


An epigenetic human cytomegalovirus infection score predicts viremia risk in seropositive lung transplant recipients

Fei-Man Hsu^{a,b}, Rashmi P. Mohanty^c, Liudmilla Rubbi^a, Michael Thompson^a, Harry Pickering^d, Elaine F. Reed^d, John R. Greenland^c, Joanna M. Schaeleman^e, and Matteo Pellegrini ^{a,b}

^aDepartment of Molecular, Cell and Developmental Biology, University of California Los Angeles, Los Angeles, CA, USA; ^bInstitute for Quantitative and Computational Biosciences – The Collaboratory, University of California Los Angeles, Los Angeles, CA, USA; ^cDepartment of Medicine, University of California San Francisco, San Francisco, CA, USA; ^dDepartment of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA; ^eDepartment of Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA

ABSTRACT

Cytomegalovirus (CMV) infection and reactivation in solid organ transplant (SOT) recipients increases the risk of viremia, graft failure and death. Clinical studies of CMV serostatus indicate that donor positive recipient negative (D^+/R^-) patients have greater viremia risk than D^-/R^- . The majority of patients are R^+ having intermediate serologic risk. To characterize the long-term impact of CMV infection and assess viremia risk, we sought to measure the effects of CMV on the recipient immune epigenome. Specifically, we profiled DNA methylation in 156 individuals before lung or kidney transplant. We found that the methylome of CMV positive SOT recipients is hyper-methylated at *loci* associated with neural development and Polycomb group (PcG) protein binding, and hypo-methylated at regions critical for the maturation of lymphocytes. In addition, we developed a machine learning-based model to predict the recipient CMV serostatus after correcting for cell type composition and ancestry. This CMV episcore measured at baseline in R^+ individual stratifies viremia risk accurately in the lung transplant cohort, and along with serostatus the CMV episcore could be a potential biomarker for identifying R^+ patients at high viremia risk.

ARTICLE HISTORY

Received 21 March 2024
Revised 28 August 2024
Accepted 4 September 2024

KEYWORDS

Cytomegalovirus; DNA methylation; epigenetics; kidney transplantation; lung transplantation; biomarker



Introduction


Human cytomegalovirus (CMV) is a β -herpesvirus that typically resides in the host in a latent form without causing overt symptoms [1]. CMV infects around 60–90% of adults worldwide [2], and the global CMV seroprevalence rate ranges from 45 to 100% among women of reproductive age [1]. In certain conditions, such as a weakened immune system, CMV can reactivate into the lytic phase and cause viremia, leading to symptomatic infection [2].

The characteristic owl's eye inclusions caused by CMV infection were first seen in stillbirths in 1910 and later among patients undergoing solid organ transplantation (SOT) [3]. Seropositive organ donors (D^+) have approximately a 78% chance to transmit CMV to seronegative recipients (R^-) [4]. About 40% of seropositive organ recipients (R^+) reactivate latent CMV during immunosuppression therapy post-

transplantation, and those with seropositive donors can also be reinfected with new strains of CMV [4,5]. CMV reactivation can cause viremia, allograft rejection, and end-organ diseases post-transplantation [4,6,7]. Serologic risk groups are determined based on the CMV serostatus of recipient and donor combinations. A study of 653 renal transplant patients indicated that D^+/R^- (high sero-risk) has a CMV viremia odds ratio 87.46 compared to the D^-/R^- (low sero-risk), and R^+ (intermediate sero-risk) slightly increases the risk but does not arrive at statistical significance [8]. Given the fact that more than half of the human population are CMV seropositive with intermediate sero-risk, a comprehensive CMV viremia risk assessment method would be advantageous to most of the SOT patients.

Peripheral blood DNA methylation is one of the most studied epigenetic modifications and

CONTACT Matteo Pellegrini  matteop@mcdb.ucla.edu  Department of Molecular, Cell and Developmental Biology, University of California Los Angeles, Los Angeles, CA 90095, USA

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/15592294.2024.2408843>

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

reflects the cumulative record of lifetime exposures that are associated with several non-commutable diseases, including Alzheimer's, multiple sclerosis, type 2 diabetes, systemic lupus erythematosus and cardiovascular disease. In many cases methylation signatures have high predictive value and could be used to predict health outcomes [9,10]. The epigenetic clock [11–13] and epigenetic pacemaker [14] have been widely studied and predict chronological ages. However, less is known about the association of DNA methylation with infections. Recently, the Milieu Intérieur Consortium reported that latent CMV infection drives DNA methylation variation in blood through the regulation of host transcription factors in a cell-composition-mediated manner [15]. This raises the possibility that end-organ diseases post-transplantation and other sequelae of SOT are also affected by alterations in the DNA methylome of immune cells.

To investigate the DNA methylation perturbation due to CMV infection, we leveraged longitudinal pre- and post-transplant biorepository samples from research participants undergoing kidney and lung transplant at UCLA and UCSF and applied targeted-bisulfite sequencing (TBS-seq) to profile their peripheral blood mononuclear cells (PBMCs). In contrast to previous epigenome-wide association studies (EWAS) that used the Infinium methylation array, our approach allows us to select regions of interest that are more likely to be impacted by CMV infection. We built a model from the pre-transplant (PreTx) samples that can predict CMV serostatus even post-transplantation (PostTx). We identified CMV-associated CpG sites after correcting for the effect of cell composition and ancestry. Finally, we demonstrated that our epigenetic CMV measure named CMV episcore serves as an epigenetic biomarker to estimate time to viremia following SOT.

Materials and methods

Human subjects

Kidney transplantation cohort

The study procedures, informed consent, and data collection documents were reviewed and approved by the Institutional Review Board of the UCLA

(IRB#11-001,387). Chart review was performed to acquire demographic and clinical data including results of CMV PCR surveillance. Participants provided blood samples on the day of transplantation (PreTx) and 3 months after (PostTx).

Lung transplantation cohort

Lung transplant candidates were recruited to participate in a longitudinal database and biorepository, as previously described [16]. The study procedures, informed consent, and data collection documents were reviewed and approved by the Institutional Review Board of the UCSF (IRB#13-10,738). Participants provided blood samples prior to induction immunosuppression during the time of transplantation (PreTx) and at a clinically indicated bronchoscopy around 12 months after (PostTx).

Blood samples

8 ml of blood was drawn into the ACD tube. After Ficoll density gradient centrifugation, PBMCs were separated isolated and cryopreserved in FCS/DMSO.

Targeted bisulfite sequencing (TBS-seq)

Probe design

The probe panel design is based on the criteria to include CpG *loci* that 1) covers DNA methylation clock age estimators [12,13], 2) has cell-type specificity, and 3) locates in the promoter regions (–1000 to +250bp from TSS) of viral response genes [17]. Biotinylated probes covering the selected CpG *loci* were synthesized by IDT (Integrated DNA Technologies). The coordinates of the targeted regions (GRCh38) are listed in Supplementary Table S1.

Library preparation

Genomic DNA was extracted from PBMCs using phenol-chloroform method [18]. 500ng genomic DNA was sheared and subject to end-repair, A-tailing and ligated with methylated adaptors. Purified libraries were hybridized to biotinylated probes and subjected for bisulfite conversion (Zymo Cat# D5030). Captured DNA was PCR amplified with KAPA HiFi HotStart Uracil⁺

(Cat# KK2801) into a final TBS-seq library. Library quality was evaluated using TapeStation with the high-sensitivity D1000 tape (Agilent Cat#5067–5584). A comprehensive TBS-seq protocol is demonstrated in [17].

TBS-seq data processing

Adapter sequences were trimmed off from the raw reads using Cutadapt [19] and only reads with minimum 30bp were kept for downstream analysis. Reads were aligned to GRCh38 reference genome using *bsbolt align* function and the duplicated reads were marked with *samtools markdup* function before calling methylation using *bsbolt call-methylation* function [20]. CGmaps from all samples were aggregated into one methylation count matrix using *bsbolt aggregatematrix* function with parameters *-min-coverage 10 -min-sample 1.0*.

