

Accepted Manuscript

Title: A new perspective of the structural complexity of HCMV-specific T-cell responses (Revision 1)

Author: Andrew Sylwester Kate Z. Nambiar Stefano Caserta
Paul Klenerman Louis J. Picker Florian Kern



PII: S0047-6374(16)30023-9
DOI: <http://dx.doi.org/doi:10.1016/j.mad.2016.03.002>
Reference: MAD 10825

To appear in: *Mechanisms of Ageing and Development*

Received date: 6-11-2015
Revised date: 17-2-2016
Accepted date: 3-3-2016

Please cite this article as: Sylwester, Andrew, Nambiar, Kate Z., Caserta, Stefano, Klenerman, Paul, Picker, Louis J., Kern, Florian, A new perspective of the structural complexity of HCMV-specific T-cell responses (Revision 1). *Mechanisms of Ageing and Development* <http://dx.doi.org/10.1016/j.mad.2016.03.002>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

A new perspective of the structural complexity of HCMV-specific T-cell responses (Revision 1)

Short title: A blueprint of T-cell immunity to HCMV

Andrew Sylwester^a, Kate Z. Nambiar^b, Stefano Caserta^b, Paul Klenerman^c, Louis J. Picker^a, Prof. Florian Kern^{b*} f.kern@bsms.ac.uk

^aVaccine & Gene Therapy Institute, Oregon Health & Science University West Campus, Beaverton, OR 97006, USA.

^bDivision of Medicine, Pathogen Host Interaction (PHI), Brighton and Sussex Medical School, Brighton, BN1 9PX, United Kingdom.

^cPeter Medawar Building for Pathogen Research, University of Oxford, Oxford, OX1 3SY, United Kingdom.

***Corresponding author at:** Brighton & Sussex Medical School, Medical Research Building, Biology Road, Brighton, BN1 9PX, United Kingdom, Tel.: 01273 877671.

Highlights

- The T-cell response to Cytomegalovirus is complex and generally consists in multiple, finely graded responses with unpredictable target hierarchies
- Some 10% of responses are statistically speaking outliers by size
- A broad base of small and medium size responses provides robustness and flexibility to the recognition of epitopes and forms the underbelly of large, expanded responses
- Meaningful monitoring the T-cell response to Cytomegalovirus needs to address the complexity of this response

Abstract

Background: In studies exploring the effects of HCMV infection on immune system ageing ('immunosenescence'), after organ transplantation or in other settings, HCMV-specific T-cell responses are often assessed with respect to purportedly 'immunodominant' protein subunits. However, the response structure in terms of recognized antigens and response hierarchies (architecture) is not well understood and actual correlates of immune protection are not known.

Methods: We explored the distribution of T-cell response sizes and dominance hierarchies as well as response breadth in 33 HCMV responders with respect to >200 HCMV proteins.

Results: At the individual responder level HCMV-specific T-cell responses were generally arranged in clear dominance hierarchies; interestingly, the number of proteins recognized by an individual correlated closely with the size of their biggest response. Target-specificity varied considerably between donors and across hierarchy levels with the presence, size, and hierarchy position of responses to purportedly 'immunodominant' targets being unpredictable.

Conclusions: Predicting protective immunity based on isolated HCMV subunit-specific T-cell response is questionable in light of the complex architecture of this response. Our findings have important implications for T-cell monitoring, intervention strategies, as well as the application animal models to the understanding of human infection. (194 words)

Keywords: Cytomegalovirus; T-cells; host response; T-cell memory-inflation

1. Introduction

Large human CMV (HCMV)-specific T-cell responses have been described in numerous published reports, particularly in the transplantation setting and older life (Kern et al., 2002; Ouyang et al., 2003; Ouyang et al., 2004; Sester et al., 2002; Sylwester et al., 2005; Vescovini et al., 2007). These conspicuous responses are often thought to spearhead T-cell immunity to HCMV and many researchers believe they grow so large because this is a requirement for protection. Another reason, however, might be the life-long persistence of latent HCMV infection with frequent episodes of reactivation that gradually but steadily increase response size (Fletcher et al., 2005; Moss, 2010). Expanded CMV-specific T-cell populations in humans have been modeled in the mouse using murine CMV (MCMV) infection, where longitudinal studies have observed the phenomenon termed memory ‘inflation’ (Holtappels et al., 2000; Lang et al., 2009; Lang and Nikolich-Zugich, 2011; O’Hara et al., 2012; Snyder et al., 2011). ‘Inflation’ in this context, although not formally defined, refers to (delayed) expansion and long-term maintenance of specific T-cell populations after initial infection. Interestingly, however, removal of a highly dominant MCMV-specific T-cell response in the BALB/c model has shown that these responses are not necessarily required to control infection but that smaller responses may take on this role even without expansion (Holtappels et al., 2008). It still remains unclear to what extent these observations are applicable to humans.

Since in humans the duration of HCMV infection is usually not known, monitoring the development of T-cell responses beginning at the time of infection is not possible except in the not entirely compatible setting of organ transplantation (Hertoghs et al., 2010). Numerous human studies have used cross-sectional designs using different age groups to investigate changes of HCMV specific immunity over time, however, this approach cannot easily distinguish the effects

of immune ageing from those of the duration of HCMV infection (Alp et al., 1991; Borysiewicz et al., 1988; Fletcher et al., 2005; Gibson et al., 2004; Khan et al., 2004; Lachmann et al., 2012; Ouyang et al., 2003; Ouyang et al., 2004; Weekes et al., 1999). Interestingly, very large HCMV-specific responses were described both in older and young populations (Komatsu et al., 2006; Komatsu et al., 2003), indicating that response inflation is not a direct function of chronological age.

Most human studies have investigated CD8 T-cells specific to viral subunits, such as the HCMV UL83 ('pp65') and UL123 ('IE1') proteins (Bunde et al., 2005; Kern et al., 1999; Lachmann et al., 2012; Moss and Khan, 2004). Often select UL83-derived peptides were used, such as *NLVPMVATV*, presented by HLA-A*0201 or *TPRVTTTGAM*, presented by HLA-B7*0702 (Gibson et al., 2004; Khan et al., 2004; Ouyang et al., 2003; Ouyang et al., 2004). Both peptides frequently induce robust, sometimes exceptionally large responses, and the high frequencies of HLA-A*0201 and HLA-B7*0702 are very convenient for studies in white American and European populations (Cao et al., 2001).

