

Article

Genomic Markers Associated with Cytomegalovirus DNAemia in Kidney Transplant Recipients

Guy Shapira ^{1,2}, Hadas Volkov ^{1,2}, Itai Fabian ¹, David W. Mohr ³, Maria Bettinotti ⁴, Noam Shomron ^{1,2}, Robin K. Avery ⁵ and Ravit Arav-Boger ^{6,*}

¹ Faculty of Medicine, Tel Aviv University, Tel Aviv 69978, Israel; guypersonal@gmail.com (G.S.)

² Edmond J. Safra Center for Bioinformatics, Tel Aviv University, Tel Aviv 69978, Israel

³ Johns Hopkins Genetic Resources Core Facility, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA

⁴ Immunogenetics Laboratory, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA; mbettinotti@jhmi.edu

⁵ Department of Medicine, Division of Infectious Diseases, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA; ravery4@jhmi.edu

⁶ Department of Pediatrics, Division of Infectious Diseases, Medical College of Wisconsin, Milwaukee, WI 53226, USA

* Correspondence: rboger@mcw.edu

Abstract: Human cytomegalovirus (CMV) is a major pathogen after solid organ transplantation, leading to high morbidity and mortality. Transplantation from a CMV-seropositive donor to a CMV-seronegative recipient (D+/R−) is associated with high risk of CMV disease. However, that risk is not uniform, suggesting a role for host factors in immune control of CMV. To identify host genetic factors that control CMV DNAemia post transplantation, we performed a whole-exome association study in two cohorts of D+/R− kidney transplant recipients. Quantitative CMV DNA was measured for at least one year following transplantation. Several CMV-protective single-nucleotide polymorphisms (SNPs) were identified in the first cohort (72 patients) but were not reproducible in the second cohort (126 patients). A meta-analysis of both cohorts revealed several SNPs that were significantly associated with protection from CMV DNAemia. The copy number variation of several genes was significantly different between recipients with and without CMV DNAemia. Amongst patients with CMV DNAemia in the second cohort, several variants of interest ($p < 5 \times 10^{-5}$), the most common of which was NLRC5, were associated with peak viral load. We provide new predictive genetic markers for protection of CMV DNAemia. These markers should be validated in larger cohorts.

Keywords: human cytomegalovirus; kidney transplantation; genetic susceptibility; whole-exome sequencing; CMV DNAemia



Citation: Shapira, G.; Volkov, H.; Fabian, I.; Mohr, D.W.; Bettinotti, M.; Shomron, N.; Avery, R.K.; Arav-Boger, R. Genomic Markers Associated with Cytomegalovirus DNAemia in Kidney Transplant Recipients. *Viruses* **2023**, *15*, 2227. <https://doi.org/10.3390/v15112227>

Academic Editor: Davide Abate

Received: 5 October 2023

Revised: 29 October 2023

Accepted: 1 November 2023

Published: 8 November 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Infection with human cytomegalovirus (CMV), a member of the herpesvirus family, is common in humans. Seroprevalence rates increase with age, reaching 90% in individuals older than 80 years [1]. CMV establishes lifelong persistent infection, and individuals typically remain asymptomatic. In immunocompromised hosts, CMV causes significant morbidity and mortality [2–5]. Over 75% of solid organ transplant recipients are newly infected or reactivate latent CMV after transplantation. Kidneys are the most commonly transplanted solid organs, and CMV-seronegative recipients from a CMV-seropositive donor (D+/R−) are the subgroup at highest risk for CMV infection and disease [6]. Without prophylaxis, infection is diagnosed in 50–60% of kidney transplant recipients [7]. Prophylactic antiviral therapy has decreased the incidence of CMV infection and disease in the early post-transplant period, but these agents have significant toxicities [8]. Moreover, late-onset CMV disease is associated with allograft failure and mortality [9].

The CMV double-stranded DNA genome is around 235 kbp and has the largest genome among herpesviruses. Multiple CMV-encoded gene products are devoted to host immune evasion, among which are chemokines, chemokine receptors, and cytokines, allowing CMV to modify and interfere with host immune responses [10–13]. The role of host immune response in CMV reactivation following organ transplantation has been difficult to study, partially because of sample availability, the complexity of CMV interaction with multiple cellular pathways, and the ability to identify a defined outcome measure.

CMV infection following organ transplantation was associated with variants in several human genes: those encoding toll-like receptors (TLRs) [14], programmed death-1 (PD-1), and interleukin-12p40 (IL12B). Polymorphisms in the interferon lambda 3/4 (IFNL3/4) region also influenced susceptibility to CMV replication in solid organ transplant recipients. However, no association was found between ten genetic variants in TLR4, TNF- α , IL10, IFN- γ , and IL37 in CMV-positive renal allograft recipients and active infection in a sub-analysis (116 blood samples) of a prospective randomized VIPP study (NCT00372229) [15], demonstrating the limitation of the small cohorts and the need for additional genetic epidemiological studies with large cohorts to elucidate the genetic mechanisms of CMV infection and reactivation following solid organ transplantation.

A one-year retrospective study of CMV reactivation (based on CMV antigenemia) in 200 kidney transplant recipients investigated the relationship between CMV infection and 59 HLA alleles. Recipients with HLA-B44 were more commonly infected with CMV compared with patients without this allele ($p = 0.024$). In contrast, recipients with HLA-DR1 were less likely to have CMV reactivation than patients without this allele (31% vs. 55%, respectively, $p = 0.02$) [16]. A positive correlation between the presence of the HLA-E*01:03 allele in living-donor kidney recipients and CMV reactivation during the first year after transplantation was reported, suggesting that HLA-E genotyping may help identify CMV replication-prone patients [17].

The copy number variation (CNV) is another genomic variation, which may play an important role in the susceptibility to infectious diseases [18]. For example, a relationship between CCL3L1 dose and susceptibility to HIV/AIDS was reported. Possession of a CCL3L1 copy number lower than the population average was associated with markedly enhanced susceptibility to HIV [19]. CNV was reported in CMV reactivation. A significant association was found between donor NKG2C copy number and protection against CMV reactivation after double cord blood transplantation [20].

Here, we aimed to investigate genetic markers of CMV DNAemia in two cohorts of kidney transplant recipients with a well-defined phenotype. The recipients were all CMV-seronegative at the time of transplantation, and their donors were all CMV-seropositive.

2. Materials and Methods

2.1. Patients and Samples

Serum samples were collected from kidney transplant recipients and stored in the Immunogenetics Laboratory at Johns Hopkins (Baltimore, MD, USA), under IRB approval. CMV infection was monitored for at least one year after transplantation. Data collected from recipients of a kidney transplant included: patient age at transplant, gender, race, any positive CMV PCR (date and viral load), peak viral load, immunosuppression therapy, and CMV prophylaxis.

2.2. Sequencing and Genotyping

Samples were sequenced and genotyped at the Johns Hopkins University Genetic Resources Core Facility DNA Services (Baltimore, MD, USA). Illumina InfiniumQCArray-24v1-0 array (Illumina, San Diego, CA, USA) was used to confirm gender and assess relatedness. Exome capture was performed using Agilent SureSelectXT HumanAllExon (V6 S07604514) (Agilent, Santa Clara, CA, USA) for cohort A and Twist Human Core Exome (Twist Bioscience, South San Francisco, CA, USA) for cohort B. Cohort A was sequenced using the HiSeq2500 platform (Illumina, San Diego, CA, USA), 125 bp paired-end reads.

