

Discrimination of Primary and Chronic CMV Infection based on Humoral Immune Profiles in Pregnancy

Andrew P. Hederman¹, Christopher J. Remmel¹, Shilpee Sharma², Harini Natarajan³, Joshua A. Weiner¹, Daniel Wrapp⁴, Catherine Donner⁵, Marie-Luce Delforge⁶, Piera d'Angelo⁷, Milena Furione⁷, Chiara Fornara^{7,8}, Jason S. McLellan⁴, Daniele Lilleri⁷, Arnaud Marchant², Margaret E. Ackerman^{1,3,#}

¹Thayer School of Engineering, Dartmouth College, Hanover, NH, USA

²European Plotkin Institute for Vaccinology, Université libre de Bruxelles, Brussels, Belgium

³Department of Microbiology and Immunology, Geisel School of Medicine, Hanover, NH, USA

⁴Department of Molecular Biosciences, The University of Texas, Austin, TX, USA

⁵Université Libre de Bruxelles (ULB), Hôpital Universitaire de Bruxelles (H.U.B.), CUB Hôpital Erasme, Department of Obstetrics and Gynecology, Brussels, Belgium

⁶Université Libre de Bruxelles (ULB), Hôpital Universitaire de Bruxelles (H.U.B.), CUB Hôpital Erasme, National Reference Center for Congenital Infections, Brussels, Belgium

⁷Microbiology and Virology, Fondazione IRCCS Policlinico San Matteo, 27100 Pavia, Italy

⁸Current affiliation: General Clinical Laboratory, with specialist areas of Clinical Pathology, Microbiology and Virology, Istituti Clinici Scientifici Maugeri IRCCS, 27100 Pavia, Italy

Corresponding Author

Margaret E. Ackerman

14 Engineering Drive

Hanover, NH 03755

margaret.e.ackerman@dartmouth.edu

(ph) 603 646 9922

(fax) 603 646 3856

Abstract

Most humans have been infected by Cytomegalovirus (CMV) by the time they reach forty years of age. Whereas most of these CMV infections are well controlled by the immune system, congenital infection can lead to serious health effects and death for the fetus and neonate. With clear evidence that risk to the fetus is lower during chronic maternal infection, and varies in association with gestational age at the time of primary maternal infection, further research on humoral immune responses to primary CMV infection during pregnancy is needed. Here, systems serology tools were applied to characterize antibody responses to CMV infection among pregnant and non-pregnant women experiencing either primary or chronic infection. Whereas strikingly different antibody profiles were observed depending on infection status, more limited differences were associated with pregnancy status. Beyond known differences in IgM responses that are used clinically for identification of primary infection, distinctions observed in IgA and FcγR binding antibody responses and among viral antigen specificities accurately predicted infection status in a cross-sectional cohort. Leveraging machine learning, longitudinal samples were also used to define an immunological clock of CMV infection from primary to chronic states and predict time since primary infection with high accuracy. Humoral responses diverged over time in an antigen-specific manner, with IgG3 responses toward tegument decreasing over time as is typical of viral infections, while those directed to pentamer and glycoprotein B were lower during acute and greatest during chronic infection. In sum, this work provides new insights into the antibody response associated with CMV infection status in the context of pregnancy, revealing aspects of humoral immunity that have the potential to improve CMV diagnostics and to support clinical trials of interventions to reduce mother-to-fetus transmission of CMV.

Introduction:

A member of the herpesvirus family, human cytomegalovirus (CMV) commonly manifests as a mild or asymptomatic infection. However, in very young and immunocompromised individuals, CMV can cause severe disease; it is the leading cause of congenital infection among newborns¹. In the United States alone it is estimated that 40,000 children are born with congenital CMV (cCMV) infection every year¹. This high burden is still likely an underestimate since many cases are asymptomatic at birth. Many of the diagnosed infections are severe, leading to an estimated 400 deaths and an additional 8,000 cases presenting with permanent disabilities, including speech and language impairment, hearing loss, mental disability, cerebral palsy, and vision impairment, annually^{1,2}. Fetal infection results from intrauterine transmission and is most likely to occur when a mother experiences primary CMV infection during pregnancy^{3,4}. The difference in fetal infection risk between primary and reactivated maternal CMV infection is striking, with approximately one third of primary infections leading to CMV infection of the fetus as compared to < 3.5% estimated to result from CMV reactivation or superinfection⁵⁻⁹. New insights into CMV infection during pregnancy that could contribute to identification of pregnancies at greatest risk, and efficient testing of new therapeutic interventions are urgently needed¹⁰.

Pregnant women present a population of subjects who are not immunodeficient, but in a unique state of immune regulation in order to ensure fetal tolerance, and can transmit CMV to their fetus during pregnancy. While the differing risk of transmission associated with maternal seropositivity provides strong evidence that pre-conception immunity plays a protective role, studies evaluating the clinical potential of CMV-hyperimmune globulin to improve neonatal outcomes have yielded mixed results¹¹⁻¹³. It has been speculated that these differing outcomes may relate to gestational age and the timing of the intervention following maternal infection and therefore be impacted by both the reliability of the diagnostic approach and strictness in the definition of primary infection cases. Because the suitability of these approaches for application during pregnancy can be supported by comparison of responses in pregnant with non-pregnant individuals during primary infection, developing a deeper understanding of humoral immune profiles in this unique immune state may have implications for the clinical development of diverse small molecule and biologic antiviral interventions.

Short of these goals, given the widely diverging congenital infection risks associated with primary and chronic infection, confident discrimination between these two states is crucial for identifying newborns with highest risk of cCMV infection. Complementing virological assessment, discrimination of these two states using serology may allow antiviral treatment in pregnant women and hearing and learning interventions in newborns to be initiated early^{14,15}. Current diagnostic assays for pregnant women are performed using serology measurements for IgM and IgG avidity, while a PCR based urine test, often following a saliva PCR screening test, is the standard for newborns^{14,16-18}. However, although CMV is a common infection, there is no universal standard diagnostic assay across countries and healthcare centers¹⁹, introducing variability in care and posing challenges to clinical trial design for testing of novel interventions. Additionally, many countries, including the United States, do not routinely screen for CMV

in pregnant women, whereas some European countries do perform routine screening, adding another layer of complexity into clinical data regarding the impact of maternal CMV status, antibody responses, and timing of infection as assessed by differing diagnostic measures and across populations. Irrespective of these details, following a primary CMV diagnosis during pregnancy, clinicians typically counsel pregnant women about the future risks to their fetus.

The dichotomy in fetal risk profiles of primary and non-primary CMV infection has important implications in transmission risk for pregnant women. Beyond this classification, the specific timing of CMV infection is also a crucial factor in potential of severe congenital disease, with a primary infection in the first term of pregnancy leading to higher likelihood of severe cCMV infections as compared to a primary infection in the second or third trimester²⁰. Further insights into relative risk profiles are challenged by imperfect precision in defining and estimating the timing of primary infection, particularly during pregnancy, which has been associated with varying and dynamic immunological state changes²¹. As a result, more accurate and confident diagnosis of primary and non-primary CMV infection would benefit the medical community and affected birthing parents by providing clearer and more definitive insights into their CMV infection status and associated risk to the fetus.

However, previous work in a variety of infectious disease settings has shown that antibody responses evolve over time, exhibiting complex patterns in response magnitude and characteristics²²⁻²⁴, and that they can also differ in association with pregnancy²⁴⁻²⁶. Pregnant women with primary CMV infection represent a unique intersection of these two complex antibody response scenarios with established clinical significance. Here, leveraging machine learning and highly multiplexed assays to capture a wide range of antibody response attributes over time, we evaluate how responses to primary or non-primary CMV infection during pregnancy and over time may vary, providing insight into the role pregnancy may play in modifying these responses and new means of predicting and reducing risk of cCMV infection for affected pregnancies.

