN genotypic frequencies are reported in each column.*

Method	N° cases	Population	gN-1 (%)	gN-2 (%)	gN-3 (%)			gN-4 (%)			Reference
Nested PCR Sequencing	63	32 congenital and 31 post-natal infections	12.7	6.4	3a 19	3b 11.1		50.8			Yan *et al.*, 2008
PCR-RFLP	48	Chinese (clinical data not specified)	29.2	22.9	4			5.5			Chen *et al.*, 2008
Cloning	16	HCMV-seropositive urban, low income, African American women	35.3	8.8	41.2			14.7			Novak *et al.*, 2008
Nested PCR or Touchdown PCR, RFLP	23	HCMV seropositive healthy blood donors	87	-	-			13			Pignatelli *et al.*, 2006
PCR Sequencing or RFLP	13	Lung transplant recipients	26.6	-	40			4a 13.3	4b 13.3	4c 6.6	Puchhammer-Stöckl *et al.*, 2006
Touchdown PCR Sequencing	74	Solid organ transplant recipients	24.3	5.4	3a 4	3b 5.5		4a 19	4b 12	4c 30	Rossini *et al.*, 2005
PCR Sequencing	46	19 patients with Kaposi's sarcoma and relatives, 51% HIV+	65.2	-	3a 28.3	3b 4.3	3c 2.2	-			Beyari *et al.*, 2005
PCR Sequencing	223	Congenital (97) and pediatric (20) infections; 29 AIDS; 50 transplant; 10 BMT; 17 NA	27.3	2.2	15			55.5			Pignatelli *et al.*, 2003a
PCR Sequencing*	29	AIDS patients	48.3	-	10.3			4a 10	4b 7	4c 21	Pignatelli *et al.*, 2003b
PCR Sequencing*	50	Solid organ transplant recipients	28	8	18			46			Pignatelli *et al.*, 2003b
PCR Sequencing*	93	HCMV congenital infections	23.6	1.1	12.9			62.4			Pignatelli *et al.*, 2003c
PCR Sequencing*	43	Lab strains + 38 patients (transplant recipients, HIV+, congenital infections)	16.7	12	16.7			4a 19	4b 16.7	4c 19	Pignatelli *et al.*, 2001

Abbreviations: BMT: bone marrow transplant recipients; NA: not available; PCR: polymerase chain reaction; RFLP: restriction fragment length polymorphism.
*clinical isolates were passaged on human embryonic lung fibroblasts

different in enzyme usage, upper primer sequence and concentration, thermal cycler and amount of starting DNA. All these discrepancies may be responsible for the different sensitivity and specificity of these two PCRs and yield of amplification products, thereby explaining why Chen *et al.*, 2008 partly failed to amplify gN-2 and gN-3 isolates using the "mismatched" gN-up primer. Indeed, there is no clear indication that at average stringency conditions PCR should not entirely work with one or two mismatches at upper primer level.

To overcome any other confusion about operating with all the recommended UL73-gN primers, a clearer picture of the alignment of the seven gN genotypes is shown in Figure 1, also displaying the relative position of the upper primers used in each of the studies summarized in Table 1.

Furthermore, a high gN-2 (22.9%) and gN-3 (14,6%) genotypic frequency was reported by Chen *et al.* (2008), significantly differing from other studies analysing a Chinese population. Again, the authors claim that primer design could (obviously) influence PCR results, leading to an underestimation of these variants due to a mismatch in the last nucleotides of the annealing region. This is quite plausibile. However, other important factors inferable from other studies reported in Table 2 and mentioned in the following paragraph must be addressed.