Despite many advances, a breakthrough in terms of HCMV-specific T-cell response monitoring beyond the detection of post-transplantation immune reconstitution has not been achieved to date (Borchers et al., 2014). While the technological basis for monitoring HCMV-specific T-cells has existed since roughly the mid 1990s (Altman et al., 1996; Picker et al., 1995; Waldrop et al., 1997), correlates of protection remain undefined. This is owed largely to the fact that we lack a clear understanding of the structural composition of this response.

A study published by us in 2005 showed a surprisingly wide spectrum of HCMV-specific T-cell target selection across a population sample of 33 CMV-infected healthy donors; CD8 and CD4 T-cell responses to 213 HCMV proteins were tested. This study provided novel, fundamentally

important information such as a complete list of proteins recognized by CD8 and CD4 T-cells, the most frequently recognized ones, and the most dominant ones (by summated response size) (Sylwester et al., 2005). However, this dataset contained a wealth of additional information that was not extracted at the time of publication. In particular it was uniquely suited for studying the architecture of the HCMV-specific T-cell response in terms of the relationships between response numbers, sizes, hierarchy levels, and target specificities, both in individuals and across the population. Here we present a new, differential analysis of this dataset devised to improve our understanding of the blueprint of this response.

Our new analysis shows that in the vast majority of individuals multiple T-cell responses to HCMV are organized in finely graded dominance hierarchies with highly variable target specificity at each hierarchy level. In addition there is a striking positive association between the number of recognized target proteins in each individual and the size of their biggest response, associating the presence of many responses with the presence of large responses. Despite overall structural complexity, and hugely different levels of response size, the basic response architecture was surprisingly similar in different individuals. This novel insight should inform the measurement and interpretation HCMV-specific T-cell responses in research and clinical settings and it should be interesting to establish if it holds true for T-cell responses to other types of antigens.

2. Materials and Methods

2.1. Participants

The original study was carried out in Portland/Oregon, USA and approved by the Institutional Review Board of Oregon Health and Science University. The study was conducted according to

the Declaration of Helsinki; all participants gave written informed consent. It included 33 CMV⁺ healthy individuals of various ethnic backgrounds (white Caucasian, black African and Asian), between 19 and 53 years of age. Additional relevant details are published (Sylwester et al., 2005).

2.2. Blood samples and activation assays.

Blood samples were obtained by apheresis and anticoagulated with sodium-heparin. PBMC were prepared according to standard protocols and cryopreserved as reported in detail elsewhere (Sylwester et al., 2005). Briefly, thawed or fresh PBMC were incubated with sets of CMV protein-spanning peptide pools for 1 hour before addition of Brefeldin A ('BFA', Sigma, St. Louis, MO), then incubated at 37C (standard incubator) for additional 5 hours (final volume of 1 ml) in the presence of co-stimulatory antibodies, CD28 and CD49d (BD Biosciences). Following incubation, cells were harvested in PBS and surface stained, fixed and lysed with FACS Lysing solution (BD Biosciences), permeabilized with BD Permeabilizing solution (BD Biosciences), washed and stained with fluorochoime-labeled antibodies for intracellular IFN- γ and/or other activation markers.

2.3. Protein selection and peptide pools

A total of 191 ORF sequences were sourced from the AD 169 laboratory strain sequence; an additional 22 sequences were added from the Toledo and Towne strains. Sequence data were obtained from GenBank/ EMBL/DDBJ (protein database) site and from the SWISS-PROT database. Consecutive 15mer peptides, overlapping by 10 or 11 amino acids, were synthesized for each protein sequence using solid-phase peptide synthesis methods, employing an Fmoc

synthesis strategy (Mimotopes Pty., Ltd.). All peptides had free acid C-termini and free amine N-termini. Peptides associated with each CMV protein were pooled and stored in freeze-dried aliquots at -80°C . Prior to usage, freeze-dried aliquots were dissolved in DMSO. Final peptide concentrations were $2 \mu\text{g/ml}$. Additional details were published previously (Sylwester et al., 2005). The assignment of protein kinetic classes was adopted from the historical study (Sylwester et al., 2005) and indicates the time after viral reactivation at which certain proteins are expressed, including immediate early (IE), early (E), early-late (EL), late (L), and non-classified (NC) (Chambers et al., 1999).

2.4. Flow-cytometric analysis

Data files were analyzed using the PAINT-A-GATE Plus software program (BD Biosciences), as described previously (Sylwester et al., 2005). Briefly, T-cell populations (CD4+ or CD8+) were gated from lymphocytes defined by scatter parameters, and viewed against IFN- γ expression in 2D dotplots in order to identify positive events. Responses were expressed in percent of the respective reference population. Background determined from unstimulated samples (no peptide added) was subtracted. To rule out any spurious responses, only net CD4 responses $>0.06\%$ and net CD8 responses $>0.08\%$ were counted as positive; the algorithm used to establish these thresholds is described in detail elsewhere (Sylwester et al., 2005).

2.5. Statistical analysis

Tests for the normality of distribution, parametric and non-parametric tests, and multiple regression analysis were performed using SPSS 20.0 (IBM, Portsmouth, UK). T-cell response standardization was performed on log-transformed data. Z-scores, also known as standard scores

(z) were calculated as $z=(X-\mu)/\sigma$, where X is a measured value, μ is the arithmetic mean of all measured values, and σ is the standard deviation. Definitions used for statistical outliers were as follows: i) values greater than the upper quartile + 1.5 x the inter-quartile range (non-parametric) and ii) the arithmetic mean+2 standard deviations (SD) (parametric); when using parametric definitions T-cell percentages were log-transformed.