Results

Antibody Profiles Distinguish Primary from Chronic CMV Infection

Antibody responses were profiled among both pregnant and non-pregnant individuals with either primary or chronic CMV infection (**Table 1**), with primary infection strictly defined by CMV-specific IgG seroconversion, CMV-specific IgM antibody detection, low IgG avidity index, and/or CMV DNAemia. Conversely, chronic infection was defined by seropositivity in the absence of these diagnostic measures. To more comprehensively understand how antibody profiles vary among pregnant primary (n=114), non-pregnant primary (n=29), pregnant chronic (n=30), non-pregnant chronic (n=30) CMV-infected individuals, in addition to naïve subjects (n=9), antibodies specific for CMV tegument protein²⁷ (CG1 and CG2), glycoprotein B (gB), and the pentamer complex were characterized for isotype, subclass, and Fc Receptor (FcR) binding capacity (**Supplemental Figure 1, Supplemental Table 1**). We profiled the antibody response examining a diverse set of antibody features and CMV antigens, including gB and

pentamer complexes from several different sources, with the aim of gaining a more holistic view of the humoral immune response to primary or chronic CMV infection. Glycoproteins tested included several that are commercially available (Sino (Sino Biologicals), and NA (Native Antigen)), as well as others that have been characterized in structural studies conducted with an eye toward vaccine development (GSK)^{28,29}, and by academic groups (UT)^{30,31}, including a novel modified form of gB with engineered proline mutations (3p).

To define differences in antibody profiles among subject groups that go beyond the measures typically used in clinical diagnosis, we performed UMAP analysis on CMV-specific antibody features excluding IgM (**Figure 1A**). This unsupervised analysis revealed that the main aspects of differentiation among subjects related to infection status; strong distinctions in the antibody response were observed between individuals with primary as opposed to chronic infection. In contrast, limited differences in the global response profiles were observed between pregnant and non-pregnant individuals. While univariate analysis between pregnant and non-pregnant individuals did reveal some nominally statistically significant differences across or within primary and chronic infection groups (**Supplemental Figure 2**), we could not exclude the possibility that these differences resulted from a difference in timing of diagnosis or were impacted by differences in age between cohorts. The former may be likely, as similarly subtle differences were observed between primary subjects in these groups and another cohort (Erasmus Hospital) of pregnant women with primary CMV infection (**Supplemental Figure 3**). Given these observations and the lack of robust differences associated with pregnancy status, this variable was not considered in subsequent analyses.

We next investigated which individual features of the immune response might contribute to clustering of subjects in primary and chronic infection groups. Visualizing the degree and confidence of differences in antibody response features between primary and chronic infection status groups revealed aspects of the humoral response that consistently differed (**Figure 1B-C**). Distinct levels of CMV-specific IgA, IgM, and FcR binding antibodies were observed between primary and chronic infection groups across multiple CMV antigens. IgM responses were elevated across all antigens for the primary infection group, which was expected as IgM titers were used to define infection status. Interestingly, total CMV-specific IgA and both IgA1 and IgA2 subclasses were elevated in primary infection subjects. Somewhat surprisingly, total IgG had very modest differences between primary and chronic infection groups.

Given the striking differences among groups we next wanted to further examine individual responses for IgM, IgA, and IgG across the panel of CMV antigens. As expected, IgM responses were strongly and significantly elevated among subjects with primary infection, as were IgA responses across diverse antigen specificities (**Figure 1C**). Interestingly, and in contrast with total IgG response levels (**Figure 1C**), which exhibited statistically significant but relatively small differences in response magnitude, FcγR-binding antibodies were both significantly and strongly elevated among subjects with chronic CMV infection (**Figure 1B, Supplemental Figure 4**), suggesting the presence of qualitative differences in the antibodies present during primary and chronic infection.

IgG Subclasses and CMV Antigens Display Distinct Profiles in Primary and Chronic Infection

The observation that FcγR binding antibodies but not total IgG were reliably elevated in chronic subjects was intriguing, and suggested that differences in induction of IgG subclasses, which differ dramatically in their FcR binding capacity, may exist. However, with the exception of IgG2 responses to tegument, IgG2 and IgG4 responses to CMV antigens were uncommon, and therefore differences in the levels of these more functionally inert subclasses were not observed (**Supplemental Figure 5**). Humoral responses to viral infections are generally dominated by IgG1 and IgG3 antibodies, which both bind well to activating FcγR and have the potential to elicit the potent antiviral activities of the complement cascade and innate immune effector cells. While statistically significant differences were observed for a subset of antigens tested, IgG1 responses, which are typically dominant following acute viral infections^{32,33}, were generally similar in primary and chronic infection (**Figure 2A**). Only one of each of the gB, pentamer, and tegument proteins tested showed a statistically significant difference, and each of these demonstrated only a small increase among the chronic infection group. In contrast, IgG3 responses were more uniformly and strikingly distinct between groups. Perhaps most surprisingly, the direction of these differences varied by antigen-specificity. Antibody responses to tegument proteins exhibited elevated levels in primary infection, but elevated levels of gB and pentamer-specific IgG3 were observed among the chronic infection group (**Figure 2A**). While all pentamer complexes tested showed this profile, responses to recombinant gB proteins were more variable, with two gB preparations showing elevated levels among chronic subjects, and one preparation showing elevated levels in primary infection. While some distinctions in IgG subclass composition in the context of primary and latent herpesvirus infections have been reported³⁴⁻³⁸, how well they may support discrimination between primary and chronic CMV infection is not known. In sum, while IgG1 responses tended to either persist or increase between primary and latent infection classes, IgG3 responses differed dramatically by antigen-specificity; responses to tegument were higher among subjects with primary infection, but responses to gB and pentamer were instead elevated in chronic infection.

While UT pentamer-specific IgG1 and IgG3 levels were correlated with each other among subjects with primary infection ($R_p=0.33$, $p<0.0001$), they were not well correlated among individuals with chronic infection ($R_p=0.20$, $p=0.08$). This difference in degree of correlation between IgG1 and IgG3 responses between infection status states was consistent across pentamer proteins evaluated. Additionally, the strength of correlation observed between these IgG subclasses and the FcγRIIIaV-binding capacity of pentamer-specific antibodies changed over time, with stronger correlations between IgG1 and FcγRIIIaV observed in primary infection, but IgG3 and FcγRIIIaV binding in chronic infection (**Figure 2C**). These patterns were consistent across FcγRs and pentamer proteins tested (**Figure 2D**), suggesting that the pool of antibodies most capable of eliciting FcγR-dependent effector functions changes in composition over the course of infection. These temporal differences may have important implications to viral pathogenesis and host defense.

Prediction of CMV infection status using machine learning

Next, because unsupervised analysis of antibody profiles revealed clear distinctions between primary and chronic CMV infection, we applied supervised machine learning to explore the ability of a model to accurately discriminate primary and chronic CMV infection (**Figure 3A**). We employed a logistic regression framework with regularization to classify primary or chronic CMV infection based on antibody profiling data while reducing the risk of overfitting associated with high-dimensional data. Again, IgM responses were excluded because they were used in clinical class assignment of the study groups. The model was trained on 80% of primary and chronic subject profiles while the remaining 20% was used for testing. Furthermore, repeated five-fold cross validation was employed so each subject would be part of the test set and representative accuracy across different folds could be defined.

Model predictions were highly accurate; across 100 repeated five-fold cross validation runs, the median model accuracy for predicting primary or chronic CMV infection was 94% (**Figure 3B**). In contrast, prediction results were essentially random when training and testing was performed after permutation of class labels, which serves as a means to assess model robustness by measuring the potential for overfitting. Misclassifications were not biased toward one or the other class, as shown for the cross validation run repeat presenting median accuracy as a representative confusion matrix (**Figure 3C**). Classification calls for actual but not permuted class labels were both typically correct and assigned to respective classes with high probability (**Figure 3D**). The relatively few incorrect classifications were typically close to the decision boundary. Given this evidence of model accuracy and robustness, a final model was trained on all subjects in the discovery cohort. When applied to an independent set of 57 primary infection and 40 chronic infection samples serving as a validation cohort, this model resulted in similarly high confidence and perfect classification accuracy (**Figure 3D**). Given this validation, the top features employed in the final model were examined (**Figure 3E**). The features with largest positive coefficients, which serve to identify primary subjects, were primarily IgA-related antibody responses directed to tegument antigens (4/5 features). Conversely, features with large negative coefficients, useful to identify chronic infection, were typically related to IgG3 and FcγRIII-binding capacity of response to gB and pentamer (4/5 features). Collectively, these modeling results point to novel antibody response attributes, distinct from traditionally considered parameters such as IgM and IgG avidity, as being excellent candidate markers for distinguishing primary and chronic infection status.