In particular, the overall gN-3 genotypic frequency reported in most papers ranges from 9.5 to 18%, not significantly different from that reported for the Chinese, Australian, European and Northern American populations, but showing two extra peaks in Novak *et al.*, 2008 (41.2%) and Beyari *et al.*, 2005 (34.8%). The explanation for this overestimation of gN-3 in these last two studies, if any, cannot be sought only in the methodological approaches, since they are strongly different (PCR-cloning *vs* PCR-sequencing) and Beyari *et al.* even used the same PCR reported by Pignatelli *et al.*, 2003a (with the upper primer mismatch). It would be better to consider the geographical origin of clinical strains: for example, the very small Chinese sample analysed by Pignatelli *et al.*, 2003a came from Shenyang, a region far away from Taiwan, where the samples examined by Chen *et al.*, 2008 were collected; thus a local selection of HCMV strains might have occurred.

To confirm that not only primer mismtaches may explain discrepancies in genotypic frequencies, it should be useful to report these data on a gN-2 variant. Our latest paper on HCMV polymorphisms among congenitally infected newborns (Pignatelli *et al.* 2008, submitted for publication) describes the following general distribution of the detected gN variants: gN-1 20.3%; gN-2 8.1%; gN-3a 8.1%; gN-3b 6.8%; gN-4a 20.3%; gN-4b 13.5%; gN-4c 22.9%.

The genotypic frequencies of the rarest variants (gN-2 and gN-3) did not significantly vary from previous papers, although they were obtained using different sets of primers and genotyping methods (Rossini *et al.*, 2005; Puchhammer-Stöckl *et al.*, 2006; Novak *et al.*, 2008). For instance, as reported in Table 2, Puchhammer-Stöckl *et al.* (2006) and Novak *et al.* (2008) found the following gN genotypes distributions: gN-1 26.6% and 35.3%; gN-2 0% and 8.8%; gN-3 40% and 41,2%; gN-4 33.2% and 14.7% respectively.

A number of studies among those reported in Table 2 simply mention a gN-3 group rather than its subgroups and rarely clarify this discrepancy in the text.

The two gN-3 sub-variants are clearly discernible by both sequencing and RFLP analysis and the two nucleotide and amminoacidic sequences strongly diverge as much as they could potentially have different biological properties (e.g. antigenic, structural…). Thus, the gN-3 subgroups must be reported and analysed for their clinical relevance as determinants of strain virulence as separate strains. This should instead be overcome for the gN-4 sub-variants, especially gN-4b and gN-4c, which show a lower intragroup divergence both at nucleotide and aminoacid level (Pignatelli *et al.*, 2001 and 2003a).

Most importantly, only a small proportion of the studies mentioned in Table 1 reported in the methods section their respective PCR efficiency. If wording such as "no unsuccessful PCR was obtained" or "the PCR efficiency reaches 97-99%, with a very low failure rate", were reported in the text, any further doubt about possible underestimation of the rarest variants would be thereby eliminated.

Since the main question to be answered when comparing genotyping studies is related to PCR sensitivity, a percentage of unsuccessful gN-PCR should be taken into account. This is probably

TABLE 3 - *Comparison of the PCR-RFLP approaches by Pignatelli et al., 2003a and Chen et al., 2008.*