Cohort B was sequenced using the NovaSeq 6000 platform (Illumina, San Diego, CA, USA), with 100 bp paired-end reads. Raw sequencing data were aligned by BWA-Mem v0.7.15 [21] to the GRCh37 genome reference. Variant calling was performed according to the Genome Analysis Toolkit (GATK v4.0.1.1) best practice, including duplicate marking, base quality and score recalibration, and joint genotyping [22]. The called variants are summarized in Supplementary File S1. Filtered variant call sets were imputed using 1000 genomes phase 3 genomic data. A full description of the process can be found in the Supplementary Data. Plink was used to aggregate genotypes and to verify sex assignment (In addition to the prior verification, detailed in the Supplementary Data), and PCA (<http://pngu.mgh.harvard.edu/purcell/plink/>; accessed on 7 March 2023).

2.3. Association and Meta-Association Analysis

Association analyses were performed for each cohort using a generalized linear model under Gemma [23]. Because of the small sample size, mixed modeling and more complex approaches were not used. Outlier samples were identified using PCA and removed prior to downstream analysis. Sex and the first principle component were used as covariates (See Supplementary Figures S1 and S2 for PCA plots). Meta-association was performed using METAL, operating on results from per-cohort tests, weighted by the size of the cohort. Because of the small number of non-Caucasian participants, a separate Caucasian-only meta-analysis was performed. Variant annotations were obtained from snpEff version 5.2 [24] and processed using in-house scripts. Allele frequencies (AF) were obtained from large-scale databases, including NCBI's ALFA and gnomAD.

2.4. Copy Number Variant Analysis

Copy number variants were called using CNVKit, with the target coordinates for the capture kit used for sequencing each cohort (7485 segments in total). Each segment was matched with overlapping genes and counted only if one or more copies were gained or lost. The significance was measured by the uniformity of copy number alteration, using binomial testing with false discovery rate (FDR)-adjusted *p*-values.

2.5. Extra Visualization and Downstream Analysis

Locus zoom was used to visualize results loci [25]. Global Manhattan plots and the rest of the figures presented in this paper were generated using the ggplot2 package in R. Gene set enrichment was performed using hypergeometric tests, with FDR correction applied to each set family (KEGG, GO, etc.).

3. Results

3.1. Study Population

Exome variant association analysis was performed on two cohorts. Cohort A included 72 patients who received a kidney transplant between 2013 and 2016, and cohort B—126 patients who received a kidney transplant between 2005 and 2012 (Table 1) at Johns Hopkins University School of Medicine. All patients were CMV-seronegative with CMV-seropositive donors at the time of transplant. The majority of participants were Caucasians, with 16 African, one Asian, and one Hispanic in cohort A. The non-Caucasian portion of cohort A was too small to uncover reliable ethnicity-specific associations; we therefore generated an additional Caucasian-only analysis. Sex and the first principal component were used as covariates, to account for ancestry differences. The monitoring protocol for D+/R− kidney transplant recipients after completion of valganciclovir prophylaxis was CMV PCR every two weeks for three months, then CMV PCR monthly x three months. Usually, this corresponded to Months 7–9 and Months 10–12 post-transplant.

Table 1. Descriptive statistics of the cohorts analyzed in the study.

Cohort	A	B	Total
Participants (%)	72 (36.3%)	126 (63.6%)	198 (100%)
Male (%)	41 (56.9%)	75 (59.5%)	116 (58.5%)
CMV DNAemia (%)	36 (50%)	60 (47.6%)	96 (48.4%)
Caucasian (%)	54 (75%)	126 (100%)	180 (90.9%)
Age at time of transplant (Mean, SE)	54.5 (14)	51.2 (1.4)	51.2 (1.4)
Peak viral load (Mean, SE)	2.4×10^6 (1.4×10^6)	6.5×10^5 (2.6×10^5)	1.3×10^6 (5.5×10^6)

3.2. Induction and Maintenance Immunosuppression Regimen

Of the 198 patients, all received standard induction immunosuppression including thymoglobulin, except for 18 patients who received basiliximab and 4—daclizumab. Three patients received no induction. In addition to thymoglobulin, 38 patients also received plasmapheresis +/- rituximab. Of those, 24 were in the no-CMV DNAemia group and 14 in the CMV DNAemia group. All patients received maintenance immunosuppression with prednisone/tacrolimus/mycophenolate, except for 10 patients who received a sirolimus-based regimen (six in the non-CMV DNAemia group, four in the CMV DNAemia) and five patients who received an everolimus-based regimen (four in the no-CMV DNAemia and one in the CMV DNAemia group). There were eight patients who received belatacept, six—cyclosporine, six—alemtuzumab, one—eculizumab, and one—daclizumab for maintenance. All 72 patients in cohort A and 126 patients in cohort B received six months of valganciclovir prophylaxis at 900 mg daily (or adjusted for renal function).

3.3. CMV DNAemia Variant Association Analysis

Association analysis was performed for each cohort independently, followed by meta-analysis. The reproducibility between the cohorts was generally low, with only a few single-nucleotide polymorphisms (SNPs) associated with the presence or absence of CMV DNAemia in the meta-analysis (Figure 1; Supplementary Files S2–S4). The results of the exome sequencing consisted of very common variants (minor allele frequency, MAF > 10%) with low linkage disequilibrium (LD). We defined a $p < 5 \times 10^{-5}$ as the threshold for variants of interest. Although genome-wide association studies use strict criteria for significance, it is necessary to account for the lower number of variants called from our targeted sequencing [26]. Of the five variants that passed the threshold in the meta-analysis, three were based on results from both cohorts, and an additional two were sequenced in only one of the cohorts, one from each (Table 2). All five variants are common in the general population (MAF > 10%; gnomAD). The most significant variant was a frameshift and missense mutation of Dynein Heavy-Chain Domain 1 (DNHD1), with a protective effect (meta: $p = 1 \times 10^{-5}$; $Z = -4.4$; Figure 2). Intronic variants detected in Nephrocystin 4 (NPHP4) and Latent Transforming Growth Factor Beta Binding Protein 4 (LTBP4) were moderately associated with DNAemia in both cohorts. A SNP in NPHP4 was associated with susceptibility to CMV DNAemia, while a SNP in LTBP4 was associated with protection from CMV DNAemia. An intronic variant in PRR5—ARHGAP8 was found only in cohort B, and a SNP upstream of HLA-DRB1 was exclusive to cohort A. A flavin-containing dimethylaniline monooxygenase 9 (FMO9P) splice donor variant in a pseudo-gene was genotyped in both cohorts and narrowly missed the threshold (meta: $p = 5.1 \times 10^{-5}$; $Z = 4$). Gene set analysis of the 100 most significant genes in the DNAemia meta-analysis found significant enrichment of genes associated with autoimmune disease of skin and connective tissue (DOID:0060039: FDR = 0.02; TG, DSG1, DST, LAMA3, HLA-DRB1).

Table 2. The top 10 variants most significantly associated with CMV DNAemia or protection from DNAemia in a meta-analysis of cohorts A and B. P and beta values are presented for both cohorts, with meta-analysis z-score and *p*-value added. The beta values represent the magnitude and direction of the effect each variant has on the DNAemia outcome. A negative beta value indicates protection, while positive beta values indicate susceptibility to CMV DNAemia.