3. Results

3.1. CMV protein immunodominance is unpredictable in healthy individuals

To begin, we sought to identify immunodominant HCMV antigens that consistently generated large or very large responses in healthy HCMV⁺ people. The distribution of CD8 and CD4 T-cell responses to 213 HCMV proteins in 33 donors was visualized in a heat map-like format in order to provide an overview of response numbers, sizes, and target selection (**Supplementary Figures 1 and 2**). All responses classified positive as per the original report (Sylwester et al., 2005) were used in this and subsequent analyses. The most dominant antigens varied considerably between donors and in regards to kinetic class. Responses ranged from barely detectable to double figure percentages, with the full spectrum of response sizes often present in the same individuals. There were no significant gender differences with respect to the number of recognized target proteins or summated responses.

Most responders recognized 5-10 CMV proteins within each T-cell compartment. Mapping all responses down to the level of single peptides was beyond the scope of the study, however, in light of our previous observation that protein-specific responses are usually dominated by single epitopes, this would suggest that responders recognize 5-10 major epitopes

(Kirchner et al., 2008). In order to address the size of these responses systematically we first compared T-cell responses to the same protein in different responders, separately for CD8 and CD4 T-cells.

For CD8 T-cells this analysis was limited to 33 proteins recognized by at least four responders in order to allow the application of statistical outlier definitions as a working model for ‘inflated’ responses (i.e. identifying expansions outside the expected range of responses). In this context, "inflation" is taken to represent those responses that show an extreme phenotype amongst the population. In studies of mice, a broader definition may be used based on longitudinal measurements post-infection, assessment of surface marker expression and tissue distribution, which might include a number of different patterns of responsiveness and a wide range of frequencies (Munks et al., 2006a).

CD8 T-cell responses to the same protein varied widely in size between donors and the threshold for calling a large response an outlier (i.e. ‘inflated’) ranged from 0.33 to 10.57% of CD8 T-cells, depending on the protein (**Fig. 1**). Such outliers accounted for 10.4 % of all responses (**Table 1**). We then reversed this approach by analyzing the responses to the proteins recognized in each individual (same responder, different proteins). Here, the threshold for calling large responses outliers ranged from 0.52 to 4.82% of CD8 T-cells, depending on the responder (28 individuals recognizing four or more proteins were included). The proportion of outliers was in the same order of magnitude as in the first analysis (**Table 1**), but more importantly, the proteins recognized by the largest T-cell responses in each responder varied considerably. This confirmed that HCMV protein immunodominance is not predictable in healthy humans, and agrees with our more limited, earlier finding that UL83 and UL123 epitope immunodominance cannot be predicted even if the individual HLA-type is considered (Kirchner et al., 2008). The

analysis of CD4 T-cell responses provided very similar results (**Table 1**). The threshold for calling large CD4 T-cell responses outliers ranged from 0.22 to 13.93% for different proteins (across all responders), and from 0.41 to 4.31% for different responders (across the proteins recognized by each). Additional details are provided in **Table 2**.

3.2. Different individuals display similar response distributions but different response levels

In the same individuals, the maximum response sizes of CD8 and CD4 T-cells were often very different, up to an order of magnitude (**Fig. 2 A**). The response levels in each T-cell compartment were highly variable across the population. We, therefore, standardized response size in order to compare response distributions. Z-scores were calculated for all responses in all individuals (i.e. for each response its deviation from the mean was expressed in multiples of the SD). Interestingly, increasing response counts were significantly associated with higher Z-scores and increasing numbers of outliers (i.e. Z-scores >2) (**Fig. 2 B**). A more detailed view of Z-scores by protein target is provided in **Fig. 3**. To visualize an ‘overall response distribution’ (i.e. at the population level) we binned responses of similar Z-score, revealing a clear predominance of small responses and a sharp decline of response frequencies with increasing response size in both CD4 and CD8 T-cells (**Fig. 4A**).

To further explore an association between the number of proteins recognized in each individual and absolute response size, we considered summated and average responses (sum of all protein-specific responses in each individual) as well as maximum responses. Summated response correlated significantly with the number of recognized proteins, however, average responses were unexpectedly similar across the donor population (**Fig. 4B**). At the same time

there was a striking and significant correlation between the number of recognized proteins in each participant and the size of their maximum response (**Fig. 4C**).

Of note, the historical study identified a subset of 15 ‘most dominant’ proteins for CD8 T-cells and 6 ‘most dominant’ proteins for CD4 T-cells, the summed response to which correlated highly with the respective summed response to all proteins. Interestingly, the association between the number of recognized proteins and the size of the maximum response was preserved for these subsets of proteins, albeit somewhat weaker ($r_s=0.554$, $p=0.001$ for CD8 T-cells, 15 proteins; $r_s=0.508$, $p=0.003$ for CD4 T-cells, 6 proteins).

3.3. HLA-type is significantly related to response levels

Because of the great differences in response levels between individuals, we explored the effect of HLA type on T-cell response size using the available typing data (2-letter code). The number of observations was relatively small for this type of however, associations between the presence of the most frequent HLA molecules in the cohort and summed response size were explored. Logistic regression using stepwise inclusion of HLA types with at least two respondents revealed that HLA-B7 was associated with the biggest summed responses ($p<0.001$) (**Supplementary Fig. 3**). While these results should be considered preliminary because of the small response size, the findings are in agreement with our previous observation that individuals who were HLA-B7, HLA-B8, or HLA-A1 positive had significantly larger responses to UL123 than individuals with none of these HLA allomorphs (Kern et al., 2000). In addition, individuals having the following phenotypes showed significantly greater maximum responses than the rest, HLA-B7 ($p=0.002$), HLA-B56 ($p=0.002$), HLA-A1 ($p=0.006$), and HLA-A23 ($p=0.002$) (not

shown). **Supplementary Table 1** provides an overview of summed response sizes by donor and all HLA-allomorphs present in at least two people.

4. Discussion

Our analysis integrates individual population-based profiles of response counts, sizes, hierarchies, and target specificities, to create a uniquely comprehensive picture of the structural ‘scaffold’ behind HCMV-specific T-cell immunity. Statistical outlier definitions were used as an objective way to identify ‘inflated responses’ whereas the normalization of response sizes allowed us to compare their distribution across the population despite different individual response levels.