Cell type deconvolution

A reference-based cell type deconvolution approach was used to estimate cell type composition with DNA methylation profiles [21]. To recapitulate cell type composition of PBMC, WGBS dataset from 6 cell types: B cell, CD4 T cell, CD8 T cell, NK cell, naïve T cell and monocyte were acquired from GSE186458 [22], and neutrophil band cell's methylation profiles were from the BluePrint database [23]. In total, 34 WGBS profiles from each cell type with replicates were included (Supplementary Table 2). Cell type-specific differentially methylated regions (DMRs) were identified by one-vs-all comparisons using *metilene* [24] with the criteria to find DMRs that are 1) at least 500bp, 2) with the delta methylation level $< -30\%$, and 3) with a false discovery rate (FDR) < 0.05 . Cell type-specific CpG sites were further subtracted from each TBS-seq sample with *bedtools intersect* function and used as input files to deconvolute. A non-negative least square approach was applied to every TBS-seq profile and regress to the WGBS references for coefficient estimation. The detailed deconvolution data could be found in Supplementary Figure S1.

Methylation modeling

Multivariate multiple linear regression (MMLR)

Per recruit i , the methylation status of targeted locus j is denoted as M_{ij} . Suppose every M_{ij} is described by k methylation-associated traits, *i.e.* multivariate, of each recruit T_{ik} , that are weighted by per site coefficient C_{kj} , the methylation model is formulated as Equation 1

$$M_{ij} = T_{ik} \times C_{kj} \begin{cases} i \in \text{number of recruits} \\ j \in \text{number of CpG loci} \\ k \in \text{number of traits} \end{cases} \quad (1)$$

This model represents a system of equations in which T_{ik} and M_{ij} are known variables. Our goal is to derive critical CpG sites for each trait, *i.e.* the unknown C_{kj} that represents characteristics of sites, and could be achieved by solving Equation 2 as,

$$C_{kj} = T_{ik}^\dagger \times M_{ij} \quad (2)$$

Here, T_{ik}^\dagger is derived through Moore-Penrose pseudoinverse of T_{ik} .

Leave-one-out cross validation (LOOCV)

To avoid over-fitting, for each biological sample, a separate MMLR model was trained with the rest samples to derive C , and the trait prediction is made as,

$$T_{ik} = M_{ij} \times C_{kj}^\dagger \quad (3)$$

Here, C_{kj}^\dagger is the Moore-Penrose pseudoinverse of C_{kj} .

Identification of CMV serostatus-associated CpG sites

To identify statistically significant associations between CMV serostatus and the methylation per site, for each locus we estimated the significance of the coefficients described below,

$$y_m = \beta_0 + \beta_1 x_{m1} + \beta_2 x_{m2} + \dots + \beta_n x_{mn} + \varepsilon \quad (4)$$

Here, y_m is the methylation level at locus m , x_n are the explanatory variables including age, sex, CMV serostatus, cohort, cell type PCs, and ancestry PCs. β_0 is the y-intercept, β_n is the coefficient for each explanatory variable, and ε is the error.

For each CpG site y_m , p values from the model per explanatory variable x_n were derived and adjusted for multiple hypothesis testing with Benjamini-Hochberg correction. CpG sites with adjusted p value <0.05 regarding CMV serostatus as x were determined as CMV associated sites.

Cox proportional hazards (coxph) model

The sero-risk per subject was categorized into high (D^+/R^-), intermediate (R^+) and low (D^-/R^-). Both sero-risk and the CMV episcore derived from the MMLR prediction of CMV status were treated as covariates of the Coxph model to estimate the rate of CMV viremia over the tracking time, *i.e.* 3 months for the kidney cohort and 12 months for the lung cohort. The Coxph regression analysis was performed with R package ‘survival’ and plotted with R package ‘ggsurvplot.’

Functional enrichment analysis

Site-level GO enrichment analysis was performed using GREAT [25] with CMV associated sites’ coordinates as foreground and all TBS-seq captured CpG sites as background. Gene-level GO enrichment analysis was conducted with Enrichr [26]. TFBS enrichment analysis were performed using Cistrome [27].

RNA-seq

Library preparation

RNA was extracted from PBMC. mRNA libraries were constructed using KAPA RNA HyperPrep Kit following the manufacturer’s instruction (Cat#KK8540). RNA was sheared and primed with oligos for cDNA synthesis. After adapter ligation and PCR amplified, the final library was quantified and accessed quality with TapeStation.

RNA-seq data processing

Reads were aligned to reference genome GRCh38 using STAR using default parameters [28]. A gene count table was generated using featurecount [29]. Gene count matrix was first normalized and DEGs were identified using R package DESeq2 (Love *et al.*, 2014). Genes with adjusted p value <0.05 and at least 2-fold difference between CMV serostatus were considered differentially expressed.

Results

Latent CMV infection alters the host blood methylome

The study design is illustrated in Figure 1. Participants in the kidney cohort underwent transplantation between April 2015 and September 2021, and the lung cohort was recruited between September 2015 and November 2020. Blood samples were obtained on the day of transplantation from 84 lung and 72 kidney participants, and a subset of 57 participants also provided blood samples after SOT at about 12 and 3 months. The participant characteristics are shown in Table 1 and the correlation between traits is shown in Supplementary Figure S2a.

To profile DNA methylation we used targeted bisulfite sequencing (TBS-seq). We designed the TBS-seq probe panel to capture sites that are associated with age [12,13] along with cell type specific regions and the promoters of genes that mediate responses to viral infections (Supplementary Table S1). Our assay captured 37,379 CpG sites with a minimum coverage of 10X across all samples (Supplementary Figure S2b).

To identify the factors that drive variation in methylation across our cohort we used Uniform Manifold Approximation and Projection (UMAP), to visualize our samples in two dimensions. In PreTx samples, we found that in addition to sex, which is the main factor driving clustering of samples, CMV serostatus further separates individuals (Figure 2a). Principal component analysis (PCA) of DNA methylation also shows correlation with CMV serostatus to PC1 (Supplementary Figure S2c). These results suggest that latent CMV infection has a significant impact on the host methylome.

CD8 T cell composition is associated with CMV serostatus

It is well established that a significant driver of changes in DNA methylation in PBMCs is cell type composition. To account for this effect, we carried out cell type deconvolution using cell-type specific DNA methylation sites (see Materials and Methods and Supplementary Figure S1) to estimate the percentage of B cells, CD4, CD8 and

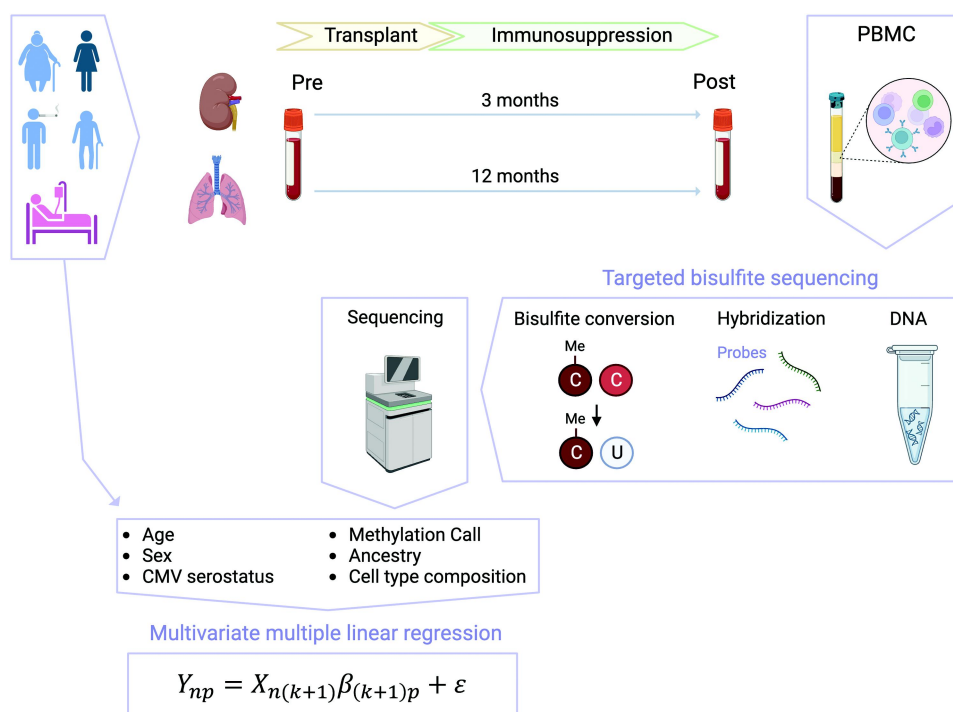


Figure 1. Schematic of this CMV study.