Modeling Longitudinal Responses to CMV Infection Reveals a Molecular Clock of Antibody Responses

The excellent discrimination of primary and chronic CMV infection in the validation cohort led us to next explore how class predictions related to longitudinal development of humoral immune responses to infection. To this end, subsequent samples were available for the validation cohort over a series of up to four visits that extended out to half a year after initial sampling or infection onset. The model trained on the initial cohort was used to predict class for longitudinal samples from the validation cohort (**Figure 4A**).

Whereas chronic subjects were consistently classified as such with similarly high confidence at all subsequent visits, subjects with primary infection at the initial timepoint became less confidently classified as primary at subsequent visits (**Figure 4B**). By visit four, the majority of samples from subjects with primary infection at their initial visit exhibited class probabilities below the lowest observed at the initial sampling. However, because visits were not consistently spaced in time between subjects, these longitudinal profiles can be more meaningfully compared over time post symptom onset (**Figure 4C**). Despite the imperfect reliability of projected timing of infection, the classification model probabilities presented a clear relationship with time. Samples fell below the midpoint on the classification scale as early as 90 days post infection onset, but failed to reach values typically observed among chronic subjects even out at 250 days post infection, suggesting a relatively prolonged transition to reaching the chronic state profile. As expected, the individual antibody response features making the greatest contributions to the classification model also demonstrated clear changes among subjects with primary, but not chronic infection, over time (**Figure 4D**).

To this point, machine learning models have only been concerned with making predictions on the probability of a sample belonging to either the primary or chronic infection class. However, given the clear ability for these binary classification models to provide insight into time since infection, we next sought to evaluate models explicitly trained for this specific purpose. For this purpose, longitudinal profiles of the primary infection cases across visits were used to train a model to predict time since symptom onset as a continuous variable (**Figure 4E**). The resulting linear regression model, which showed good robustness in the context of five-fold cross-validation (**Supplemental Figure 6A**), was then used to predict days post infection for the cross-sectional primary samples. This model, which relied primarily on IgG3 features (**Supplemental Figure 6B**) showing strong time-dependence (**Supplemental Figure 6C**), was then applied to the cross-sectional cohort. Predictions of time since symptom onset in the unseen validation cohort exhibited excellent accuracy (**Figure 4F**). Overall, this analysis demonstrated that antibody profiles in CMV-infected individuals exhibit generalizable temporal patterns in their dynamic antibody responses during primary infection.

Discussion

Presently, discrimination of primary and non-primary CMV infection in the context of pregnancy is used in counseling as to the risk of congenital CMV infection based on the lower risk of congenital infection associated with non-primary infection and the prescription of antiviral therapies. Beyond this value, in the absence of an effective vaccine, a deeper understanding of how immune responses differ in association with transmission risk has the potential to contribute to the development of new interventions. Here, high dimensional antibody profiles beyond IgM levels and IgG avidity were developed from a commonly used multiplex assay format and supervised and unsupervised machine learning was used to differentiate primary and chronic infection and pregnancy status.

Whereas limited or no differences in humoral responses were associated with pregnancy status, our study showed clear distinctions in antibody profiles between primary and chronic infection cases. Both composite profiles individual antibody features related to antigen-specificity and immunoglobulin isotype, subclass, and binding to Fc receptors demonstrated these distinctions. Importantly, these differences extended beyond those previously known to exist and which are presently applied to support clinical diagnosis. The number and longitudinal profiles of the features that distinguish infection history suggested that a sort of humoral clock could be defined in order to time the onset of primary CMV infection. Indeed, supervised machine learning models that reliably captured how antibody responses to CMV infection varied over time were learned and validated on independent samples.

Beyond potential clinical utility, the defining features of how responses to primary CMV infection transition over time is relevant to understanding the evolution of the immune response to this member of the notoriously immune-evasive Herpesviridae family^{39,40}. Current diagnostic methods rely on detection/levels of IgM and of IgG avidity. The presence of IgM alone is insufficient to diagnose a primary CMV infection: poor correlation between commercial tests have been reported⁴¹, and false positive primary status calls can result from both persistence of IgM as well as boosting in response to reactivation^{42,43}. Likewise, a positive test for CMV-specific IgG indicates an individual is positive for CMV but provides little information into the time since infection^{14,44}. However, the combination of IgM positivity and low IgG avidity is generally considered to be a reliable indicator of primary CMV infection, though interpretation of IgG avidity tests are confounded by low levels of CMV-specific IgG, intermediate responses raise classification issues, and these indicators are supported by clinical studies of small size. Coupled to the lack of an international standard serum panel of samples from primary infection, the difficulty inherent to establishing a uniform, accurate, and robust diagnostic method is clear⁴⁵. Our data demonstrates that there are other changes in the secreted antibody repertoire that reliably occur over consistent time periods during primary infection.

Prior studies of IgG subclasses of CMV-specific antibodies have consistently reported the IgG1 and IgG3 bias typical of anti-viral antibodies but noted that subclass responses and total IgG titers can be discordant^{35,46}. Further, they have attributed relatively greater neutralization potency to this minor portion to IgG3³⁶. To that end, we were surprised to observe increasing levels of IgG3 specific for viral glycoproteins over time. Whereas IgG3 responses are usually associated with acute infections and typically wane over time^{33,47-49}, they were observed to increase across both cross-sectional and longitudinal cohorts, but to do so in an antigen-specific manner. Namely, whereas IgG3 responses to tegument proteins showed the typical pattern and decreased while IgG1 responses were stable or increased, IgG3 responses to pentamer increased while IgG1 responses remained largely stable. Memory B cells (MBC) specific for gB are known to exhibit phenotypic states distinct from those specific for tegument antigens⁵⁰. The reduced frequency of MBCs with effector potential specific for glycoprotein as compared to tegument observed in this prior study may relate to the altered kinetics of IgG subclasses observed here. While the mechanistic underpinning and the biological implications of this unusual pattern

have yet to be determined, CMV expresses two Fc binding proteins, gp34 and gp68⁵¹. Unlike the viral FcR of HSV, these proteins have been reported to bind to all human IgG subclasses^{52,53}, though affinities have not been reported. Additionally, how these two surface proteins contribute to viral pathogenesis is incompletely understood⁵⁴.

Limitations of this study include the fact that the majority of samples were sourced from a single geographic region, leaving open the possibility that differences associated with host genetics and environmental and infectious disease exposure history impact these observations. While relationships between maternal immune responses and transmission of cCMV are of exceptionally high clinical relevance, this study focused on CMV infection history. However, while risk is considerably greater during primary infection, there are certainly differences in virologic, innate immune, and other factors that contribute to these differing risk profiles⁵⁵⁻⁵⁸. The influence of maternal antibody responses is unclear, particularly given the conflicting results in studies of passive antibody therapy in the context of primary maternal infection^{59,60}. While the dosing and frequency of hyperimmune globulin also differed in these studies, a positive effect was only observed in women with very early infection, pointing to the potential importance of accuracy in the dating of infection recency. Recent studies have started to explore the role that antibody responses play in CMV transmission in a rhesus macaque model^{61,62}. However, data in humans is limited and often confounded by the differing risk of infection associated with primary as compared to non-primary infection⁶³. Additionally, the panel of antigens tested in our investigation was not exhaustive, and the reactivity patterns sometimes differed in association with different sequences, conformational states, expression hosts, and other factors, the contributions and importance of which have yet to be defined. These and other factors may represent worthwhile directions for a future study.

Overall, many open questions regarding the role of humoral immunity in the context of CMV infection, transmission, latency, and reactivation remain. Higher resolution and more comprehensive analysis of antibody responses using systems serology approaches has the potential to improve our understanding of the complex virus-host interactions at play. Here, by analyzing highly curated cohorts, we report and validate phenotypic signatures of gB, pentamer, and tegument-specific antibody responses that not only robustly classify primary infection status, but also provide insights into time of infection. It remains to be seen if the atypical dynamic profile of IgG3 responses to envelope glycoproteins elicited by CMV is antigen-intrinsic and might be recapitulated when these antigens are delivered by other means, or if it may represent an evasion strategy dependent on other viral genes or aspects of the innate response to viral infection. In the meantime, this work stands to define novel hallmarks of primary CMV infection and time of infection that may present new opportunities to streamline primary infection diagnosis, potentially impacting current clinical practice and enrolment of pregnant women with primary infection in interventional trials, thereby providing new insights into relative cCMV risk and management strategies.