With many responses per individual, generally organized in finely graded hierarchies, the response to CMV in humans appeared clearly less polarized than shown in mouse models (Munks et al., 2006a; Northfield et al., 2005; Snyder et al., 2011). This is probably due to the outbred human MHC allowing the presentation of many more peptides resulting in different degrees of expansion in parallel and more than one inflated response. Response hierarchies had a similar architecture throughout the responder population, the smallest response being generally just above the detection threshold, irrespective of the size of the maximum response. Response size standardization (Z-scores) allowed us to compare response size distributions across the population, irrespective of individual response level. The observed distribution of response sizes corroborated the finding that, generally, the biggest responses are underpinned by numerous very small, small, and medium size responses rather than existing as isolated peaks. This fitted well with the observation that the size of the maximum response could be different by an order of magnitude between some but the average response size was almost the same in all individuals

(with many small responses reducing the effect of very large responses on the average). The fact that the size of the maximum response correlated with the number of protein-specific responses in each individual also fits well with this explanation. Assuming that HCMV-specific T-cell responses expand as a function of time after infection, this might suggest that very small responses, initially below the detection threshold, become detectable over time. However, this would have to be addressed in a separate, longitudinal study.

The overall response architecture featuring a broad base of small and medium responses might provide robustness and flexibility to the recognition of epitopes (Naumov et al., 2003) and is more likely to be a guarantor of sustained, protective immunity than the presence of isolated inflated responses. It would be very interesting to study if a similar response architecture is present in T-cell responses to other infectious agents, tumour antigens, or self antigens, and whether there is an association between certain structural features and protective immunity.

Two reports by Ouyang et al. from 2003 and 2004 described large, isolated, epitope-specific responses in healthy older people and argued that such responses might be critical for controlling HCMV and continue to grow bigger as a result of diminished response ‘quality’ (suggesting more of these cells are required to do the same job) (Ouyang et al., 2003; Ouyang et al., 2004). By contrast, our analysis shows that individuals with very large responses usually have many additional responses, underscoring the advantage of studying multiple specificities simultaneously. The individuals studied in the aforementioned reports were on average a lot older than the oldest individuals in our study, leaving the possibility that significant changes might affect the TCR repertoire during older age. Nevertheless, Ouyang et al. did not look for other responses and our own current research suggests that many responses can be found even in very old people (manuscript in preparation). It is interesting in this context that experiments in

BALB/c mice showed that the most immune-dominant responses are not required for protection (Holtappels et al., 2008).

Finally, our preliminary data with respect to MHC types and response size argue that there is a contribution of certain HLA-allomorphs to determining the magnitude of responses, confirming previous observations limited to UL83 and UL123 (Kern et al., 2000). There is limited scope for addressing this question in inbred mouse models. On the other hand, it is known that genetic factors other than MHC type contribute significantly to T-cell response size in MCMV models. A comparison between C57BL/6 mice and two other strains, 129/SvJ and BALB.B (both share the haplotype, H-2b, with the former but differ in their non-MHC genes) found that all strains responded to the same set of epitopes, but response sizes and hierarchies differed significantly between them (Munks et al., 2006b). Such an experiment would be difficult to conduct in an outbred human or animal population and underlines a strength of inbred mouse models.

In summary, our data demonstrates that HCMV-specific T-cell immunity cannot be assessed representatively by testing a few, arbitrarily selected, epitope-specific responses, for example by using single antigenic peptides for stimulation or staining with MHC-multimers based on the presence of certain, common HLA-types. A meaningful quantitative comparison of HCMV-specific T-cell responses between different individuals (even if sharing one or more HLA alleles) would clearly have to cover many target proteins. More diverse, outbred animal models might better reflect human diversity in future, and the measurement of broader responses within such models will help dissect the features underlying T-cell response evolution.

Funding: This work was supported by National Institutes of Health grants R01-AI47606, R01-AI21640, MO1-RR000334, and P51-RR00163 (JLP).

Conflict of interest/competing financial interest statement

Andrew Sylwester: no COI; Kate Nambiar: no COI; Stefano Caserta: no COI ; David Thomas: no COI; Paul Klenerman: no COI; Louis Picker: no COI; Florian Kern: FK is co-owner of a patent describing the use of overlapping peptide pools for T-cell stimulation (WO 01/63286 A2) as used in this publication. FK is also works as head of product development for JPT Peptide Technologies, Berlin, Germany.

References

- Alp, N.J., Allport, T.D., Van Zanten, J., Rodgers, B., Sissons, J.G., Borysiewicz, L.K., 1991. Fine specificity of cellular immune responses in humans to human cytomegalovirus immediate-early 1 protein. *J Virol* 65, 4812-4820.
- Altman, J.D., Moss, P.A.H., Goulder, P.J.R., Barouch, D.H., McHeyzer-Williams, M.G., Bell, J.I., McMichael, A.J., Davis, M.M., 1996. Phenotypic analysis of antigen-specific T lymphocytes [published erratum appears in *Science* 1998 Jun 19;280(5371):1821]. *Science* 274, 94-96.
- Borchers, S., Ogonek, J., Varanasi, P.R., Tischer, S., Bremm, M., Eiz-Vesper, B., Koehl, U., Weissinger, E.M., 2014. Multimer monitoring of CMV-specific T cells in research and in clinical applications. *Diagnostic microbiology and infectious disease* 78, 201-212.
- Borysiewicz, L.K., Hickling, J.K., Graham, S., Sinclair, J., Cranage, M.P., Smith, G.L., Sissons, J.G., 1988. Human cytomegalovirus-specific cytotoxic T cells. Relative frequency of stage-specific CTL recognizing the 72-kD immediate early protein and glycoprotein B expressed by recombinant vaccinia viruses. *J Exp Med* 168, 919-931.
- Bunde, T., Kirchner, A., Hoffmeister, B., Habedank, D., Hetzer, R., Cherepnev, G., Proesch, S., Reinke, P., Volk, H.D., Lehmkuhl, H., Kern, F., 2005. Protection from cytomegalovirus after transplantation is correlated with immediate early 1-specific CD8 T cells. *J Exp Med* 201, 1031-1036.
- Cao, K., Hollenbach, J., Shi, X., Shi, W., Chopek, M., Fernandez-Vina, M.A., 2001. Analysis of the frequencies of HLA-A, B, and C alleles and haplotypes in the five major ethnic groups of the United States reveals high levels of diversity in these loci and contrasting distribution patterns in these populations. *Hum Immunol* 62, 1009-1030.
- Chambers, J., Angulo, A., Amaratunga, D., Guo, H., Jiang, Y., Wan, J.S., Bittner, A., Frueh, K., Jackson, M.R., Peterson, P.A., Erlander, M.G., Ghazal, P., 1999. DNA microarrays of the complex human