Renal transplant candidates at UCLA were recruited to participate before kidney transplantation, while lung transplant candidates listed for transplantation were recruited at UCSF. Participants provided blood on the day of transplantation and followed up after 3 months (kidney) and 12 months (lung) for PostTx sampling. PBMCs were isolated and genomic DNA was extracted for TBS-seq. After sequencing, the DNA methylation profiles were analyzed and modeled with the multivariate multiple linear regression (MMLR) model. The illustration is created with BioRender.

Table 1. Participant characteristics.

	Kidney	Lung
N	72	84
Age median [IQR]	52.3 [40.5, 59.0]	63.8 [58.4, 67.3]
Sex = Male, N (%)	43 (59.7)	34 (40.5)
Ethnicity, N (%)		
Asian	7 (9.9)	4 (4.8)
Black	10 (14.1)	6 (7.1)
Hispanic	38 (53.5)	21 (25.0)
Other	2 (2.8)	2 (2.4)
White	14 (19.7)	51 (60.7)
Double transplant, N (%)	59 (83.1)	80 (95.2)
Cytomegalovirus serostatus, N (%)		
D ⁻ /R ⁻	6 (8.3)	12 (14.3)
D ⁻ /R ⁺	18 (25.0)	13 (15.5)
D ⁺ /R ⁻	17 (23.6)	19 (22.6)
D ⁺ /R ⁺	31 (43.1)	33 (39.3)
CMV viremia after Tx, N (%)	18 (25.0)	27 (32.1)
PostTx sequenced, N (%)	18 (25.0)	2 (2.4)
Post-transplant/Pre-transplant, N (%)	50/72 (69.4)	7/84 (8.3)

naïve T cells, along with NK cells, monocytes, and neutrophils in each sample. Among these cell types, only CD8⁺ T cells were significantly increased in CMV⁺ patients (Figure 2b).

DNA methylation serves as a predictor variable of CMV serostatus

We next asked if DNA methylation could be used to predict CMV serostatus. We trained a Leave One Out

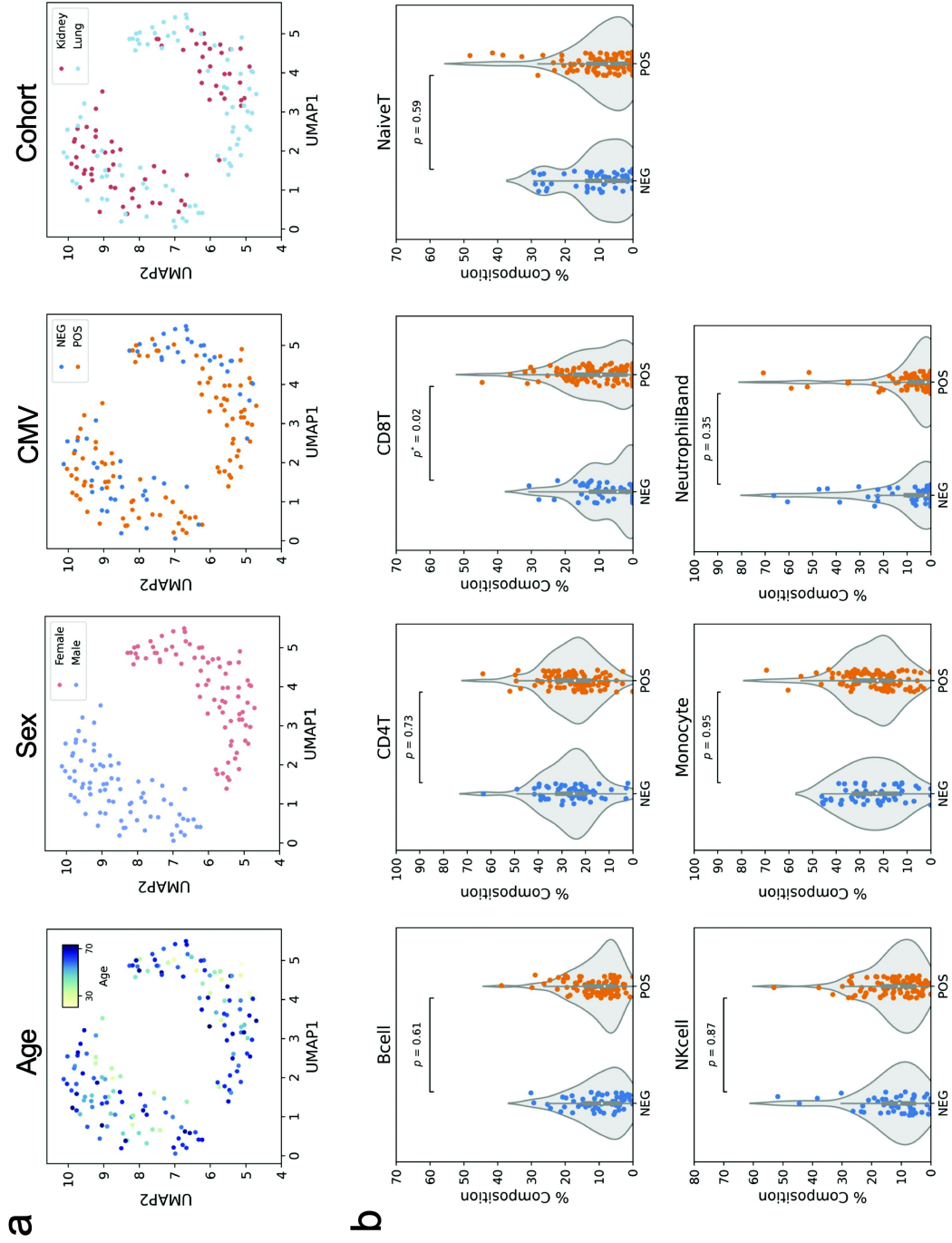


Figure 2. CMV latent infection alters the blood methylome and cell-type composition. (a) UMAPs show the naive clustering of DNA methylation groups with sex as the major and CMV serostatus as minor factor differentiating samples. (b) Cell-type composition estimated from DNA methylation suggests CD8 T cells increased in CMV seropositive subjects.

Cross Validated (LOOCV) penalized logistic regression model in a cohort of 156 PreTx samples (Supplementary Figure S3a). The trained DNA methylation model excluding the test individual was first cross-validated to optimize the hyper parameter and then correctly predicted recipient CMV status for 115 participants (74.7%) (p value $< 0.001^{***}$, Spearman's rank correlation), corresponding to an area under curve (AUC) of 0.68.

To determine whether the effects of CMV are mediated by changes in cell type composition and other covariates we used LOOCV to assess a multivariate multiple linear regression model (MMLR) with DNA methylation as the response variable and multiple scaled traits including age, sex and CMV serostatus as dependent variables (Figure 3). We also included into the MMLR model our estimated cell type PCs (Supplementary Figure S3b) and ancestry PCs derived from genotypes inferred from the targeted bisulfite sequence data. By inverting the model, we could also predict the values of each trait for each individual from their DNA methylation profiles (see Materials and Methods).

Figure 3a shows the spearman correlation between the predicted and actual values of traits in our MMLR

model. We see that the predicted levels of CMV serostatus have a 0.47 correlation coefficient with the actual value (p value $< 0.001^{***}$, Figure 3b). We named the predicted CMV serostatus from the MMLR model the 'CMV episcore' and find that the associated AUC is 0.78 (Figure 3c). These results indicate that DNA methylation can be used to predict CMV serostatus even when other covariates are considered.

We next asked if CMV serostatus of individuals could be predicted after SOT. We used the PreTx LOOCV MMLR model to predict the CMV serostatus of the 56 PostTx individuals, and the predict-actual correlation coefficient is 0.46 with AUC 0.81 and p value $< 0.01^{**}$ (Supplementary Figure S3c). To avoid overfitting in this analysis, we left out the PreTx individual when predicting their paired PostTx sample. This result suggests that CMV episcores are stable across 3–12 months after organ transplantation.