ID	Variant			CADD	Meta	<i>p</i> -Value		Z-Meta	Beta	
	GENE	EFFECT	EFFECT			A	B		A	B
11_6569896_G_A	DNHD1	Frameshift, missense	23.5	1.00×10^{-5}	1.20×10^{-2}	4.90×10^{-4}	-4.41	-0.25	-0.28	
1_5937391_C_T	NPHP4	Intron	0.184	2.30×10^{-5}	5.40×10^{-3}	1.10×10^{-3}	4.23	0.3	0.24	
19_41119273_G_C	LTBP4	Intron	6.661	2.80×10^{-5}	3.20×10^{-3}	1.70×10^{-3}	-4.19	-0.26	-0.24	
22_45205158_G_GT	PRR5—ARHGAP8	Intron	2.397	2.80×10^{-5}	N/A	4.70×10^{-5}	4.19	N/A	0.38	
6_32557621_T_G	HLA-DRB1	Upstream	N/A	4.60×10^{-5}	4.60×10^{-5}	N/A	4.07	0.51	N/A	
1_166591271_T_C	FMO9P	Splice donor and intron	23.2	5.20×10^{-5}	9.60×10^{-3}	1.30×10^{-3}	4.05	0.34	0.25	
21_27840567_C_T	CYYR1	3' prime UTR	7.693	6.20×10^{-5}	6.20×10^{-5}	N/A	-4.01	-0.58	NA	
8_126085586_G_A	WASHC5	Intron	1.184	6.40×10^{-5}	5.90×10^{-3}	2.70×10^{-3}	-4	-0.42	-0.25	
8_126068873_T_G	WASHC5	Intron	0.375	8.50×10^{-5}	2.40×10^{-2}	1.10×10^{-3}	-3.93	-0.43	-0.36	
10_12195881_G_A	SEC61A2	Intron	3.243	8.90×10^{-5}	8.90×10^{-5}	N/A	-3.92	-0.37	N/A	

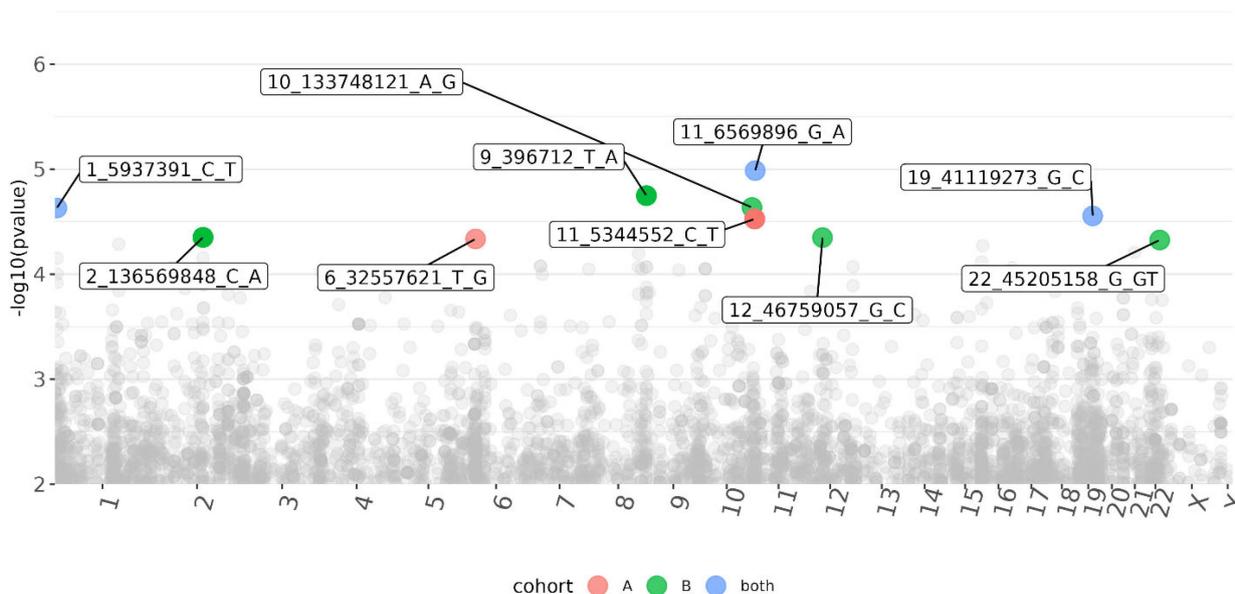


Figure 1. Manhattan plot of variants associated with CMV DNAemia in both cohorts of kidney transplant recipients. Each dot represents a variant, placed according to its genomic position (*x*-axis, labeled chromosome name) and the significance of the association (*y*-axis; $-\log_{10}(p\text{-value})$); higher indicates higher significance). Associations that passed the significance threshold ($p < 5 \times 10^{-5}$) are labeled and colored (see Supplementary Methods S1 and S2).