Methods

Clinical Samples

Serum samples were gathered from subject cohort groups of pregnant primary infection, pregnant latent infection, non-pregnant primary infection, non-pregnant latent infection, as well as CMV negative patient cohort as a negative control group (**Table 1**). Human subjects were recruited from Fondazione IRCCS Policlinico San Matteo, Pavia, Italy, and included healthy, primary, and latent CMV infected subjects, as well as pregnant and non-pregnant subjects. An additional cohort of pregnant women with primary infection was recruited from Erasme Hospital. Diagnosis of primary HCMV infection was based on two or more of the following criteria: CMV-specific IgG seroconversion, CMV-specific IgM antibody detection, low IgG avidity index and CMV DNAemia. Chronic infection was defined by the presence of CMV-specific IgG and absence of CMV-specific IgM antibody, and no detection of CMV DNA in blood, saliva, urine and genital secretion during pregnancy. In 89/114 (78%) pregnant women and 28/29 (95%) non-pregnant subjects, onset of primary infection was defined by the appearance of symptoms, while in 19/114 (17%) asymptomatic pregnant women onset of infection was estimated on seroconversion (i.e. in the mid between the last IgG negative and the first IgG positive result), occurring within a ≤ 6 -week interval. Finally, in six asymptomatic pregnant women (5%) and one (3%) non-pregnant subject, onset of infection was estimated on the kinetics of CMV-specific IgM and IgG avidity index. For a subgroup of 40 pregnant women and 28 non-pregnant subjects with primary infection, 2-4 sequential serum samples collected until 6 months after onset of infection were available. Samples from pregnant women with chronic infection were collected at 10, 20, and 30 weeks of gestation and at delivery. Sample collection was approved by Fondazione IRCCS Policlinico San Matteo, and each participant gave written informed consent.

Antibody Profiling Experiments

Fc receptors were expressed as previously described⁶⁴. A complete list of Fc detection reagents and antigens used is provided in **Supplemental Table 1**. Pentamer and glycoprotein B were sourced both commercially, through industry partners, and expressed in-house. For the latter, HCMV Pentamer (UT, Towne strain) was produced by mixing plasmids encoding for residues 24 to 718 of gH with a C-terminal 6x HisTag, residues 31 to 278 of gL, residues 21 to 171 of UL128, residues 26 to 214 of UL130, and residues 19 to 129 of UL131A, all with artificial signal sequences, at an equimolar ratio³¹. This mixture was then used to transiently transfect FreeStyle293-F cells via polyethylenimine. Plasmids encoding for postfusion HCMV gB (AD169 strain) residues 32–692 with an artificial N-terminal signal sequence and a C-terminal HRV3C protease cleavage site, 8xHisTag, and a TwinStrep tag. One postfusion gB construct (UT, or JSM-1074) contained substitutions Y155G, I156H, Y157R, Y206H, S238N, W240T, L241T, Y242H and C246S, which have been reported previously³⁰ and the other construct (3p, or JSM-956) contained substitutions Y155G, I156H, Y157R, Y206H, W240A, L241T, Y242H, C246S, R456A, R459G, M472P, R491P, G492P and a C-terminal T4 fibrin trimerization motif. Plasmids were transiently transfected into FreeStyle 293-F cells via polyethylenimine. Pentamer was purified using NiNTA resin (Thermo Scientific) and postfusion gB proteins were purified using Strep-Tactin

affinity resin (IBA Lifesciences). Affinity-purified proteins were further purified by size-exclusion chromatography using a Superose6 10/300 column (Cytiva) in a buffer composed of 2 mM Tris pH 8.0, 200 mM NaCl, 0.02% NaN₃.

Characterization of antibody profiles was performed using the Fc array assay^{65,66}. Antigens were covalently coupled to magnetic microspheres (Luminex Corporation) using carbodiimide chemistry. Serum dilutions used in assays ranged from 1:250-1:5000 based on initial pilot experiments and previous experience. Detection of antigen specific antibodies was done using R-phycoerythrin-conjugated secondary reagents specific to human immunoglobulin isotypes and subclasses and by Fc receptor tetramers. Median fluorescent intensity data was acquired on a FlexMap 3D array reader (Luminex Corporation). Samples were run in technical duplicates and results were averaged.

Classification of CMV infection status

A binomial logistic regression model with least absolute shrinkage and selection operator (LASSO) regularization was used to predict infection status. Model training was performed using the scikit-learn (version 1.3) in Python (version 3.9) with default options. The regularization parameter was chosen using the option that gave the lowest classification error. Model accuracy was determined by the test set label predictions compared with true labels. Accuracy was assessed over 100 repetitions of five-fold cross validation. Permutation testing was done to measure model robustness by doing the same procedure as described but on randomly shuffled data. Feature importance was determined from a final model that included all subjects.

Prediction of time since infection

The same machine learning model as described for predicting CMV infection status was used, this time minimizing mean squared error in time since infection. The model was trained on cross sectional data and tested on all longitudinal samples.

Statistics

Statistical analysis was performed in GraphPad Prism (version 9.7). UMAP plots were generated in Python (version 3.9) using the umap-learn package (version 0.4)⁶⁷ and then plotted using Prism. Volcano plots were generated in R (version 4.3) using ggplot2. Statistical tests are described in the relevant figure legends.

Study approval

The study was approved by the IRBs at Fondazione IRCCS Policlinico San Matteo and Erasme Hospital for sample collection, and Dartmouth College for sample testing and analysis.

Data availability

Data used in the study is available as **Supplemental Data File 1**.

Acknowledgements

This study was supported in part by Infect-ERA, GSK and Sanofi via the CYMAF consortium, as well as by NIH U19AI45825. Pentamer and gB proteins were provided by GSK. A.M. is Research Director at the F.R.S., FNRS, Belgium.

Conflict of Interest

A patent application for Systems and methods for identifying and treating primary and latent infections and/or determining time since infection, WO2023154857A2, has been submitted by A.P.H, D. L., A.M, and M.E.A..

References:

- 1 Cannon, M. J. & Davis, K. F. Washing our hands of the congenital cytomegalovirus disease epidemic. *BMC Public Health* **5**, 70 (2005). <https://doi.org/10.1186/1471-2458-5-70>
- 2 Schleiss, M. R. Congenital cytomegalovirus infection: molecular mechanisms mediating viral pathogenesis. *Infect Disord Drug Targets* **11**, 449-465 (2011).
<https://doi.org/10.2174/187152611797636721>
- 3 Dietrich, M. L. & Schieffelin, J. S. Congenital Cytomegalovirus Infection. *Ochsner J* **19**, 123-130 (2019). <https://doi.org/10.31486/toj.18.0095>
- 4 Pass, R. F. & Anderson, B. Mother-to-Child Transmission of Cytomegalovirus and Prevention of Congenital Infection. *J Pediatric Infect Dis Soc* **3 Suppl 1**, S2-6 (2014).
<https://doi.org/10.1093/jpids/piu069>
- 5 Stagno, S. *et al.* Congenital cytomegalovirus infection: The relative importance of primary and recurrent maternal infection. *N Engl J Med* **306**, 945-949 (1982).
<https://doi.org/10.1056/nejm198204223061601>
- 6 Lazzarotto, T. *et al.* Congenital Cytomegalovirus Infection: A Narrative Review of the Issues in Screening and Management From a Panel of European Experts. *Frontiers in Pediatrics* **8** (2020).
<https://doi.org/10.3389/fped.2020.00013>
- 7 Simonazzi, G. *et al.* Perinatal Outcomes of Non-Primary Maternal Cytomegalovirus Infection: A 15-Year Experience. *Fetal Diagn Ther* **43**, 138-142 (2018). <https://doi.org/10.1159/000477168>
- 8 Gatta, L. A. *et al.* Clinical factors associated with cytomegalovirus shedding among seropositive pregnant women. *Am J Obstet Gynecol MFM* **4**, 100560 (2022).
<https://doi.org/10.1016/j.ajogmf.2021.100560>
- 9 Zelini, P. *et al.* Human cytomegalovirus non-primary infection during pregnancy: antibody response, risk factors and newborn outcome. *Clin Microbiol Infect* **28**, 1375-1381 (2022).
<https://doi.org/10.1016/j.cmi.2021.09.013>
- 10 Kenneson, A. & Cannon, M. J. Review and meta-analysis of the epidemiology of congenital cytomegalovirus (CMV) infection. *Rev Med Virol* **17**, 253-276 (2007).
<https://doi.org/10.1002/rmv.535>
- 11 Revello, M. G. *et al.* A randomized trial of hyperimmune globulin to prevent congenital cytomegalovirus. *N Engl J Med* **370**, 1316-1326 (2014). <https://doi.org/10.1056/NEJMoa1310214>
- 12 Hughes, B. L. *et al.* A Trial of Hyperimmune Globulin to Prevent Congenital Cytomegalovirus Infection. *New England Journal of Medicine* **385**, 436-444 (2021).
<https://doi.org/10.1056/NEJMoa1913569>
- 13 Kagan, K. O. *et al.* Outcome of pregnancies with recent primary cytomegalovirus infection in first trimester treated with hyperimmunoglobulin: observational study. *Ultrasound Obstet Gynecol* **57**, 560-567 (2021). <https://doi.org/10.1002/uog.23596>
- 14 Ross, S. A., Novak, Z., Pati, S. & Boppana, S. B. Overview of the diagnosis of cytomegalovirus infection. *Infect Disord Drug Targets* **11**, 466-474 (2011).
<https://doi.org/10.2174/187152611797636703>
- 15 Year 2007 position statement: Principles and guidelines for early hearing detection and intervention programs. *Pediatrics* **120**, 898-921 (2007). <https://doi.org/10.1542/peds.2007-2333>
- 16 Prince, H. E. & Lapé-Nixon, M. Role of cytomegalovirus (CMV) IgG avidity testing in diagnosing primary CMV infection during pregnancy. *Clin Vaccine Immunol* **21**, 1377-1384 (2014).
<https://doi.org/10.1128/cvi.00487-14>
- 17 Saldan, A. *et al.* Testing for Cytomegalovirus in Pregnancy. *Journal of Clinical Microbiology* **55**, 693-702 (2017). <https://doi.org/doi:10.1128/JCM.01868-16>

- 484 18 Hughes, B. L. & Gyamfi-Bannerman, C. Diagnosis and antenatal management of congenital
485 cytomegalovirus infection. *Am J Obstet Gynecol* **214**, B5-b11 (2016).
486 <https://doi.org/10.1016/j.ajog.2016.02.042>
- 487 19 Abdullahi Nasir, I., Babayo, A. & Shehu, M. S. Clinical Significance of IgG Avidity Testing and
488 Other Considerations in the Diagnosis of Congenital Cytomegalovirus Infection: A Review
489 Update. *Medical Sciences* **4**, 5 (2016).
- 490 20 Chatzakis, C. *et al.* Timing of primary maternal cytomegalovirus infection and rates of vertical
491 transmission and fetal consequences. *Am J Obstet Gynecol* **223**, 870-883.e811 (2020).
492 <https://doi.org/10.1016/j.ajog.2020.05.038>
- 493 21 Aghaeepour, N. *et al.* An immune clock of human pregnancy. *Science Immunology* **2**, eaan2946
494 (2017). <https://doi.org/doi:10.1126/sciimmunol.aan2946>
- 495 22 Zuniga, E. I., Macal, M., Lewis, G. M. & Harker, J. A. Innate and Adaptive Immune Regulation
496 During Chronic Viral Infections. *Annu Rev Virol* **2**, 573-597 (2015).
497 <https://doi.org/10.1146/annurev-virology-100114-055226>
- 498 23 Rouse, B. T. & Sehrawat, S. Immunity and immunopathology to viruses: what decides the
499 outcome? *Nat Rev Immunol* **10**, 514-526 (2010). <https://doi.org/10.1038/nri2802>
- 500 24 Ovies, C., Semmes, E. C. & Coyne, C. B. Pregnancy influences immune responses to SARS-CoV-2.
501 *Science Translational Medicine* **13**, eabm2070 (2021).
502 <https://doi.org/doi:10.1126/scitranslmed.abm2070>
- 503 25 Steinhoff, M. C. *et al.* Influenza Immunization in Pregnancy — Antibody Responses in Mothers
504 and Infants. *New England Journal of Medicine* **362**, 1644-1646 (2010).
505 <https://doi.org/10.1056/NEJMc0912599>
- 506 26 Atyeo, C. G. *et al.* Maternal immune response and placental antibody transfer after COVID-19
507 vaccination across trimester and platforms. *Nature Communications* **13**, 3571 (2022).
508 <https://doi.org/10.1038/s41467-022-31169-8>
- 509 27 Ripalti, A. *et al.* Construction of polyepitope fusion antigens of human cytomegalovirus ppUL32:
510 reactivity with human antibodies. *Journal of clinical microbiology* **32**, 358-363 (1994).
511 <https://doi.org/10.1128/jcm.32.2.358-363.1994>
- 512 28 Chandramouli, S. *et al.* Structure of HCMV glycoprotein B in the postfusion conformation bound
513 to a neutralizing human antibody. *Nat Commun* **6**, 8176 (2015).
514 <https://doi.org/10.1038/ncomms9176>
- 515 29 Chandramouli, S. *et al.* Structural basis for potent antibody-mediated neutralization of human
516 cytomegalovirus. *Sci Immunol* **2** (2017). <https://doi.org/10.1126/sciimmunol.aan1457>
- 517 30 Ye, X. *et al.* Recognition of a highly conserved glycoprotein B epitope by a bivalent antibody
518 neutralizing HCMV at a post-attachment step. *PLoS Pathog* **16**, e1008736 (2020).
519 <https://doi.org/10.1371/journal.ppat.1008736>
- 520 31 Wrapp, D. *et al.* Structural basis for HCMV Pentamer recognition by neuropilin 2 and neutralizing
521 antibodies. *Sci Adv* **8**, eabm2546 (2022). <https://doi.org/10.1126/sciadv.abm2546>
- 522 32 Nachbagauer, R. *et al.* Age Dependence and Isotype Specificity of Influenza Virus Hemagglutinin
523 Stalk-Reactive Antibodies in Humans. *mbio* **7**, 10.1128/mbio.01996-01915 (2016).
524 <https://doi.org/doi:10.1128/mbio.01996-15>
- 525 33 Yates, N. L. *et al.* Multiple HIV-1-specific IgG3 responses decline during acute HIV-1: implications
526 for detection of incident HIV infection. *Aids* **25**, 2089-2097 (2011).
527 <https://doi.org/10.1097/QAD.0b013e32834b348e>
- 528 34 Doerr, H. W., Rentschler, M. & Scheifler, G. Serologic detection of active infections with human
529 herpes viruses (CMV, EBV, HSV, VZV): Diagnostic potential of IgA class and IgG subclass-specific
530 antibodies. *Infection* **15**, 93-98 (1987). <https://doi.org/10.1007/BF01650204>