cytomegalovirus genome: profiling kinetic class with drug sensitivity of viral gene expression. *J Virol* 73, 5757-5766.

Fletcher, J.M., Vukmanovic-Stejic, M., Dunne, P.J., Birch, K.E., Cook, J.E., Jackson, S.E., Salmon, M., Rustin, M.H., Akbar, A.N., 2005. Cytomegalovirus-specific CD4⁺ T cells in healthy carriers are continuously driven to replicative exhaustion. *J Immunol* 175, 8218-8225.

Gibson, L., Piccinini, G., Lilleri, D., Revello, M.G., Wang, Z., Markel, S., Diamond, D.J., Luzuriaga, K., 2004. Human cytomegalovirus proteins pp65 and immediate early protein 1 are common targets for CD8⁺ T cell responses in children with congenital or postnatal human cytomegalovirus infection. *Journal of immunology* 172, 2256-2264.

Hertoghs, K.M., Moerland, P.D., van Stijn, A., Remmerswaal, E.B., Yong, S.L., van de Berg, P.J., van Ham, S.M., Baas, F., ten Berge, I.J., van Lier, R.A., 2010. Molecular profiling of cytomegalovirus-induced human CD8⁺ T cell differentiation. *J Clin Invest* 120, 4077-4090.

Holtappels, R., Pahl-Seibert, M.F., Thomas, D., Reddehase, M.J., 2000. Enrichment of immediate-early 1 (m123/pp89) peptide-specific CD8 T cells in a pulmonary CD62L(lo) memory-effector cell pool during latent murine cytomegalovirus infection of the lungs. *J Virol* 74, 11495-11503.

Holtappels, R., Simon, C.O., Munks, M.W., Thomas, D., Deegen, P., Kuhnappel, B., Daubner, T., Emde, S.F., Podlech, J., Grzimek, N.K., Oehrlein-Karpi, S.A., Hill, A.B., Reddehase, M.J., 2008. Subdominant CD8 T-cell epitopes account for protection against cytomegalovirus independent of immunodomination. *J Virol* 82, 5781-5796.

Kern, F., Bunde, T., Faulhaber, N., Kiecker, F., Khatamzas, E., Rudawski, I.M., Pruss, A., Gratama, J.W., Volkmer-Engert, R., Ewert, R., Reinke, P., Volk, H.D., Picker, L.J., 2002. Cytomegalovirus (CMV) phosphoprotein 65 makes a large contribution to shaping the T cell repertoire in CMV-exposed individuals. *J Infect Dis* 185, 1709-1716.

- Kern, F., Faulhaber, N., Frömmel, C., Khatamzas, E., Prösch, S., Schönemann, C., Kretzschmar, I., Volkmer-Engert, R., Volk, H.-D., Reinke, P., 2000. Analysis of CD8 T cell reactivity to cytomegalovirus using protein-spanning pools of overlapping pentadecapeptides. *Eur J Immunol* 30, 1676-1682.
- Kern, F., Surel, I.P., Faulhaber, N., Frommel, C., Schneider-Mergener, J., Schonemann, C., Reinke, P., Volk, H.D., 1999. Target structures of the CD8(+)-T-cell response to human cytomegalovirus: the 72-kilodalton major immediate-early protein revisited. *J Virol* 73, 8179-8184.
- Khan, N., Hislop, A., Gudgeon, N., Cobbold, M., Khanna, R., Nayak, L., Rickinson, A.B., Moss, P.A., 2004. Herpesvirus-specific CD8 T cell immunity in old age: cytomegalovirus impairs the response to a coresident EBV infection. *J Immunol* 173, 7481-7489.
- Kirchner, A., Hoffmeister, B., Cherepnev, G.G., Fuhrmann, S., Streitz, M., Lachmann, R., Bunde, T., Meij, P., Schonemann, C., Hetzer, R., Lehmkuhl, H.B., Volkmer-Engert, R., Volk, H.D., Gratama, J.W., Kern, F., 2008. Dissection of the CMV specific T-cell response is required for optimized cardiac transplant monitoring. *J Med Virol* 80, 1604-1614.
- Komatsu, H., Inui, A., Sogo, T., Fujisawa, T., Nagasaka, H., Nonoyama, S., Sierro, S., Northfield, J., Lucas, M., Vargas, A., Klenerman, P., 2006. Large scale analysis of pediatric antiviral CD8+ T cell populations reveals sustained, functional and mature responses. *Immun Ageing* 3, 11.
- Komatsu, H., Sierro, S., A, V.C., Klenerman, P., 2003. Population analysis of antiviral T cell responses using MHC class I-peptide tetramers. *Clin Exp Immunol* 134, 9-12.
- Lachmann, R., Bajwa, M., Vita, S., Smith, H., Cheek, E., Akbar, A., Kern, F., 2012. Polyfunctional T cells accumulate in large human cytomegalovirus-specific T cell responses. *J Virol* 86, 1001-1009.
- Lang, A., Brien, J.D., Nikolich-Zugich, J., 2009. Inflation and long-term maintenance of CD8 T cells responding to a latent herpesvirus depend upon establishment of latency and presence of viral antigens. *Journal of immunology* 183, 8077-8087.