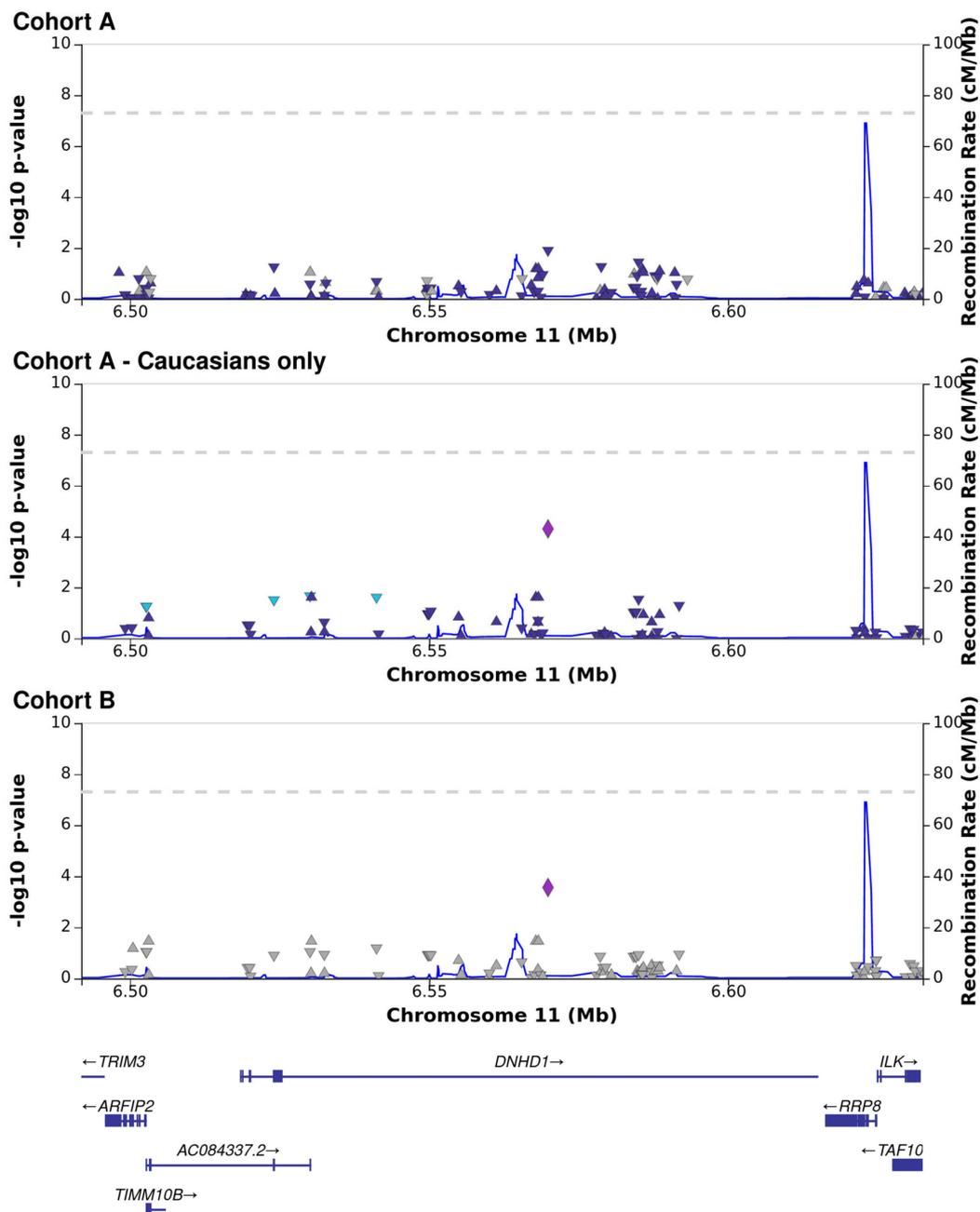


Figure 2. Association analysis between CMV DNAemia and genetic variants in the dynein heavy chain domain 1 (DNHD1) locus, presented for cohort A, Caucasians only in cohort A, and cohort B. Each shape represents a single genetic variant, placed on the X-axis according to its genomic position in relation to genes (visualized on the bottom panel) and on the Y-axis (left side) according to the log-scaled p -value of association (each of the three top panels applies to a single cohort). The recombination rate is denoted by the thin blue line and the values on the Y-axis (right side). Leading SNP results are marked by larger symbols, surrounding SNPs are colored according to linkage.

3.4. CMV DNAemia Copy Number Association Analysis

Copy number variations (CNVs) were called for the participants with DNAemia, using the participants with undetectable CMV DNA as a reference panel. Each CNV was associated with the genes overlapping the locus (see Supplementary Methods S1 and S2). Sixteen genes had copy variants in 10 viremic patients or more, 14 of which were called in participants of both cohorts. In the viremic group, the LCE3B-LCE3C locus had significant copy loss in both cohorts. There were 28 viremic individuals from both cohorts with copy

number loss in the LCE3B-LCE3C locus. Another four genes also had copy loss, albeit not as significant (Table 3, Supplementary File S5).

Table 3. Genes with significant copy number variations in viremic participants compared to the aviremic group. The number of participants with copy gain or loss for a specific gene is shown for each cohort. The adjusted significance (FDR) of the binomial test for each gene is shown.

SYMBOL	A_Loss	B_Loss	A_Gain	B_Gain	A	B	Loss	Gain	Overall	FDR
LCE3B	12	16	0	0	12	16	28	0	28	2.51×10^{-6}
LCE3C	13	15	0	0	13	15	28	0	28	2.51×10^{-6}
TAS2R43	5	13	0	0	5	13	18	0	18	1.71×10^{-3}
GSTM1	17	0	0	0	17	0	17	0	17	2.06×10^{-3}
GSTM2	17	0	0	0	17	0	17	0	17	2.06×10^{-3}
AHNAK2	9	3	0	0	9	3	12	0	12	2.99×10^{-2}

3.5. CMV DNAemia Peak Viral Load Association Analysis

In patients with CMV DNAemia, we tested their peak viral load values for genetic associations. Association analysis on viremic participants reduced the sample size (N = 36, 59 viremic participants in cohorts A and B, respectively), resulting in unreliable results for cohort A and underpowered results for cohort B (Supplementary Files S6–S8). Nineteen variants of interest ($p < 5 \times 10^{-5}$) were associated with peak viral load for cohort B (Table 4), despite the low statistical power.

Table 4. Variants of interest associated with peak viral load in CMV DNAemia in cohort B.

ID	GENE	EFFECT	p-Value	Beta	CADD
13_23808782_T_C	SGCG	Frameshift and missense	1.10×10^{-5}	5.21	5.754
11_133788869_A_G	IGSF9B	Intron	1.80×10^{-5}	−3.01	0.173
17_39394962_C_T	KRTAP9-8	3_prime_UTR intron	2.00×10^{-5}	2.86	3.321
2_112939548_T_C	FBLN7	Intron	2.10×10^{-5}	5.07	6.004
2_112940578_C_T	FBLN7	Intron	2.10×10^{-5}	5.07	1.94
6_36292007_G_A	BNIP5	Intron	2.10×10^{-5}	4.14	0.052
16_57068107_C_T	NLRC5	Frameshift and missense	2.60×10^{-5}	−4.42	1.901
16_57071209_T_C	NLRC5	Intron	3.40×10^{-5}	−3.92	1.492
16_57071226_TCC_T	NLRC5	Intron	3.40×10^{-5}	−3.92	4.507
16_57071236_C_T	NLRC5	Intron	3.40×10^{-5}	−3.92	4.909
15_59499179_G_A	LDHAL6B MYO1E	Frameshift and missense	3.50×10^{-5}	2.86	8.694
15_59500116_T_C	LDHAL6B MYO1E	frameshift and missense	3.50×10^{-5}	2.86	23.1
16_57075406_T_C	NLRC5	Frameshift and missense	3.60×10^{-5}	−4.48	8.512
16_57076018_G_C	NLRC5	Intron	3.60×10^{-5}	−4.48	0.213
16_57077523_G_A	NLRC5	Splice region and intron	3.60×10^{-5}	−4.48	2.641
16_57077581_C_T	NLRC5	Intron	3.60×10^{-5}	−4.48	0.278
16_57060213_C_T	NLRC5	Frameshift and missense	3.60×10^{-5}	−4.48	3.843
16_57060340_T_C	NLRC5	Frameshift and missense	3.60×10^{-5}	−4.48	5.063
16_57060353_T_C	NLRC5	Frameshift and missense	3.60×10^{-5}	−4.48	0.581

4. Discussion

Identification of host markers associated with CMV reactivation after transplantation may play a key role in the management and therapeutic decisions for transplant recipients. It is well known that D+/R− transplant recipients are at the highest risk for the development of symptomatic CMV disease [27], but there is a wide variation in the clinical manifestations and outcomes within this group. Some patients develop highly symptomatic CMV with high viral loads, end-organ disease, while others have a milder disease or no CMV DNAemia at all. A better understanding of host protective or risk factors could allow for the personalization of CMV prevention strategies, optimize outcomes, and minimize toxicity and cost. Identification of these host factors requires large, well-designed cohorts with a defined phenotype. Here, we studied two cohorts of CMV-seronegative kidney transplant recipients who received kidneys from CMV-seropositive donors and were treated with the same CMV prophylaxis regimen. Their induction and maintenance immunosuppression regimens were overall similar, and the anti-B cell therapy did not appear to be a risk factor for CMV DNAemia.

CMV DNA was quantified in blood for ~12 months follow-up, differentiating between patients with recurrent CMV DNAemia and those with undetectable CMV DNA. We found a few candidate variants that were associated with the risk of CMV DNAemia. Despite the relatively small sample size, our results indicate that multiple common variants (AF > 10%) significantly affected the risk of developing DNAemia in CMV-naive transplant recipients.

Of the variants associated with protection from CMV DNAemia in both cohorts, several could have a pathogenesis role in CMV (Table 2, Figures 1 and 2). These include a missense variant in dynein (DNHD1) and an intron variant in LTB4 which encodes for a protein that binds to transforming growth factor-beta (TGF- β) as it is secreted and targeted to the extracellular matrix.