- 35 Gilljam, G. & Wahren, B. Properties of IgG subclasses to human cytomegalovirus. *J Virol Methods* **25**, 139-151 (1989). [https://doi.org/10.1016/0166-0934\(89\)90028-1](https://doi.org/10.1016/0166-0934(89)90028-1)
- 36 Gupta, C. K., Leszczynski, J., Gupta, R. K. & Siber, G. R. IgG subclass antibodies to human cytomegalovirus (CMV) in normal human plasma samples and immune globulins and their neutralizing activities. *Biologicals* **24**, 117-124 (1996). <https://doi.org/10.1006/biol.1996.0015>
- 37 Joassin, L., Somze, A. & Reginster, M. Detection by enzyme-linked immunosorbent assay of specific immunoglobulin G isotypes in primary and established cytomegalovirus infections. *Journal of Clinical Microbiology* **27**, 139-144 (1989). <https://doi.org/doi:10.1128/jcm.27.1.139-144.1989>
- 38 Rodriguez, G. E. & Adler, S. P. Immunoglobulin G subclass responses to cytomegalovirus in seropositive patients after transfusion. *Transfusion* **30**, 528-531 (1990). <https://doi.org/https://doi.org/10.1046/j.1537-2995.1990.30690333484.x>
- 39 Powers, C., DeFilippis, V., Malouli, D. & Früh, K. Cytomegalovirus immune evasion. *Curr Top Microbiol Immunol* **325**, 333-359 (2008). https://doi.org/10.1007/978-3-540-77349-8_19
- 40 Jackson, S. E. *et al.* CMV immune evasion and manipulation of the immune system with aging. *Geroscience* **39**, 273-291 (2017). <https://doi.org/10.1007/s11357-017-9986-6>
- 41 Genser, B., Truschnig-Wilders, M., Stünzner, D., Landini, M. P. & Halwachs-Baumann, G. Evaluation of five commercial enzyme immunoassays for the detection of human cytomegalovirus-specific IgM antibodies in the absence of a commercially available gold standard. *Clin Chem Lab Med* **39**, 62-70 (2001). <https://doi.org/10.1515/cclm.2001.014>
- 42 Rasmussen, L. *et al.* Virus-specific IgG and IgM antibodies in normal and immunocompromised subjects infected with cytomegalovirus. *J Infect Dis* **145**, 191-199 (1982). <https://doi.org/10.1093/infdis/145.2.191>
- 43 Naumnik, B. *et al.* Comparison of serology assays and polymerase chain reaction for the monitoring of active cytomegalovirus infection in renal transplant recipients. *Transplant Proc* **39**, 2748-2750 (2007). <https://doi.org/10.1016/j.transproceed.2007.08.040>
- 44 Razonable, R. R. *et al.* Clinical Diagnostic Testing for Human Cytomegalovirus Infections. *J Infect Dis* **221**, S74-S85 (2020). <https://doi.org/10.1093/infdis/jiz601>
- 45 Revello, M. G. *et al.* Comparative evaluation of eight commercial human cytomegalovirus IgG avidity assays. *J Clin Virol* **48**, 255-259 (2010). <https://doi.org/10.1016/j.jcv.2010.05.004>
- 46 Joassin, L., Somze, A. & Reginster, M. Detection by enzyme-linked immunosorbent assay of specific immunoglobulin G isotypes in primary and established cytomegalovirus infections. *J Clin Microbiol* **27**, 139-144 (1989). <https://doi.org/10.1128/jcm.27.1.139-144.1989>
- 47 Dugast, A.-S. *et al.* Independent evolution of Fc- and Fab-mediated HIV-1-specific antiviral antibody activity following acute infection. *European Journal of Immunology* **44**, 2925-2937 (2014). <https://doi.org/https://doi.org/10.1002/eji.201344305>
- 48 Sadanand, S. *et al.* Temporal variation in HIV-specific IgG subclass antibodies during acute infection differentiates spontaneous controllers from chronic progressors. *AIDS* **32**, 443-450 (2018). <https://doi.org/10.1097/qad.0000000000001716>
- 49 Damelang, T., Rogerson, S. J., Kent, S. J. & Chung, A. W. Role of IgG3 in Infectious Diseases. *Trends Immunol* **40**, 197-211 (2019). <https://doi.org/10.1016/j.it.2019.01.005>
- 50 Dauby, N. *et al.* Limited Effector Memory B-Cell Response to Envelope Glycoprotein B During Primary Human Cytomegalovirus Infection. *J Infect Dis* **213**, 1642-1650 (2016). <https://doi.org/10.1093/infdis/jiv769>
- 51 Corrales-Aguilar, E. *et al.* Human cytomegalovirus Fcγ binding proteins gp34 and gp68 antagonize Fcγ receptors I, II and III. *PLoS Pathog* **10**, e1004131 (2014). <https://doi.org/10.1371/journal.ppat.1004131>

- 52 Atalay, R. *et al.* Identification and expression of human cytomegalovirus transcription units coding for two distinct Fcγ receptor homologs. *J Virol* **76**, 8596-8608 (2002).
<https://doi.org/10.1128/jvi.76.17.8596-8608.2002>
- 53 Antonsson, A. & Johansson, P. J. H. Binding of human and animal immunoglobulins to the IgG Fc receptor induced by human cytomegalovirus. *J Gen Virol* **82**, 1137-1145 (2001).
<https://doi.org/10.1099/0022-1317-82-5-1137>
- 54 Vezzani, G. *et al.* Human immunoglobulins are transported to HCMV viral envelope by viral Fcγ receptors-dependent and independent mechanisms. *Front Microbiol* **13**, 1106401 (2022). <https://doi.org/10.3389/fmicb.2022.1106401>
- 55 Ssentongo, P. *et al.* Congenital Cytomegalovirus Infection Burden and Epidemiologic Risk Factors in Countries With Universal Screening: A Systematic Review and Meta-analysis. *JAMA Network Open* **4**, e2120736-e2120736 (2021). <https://doi.org/10.1001/jamanetworkopen.2021.20736>
- 56 Wang, C., Zhang, X., Bialek, S. & Cannon, M. J. Attribution of congenital cytomegalovirus infection to primary versus non-primary maternal infection. *Clin Infect Dis* **52**, e11-13 (2011).
<https://doi.org/10.1093/cid/ciq085>
- 57 Picone, O. *et al.* A series of 238 cytomegalovirus primary infections during pregnancy: description and outcome. *Prenat Diagn* **33**, 751-758 (2013). <https://doi.org/10.1002/pd.4118>
- 58 Lazzarotto, T., Guerra, B., Gabrielli, L., Lanari, M. & Landini, M. P. Update on the prevention, diagnosis and management of cytomegalovirus infection during pregnancy. *Clin Microbiol Infect* **17**, 1285-1293 (2011). <https://doi.org/10.1111/j.1469-0691.2011.03564.x>
- 59 Nigro, G., Adler, S. P., La Torre, R. & Best, A. M. Passive Immunization during Pregnancy for Congenital Cytomegalovirus Infection. *New England Journal of Medicine* **353**, 1350-1362 (2005).
<https://doi.org/10.1056/NEJMoa043337>
- 60 Jückstock, J., Rothenburger, M., Friese, K. & Traunmüller, F. Passive Immunization against Congenital Cytomegalovirus Infection: Current State of Knowledge. *Pharmacology* **95**, 209-217 (2015). <https://doi.org/10.1159/000381626>
- 61 Gong, Y. *et al.* Mathematical Modeling of Rhesus Cytomegalovirus Transplacental Transmission in Seronegative Rhesus Macaques. *Viruses* **15**, 2040 (2023).
- 62 Otero, C. E. *et al.* Relationship of maternal cytomegalovirus-specific antibody responses and viral load to vertical transmission risk following primary maternal infection in a rhesus macaque model. *PLoS Pathog* **19**, e1011378 (2023). <https://doi.org/10.1371/journal.ppat.1011378>
- 63 Semmes, E. C. *et al.* ADCC-activating antibodies correlate with decreased risk of congenital human cytomegalovirus transmission. *JCI Insight* **8** (2023).
<https://doi.org/10.1172/jci.insight.167768>
- 64 Boesch, A. W. *et al.* Highly parallel characterization of IgG Fc binding interactions. *mAbs* **6**, 915-927 (2014). <https://doi.org/10.4161/mabs.28808>
- 65 Brown, E. P. *et al.* Multiplexed Fc array for evaluation of antigen-specific antibody effector profiles. *J Immunol Methods* **443**, 33-44 (2017). <https://doi.org/10.1016/j.jim.2017.01.010>
- 66 Brown, E. P. *et al.* Optimization and qualification of an Fc Array assay for assessments of antibodies against HIV-1/SIV. *J Immunol Methods* **455**, 24-33 (2018).
<https://doi.org/10.1016/j.jim.2018.01.013>
- 67 McInnes, L., Healy, J. & Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426* (2018).