- Lang, A., Nikolich-Zugich, J., 2011. Functional CD8 T cell memory responding to persistent latent infection is maintained for life. *Journal of immunology* 187, 3759-3768.
- Moss, P., 2010. The emerging role of cytomegalovirus in driving immune senescence: a novel therapeutic opportunity for improving health in the elderly. *Curr Opin Immunol* 22, 529-534.
- Moss, P., Khan, N., 2004. CD8(+) T-cell immunity to cytomegalovirus. *Hum Immunol* 65, 456-464.
- Munks, M.W., Cho, K.S., Pinto, A.K., Sierro, S., Klenerman, P., Hill, A.B., 2006a. Four distinct patterns of memory CD8 T cell responses to chronic murine cytomegalovirus infection. *Journal of immunology* 177, 450-458.
- Munks, M.W., Gold, M.C., Zajac, A.L., Doom, C.M., Morello, C.S., Spector, D.H., Hill, A.B., 2006b. Genome-wide analysis reveals a highly diverse CD8 T cell response to murine cytomegalovirus. *Journal of immunology* 176, 3760-3766.
- Naumov, Y.N., Naumova, E.N., Hogan, K.T., Selin, L.K., Gorski, J., 2003. A fractal clonotype distribution in the CD8+ memory T cell repertoire could optimize potential for immune responses. *J Immunol* 170, 3994-4001.
- Northfield, J., Lucas, M., Jones, H., Young, N.T., Klenerman, P., 2005. Does memory improve with age? CD85j (ILT-2/LIR-1) expression on CD8 T cells correlates with 'memory inflation' in human cytomegalovirus infection. *Immunol Cell Biol* 83, 182-188.
- O'Hara, G.A., Welten, S.P., Klenerman, P., Arens, R., 2012. Memory T cell inflation: understanding cause and effect. *Trends in immunology* 33, 84-90.
- Ouyang, Q., Wagner, W.M., Wikby, A., Walter, S., Aubert, G., Dodi, A.I., Travers, P., Pawelec, G., 2003. Large numbers of dysfunctional CD8+ T lymphocytes bearing receptors for a single dominant CMV epitope in the very old. *J Clin Immunol* 23, 247-257.

- Ouyang, Q., Wagner, W.M., Zheng, W., Wikby, A., Remarque, E.J., Pawelec, G., 2004. Dysfunctional CMV-specific CD8(+) T cells accumulate in the elderly. *Exp Gerontol* 39, 607-613.
- Picker, L.J., Singh, M.K., Zdraveski, Z., Treer, J.R., Waldrop, S.L., Bergstresser, P.R., Maino, V.C., 1995. Direct demonstration of cytokine synthesis heterogeneity among human memory/effector T cells by flow cytometry. *Blood* 86, 1408-1419.
- Sester, M., Sester, U., Gartner, B., Kubuschok, B., Girndt, M., Meyerhans, A., Kohler, H., 2002. Sustained high frequencies of specific CD4 T cells restricted to a single persistent virus. *J Virol* 76, 3748-3755.
- Snyder, C.M., Cho, K.S., Bonnett, E.L., Allan, J.E., Hill, A.B., 2011. Sustained CD8+ T cell memory inflation after infection with a single-cycle cytomegalovirus. *PLoS pathogens* 7, e1002295.
- Sylwester, A.W., Mitchell, B.L., Edgar, J.B., Taormina, C., Pelte, C., Ruchti, F., Sleath, P.R., Grabstein, K.H., Hosken, N.A., Kern, F., Nelson, J.A., Picker, L.J., 2005. Broadly targeted human cytomegalovirus-specific CD4+ and CD8+ T cells dominate the memory compartments of exposed subjects. *J Exp Med* 202, 673-685.
- Vescovini, R., Biasini, C., Fagnoni, F.F., Telera, A.R., Zanlari, L., Pedrazzoni, M., Bucci, L., Monti, D., Medici, M.C., Chezzi, C., Franceschi, C., Sansoni, P., 2007. Massive load of functional effector CD4+ and CD8+ T cells against cytomegalovirus in very old subjects. *J Immunol* 179, 4283-4291.
- Waldrop, S.L., Pitcher, C.J., Peterson, D.M., Maino, V.C., Picker, L.J., 1997. Determination of antigen-specific memory/effector CD4+ T cell frequencies by flow cytometry: evidence for a novel, antigen-specific homeostatic mechanism in HIV-associated immunodeficiency. *J Clin Invest* 99, 1739-1750.
- Weekes, M.P., Wills, M.R., Mynard, K., Carmichael, A.J., Sissons, J.G., 1999. The memory cytotoxic T-lymphocyte (CTL) response to human cytomegalovirus infection contains individual peptide-specific CTL clones that have undergone extensive expansion in vivo. *J Virol* 73, 2099-2108.

Figure Captions

Fig. 1: CD8 T-cell immunity to four exemplar proteins shows multiple ‘inflated’ responses

(statistical outliers) across the population. Large CD8 T-cell expansions are found with respect to most CMV target proteins. Dotted horizontal lines indicate the upper limit above which responses are considered outliers (upper quartile + 1.5 x inter quartile range, non-parametric definition). Note the differences in y-axis scaling in each diagram. IFN- γ responses are shown. The examples illustrate considerable differences in response ‘levels’ in regards to different proteins and different donors. Donors are shown as ‘Dx’ where x identifies identical donors in all figures.

Fig. 2: The repertoire of targeted proteins in different individuals is highly variable.