Dyneins are broadly associated with viral transport and assembly in the host cell [28] and are considered essential for CMV infection [29]. CMV-infected cells abuse Dynein to maintain favorable mTORC1 activity under stress [30]. Dynein is integral to the formation of the assembly complex and the characteristic large, kidney-shaped nucleus in CMV-infected cells [31]. More recently, it was shown that inhibition of Dynein reduced the number of virions transported to the nucleus and protein synthesis of herpes simplex virus type 1 (HSV1) [32].

An intronic variant in LTBP4 was moderately associated with protection from CMV DNAemia in both cohorts (Table 2). HCMV miRNAs produced during latency induced the expression of TGF- β while protecting the infected cell from TGF- β signaling for efficient viral latency [33]. Transforming growth factor-beta 1 (TGF- β 1) was originally reported to stimulate CMV replication [34]. The TGF- β 1 promoter was then found to be activated independently by the CMV-encoded immediate-early proteins and transactivated early after infection [35]. TGF- β 1 mRNA increased during the early phase of infection. It has been suggested that induction of TGF- β 1 by CMV could modify infected cells and systemic immune reactions to benefit virus replication by both inducing CMV replication and downregulating host immune responses. Urinary excretion of TGF- β is reportedly increased in kidney transplant recipients during CMV infection [36]. Persistent CMV infection in kidney allografts was associated with increased expression of TGF- β [37]. Human renal tubular epithelial cells infected with CMV and exposed to TGF- β 1 underwent both morphologic and transcriptional changes of epithelial to mesenchymal transition, similar to uninfected renal epithelial cells. Infected cells also activated extracellular latent TGF- β 1 [38].

A SNP in the upstream gene of HLA-DRB1 was associated with susceptibility to CMV DNAemia. HLA-DRB1 was found to be a risk allele after bone marrow transplantation, where in contrast to solid organ transplantation, recipient CMV seropositivity is the highest risk for CMV reactivation. HLA-DRB1*09 was associated with an increased incidence of CMV infection and disease in a cohort of 60 allogeneic hematopoietic stem cell transplant

(HSCT) recipients [39]. Among kidney transplant recipients, HLA-DQ3 was identified as an independent predictor of CMV infection in 129 CMV D+/R− patients [40].

CNV is a less-studied type of genomic variation, involved in human disease pathogenesis [41,42]. We detected multiple genes that underwent CNV in a sizable share of both cohorts and in a uniform direction (Table 3). The reproducibility and functional annotations of gene CNVs are suggestive of a substantial effect on the degree of CMV DNAemia. Large duplications/deletions of gene clusters affected multiple genes of the same families, including glutathione S-transferase and Late Cornified Envelope Protein 3D (LCE3D, Table 3). Virus-like vesicles induced the expression of a panel of epithelial differentiation genes, especially genes belonging to the epidermal differentiation complex (SPRR2C, SPRR2D, SPRR3, LCE3D, and SCEL) [43]. The LCE gene cluster members LCE3D and LCE3E were downregulated in an EBV-HPV coinfection model of cervical intraepithelial neoplasia, suggesting that coinfection of EBV and HPV increased the effect of HPV on epithelial differentiation and development [44].

LCE1 genes, located in the LCE gene clusters encoding multiple well-conserved stratum corneum proteins, are reported downstream targets of p53 and regulate protein arginine methyltransferase 5 (PRMT5) activity [45].

Lastly, variants of interest were identified which may be associated with higher peak CMV viral load, although a larger sample size would be needed to confirm these findings (Table 4). Insofar as the magnitude of the CMV peak viral load often reflects the severity of CMV disease, further investigation of these candidate variants would be of interest. NLRC5 is reportedly upregulated in CMV-infected human fibroblasts and plays a role in the JAK/STAT-mediated autocrine signaling loop involving IFN-gamma. Overexpression of NLRC5 protein resulted in activation of the IFN-responsive regulatory promoter elements, IFN-gamma activation sequence, and IFN-specific response element and upregulation of antiviral target genes (e.g., IFN-alpha, OAS1, and PRKRIR) [46].

Our study includes several limitations. The association analysis was underpowered because of the small cohorts. Antiviral immunity is a highly complex, polygenic trait, often studied using cohorts of several thousand individuals [47,48]. The small size of our cohort limits our results to common alleles with relatively high effect sizes and under-estimated significance. The need for larger cohorts is also supported by a study of genetic variants associated with CMV infections after hematopoietic stem cell transplantation [49], showing that most genomic variants previously associated with CMV phenotypes did modify the risk for CMV reactivation or disease after transplantation.

Another limitation is the mostly mono-ethnic composition of our cohort. While the significance of host ethnicity in CMV infection is yet to be determined [50], it is a highly significant factor in the susceptibility and severity of many viral infections [51–53]. Ethnic differences may lead to differences in disease severity on a national level [54] or alter the effect of common variants depending on race [55]. Additional advantages include the discovery of rare ethnicity-specific variants [56] and improving the specificity of causal variant detection [57].

Finally, the use of two different sets of exome target sets reduced reproducibility due to lack of overlap. An additional disadvantage of exome sequencing is the decreased sensitivity and accuracy of copy number variant calling [58].

Summarized, despite several limitations, our study provides new insights into complex genetic variants that may play a role in CMV reactivation following kidney transplantation. Larger cohorts as well as future biological systems should validate those markers for future clinical use.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/v15112227/s1>, Results and summary statistics for all tests described in this study are deposited at <https://zenodo.org/record/8243954>. Summary statistics for association tests contain only the 10,000 most significant results. Figure S1A,B: PCA plots of cohort A genotypes, colored according to racial assignment (A) or CMV DNAemia status (B) and shaped according to sex; Figure S2A,B: PCA plots of cohort A genotypes before (A) and after removing outliers

(B,C), colored according to racial assignment (A,B) or CMV DNAemia status (C) and shaped according to sex. Supplementary File S1: “variant_calling_summary_stats.xlsx” contains summary statistics for all variants called in cohorts A and B. Supplementary File S2: “dnaemia.cohort1.assoc.txt” genomic associations with DNAemia in Cohort A. Supplementary File S3: “dnaemia.cohort2.assoc.txt” genomic associations with DNAemia in Cohort B. Supplementary File S4: “dnaemia.meta_assoc.txt” genomic meta-associations with DNAemia in Cohorts A and B. Supplementary File S5: “cnv.results.tsv” contains gene-level results associating copy number variants with DNAemia status. Supplementary File S6: “peak_viral_load.cohort1.assoc.txt” genomic associations with peak viral load in Cohort A. Supplementary File S7: “peak_viral_load.cohort2.assoc.txt” genomic associations with peak viral load in Cohort B. Supplementary File S8: “peak_viral_load.meta_assoc.txt” genomic meta-associations with peak viral load in Cohorts A and B. Supplementary Methods S1 and S2: Extended descriptions of the methodology used for cohort A and B, respectively. References [22,59–69] are cited in the Supplementary Methods S1 and S2.

Author Contributions: G.S. conducted data analysis and drafted the manuscript. H.V. and I.F. conducted data analysis. D.W.M. performed all the assays and data collection. M.B. contributed to the study design and implementation. N.S. provided consultation on data analysis. R.K.A. contributed to the study design, data collection, implementation, and manuscript writing. R.A.-B. conceptualized the study, supported data collection and analysis, and manuscript writing. All authors have read and agreed to the published version of the manuscript.

Funding: Department of Pediatrics innovation grant Johns Hopkins, Department of Pediatrics Medical College of Wisconsin (RAB).

Institutional Review Board Statement: The Johns Hopkins IRB approved this study (IRB00139199).

Informed Consent Statement: Not applicable.