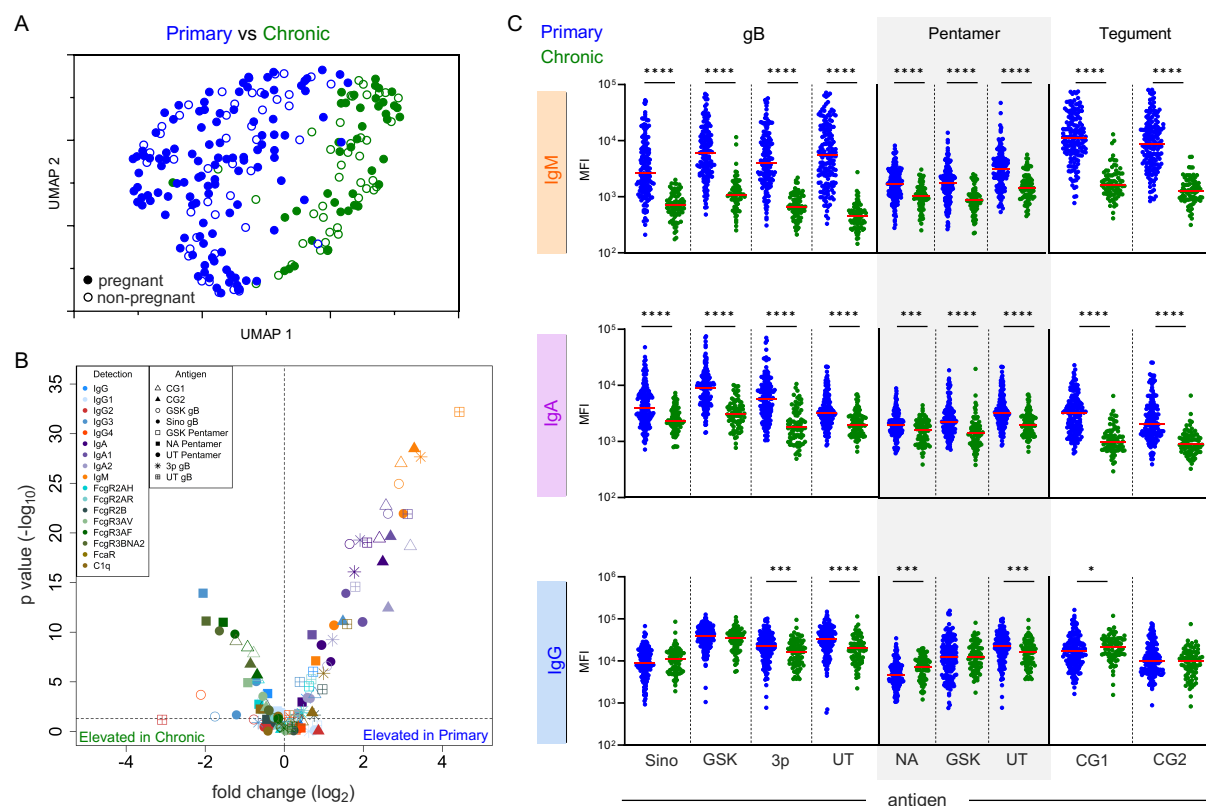


Figure 1: Antibody features distinguish primary from Chronic CMV infection but not pregnancy status. **A.** Uniform manifold approximation (UMAP) biplot of antibody features excluding IgM. Distinct clusters of subjects with primary (blue) and chronic (green) infection but not pregnancy status (hollow and filled symbols) are observed. **B.** Volcano plot of each CMV-specific antibody feature assessed. Volcano plot represents the \log_2 fold change (x-axis) against the $-\log_{10}$ p value (Mann Whitney test: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, and **** $p < 0.0001$). Antibody specificities (Antigen) are indicated by shape and Fc characteristics (Detection) indicated by color. **C.** IgM, IgA, and IgG binding to CMV antigens (further described in Supplemental Table 1). Data are the mean median fluorescent intensity (MFI) values of technical replicates. Solid red line indicates median.

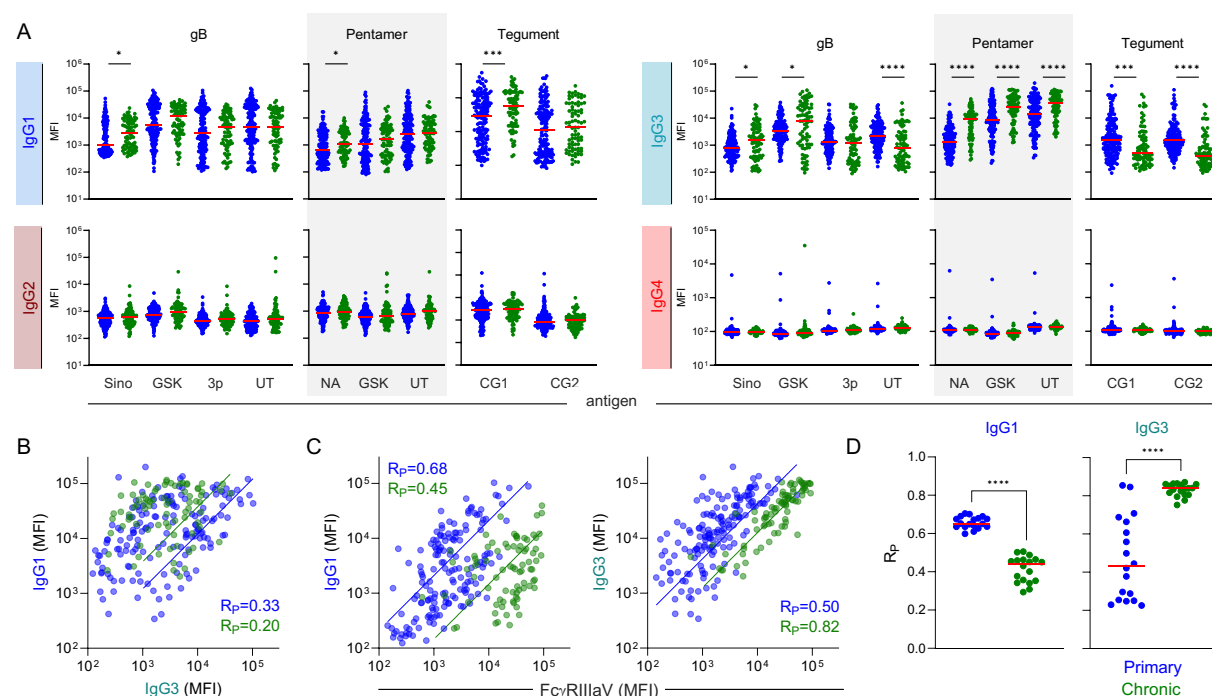


Figure 2: IgG Subclasses display distinct profiles in primary and chronic infection depending on antigen specificity. **A.** Levels of gB-, pentamer-, and tegument-specific IgG1, IgG2, IgG3, and IgG4 antibodies in individual with primary (blue) or chronic (green) CMV infection. **B.** Scatterplot and fit lines of levels of (UT) pentamer-specific IgG1 and IgG3 responses by infection status. Pearson correlation coefficient (R_p) indicated in inset. **C.** Scatterplot and fit line of levels of (UT) pentamer-specific IgG1 (left) and IgG3 (right) versus (UT) pentamer-specific FcγRIIIaV-binding antibodies. Pearson correlation coefficient (R_p) indicated in inset. **D.** Pearson correlation coefficients (R_p) for each pentamer antigen tested across FcγR by infection status. Solid lines denote group medians; differences between groups were assessed by Mann-Whitney test (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, and **** $p < 0.0001$); values presented are median fluorescent intensities (MFI).