(A) CD8 and CD4 T-cell responses in two representative donors, D25 (donor 25) and D29 (donor 29) are shown. Multiple target proteins are recognized at different dominance hierarchy levels. Note that in D25 the CD8 response is small compared to the CD4 response but the opposite is true in D29. Large differences in response size between the two subsets were frequently observed. Columns show non log-transformed background-subtracted responses. Donors are shown as ‘Dx’ where x identifies identical donors in all figures. (B) Following standardization of response sizes using the Z-score calculated as $Z\text{-score } (z) = (X - \mu) / \sigma$, where X is a measured value, μ is the arithmetic mean of all measured values, and σ is the standard deviation. The maximum Z-score (‘Max Z-score’) for each individual was plotted against their corresponding response count. Values outside the mean+2 standard deviations (SD) were defined as ‘inflated responses’ or outliers with respect to individual response levels according to our definition. The highly significant correlation between these parameters confirms that maximum deviation from the

mean is a function of response breadth, i.e. the number of recognized HCMV proteins in a given individual, for both CD8 and CD4 T-cells. Lines of best fit (linear) are shown in (B) and (C) and the Pearson correlation coefficient (r) is provided (normalized data). The significance threshold was set at $p < 0.05$.

Fig. 3: The bigger the response count the higher is the frequency of responses representing statistical outliers. Response counts were plotted against the Z-scores of log-transformed CD4 and CD8 T-cell responses in each individual. A positive correlation between sample size and number of outliers is expected if the response distribution is based on a normal distribution. Protein kinetic classes are color-coded (IE=Immediate Early, E=Early, E-L=Early-Late, L=Late, NC=not classified). Proteins of Immediate Early and Late kinetic classes appear to be the preferred targets of CD8 T-cell response outliers, Early, Early-Late and Late proteins of CD4 T-cell response outliers. HCMV protein kinetic classes were assigned as previously published. Donors are shown as 'Dx' where x identifies identical donors in all figures.

Fig. 4: Maximum response size is related to the number of recognized proteins.

(A) Response sizes across the whole population sample (log-transformed percentages) were standardized using the Z-score $(z) = (X - \mu) / \sigma$, where X is a measured value, μ is the arithmetic mean of all measured values, and σ is the standard deviation. Z-scores were categorized (48 bins) to facilitate the visualization of response size distributions. Diagrams show a large predominance of small and medium sized responses (smaller Z-scores), suggesting these responses constitute the backbone of HCMV-specific T-cell immunity. A total of 377 Z-scores

were included for CD8, and 499 Z-scores for CD4 responses. (B) Both CD8 (left) and CD4 (right) response sums (in % of the reference subset, filled circles) correlate significantly with the number of responses in each individual ('response count'). However, the average response size (empty circles) in each compartment (all proteins, all donors) was very similar across individuals (compare Fig. 3B and Fig. 4). (C) The size of the biggest CD8 (left) and CD4 (right) response in each individual ('Max. response') also correlated significantly with the response count. Lines of best fit (linear) are shown in (B) and (C) and the Spearman-Rank correlation coefficient (r_s) is provided. The significance threshold was set at $p < 0.05$.