Data Availability Statement: Data for individual participants is protected under patient confidentiality laws and cannot be shared. See Supplementary Materials section for shared top results.

Conflicts of Interest: The authors report no conflicts of interest. Robin Avery has been an investigator on studies funded by Aicuris (Wuppertal, Germany), Astellas (Tokyo, Japan), Astra-Zeneca (Tokyo, Japan), Chimerix (Durham, NC, USA), Merck (Darmstadt, Germany), Oxford Immunotec (Oxfordshire, UK), Qiagen (Hong Kong, China), Regeneron (New York, NY, USA), and Takeda (Tokyo, Japan). No personal financial remuneration was received; study/grant support only.

References

1. Staras, S.A.; Dollard, S.C.; Radford, K.W.; Flanders, W.D.; Pass, R.F.; Cannon, M.J. Seroprevalence of cytomegalovirus infection in the United States, 1988–1994. *Clin. Infect. Dis.* **2006**, *43*, 1143–1151. [[CrossRef](#)] [[PubMed](#)]
2. Griffiths, P.D.; Clark, D.A.; Emery, V.C. Betaherpesviruses in transplant recipients. *J. Antimicrob. Chemother.* **2000**, *45* (Suppl. T3), 29–34. [[CrossRef](#)] [[PubMed](#)]
3. Kovacs, A.; Schluchter, M.; Easley, K.; Demmler, G.; Shearer, W.; La, R.P.; Pitt, J.; Cooper, E.; Goldfarb, J.; Hodes, D.; et al. Cytomegalovirus infection and HIV-1 disease progression in infants born to HIV-1-infected women. Pediatric Pulmonary and Cardiovascular Complications of Vertically Transmitted HIV Infection Study Group. *N. Engl. J. Med.* **1999**, *341*, 77–84. [[CrossRef](#)] [[PubMed](#)]
4. Razonable, R.R.; Humar, A. Cytomegalovirus in solid organ transplant recipients—Guidelines of the American Society of Transplantation Infectious Diseases Community of Practice. *Clin. Transpl.* **2019**, *33*, e13512. [[CrossRef](#)] [[PubMed](#)]
5. Kotton, C.N.; Kumar, D.; Caliendo, A.M.; Huprikar, S.; Chou, S.; Danziger-Isakov, L.; Humar, A.; The Transplantation Society International CMV Consensus Group. The Third International Consensus Guidelines on the Management of Cytomegalovirus in Solid-organ Transplantation. *Transplantation* **2018**, *102*, 900–931. [[CrossRef](#)] [[PubMed](#)]
6. De Keyser, K.; Van Laecke, S.; Peeters, P.; Vanholder, R. Human cytomegalovirus and kidney transplantation: A clinician’s update. *Am. J. Kidney Dis.* **2011**, *58*, 118–126. [[CrossRef](#)] [[PubMed](#)]
7. De Matos, S.B.; Meyer, R.; Lima, F.W.M. Cytomegalovirus Infection after Renal Transplantation: Occurrence, Clinical Features, and the Cutoff for Antigenemia in a University Hospital in Brazil. *Infect. Chemother.* **2017**, *49*, 255–261. [[CrossRef](#)]
8. Hodson, E.M.; Ladhani, M.; Webster, A.C.; Strippoli, G.F.; Craig, J.C. Antiviral medications for preventing cytomegalovirus disease in solid organ transplant recipients. *Cochrane Database Syst. Rev.* **2013**, *2*, CD003774. [[CrossRef](#)] [[PubMed](#)]
9. Arthurs, S.K.; Eid, A.J.; Pedersen, R.A.; Kremers, W.K.; Cosio, F.G.; Patel, R.; Razonable, R.R. Delayed-onset primary cytomegalovirus disease and the risk of allograft failure and mortality after kidney transplantation. *Clin. Infect. Dis.* **2008**, *46*, 840–846. [[CrossRef](#)]