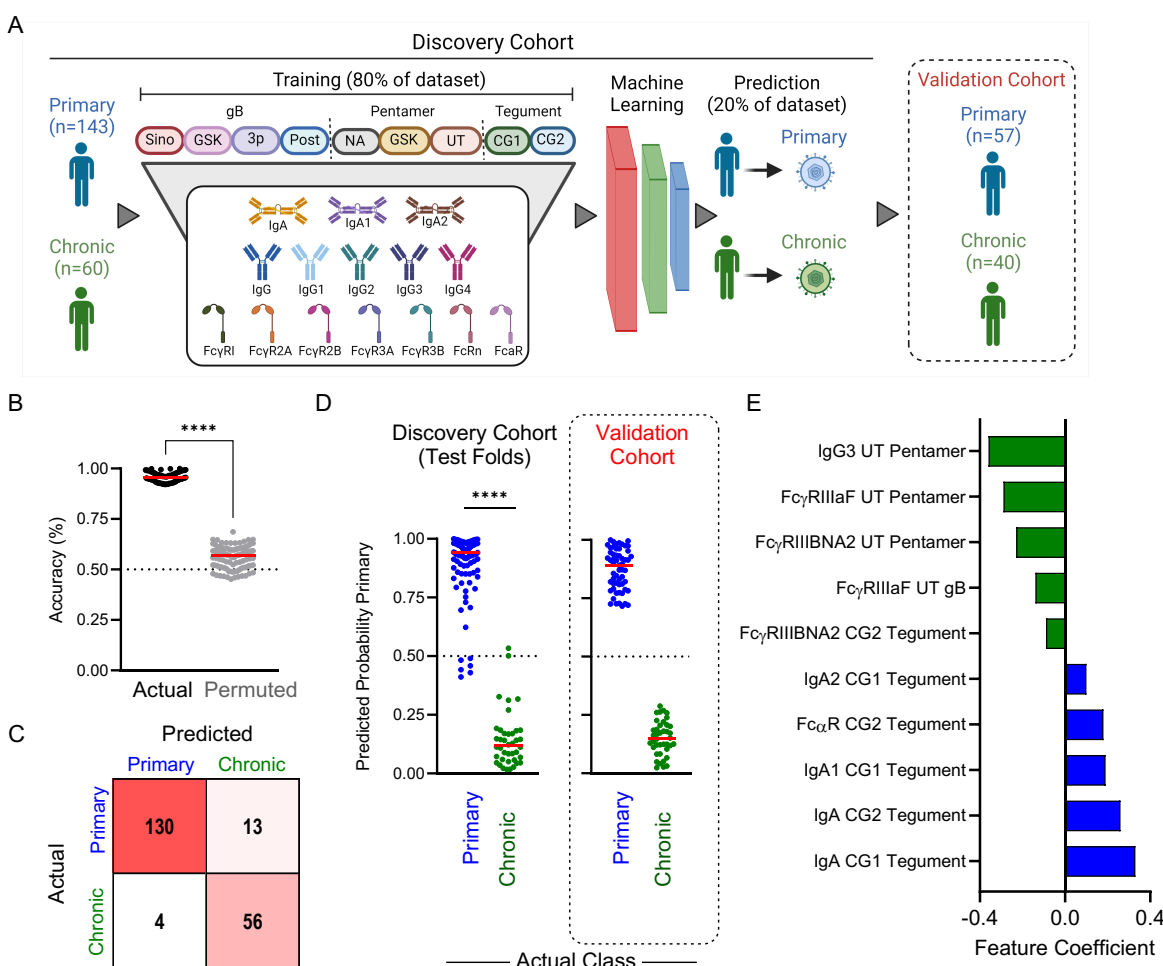


Figure 3: Machine learning accurately predicts primary or chronic CMV infection status. A.

Schematic overview of cross-validated machine learning workflow employing antibody profiling data to discriminate between primary and chronic infection status in discovery and validation cohorts. **B.** Prediction accuracy for 100 repeated 5-fold cross-validation runs on actual (black) and permuted (gray) class labels. **C.** Confusion matrix of predicted versus actual class labels in the median model for 5-fold cross validation. **D.** Class probabilities of each sample in the discovery set when evaluated as a test sample in the cross-validation run exhibiting median performance (left) and for the validation cohort using the final model (right). **E.** The identities and coefficients of the features making the largest positive (n=5) or negative (n=5) contributions to the final model. Solid lines denote group medians; differences between groups or conditions were assessed by Mann-Whitney test (*p < 0.05, **p < 0.01, ***p < 0.001, and ****p < 0.0001).

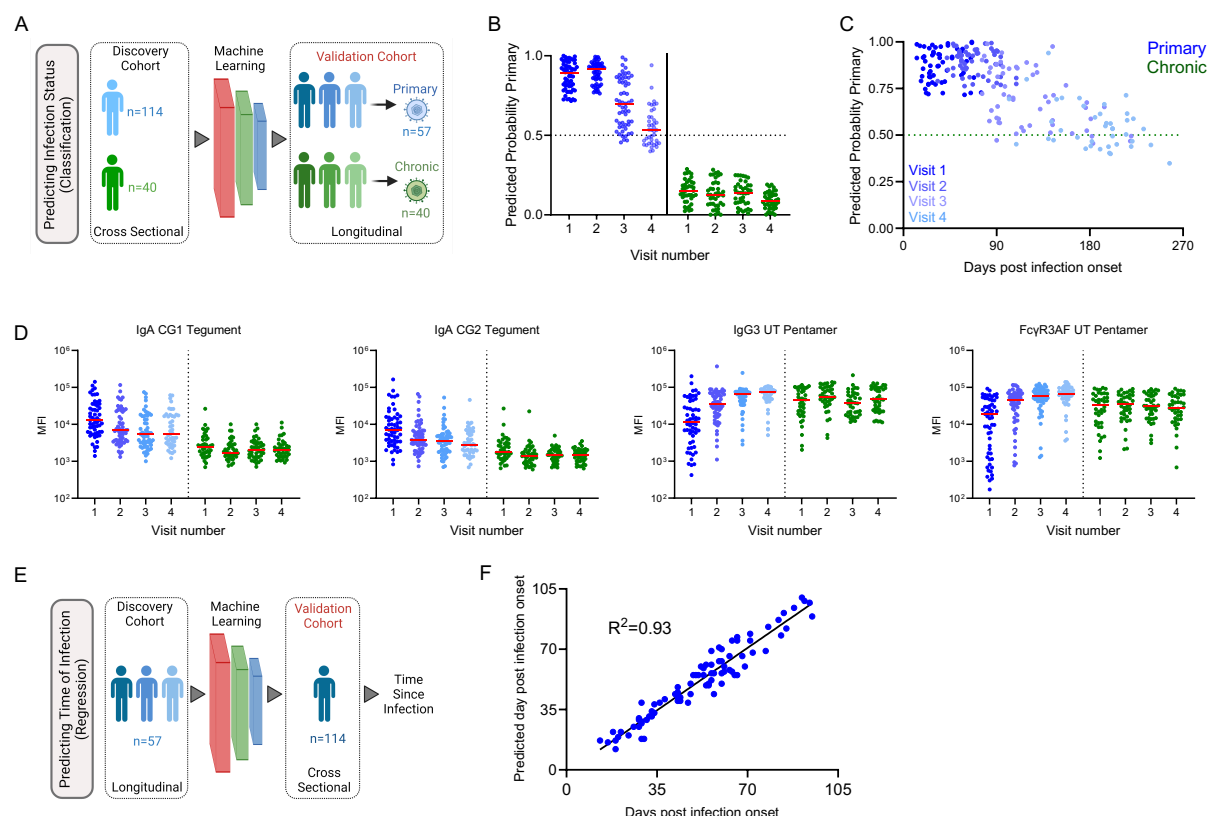


Figure 4: Longitudinal models define a molecular clock of CMV primary infection. **A.** Analysis overview. The infection status classification model trained on the cross-sectional cohort was applied to longitudinal samples available from the validation cohort. **B.** Class probabilities of each sample in longitudinal cohort over sample collection visits for individuals with primary (blue) and chronic (green) infection. **C.** Scatterplot of class probabilities for subjects defined as having primary infection at visit 1 over time. **D.** Scatterplots of features employed by classification model to predict infection status over time in the longitudinal cohort. **E.** Analysis overview. Primary infection samples from the longitudinal cohort samples were used to train a regression model to predict time since infection (days post symptom onset) that was applied to the primary samples from the cross-sectional cohort. **F.** Scatterplot of model predictions of time since infection when primary samples used for predicting days post symptom onset. Data shows the measure of the predicted label and its closeness to the true label.

Table 1. Cohort characteristics

Values that were not available are listed as NA.

[illegible]