Characterization of CpG sites that are associated with CMV latent infection

We next examined CpG sites associated with the CMV serostatus. We modeled the methylation level

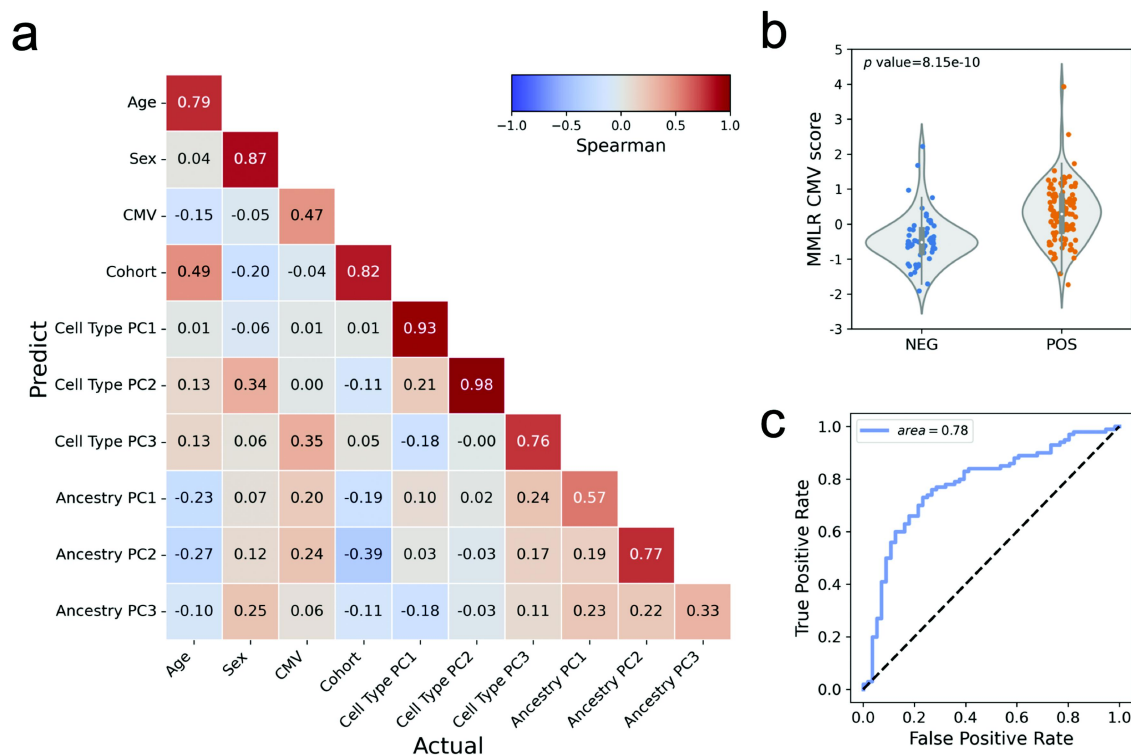


Figure 3. Multivariate multiple linear regression methylation model.

(a) Spearman's correlation matrix of the actual (x-axis) and predicted (y-axis) traits. (b) Violin plot with data points shows the predicted value, CMV episcore, has a significant difference in distribution between CMV seropositive and seronegative recipients. (c) ROC curve of the MMLR prediction model for CMV serostatus.

per site using a multiple linear regression model and characterized CpG *loci* with adjusted p value <0.05 (see Materials and Methods). The sign (+ or -) of the coefficient of the CMV serostatus determines whether the sites are hyper- and hypo-methylated in CMV seropositive recipients. We found 2,217 CpGs that are hyper-methylated in CMV seropositive recipients and 1,535 CpGs that are hypo-methylated (Supplementary Figure S4a). Genes that are proximal to these sites are listed in Supplementary Table 3 (hyper-methylated) and Supplementary Table S4 (hypo-methylated). Hyper-methylation of genes in promoters typically reduces gene expression. The hyper-methylated CpGs are enriched in neural functions such as synapse assembly and neuroactive ligand-receptor interaction (Figure 4a). Gene ontology analysis of hypo-methylated genes in CMV seropositive patients, on the other hand, showed enrichment in hematopoietic cell lineage, T cell receptor complex, major histocompatibility complex (MHC) protein binding, and T cell activation. Interestingly, we also saw both hyper- and hypo-methylation of the ACE2 locus, the receptor for the spike glycoprotein of the human coronavirus SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2), as shown in Supplementary Figure S4b.

To further characterize these differentially methylated sites we asked if they were enriched for the binding of specific transcription factors that might be regulating their methylation levels. This analysis was performed using Cistrome [27] (Figure 4b). CMV latent infection associated hyper-methylated CpG sites are enriched in PcG proteins including RNF2, EZH2, KDM2B and JARID2. By contrast, CpG sites that are hypo-methylated in CMV seropositive cases mainly reside in the transcription factor binding sites (TFBSs) of 1) the NF κ B pathway, including RELA and RELB, and 2) AP-1 transcription factor members including BATF, BATF3, JUNB, JUND and FOSL1.

Gene expression related to T cell receptorsignaling and activation is up-regulated in CMV seropositive patients

To determine whether the DNA methylome alterations we observed to be associated with

CMV seropositivity are accompanied by gene expression changes, we performed bulk RNAseq in a subset of 23 CMV seropositive and 14 CMV seronegative individuals in the kidney transplant cohort. 194 up- and 121 down-regulated genes were identified between CMV positive and negative individuals (Figure 5a, Supplementary Tables S5&6). The CMV seropositive up-regulated genes are enriched in the antigen receptor mediated signaling pathway, T cell receptor signaling pathways, and T cell activation (Figure 5b), and these GO terms are similar to those we found enriched among the hypo-methylated genes (Figure 5b). The down-regulated genes are enriched in cellular response to increased oxygen level, regulation of autophagy, and intracellular pH reduction (Figure 5b).

As shown in the PCA, CMV seronegative status is separated from CMV seropositivity by the PC1 axis, suggesting the latent CMV infection also impacts the human blood transcriptomes (Figure 5c). We identified 5 genes that are both differentially methylated and expressed based on CMV serostatus. One of these, BRD4 (Bromodomain-containing protein 4), is a histone acetylation reader that regulates CMV latency and reactivation [30], and whose methylation is significantly increased in CMV seropositive patients [15]. We found 1 CpG site downstream of BRD4 to be hyper-methylated, and BRD4 is down-regulated in CMV seropositive patients (Figure 5d). Interestingly, we found CD8A and CD8B, subunits of CD8 antigen receptor complex on T cell surfaces, are hypo-methylated and up-regulated in CMV seropositive participants (Figures 5e,f).

CMV epigenetic score as a biomarker to predict CMV viremia

Traditional serologic risk assessment trichotomized SOT patients into high (D^+/R^-), intermediate (R^+) and low (D^-/R^-) based on the serostatus in donor and recipient to avoid CMV viremia and its sequelae such as graft rejection and death. We therefore asked if the CMV score estimated from the MMLR model (CMV episcore) could be used to improve risk assessment for CMV viremia. Specifically, we hypothesized that increasing CMV-specific immune