10. Vink, C.; Smit, M.J.; Leurs, R.; Bruggeman, C.A. The role of cytomegalovirus-encoded homologs of G protein-coupled receptors and chemokines in manipulation of and evasion from the immune system. *J. Clin. Virol.* **2001**, *23*, 43–55. [[CrossRef](#)]
11. Wang, D.; Bresnahan, W.; Shenk, T. Human cytomegalovirus encodes a highly specific RANTES decoy receptor. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 16642–16647. [[CrossRef](#)] [[PubMed](#)]
12. Penfold, M.E.; Dairaghi, D.J.; Duke, G.M.; Saederup, N.; Mocarski, E.S.; Kemble, G.W.; Schall, T.J. Cytomegalovirus encodes a potent alpha chemokine. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 9839–9844. [[CrossRef](#)] [[PubMed](#)]
13. Cha, T.A.; Tom, E.; Kemble, G.W.; Duke, G.M.; Mocarski, E.S.; Spaete, R.R. Human cytomegalovirus clinical isolates carry at least 19 genes not found in laboratory strains. *J. Virol.* **1996**, *70*, 78–83. [[CrossRef](#)] [[PubMed](#)]
14. Kijpittayarit, S.; Eid, A.J.; Brown, R.A.; Paya, C.V.; Razonable, R.R. Relationship between Toll-like receptor 2 polymorphism and cytomegalovirus disease after liver transplantation. *Clin. Infect. Dis.* **2007**, *44*, 1315–1320. [[CrossRef](#)] [[PubMed](#)]
15. Mazzola, P.; Schaeffeler, E.; Witzke, O.; Nitschke, M.; Kliem, V.; Zortel, M.; Wagner, E.-M.; Schwab, M.; Hauser, I.A. No association of genetic variants in TLR4, TNF-alpha, IL10, IFN-gamma, and IL37 in cytomegalovirus-positive renal allograft recipients with active CMV infection-Subanalysis of the prospective randomised VIPP study. *PLoS ONE* **2021**, *16*, e0246118. [[CrossRef](#)] [[PubMed](#)]
16. Futohi, F.; Saber, A.; Nemati, E.; Einollahi, B.; Rostami, Z. Human Leukocyte Antigen Alleles and Cytomegalovirus Infection After Renal Transplantation. *Nephrourol. Mon.* **2015**, *7*, e31635. [[CrossRef](#)] [[PubMed](#)]
17. Guberina, H.; da Silva Nardi, F.; Michita, R.T.; Dolff, S.; Bienholz, A.; Heinemann, F.M.; Wilde, B.; Trilling, M.; Horn, P.A.; Kribben, A.; et al. Susceptibility of HLA-E*01:03 Allele Carriers to Develop Cytomegalovirus Replication After Living-Donor Kidney Transplantation. *J. Infect. Dis.* **2018**, *217*, 1918–1922. [[CrossRef](#)] [[PubMed](#)]
18. Pos, O.; Radvanszky, J.; Buglyo, G.; Pos, Z.; Rusnakova, D.; Nagy, B.; Szemes, T. DNA copy number variation: Main characteristics, evolutionary significance, and pathological aspects. *Biomed. J.* **2021**, *44*, 548–559. [[CrossRef](#)]
19. Gonzalez, E.; Kulkarni, H.; Bolivar, H.; Mangano, A.; Sanchez, R.; Catano, G.; Nibbs, R.J.; Freedman, B.I.; Quinones, M.P.; Bamshad, M.J.; et al. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* **2005**, *307*, 1434–1440. [[CrossRef](#)]
20. Cao, K.; Marin, D.; Sekine, T.; Rondon, G.; Zhao, W.; Smith, N.T.; Daher, M.; Wang, Q.; Li, L.; Saliba, R.M.; et al. Donor NKG2C Copy Number: An Independent Predictor for CMV Reactivation After Double Cord Blood Transplantation. *Front. Immunol.* **2018**, *9*, 2444. [[CrossRef](#)]
21. Li, H.; Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **2010**, *26*, 589–595. [[CrossRef](#)]
22. Van der Auwera, G.A.; Carneiro, M.O.; Hartl, C.; Poplin, R.; Del Angel, G.; Levy-Moonshine, A.; Jordan, T.; Shakir, K.; Roazen, D.; Thibault, J.; et al. From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinform.* **2013**, *43*, 11.10.1–11.10.33. [[CrossRef](#)]
23. Zhou, X.; Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **2012**, *44*, 821–824. [[CrossRef](#)] [[PubMed](#)]
24. Sham, P.C.; Purcell, S.M. Statistical power and significance testing in large-scale genetic studies. *Nat. Rev. Genet.* **2014**, *15*, 335–346. [[CrossRef](#)] [[PubMed](#)]
25. Boughton, A.P.; Welch, R.P.; Flickinger, M.; VandeHaar, P.; Taliun, D.; Abecasis, G.R.; Boehnke, M. LocusZoom.js: Interactive and embeddable visualization of genetic association study results. *Bioinformatics* **2021**, *37*, 3017–3018. [[CrossRef](#)] [[PubMed](#)]
26. Fadista, J.; Manning, A.K.; Florez, J.C.; Groop, L. The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants. *Eur. J. Hum. Genet.* **2016**, *24*, 1202–1205. [[CrossRef](#)] [[PubMed](#)]
27. Razonable, R.R.; Humar, A. Cytomegalovirus in solid organ transplantation. *Am. J. Transplant.* **2013**, *13* (Suppl. 4), 93–106. [[CrossRef](#)] [[PubMed](#)]
28. Merino-Gracia, J.; Garcia-Mayoral, M.F.; Rodriguez-Crespo, I. The association of viral proteins with host cell dynein components during virus infection. *FEBS J.* **2011**, *278*, 2997–3011. [[CrossRef](#)] [[PubMed](#)]
29. Buchkovich, N.J.; Yu, Y.; Pierciey, F.J., Jr.; Alwine, J.C. Human cytomegalovirus induces the endoplasmic reticulum chaperone BiP through increased transcription and activation of translation by using the BiP internal ribosome entry site 11. *J. Virol.* **2010**, *84*, 11479–11486. [[CrossRef](#)]
30. Clippinger, A.J.; Alwine, J.C. Dynein mediates the localization and activation of mTOR in normal and human cytomegalovirus-infected cells. *Genes Dev.* **2012**, *26*, 2015–2026. [[CrossRef](#)]
31. Alwine, J.C. The human cytomegalovirus assembly compartment: A masterpiece of viral manipulation of cellular processes that facilitates assembly and egress. *PLoS Pathog.* **2012**, *8*, e1002878. [[CrossRef](#)]
32. Dohner, K.; Wolfstein, A.; Prank, U.; Echeverri, C.; Dujardin, D.; Vallee, R.; Sodeik, B. Function of dynein and dynactin in herpes simplex virus capsid transport. *Mol. Biol. Cell* **2002**, *13*, 2795–2809. [[CrossRef](#)] [[PubMed](#)]
33. Hancock, M.H.; Crawford, L.B.; Pham, A.H.; Mitchell, J.; Struthers, H.M.; Yurochko, A.D.; Caposio, P.; Nelson, J.A. Human Cytomegalovirus miRNAs Regulate TGF-beta to Mediate Myelosuppression while Maintaining Viral Latency in CD34(+) Hematopoietic Progenitor Cells. *Cell Host Microbe* **2020**, *27*, 104–114.e4. [[CrossRef](#)] [[PubMed](#)]
34. Alcami, J.; Paya, C.V.; Virelizier, J.L.; Michelson, S. Antagonistic modulation of human cytomegalovirus replication by transforming growth factor beta and basic fibroblastic growth factor. *J. Gen. Virol.* **1993**, *74 Pt 2*, 269–274. [[CrossRef](#)] [[PubMed](#)]