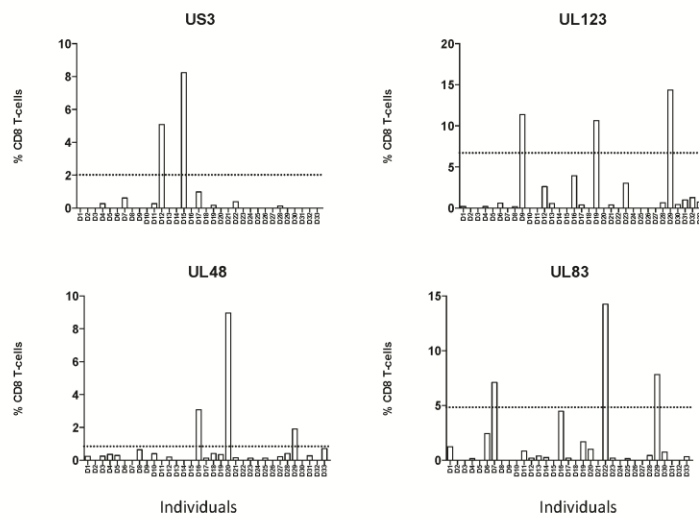


Fig. 1

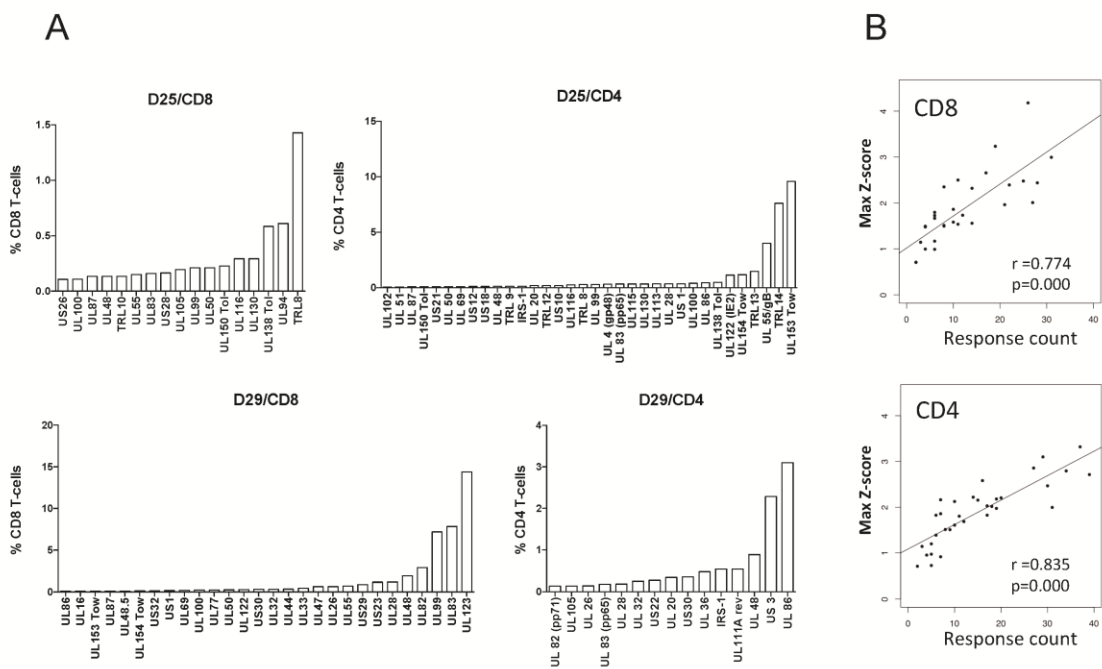


Fig. 2



Fig. 3

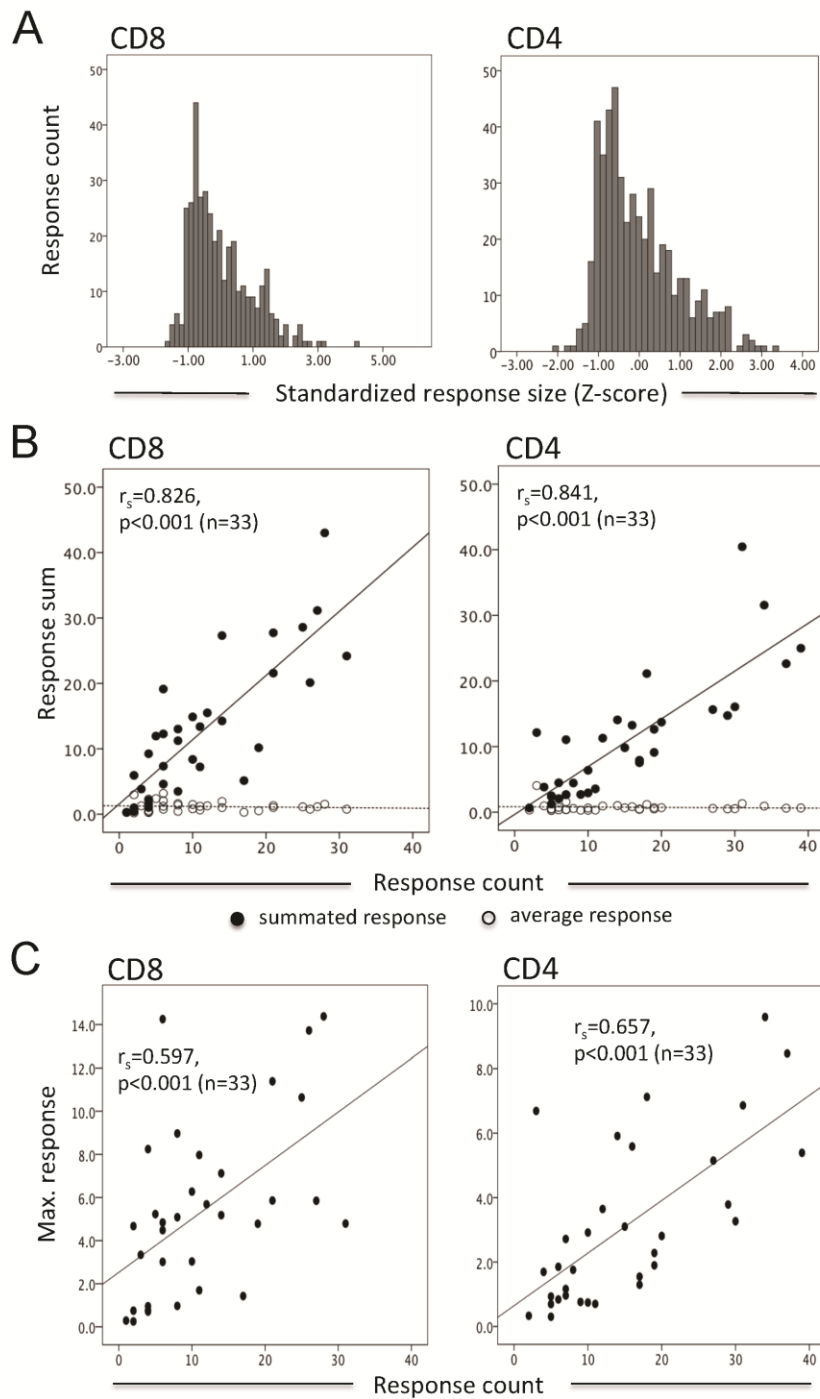


Fig. 4

Tables

Table 1: Non-inflated versus inflated CMV T-cell responses shown for proteins inducing responses in at least four donors^a and for donors responding to at least four proteins^a

T-cell Subset	Analysis by	No. responses analyzed	No. of donors in analysis	No. proteins in analysis	No. of inflated responses (outliers in %)
CD8	protein	250	33	33	26 (10.4%)
	donor	368	28	107	44 (11.9%)
CD4	protein	230	33	20	29 (12.6%)
	donor	494	31	125	51 (10.3%)

^a minimum number of responses for applying outlier definition

Table 2: Breakdown of inflated T-cell responses (outliers^a) by protein kinetic class, shown for proteins inducing responses in at least four donors^b and for donors responding to at least four proteins^b

	Protein kinetic class	Nr. of proteins with outliers	Nr. of outliers	Found in (Nr. of individuals)	average/max size of outliers (% of T-cells)
CD8/per protein	IE	2	5	5/33	10.0/14.4
	E	3	3	3/33	1.9/2.2
	E-L	5	6	4/33	2.6/3.5
	L	4	9	7/33	6.1/14.3
	Not assigned	4	3	3/33	3.2/4.8
CD8/per individual	IE	<u>5</u>	14	12/28	6.4/14.4
	E	<u>3</u>	4	4/28	2.7/6.1
	E-L	<u>5</u>	7	7/28	2.1/3.5
	L	9	15	11/28	5.0/14.3
	Not assigned	3	4	4/28	6.4/13.7
CD4/per protein	IE	2	2	1/33	2.7/2.9
	E	6	9	9/33	3.7/8.5
	E-L	5	9	8/33	2.7/7.1
	L	5	7	5/33	3.1/6.9
	Not assigned	2	2	2/33	2.3/2.8
CD4/per individual	IE	3	3	3/31	1.7/2.3
	E	5	16	14/31	2.9/8.5
	E-L	<u>5</u>	10	7/31	3.9/9.6
	L	6	21	14/31	2.8/7.6
	Not assigned	1	1	1/31	1.9/1.9

^aoutliers were defined as values > (upper quartile + (1.5 x inter-quartile range))

^b minimum number of responses for applying outlier definition