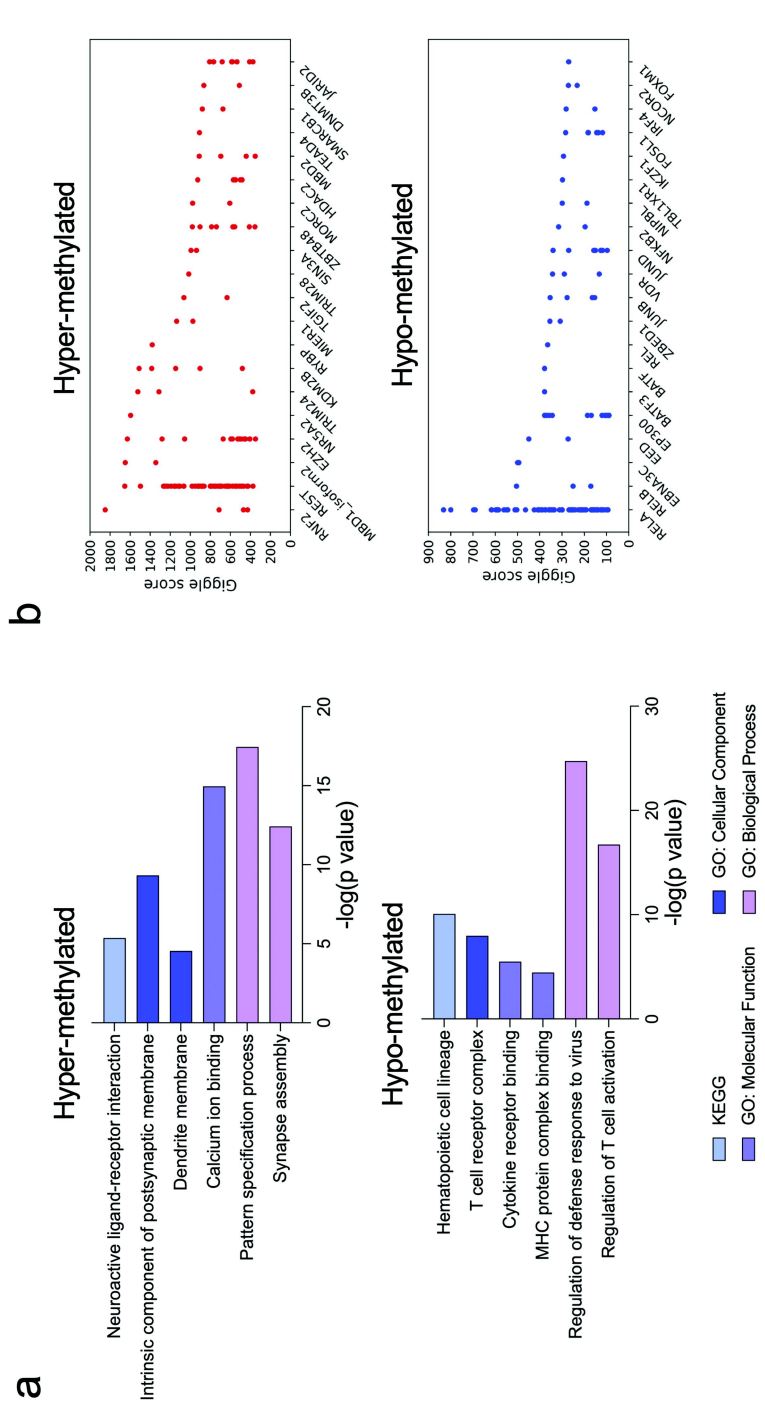


Figure 4. CMV latent infection is associated with hypermethylation of neural system genes and hypomethylation of immune response genes. (a) Gene ontology of hyper-methylated (top) and hypo-methylated (bottom) CpG sites. (b) Enrichment of TFBSs of hyper-methylated (top) and hypo-methylated (bottom) CpG sites.

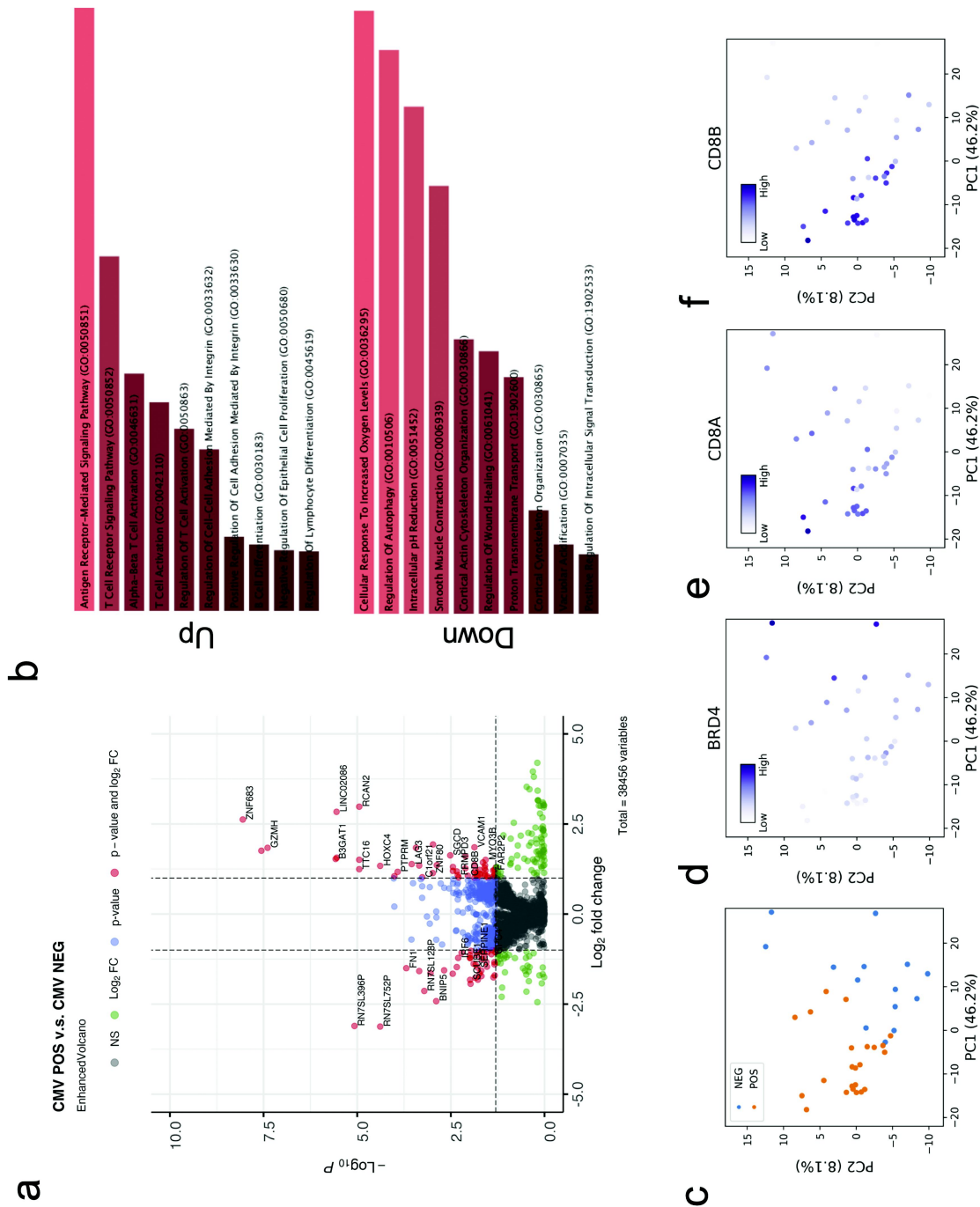


Figure 5. Transcriptomic impact of CMV latent infection.

(a) Volcano plot of differentially expressed genes regarding CMV serostatus. (b) Gene ontology of up- and down-regulated genes in CMV seropositive subjects.

(c) PCA of transcriptomes shows separation according to CMV serostatus. (d-f) Feature plots of BRD4, CD8A and CD8B that are both differentially methylated and expressed.

responses, as captured by this methylation score, would be associated with decreased risk of viremia. In both transplantation cohorts, we measured the hazard ratio of CMV viremia over 12 months (kidney) and 5 years (lung). The Cox proportional hazard regression models show that the CMV episcore (HR 0.74, 95% CI 0.57–0.96, $p^* = 0.02$) is a significant covariate in the lung cohort. Graphically, we showed that when the R^+ group is stratified by CMV episcore, the CMV episcore low group had worse survival than the high episcore group (Figure 6a). In the kidney model, the CMV episcore was not statistically associated with time-to-viremia (Figure 6b), but still from the cumulative survival plot the CMV episcore low group bears the worst hazard ratio. The differences in statistic power between cohorts could be due to the duration of the follow-up window and the immunosuppression therapies (see Discussion). In conclusion, the CMV episcore is a potential biomarker to predict CMV viremia and is especially useful for the R^+ patients.

Discussion

We identified a significant impact of CMV latent infection on the methylome of PBMC in two centers and across individuals before receiving lung or kidney transplant. Differential methylation patterns went beyond what could be explained by inferred cell compositional differences, affecting genes involved in neural programming and T cell differentiation. We also identified CMV associated changes in RNA expression. Interestingly, a DNA methylation-based CMV episcore was able to predict CMV viremia risk in intermediate sero-risk (R^+) lung transplant recipients, and the hazard ratio of patients with low CMV episcores even surpass the high sero-risk D^+/R^- group. In kidney transplant recipients with intermediate sero-risk (R^+), lower CMV episcores also result in shorter time to viremia, while not showing statistical significance. Since the CMV episcore derived PreTx is associated with CMV viremia risk after SOT, a patient with a low CMV episcore has a high CMV viremia risk independently of the matching donor's CMV serostatus and the resulting CMV prophylaxis should be calibrated accordingly.