gN-type		Primers		PCR-product size		SacI (n° cut sites)		ScaI (n° cut sites)		SalI (n° cut sites)	
Pignatelli, 2003a	*Chen, 2008*	*Pignatelli, 2003a*	*Chen, 2008*	*Pignatelli, 2003a*	*Chen, 2008*	*Pignatelli, 2003a*	*Chen, 2008*	*Pignatelli, 2003a*	*Chen, 2008*	*Pignatelli, 2003a*	*Chen, 2008*
1	1	gN-up / gN-lw	UL73-105725F UL73-106122R	420 bp	398 bp	297, 123 (1)	302, 96 (1)	420 (-)	398 (-)	420 (-)	398 (-)
2	2	gN-up / gN-lw	UL73-105725F UL73-106122R	417 bp	395 bp	229, 123, 65 (2)	229, 96, 70 (1)	417 (-)	395 (-)	296, 121 (1)	269, 126 (1)
3a	3 [3a]	gN-up / gN-lw	UL73-105725F UL73-106122R	420 bp	398 bp	420 (-)	398 (-)	420 (-)	398 (-)	420 (-)	398 (-)
3b	Not reported	gN-up / gN-lw	UL73-105725F UL73-106122R	420 bp	[398]	420 (-)	[398 (-)]	221, 172, 27 (2)	[194, 172, 32 (2)]	420 (-)	[398 (-)]
4a	4a	gN-up / gN-lw	UL73-105725F UL73-106122R	414 bp	392	291, 123 (1)	296, 96 (1)	221, 166, 27 (2)	198, 194 (1) [194, 166, 27 (2)]	341, 73 (1)	314, 78 (1)
4b	Reported Simply as 4	gN-up / gN-lw	UL73-105725F UL73-106122R	414 bp / 414 bp	[392] / 392	414 (-) / 414 (-)	[392 (-)] / 392	414 (-) / 387, 27 (1)	[392 (-)] / 360, 32 (1)	341, 73 (1) / 341, 73 (1)	[314, 78 (1)] / 314, 78 (1)
4c	Not reported	gN-up gN-lw	UL73-105725F UL73-106122R	411 bp	[389]	411 (-)	[389 (-)]	239, 172 (1)	[212, 177 (1)]	338, 73 (1)	[311, 78 (1)]

Simulation of the UL73 PCR-product digestion was performed using the Rebase software (http://www.rebase.neb.com/NEBcutter/). The PCR-primers, PCR-product size and the expected restriction fragment length for each enzyme in each study are reported. The number of restriction sites for each sequence are shown in parentheses. Dashes indicate the absence of cut sites. Data referred to genotypes gN-3 and gN-4 and their subgroups elicited by Pignatelli *et al.*, 2003a and rendered from Chen *et al.*, 2008 are reported in square brackets.

due to sample collection, DNA extraction, purity and concentration and amplification conditions. Another method widely used to genotype HCMV strains is PCR followed by RFLP analysis. For UL73 this approach has been proposed by Pignatelli *et al.*, 2003a and was based on the use of a combination of the 3 restriction endonucleases (ScaI, SalI and SacI). This method was subsequently applied by others (Chen *et al.*, 2008), introducing some misleading information, such as the failure to detect the gN-3b subgroup (gN-3 detected by the absence of cut by all enzymes corresponds to the gN-3a variant) and some cases belonging to the gN-4b and gN-4c subgroups, although they were not in discussion for primer design and grouping for these subvariants are conceptually more correct for the sequence similarity they display among HCMV strains.

A correct interpretation of gN genotypes by the analysis of restriction fragment length using different sets of primers is reported in Table 3. This

comparative summary should better clarify this point and be considered a reference for future researchers engaged in gN genotyping by PCR-RFLP analysis.

The latest papers reported in Table 1 adopted more sensitive PCR approaches (Pignatelli *et al.* 2001; Rossini *et al.*, 2005; Pignatelli *et al.*, 2006; Puchhammer-Stöckl *et al.*, 2006; Novak *et al.*, 2008; Yan *et al.*, 2008), such as Nested and Touchdown PCRs, cloning approaches (Novak, 2008) using new sets of primers, leading to the most recently optimized multiplex PCR, which uses genotype specific primers (Pignatelli *et al.*, manuscript in preparation).

The most interesting feature of the manuscript by Chen *et al.*, 2008 for future researchers in this field is that the proposed primers (UL73-105725F; UL73-106122R) we also tested in our laboratory with a percentage of unsuccessful gN-PCR of 1.3%, generated very good yields of UL73-gN PCR products, which may be successfully employed in any subsequent genotyping assay (sequencing, RFLP, cloning). In addition, they "seem" to have no mismatch with any gN variants, although this property was presumabily derived only from our previously published results based on the use of a more external upper primer (Forward-U73) (Pignatelli *et al.*, 2001). However, from that moment on the region around the ATG of the UL73 ORF was no longer available due to use of ATG-contiguous primers and the sequencing procedure, which usually fails to give sequence information on the 20-30 nucleotides adjacent to the annealing region.