35. Michelson, S.; Alcamí, J.; Kim, S.J.; Danielpour, D.; Bachelier, F.; Picard, L.; Bessia, C.; Paya, C.; Virelizier, J.L. Human cytomegalovirus infection induces transcription and secretion of transforming growth factor beta 1. *J. Virol.* **1994**, *68*, 5730–5737. [[CrossRef](#)] [[PubMed](#)]
36. Helantera, I.; Teppo, A.M.; Koskinen, P.; Tornroth, T.; Gronhagen-Riska, C.; Lautenschlager, I. Increased urinary excretion of transforming growth factor-beta(1) in renal transplant recipients during cytomegalovirus infection. *Transpl. Immunol.* **2006**, *15*, 217–221. [[CrossRef](#)] [[PubMed](#)]
37. Helantera, I.; Loginov, R.; Koskinen, P.; Tornroth, T.; Gronhagen-Riska, C.; Lautenschlager, I. Persistent cytomegalovirus infection is associated with increased expression of TGF-beta1, PDGF-AA and ICAM-1 and arterial intimal thickening in kidney allografts. *Nephrol. Dial. Transpl.* **2005**, *20*, 790–796. [[CrossRef](#)] [[PubMed](#)]
38. Shimamura, M.; Murphy-Ullrich, J.E.; Britt, W.J. Human cytomegalovirus induces TGF-beta1 activation in renal tubular epithelial cells after epithelial-to-mesenchymal transition. *PLoS Pathog.* **2010**, *6*, e1001170. [[CrossRef](#)] [[PubMed](#)]
39. Du, J.; Liu, J.; Gu, J.; Zhu, P. HLA-DRB1*09 is associated with increased incidence of cytomegalovirus infection and disease after allogeneic hematopoietic stem cell transplantation. *Biol. Blood Marrow Transpl.* **2007**, *13*, 1417–1421. [[CrossRef](#)]
40. Varga, M.; Rajczy, K.; Telkes, G.; Hidvegi, M.; Peter, A.; Rempert, A.; Korbonits, M.; Fazakas, J.; Toronyi, É.; Sárváry, E.; et al. HLA-DQ3 is a probable risk factor for CMV infection in high-risk kidney transplant patients. *Nephrol. Dial. Transpl.* **2008**, *23*, 2673–2678. [[CrossRef](#)]
41. Rice, A.M.; McLysaght, A. Dosage sensitivity is a major determinant of human copy number variant pathogenicity. *Nat. Commun.* **2017**, *8*, 14366. [[CrossRef](#)] [[PubMed](#)]
42. Ionita-Laza, I.; Rogers, A.J.; Lange, C.; Raby, B.A.; Lee, C. Genetic association analysis of copy-number variation (CNV) in human disease pathogenesis. *Genomics* **2009**, *93*, 22–26. [[CrossRef](#)] [[PubMed](#)]
43. Gong, D.; Dai, X.; Xiao, Y.; Du, Y.; Chapa, T.J.; Johnson, J.R.; Li, X.; Krogan, N.J.; Deng, H.; Wu, T.-T.; et al. Virus-Like Vesicles of Kaposi's Sarcoma-Associated Herpesvirus Activate Lytic Replication by Triggering Differentiation Signaling. *J. Virol.* **2017**, *91*, 10–1128. [[CrossRef](#)] [[PubMed](#)]
44. Feng, M.; Duan, R.; Gao, Y.; Zhang, H.; Qiao, Y.; Li, Q.; Zhao, F. Role of Epstein-Barr Virus and Human Papillomavirus Coinfection in Cervical Intraepithelial Neoplasia in Chinese Women Living With HIV. *Front. Cell. Infect. Microbiol.* **2021**, *11*, 703259. [[CrossRef](#)] [[PubMed](#)]
45. Deng, Z.; Matsuda, K.; Tanikawa, C.; Lin, J.; Furukawa, Y.; Hamamoto, R.; Nakamura, Y. Late Cornified Envelope Group I, a novel target of p53, regulates PRMT5 activity. *Neoplasia* **2014**, *16*, 656–664. [[CrossRef](#)] [[PubMed](#)]
46. Kuenzel, S.; Till, A.; Winkler, M.; Hasler, R.; Lipinski, S.; Jung, S.; Grötzinger, J.; Fickenscher, H.; Schreiber, S.; Rosenstiel, P. The nucleotide-binding oligomerization domain-like receptor NLRC5 is involved in IFN-dependent antiviral immune responses. *J. Immunol.* **2010**, *184*, 1990–2000. [[CrossRef](#)] [[PubMed](#)]
47. Tangden, T.; Gustafsson, S.; Rao, A.S.; Ingelsson, E. A genome-wide association study in a large community-based cohort identifies multiple loci associated with susceptibility to bacterial and viral infections. *Sci. Rep.* **2022**, *12*, 2582. [[CrossRef](#)] [[PubMed](#)]
48. Butković, A.; Elena, S.F. Genome-wide association studies of viral infections—A short guide to a successful experimental and statistical analysis. *Front. Syst. Biol.* **2022**, *2*, 1005758. [[CrossRef](#)]
49. Casto, A.M.; Seo, S.; Levine, D.M.; Storer, B.E.; Dong, X.; Hansen, J.A.; Boeckh, M.; Martin, P.J. Genetic variants associated with cytomegalovirus infection after allogeneic hematopoietic cell transplantation. *Blood* **2021**, *138*, 1628–1636. [[CrossRef](#)]
50. Duchowny, K.A.; Noppert, G.A. The Association Between Cytomegalovirus and Disability by Race/Ethnicity and Sex: Results From the Health and Retirement Study. *Am. J. Epidemiol.* **2021**, *190*, 2314–2322. [[CrossRef](#)]
51. Su, F.; Green, P.K.; Berry, K.; Ioannou, G.N. The association between race/ethnicity and the effectiveness of direct antiviral agents for hepatitis C virus infection. *Hepatology* **2017**, *65*, 426–438. [[CrossRef](#)] [[PubMed](#)]
52. Randolph, H.E.; Fiege, J.K.; Thielen, B.K.; Mickelson, C.K.; Shiratori, M.; Barroso-Batista, J.; Langlois, R.A.; Barreiro, L.B. Genetic ancestry effects on the response to viral infection are pervasive but cell type specific. *Science* **2021**, *374*, 1127–1133. [[CrossRef](#)] [[PubMed](#)]
53. Kist, N.C.; Lambert, B.; Campbell, S.; Katzourakis, A.; Lunn, D.; Lemey, P.; Iversen, A.K.N. HIV-1 p24Gag adaptation to modern and archaic HLA-allele frequency differences in ethnic groups contributes to viral subtype diversification. *Virus Evol.* **2020**, *6*, veaa085. [[CrossRef](#)]
54. Shapira, G.; Shomron, N.; Gurwitz, D. Ethnic differences in alpha-1 antitrypsin deficiency allele frequencies may partially explain national differences in COVID-19 fatality rates. *FASEB J.* **2020**, *34*, 14160–14165. [[CrossRef](#)] [[PubMed](#)]
55. Banday, A.R.; Stanifer, M.L.; Florez-Vargas, O.; Onabajo, O.O.; Papenberg, B.W.; Zahoor, M.A.; Mirabello, L.; Ring, T.J.; Lee, C.-H.; Albert, P.S.; et al. Genetic regulation of OAS1 nonsense-mediated decay underlies association with COVID-19 hospitalization in patients of European and African ancestries. *Nat. Genet.* **2022**, *54*, 1103–1116. [[CrossRef](#)] [[PubMed](#)]
56. Namkoong, H.; Edahiro, R.; Takano, T.; Nishihara, H.; Shirai, Y.; Sonehara, K.; Tanaka, H.; Azekawa, S.; Mikami, Y.; Lee, H.; et al. DOCK2 is involved in the host genetics and biology of severe COVID-19. *Nature* **2022**, *609*, 754–760. [[CrossRef](#)]
57. Gurdasani, D.; Barroso, I.; Zeggini, E.; Sandhu, M.S. Genomics of disease risk in globally diverse populations. *Nat. Rev. Genet.* **2019**, *20*, 520–535. [[CrossRef](#)]
58. Zhao, L.; Liu, H.; Yuan, X.; Gao, K.; Duan, J. Comparative study of whole exome sequencing-based copy number variation detection tools. *BMC Bioinform.* **2020**, *21*, 97. [[CrossRef](#)]

59. Fisher, S.; Barry, A.; Abreu, J.; Minie, B.; Nolan, J.; Delorey, T.M.; Young, G.; Fennell, T.J.; Allen, A.; Ambrogio, L.; et al. A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. *Genome Biol.* **2011**, *12*, R1. [[CrossRef](#)]
60. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* **2013**, arXiv:1303.3997v1.
61. Church, D.M.; Schneider, V.A.; Graves, T.; Auger, K.; Cunningham, F.; Bouk, N.; Chen, H.C.; Agarwala, R.; McLaren, W.M.; Ritchie, G.R.; et al. Modernizing reference genome assemblies. *PLoS Biol.* **2011**, *9*, e1001091. [[CrossRef](#)] [[PubMed](#)]
62. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **2004**, *43*, 931–945.
63. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **2010**, *20*, 1297–1303. [[CrossRef](#)] [[PubMed](#)]
64. DePristo, M.A.; Banks, E.; Poplin, R.; Garimella, K.V.; Maguire, J.R.; Hartl, C.; Philippakis, A.A.; Del Angel, G.; Rivas, M.A.; Hanna, M.; et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **2011**, *43*, 491–498. [[CrossRef](#)] [[PubMed](#)]
65. Wang, K.; Li, M.; Hakonarson, H. ANNOVAR: Functional annotation of genetic variants from next-generation sequencing data. *Nucleic Acids Res.* **2010**, *38*, e164. [[CrossRef](#)] [[PubMed](#)]
66. Exome Aggregation Consortium (ExAC), Cambridge, MA, USA. Available online: <http://exac.broadinstitute.org> (accessed on 4 October 2023).
67. 1000 Genomes Project Consortium; Abecasis, G.R.; Auton, A.; Brooks, L.D.; DePristo, M.A.; Durbin, R.M.; McVean, G.A. An integrated map of genetic variation from 1092 human genomes. *Nature* **2012**, *491*, 56–65. [[PubMed](#)]
68. Fritz, M.H.Y.; Leinonen, R.; Cochrane, G.; Birney, E. Efficient storage of high throughput DNA sequencing data using reference-based compression. *Genome Res.* **2011**, *21*, 734–740. [[CrossRef](#)]
69. Lek, M.; Karczewski, K.J.; Minikel, E.V.; Samocha, K.E.; Banks, E.; Fennell, T.; O'Donnell-Luria, A.H.; Ware, J.S.; Hill, A.J.; Cummings, B.B.; et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **2016**, *536*, 285–291. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.