We found that CMV infection was linked to increased abundance of CD8 T cells. These observations are consistent with prior studies using flow cytometry that showed the number of CD8 TEMRA is increased in CMV seropositive patients [31–34]. Since the impact of CMV infection on the epigenome could be mediated by multiple factors, we constructed a MMLR model of the methylome that accounted for factors, including age, sex, cell type composition and ancestry. This model allows us to predict CMV serostatus when controlling for these covariates. The MMLR model predicted CMV value, named CMV episcore, is associated with the actual CMV serostatus ($R = 0.47$, Figure 3a). This intermediate correlation holds for dichotomized samples (Figure 3b) and results in a classification power with an AUC of 0.78 (Figure 3c), suggesting the epigenetic perturbation by latent CMV infection is a distributed effect but not localizes to a small fraction of CpGs.

The functional enrichment analyses of CMV associated CpG sites point to hyper-methylation in neural systems genes and hypo-methylation in the immune system genes (Figure 4a). Genes that are hyper-methylated in CMV seropositive individuals include LTBP3, a TGF- β component reported in another study [15], whose mutation/loss-of-function results in hearing loss and otosclerosis [35]. KDM2B is also hyper-methylated and is related to congenital ocular defects [36]. This is consistent with the observations that CMV is one of the leading causes of congenital vision defect and hearing loss in newborns from infected mothers, but it is still unclear if these neural defects are related to DNA methylation at these *loci*.

Hyper-methylated CMV associated CpG sites are also enriched in TFBSs of PcG proteins (Figure 4b). EZH2 is a lysine methyltransferase that methylates H3K9 and H3K27 and represses gene transcription. Together with its regulators KDM2B and JARID2, this PcG complex has been reported to repress GFI1, the MIEP transcriptional repressor of CMV [37]. Hyper-methylated in PcG targeted transcription factor binding sites might release GFI1, repress MIEP, and maintain CMV latent infection from entering the lytic phase. We also found TRIM28 binding sites are hyper-

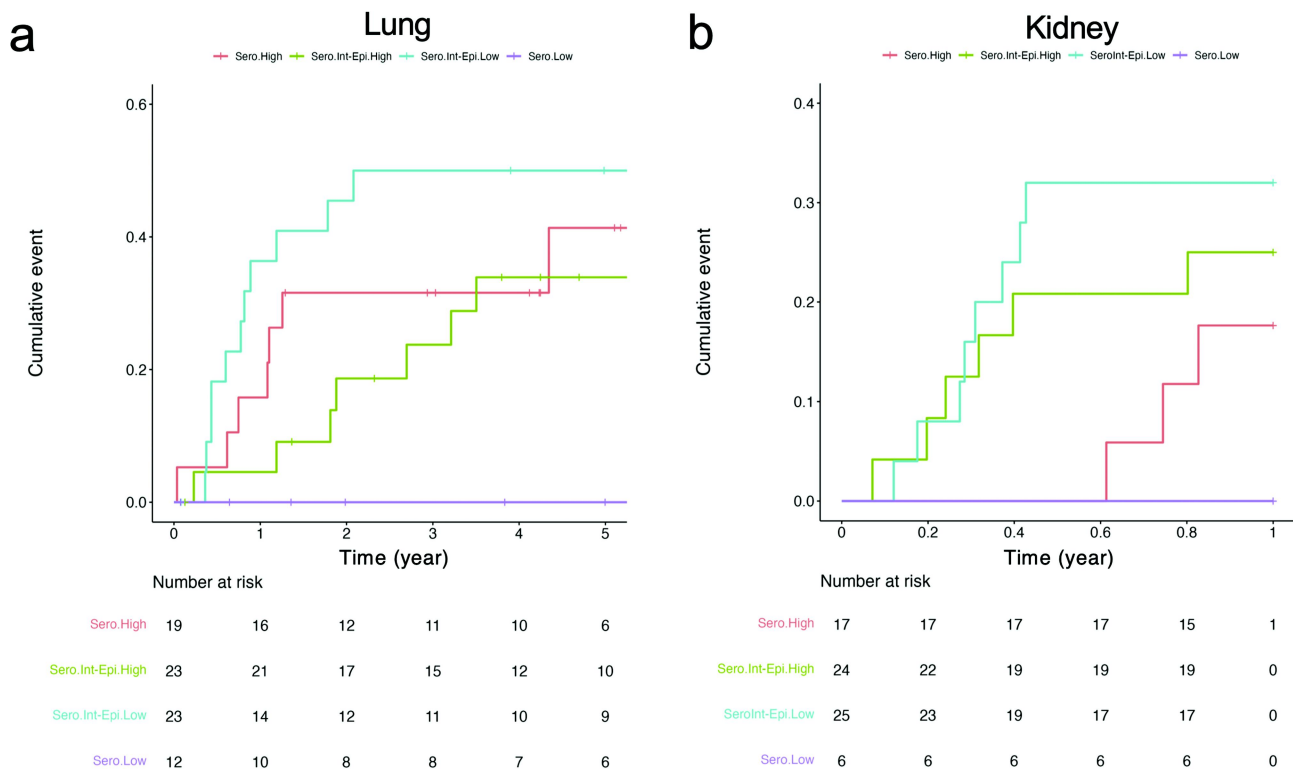


Figure 6. Time-to-viremia event analysis stratified by CMV serostatus risk group and CMV episcore.

(a) In the lung cohort ($n = 77$, 46 events), CMV episcore was associated with increased CMV viremia rates (HR 0.74, 95% CI 0.57–0.96, $p^* = 0.02$). (b) Time to CMV viremia is shown for the renal transplant recipient cohort ($n = 72$, 18 events) shows no statistical difference with CMV episcore as a covariate but still it further frames the sero-risk intermediate group into 2 layers.

methylated in CMV seropositive patients (Figure 4b). A previous study showed that CMV could establish latency in hematopoietic stem cells (HSCs) resulting in lifelong infection [38], and that TRIM28, also known as KAP1, is responsible for switching off viral genes in stem cells to maintain latency [39]. These data are consistent with the hypothesis that TRIM28 regulates CMV latency through DNA methylation.

Genes that are hypo-methylated in CMV seropositive individuals include the T cell receptor subunits CD3D and CD3G, and the co-receptors of CD8 T cells, CD8A and CD8B (Figure 4a). Hypo-methylated CMV associated CpG sites are also enriched in TFBSs of NFkB and AP-1 (Figure 4b). NFkB and AP-1 are both proinflammatory TFs triggered by cytokines and act downstream of T cell receptors (TCRs) to activate T cells [40–43]. The fact that these sites are hypo-methylated in CMV positive individuals, suggests that the corresponding immune responses might be up-regulated. Hypo-methylation at these *loci* was also associated with concordant up-

regulation of related genes/pathways (Figure 5). These results indicate that CMV latent infection may maintain the hosts' immune system primed, especially the cellular immunity, through DNA methylation.

CMV has been reported to be highly prevalent in HIV (Human immunodeficiency virus)-infected subjects, and the coinfection is associated with the risk of cardiovascular and cerebrovascular events [44]. Recent cohort studies showed CMV seropositivity is a risk factor of severity and hospitalization due to SARS-CoV-2 [45,46]. In our study we found that ACE2, the SARS-CoV-2 receptor, is hypo-methylated in CMV seropositive individuals. Although ACE2's expression is low and not increased in the PBMC samples we examined, we hypothesize that the CMV/SARS-CoV-2 superinfection in cell lines reported previously [47] could potentially be regulated by DNA methylation.

We also found that the CMV episcore could improve stratifications of CMV viremia risk in lung transplant recipients who are R⁺ with

intermediate serologic risk. Increased CMV episcore was associated with decreased viremia risk, suggesting that this score broadly assesses CMV immunity and may be a biomarker of viremia risk in this population. In the kidney cohort, a low CMV episcore in the intermediate sero-risk (R^+) group was nominally the highest risk, but this difference did not achieve statistical significance. In fact, the R^+ group has a higher viremia risk than the high sero-risk group (D^+/R^-), and this could be because R^+ patients receive 0 to 3 months CMV prophylaxes whereas the sero-risk high patients receive 6 months of CMV prophylaxis in kidney transplant recipients. We note that all lung transplant recipients were targeted for lifelong CMV prophylaxes. In addition, there were fewer CMV events and shorter follow up time in the kidney cohort, which diminished its statistical power. In conclusion from both cohorts, CMV episcore helps access the CMV viremia risk in hosts facing second infections and could represent the DNA methylation-based trained immunity [48].