Primers without any mismatch in the annealing regions are suitable to detect HCMV mixed infections. As a matter of fact, in a mixed infection with one strain predominant over the other(s), a single or double mismatch in the annealing region of the upper primer may select the predominant strain in the first cycles of amplification, due to a lower amplification efficiency. However, after the amplification step on ORF UL73, the appropriate approach to detect the presence of mixed strains in a sample should be chosen. The RFLP method has a limited capacity to detect HCMV co-infections, as reported by Novak *et al.* (2008), who used PCR-cloning to detect multiple strains in specimens. Nevertheless, as reported in Gabrielli *et al.* (2008), we also managed to detect mixed infections by PCR-RFLP, as Puchhammer-

Stöckl *et al.* also did (2006). However, to avoid any further underestimation of multiple infections, we optimized a new highly sensitive multiplex strain-specific gN-PCR (Pignatelli *et al.*, manuscript in preparation). This new assay did not show a high percentage of co-infections either in healthy patients or in congenitally infected newborns, who in all likelihood had been exposed to a single HCMV strain during pregnancy, at variance from the recent study by Novak *et al.*, (2008) who found 93.7% of multiple strains in specimens from 16 healthy urban low-income African American women by PCR-Cloning.

We emphasize that the selection of **certain** populations (e.g. transplant recipients or women with bad sexual habits) may affect the results as strongly as technical approaches, and the PCR-cloning method is not exempt from faults in (over-) estimation of multiple strains in specimens.

CONCLUSIONS

Regardless of the method used, this brief review should help both researchers currently tackling human cytomegalovirus glycoprotein N genotypes and those who will embark on this topic in the next future.

The global picture which emerges from this critical overview may be summarized as follows: multiple methods are currently available to cluster HCMV strains according to their gN genotype; their relative efficiency, sensitivity and specificity may slightly vary and the main factors influencing the results may be: primer design; patient population; viral DNA load; cloning, RFLP analysis and sequencing procedures for genotyping; the potential presence in a given sample of co-infections by multiple HCMV strains. This datum should be taken into account when studies on the clinical relevance of gN genotypes are performed. In any case, if the rarest variants are taken in consideration, such as the gN-3 and gN-2, it would be better not to address any clinical association between these variants and patients' health due to numerical limitations on further statistical analyses. In fact, as reported in a number of studies on different patient populations, which suggested an increased virulence of gN-4 group compared to the gN-1 variant, no hypotheses have been promoted for the low-frequency genotypes

gN-2 and gN-3 (Pignatelli *et al.*, 2003b and 2003c; Rossini *et al.*, 2005; Pignatelli *et al.*, 2006; Gabrielli *et al.*, 2008; Pignatelli *et al.*, submitted for publication).

Some controversial results reported by different groups and discussed in the text warn scientists they should not claim against each other and should have a better insight and criticism in their own studies and choose the appropriate approach to engage in a study on gN genotype distribution (e.g. selection of primers, population, genotyping assay, etc.).

In conclusion, we heartily recommend that scientific teams focus their efforts on the field of HCMV gene polymorphism, choosing the most appropriate technical approach, fulfilling the puzzling picture of viral biology and virulence with a potentially relevant clinical impact, and pooling their results for the purpose of practical application.

REFERENCES:

Beyari M.M., Hodgson T.A., Kondowe W., Molyneux E.M., Scully C., Porter S.R. Teo C.G. (2005). Inter- and Intra- person cytomegalovirus infection in Malawian families. *J. Med. Virol.* **75**, 575-582.

Chen H.P., Lin J.C., Yang S.P., Lan Y.C., Weng W.S., Tsai C.H., Ho D.M., Liu C.Y., Cho W.L., Chan Y.J. (2008). The type-2 variant of human cytomegalovirus glycoprotein N (gN-2) is not the rarest in the Chinese population of Taiwan: influence of primer design. *J. Virol. Meth.* **151**, 161-164.