Conclusions

Our study reveals the long-term impact of CMV latent infection on the host methylome. Besides describing the differential methylation in CMV seropositive and seronegative patients, we also quantify the PreTx DNA methylome as a CMV episcore to access the viremia risk after SOT and could be beneficial to the broad CMV seropositive organ transplant recipients. As described by the Waddington landscape of cell fate decision [49], DNA methylation may play ‘permissive’ roles in CMV reactivation, congenital disease development, and host immune system priming.

Limitations of the study

The aim of this study is to decipher the impact of CMV infection on the host DNA methylome. Our study has a few limitations. First, the clinical samples were drawn at different timepoints, and the sequencing libraries were constructed and sequenced separately. We therefore made ‘cohort’ a covariate in the MMLR model to resolve this

issue. Second, the patients received lung or kidney transplant and therefore already had chronic diseases which might affect DNA methylation. Another limitation is that the time to viremia estimation using the CMV episcore is only significant in the lung cohort. Whether it is applicable to other SOT studies remains to be addressed with more cohorts. Lastly, although this measure could be beneficial to most of SOT patients with intermediate sero-risk (R^+), applying TBS-seq to every patient might not be feasible. A more specific assay targeting fewer *loci* identified in this study and easier to apply needs to be developed in the future.

Acknowledgments

We thank the UCLA BSCRC Sequencing core for sequence the TBS-seq. This study was supported by the VA Office of Research and Development (CX002011) and the NHLBI (HL151552, JRG). JRG also receives salary support from the CFF (GREENL21AB0, HAYS19AB3, MCDYER22AB0) and NIH (HL161048, HL163294). FH is supported by UCLA QCBio Collaboratory and IDRE postdoctoral fellowships.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The work was supported by the United States Department of Veterans Affairs [JRG].

Author contributions

Formal analysis: FH, MT, MP; Library construction: LR; Probe panel design: HP; Investigation: FH, MP, JRG, JMS; Resources: HP, RPM, JMS, JRG, EFR; Writing- original draft: FH; Writing- review and editing: FH, MP, JRG, JMS, RPM, HP, EFR; Funding: MP, JRG, EFR; Supervision: MP.

Data availability statement

Sequencing data produced in this study are available in Gene Expression Omnibus upon publication. The kidney cohort is under the accession number GSE250536 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE250536>), and the lung cohort is under the accession number GSE253562 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE253562>). The code for MMLR model is deposited to GitLab (<https://gitlab.com/fmhsu0114/cmvepiscore>).

ORCIDMatteo Pellegrini  <http://orcid.org/0000-0001-9355-9564>**References**

- [1] Cannon MJ, Schmid DS, Hyde TB. Review of cytomegalovirus seroprevalence and demographic characteristics associated with infection. *Rev Méd Virol.* 2010;20(4):202–213. doi: [10.1002/rmv.655](https://doi.org/10.1002/rmv.655)
- [2] Griffiths P, Reeves M. Pathogenesis of human cytomegalovirus in the immunocompromised host. *Nat Rev Microbiol.* 2021;19(12):759–773. doi: [10.1038/s41579-021-00582-z](https://doi.org/10.1038/s41579-021-00582-z)
- [3] Hill RB, Rowlands DT, Rifkind D. Infectious pulmonary disease in patients receiving immunosuppressive therapy for organ transplantation. *N Engl J Med.* 1964;271(20):1021–1027. doi: [10.1056/NEJM19641122712001](https://doi.org/10.1056/NEJM19641122712001)
- [4] Atabani SF, Smith C, Atkinson C, et al. Cytomegalovirus replication kinetics in solid organ transplant recipients managed by preemptive therapy. *Am J Transplant.* 2012;12(9):2457–2464. doi: [10.1111/j.1600-6143.2012.04087.x](https://doi.org/10.1111/j.1600-6143.2012.04087.x)
- [5] Jane E G, Super M, Sweny P, et al. Symptomatic cytomegalovirus infection in seropositive kidney recipients: reinfection with donor virus rather than reactivation of recipient virus. *Lancet.* 1988;332(8603):132–135. doi: [10.1016/S0140-6736\(88\)90685-X](https://doi.org/10.1016/S0140-6736(88)90685-X)
- [6] Duan Z, Zhang X, Liu Y, et al. Risk factors and survival of refractory cytomegalovirus reactivation after allogeneic peripheral blood stem cell transplantation. *J Glob Antimicrob Resist.* 2022;31:279–285. doi: [10.1016/j.jgar.2022.10.009](https://doi.org/10.1016/j.jgar.2022.10.009)
- [7] Fishman JA, Emery V, Freeman R, et al. Cytomegalovirus in transplantation – challenging the status quo. *Clin Transplant.* 2007;21(2):149–158. doi: [10.1111/j.1399-0012.2006.00618.x](https://doi.org/10.1111/j.1399-0012.2006.00618.x)
- [8] Chaudhari I, Leung M, Bateni B. Characterization of cytomegalovirus viremia in renal transplant recipients. *Can J Hosp Pharm.* 2022;75(1):6–14. doi: [10.4212/cjhp.v75i1.3249](https://doi.org/10.4212/cjhp.v75i1.3249)
- [9] Thompson M, Hill BL, Rakocz N, et al. Methylation risk scores are associated with a collection of phenotypes within electronic health record systems. *NPJ Genom Med.* 2022;7(1):50. doi: [10.1038/s41525-022-00320-1](https://doi.org/10.1038/s41525-022-00320-1)
- [10] Yousefi PD, Suderman M, Langdon R, et al. DNA methylation-based predictors of health: applications and statistical considerations. *Nat Rev Genet.* 2022;23(6):369–383. doi: [10.1038/s41576-022-00465-w](https://doi.org/10.1038/s41576-022-00465-w)
- [11] Levine ME, Lu AT, Quach A, et al. An epigenetic biomarker of aging for lifespan and healthspan. *Aging Albany Ny.* 2018;10(4):573–591. doi: [10.18632/aging.101414](https://doi.org/10.18632/aging.101414)
- [12] Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol.* 2013;14(10):3156. doi: [10.1186/gb-2013-14-10-r115](https://doi.org/10.1186/gb-2013-14-10-r115)
- [13] Hannum G, Guinney J, Zhao L, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell.* 2013;49(2):359–367. doi: [10.1016/j.molcel.2012.10.016](https://doi.org/10.1016/j.molcel.2012.10.016)
- [14] Farrell C, Snir S, Pellegrini M, et al. The epigenetic pacemaker: modeling epigenetic states under an evolutionary framework. *Bioinform Oxf Engl.* 2020;36(17):4662–4663. doi: [10.1093/bioinformatics/btaa585](https://doi.org/10.1093/bioinformatics/btaa585)
- [15] Bergstedt J, Azzou SAK, Tsuo K, et al. The immune factors driving DNA methylation variation in human blood. *Nat Commun.* 2022;13(1):5895. doi: [10.1038/s41467-022-33511-6](https://doi.org/10.1038/s41467-022-33511-6)
- [16] Wang P, Leung J, Lam A, et al. Lung transplant recipients with idiopathic pulmonary fibrosis have impaired allergic immune responses. *J Hear Lung Transpl.* 2022;41(5):641–653. doi: [10.1016/j.healun.2021.11.012](https://doi.org/10.1016/j.healun.2021.11.012)
- [17] Morselli M, Farrell C, Rubbi L, et al. Targeted bisulfite sequencing for biomarker discovery. *Methods.* 2021;187:13–27. doi: [10.1016/j.ymeth.2020.07.006](https://doi.org/10.1016/j.ymeth.2020.07.006)
- [18] Guha P, Das A, Dutta S, et al. A rapid and efficient DNA extraction protocol from fresh and frozen human blood samples. *J Clin Lab Anal.* 2018;32(1):32. doi: [10.1002/jcla.22181](https://doi.org/10.1002/jcla.22181)
- [19] Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *Embnet J.* 2011;17(1):10–12. doi: [10.14806/ej.17.1.200](https://doi.org/10.14806/ej.17.1.200)
- [20] Farrell C, Thompson M, Tosevska A, et al. BiSulfite bolt: a bisulfite sequencing analysis platform. *Gigascience.* 2021;10(5):giab033. doi: [10.1093/gigascience/giab033](https://doi.org/10.1093/gigascience/giab033)
- [21] Morselli M, Farrell C, Montoya D, et al. DNA methylation profiles in pneumonia patients reflect changes in cell types and pneumonia severity. *Epigenetics.* 2022;17(12):1646–1660. doi: [10.1080/15592294.2022.2051862](https://doi.org/10.1080/15592294.2022.2051862)
- [22] Loyfer N, Magenheimer J, Peretz A, et al. A DNA methylation atlas of normal human cell types. *Nature.* 2023;613(7943):355–364. doi: [10.1038/s41586-022-05580-6](https://doi.org/10.1038/s41586-022-05580-6)
- [23] Martens JHA, Stunnenberg HG. BLUEPRINT: mapping human blood cell epigenomes. *Haematologica.* 2013;98(10):1487–1489. doi: [10.3324/haematol.2013.094243](https://doi.org/10.3324/haematol.2013.094243)
- [24] Jühling F, Kretzmer H, Bernhart SH, et al. Metilene: fast and sensitive calling of differentially methylated regions from bisulfite sequencing data. *Genome Res.* 2016;26(2):256–262. doi: [10.1101/gr.196394.115](https://doi.org/10.1101/gr.196394.115)
- [25] McLean CY, Bristol D, Hiller M, et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol.* 2010;28(5):495–501. doi: [10.1038/nbt.1630](https://doi.org/10.1038/nbt.1630)
- [26] Chen EY, Tan CM, Kou Y, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinform.* 2013;14(1):128. doi: [10.1186/1471-2105-14-128](https://doi.org/10.1186/1471-2105-14-128)
- [27] Zheng R, Wan C, Mei S, et al. Cistrome data browser: expanded datasets and new tools for gene regulatory analysis. *Nucleic Acids Res.* 2018;47(D1):D729–D735. doi: [10.1093/nar/gky1094](https://doi.org/10.1093/nar/gky1094)
- [28] Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29(1):15–21. doi: [10.1093/bioinformatics/bts635](https://doi.org/10.1093/bioinformatics/bts635)

- [29] Liao Y, Smyth GK, Shi W. The subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* 2013;41(10):e108–e108. doi: [10.1093/nar/gkt214](https://doi.org/10.1093/nar/gkt214)
- [30] Groves IJ, Jackson SE, Poole EL, et al. Bromodomain proteins regulate human cytomegalovirus latency and reactivation allowing epigenetic therapeutic intervention. *Proc Natl Acad Sci.* 2021;118(9):e2023025118. doi: [10.1073/pnas.2023025118](https://doi.org/10.1073/pnas.2023025118)
- [31] Fang Y, Doyle MF, Chen J, et al. Circulating immune cell phenotypes are associated with age, sex, CMV, and smoking status in the Framingham heart study offspring participants. *Aging (Albany NY).* 2023;15(10):3939–3966. doi: [10.18632/aging.204686](https://doi.org/10.18632/aging.204686)
- [32] Salumets A, Tserel L, Rumm AP, et al. Epigenetic quantification of immunosenescent CD8+ TEMRA cells in human blood. *Aging Cell.* 2022;21(5):e13607. doi: [10.1111/acel.13607](https://doi.org/10.1111/acel.13607)
- [33] van den BS, Derksen LY, Drylewicz J, et al. Quantification of T-cell dynamics during latent cytomegalovirus infection in humans. *PLOS Pathog.* 2021;17(12):e1010152. doi: [10.1371/journal.ppat.1010152](https://doi.org/10.1371/journal.ppat.1010152)
- [34] Patin E, Hasan M, Bergstedt J, et al. Natural variation in the parameters of innate immune cells is preferentially driven by genetic factors. *Nat Immunol.* 2018;19(3):302–314. doi: [10.1038/s41590-018-0049-7](https://doi.org/10.1038/s41590-018-0049-7)
- [35] Rämö JT, Kiiskinen T, Seist R, et al. Genome-wide screen of otosclerosis in population biobanks: 27 loci and shared associations with skeletal structure. *Nat Commun.* 2023;14(1):157. doi: [10.1038/s41467-022-32936-3](https://doi.org/10.1038/s41467-022-32936-3)
- [36] Fukuda T, Tokunaga A, Sakamoto R, et al. Fbxl10/Kdm2b deficiency accelerates neural progenitor cell death and leads to exencephaly. *Mol Cell Neurosci.* 2011;46(3):614–624. doi: [10.1016/j.mcn.2011.01.001](https://doi.org/10.1016/j.mcn.2011.01.001)
- [37] Sourvinos G, Morou A, Sanidas I, et al. The down-regulation of GFI1 by the EZH2-NDY1/KDM2B-JARID2 axis and by human cytomegalovirus (HCMV) associated factors allows the activation of the HCMV major IE promoter and the transition to productive infection. *PLOS Pathog.* 2014;10(5):e1004136. doi: [10.1371/journal.ppat.1004136](https://doi.org/10.1371/journal.ppat.1004136)
- [38] Sinclair J. Human cytomegalovirus: latency and reactivation in the myeloid lineage. *J Clin Virol.* 2008;41(3):180–185. doi: [10.1016/j.jcv.2007.11.014](https://doi.org/10.1016/j.jcv.2007.11.014)
- [39] Rauwel B, Jang SM, Cassano M, et al. Release of human cytomegalovirus from latency by a KAP1/TRIM28 phosphorylation switch. *Elife.* 2015;4:e06068. doi: [10.7554/eLife.06068](https://doi.org/10.7554/eLife.06068)
- [40] Fujioka S, Niu J, Schmidt C, et al. Nf-κB and AP-1 connection: mechanism of nf-κB-Dependent regulation of AP-1 activity. *Mol Cell Biol.* 2004;24(17):7806–7819. doi: [10.1128/MCB.24.17.7806-7819.2004](https://doi.org/10.1128/MCB.24.17.7806-7819.2004)
- [41] de Los AI, F LH-C, Turner SD, et al. The role of activator protein-1 (AP-1) family members in CD30-positive lymphomas. *Cancers (Basel).* 2018;10(4):93. doi: [10.3390/cancers10040093](https://doi.org/10.3390/cancers10040093)
- [42] Ji Z, He L, Regev A, et al. Inflammatory regulatory network mediated by the joint action of nf-κB, STAT3, and AP-1 factors is involved in many human cancers. *Proc Natl Acad Sci.* 2019;116(19):9453–9462. doi: [10.1073/pnas.1821068116](https://doi.org/10.1073/pnas.1821068116)
- [43] Tussiwand R, Lee W-L, Murphy TL, et al. Compensatory dendritic cell development mediated by BATF–IRF interactions. *Nature.* 2012;490(7421):502–507. doi: [10.1038/nature11531](https://doi.org/10.1038/nature11531)
- [44] Lichtner M, Cicconi P, Vita S, et al. Cytomegalovirus coinfection is associated with an increased risk of severe non-AIDS-Defining events in a large cohort of HIV-Infected patients. *J Infect Dis.* 2015;211(2):178–186. doi: [10.1093/infdis/jiu417](https://doi.org/10.1093/infdis/jiu417)
- [45] Weber S, Kehl V, Erber J, et al. CMV seropositivity is a potential novel risk factor for severe COVID-19 in non-geriatric patients. *PLOS ONE.* 2022;17(5):e0268530. doi: [10.1371/journal.pone.0268530](https://doi.org/10.1371/journal.pone.0268530)
- [46] Alanio C, Verma A, Mathew D, et al. Cytomegalovirus latent infection is associated with an increased risk of COVID-19-Related hospitalization. *J Infect Dis.* 2022;226(3):463–473. doi: [10.1093/infdis/jiac020](https://doi.org/10.1093/infdis/jiac020)
- [47] Perera MR, Greenwood EJD, Crozier TWM, et al. Human cytomegalovirus infection of epithelial cells increases SARS-CoV-2 superinfection by upregulating the ACE2 receptor. *J Infect Dis.* 2022;227(4):543–553. doi: [10.1093/infdis/jiac452](https://doi.org/10.1093/infdis/jiac452)
- [48] Ochando J, Mulder WJM, Madsen JC, et al. Trained immunity — basic concepts and contributions to immunopathology. *Nat Rev Nephrol.* 2023;19(1):23–37. doi: [10.1038/s41581-022-00633-5](https://doi.org/10.1038/s41581-022-00633-5)
- [49] Waddington CH. Organisers and genes. Cambridge, (UK): The Cambridge University Press; 1940.