Gabrielli L., Capretti M.G., Lazzarotto T., Lanari M., Pignatelli S., Dal Monte P., Faldella G., Landini M.P. (2008.) Risk factors of postnatal symptomatic CMV infection in preterm infants. *New Microbiol.* **31**, 131-135.

Ho M. (1990). Epidemiology of cytomegalovirus infections. *Rev. Infect. Dis.* **12**, S701.

Novak Z., Ross S.A., Kumar Patro R., Kumar Pati S., Kumbla R.A., Brice S., Boppana S.B. (2008). Cytomegalovirus strain diversity in seropositive women. *J. Clin. Microb.* **46**, 882-886.

Pignatelli S., Dal Monte P., Landini M.P. (2001). gpUL73(gN) genomic variants of human cytomegalovirus isolates are clustered into four distinct genotypes. *J. Gen. Virol.* **82**, 2777-2784.

Pignatelli S., Dal Monte P., Rossini G., Camozzi D., Toscano V., Conte R., Landini M.P. (2006). Latency-associated human cytomegalovirus (HCMV) gN genotypes in monocytes from healthy blood donors. *Transfusion* **46**, 1754-1762.

Pignatelli S., Dal Monte P., Rossini G., Chou S., Gojobori K., Hanada Guo J.J., Rawlinson W., Britt W., Mach M., Landini M.P. (2003a). Glycoprotein N (gpUL73-gN) genomic variants: identification of a novel subgroup, geographical distribution and evidence of positive selective pressure. *J. Gen Virol.* **84**, 647-655.

Pignatelli S., Dal Monte P., Rossini G., Gatto M.R., Landini M.P. (2003b). Human cytomegalovirus glycoprotein N (gN) genotypes in AIDS patients. *AIDS* **17**, 761-763.

Pignatelli S., Dal Monte P., Rossini G., Landini M.P. (2004). Genetic polymorphisms among human cytomegalovirus (HCMV) wild-type strains. *Rev. Med. Virol.* **14**, 383-410.

Pignatelli S., Dal Monte P., Rossini G., Lazzarotto T., Gatto M.R., Landini M.P. (2003c). Intrauterine cytomegalovirus infection and glycoprotein N (gN) genotypes. *J. Clin. Virol.* **28**, 38-43.

Pignatelli S., Lazzarotto T., Dal Monte P., Gabrielli L., Faldella G., Lanari M. Human cytomegalovirus gN genotypes among congenitally infected newborns: a suggestive prognostic marker for symptoms at birth and outcome? NEJM, *submitted for publication.*

Puchhammer-Stöckl E., Gorzer I., Zoufaly A., Jaksch P., Bauer C.C., Klepetko W. Popow-Kraupp T. (2006). Emergence of multiple cytomegalovirus strains in blood and lung of lung transplant recipients. *Transplantation* **81**, 187-194.

Rossini G., Pignatelli S., Dal Monte P., Camozzi D., Lazzarotto T., Gabrielli L., Gatto M.R., Landini M.P. (2005). Monitoring for human cytomegalovirus infection in solid organ transplant recipients through antigenemia and glycoprotein N (gN) variants: evidence of correlation and potential prognostic value of gN genotypes. *Microbes and Infection* **7**, 890-896.

Sinzger C., Schmidt K., Knapp J., et al. (1999). Modification of human cytomegalovirus tropism trhough propagation in vitro is associated with changes in the viral genome. *J. Gen. Virol.* **80**, 2867-2877.

Yan H., Koyano S., Inami Y., Yamamoto Y., Suzutani T., Mizuguchi M., Ushijima H., Kurane I., Inoue N. (2008). Genetic linkage among human cytomegalovirus glycoprotein N (gN) and gO genes, with evidence for recombination from congenitally and post-natally infected Japanese infants. *J. Gen. Virol.* **89**, 2275-2279.