American Journal of Interdisciplinary Research and Development
ISSN Online: 2771-8948
Website: www.ajird.journalspark.org
Volume 50, March- 2026

CUMULATIVE GENETIC RISK SCORE FOR PREDICTING MANIFEST PERSISTENT CYTOMEGALOVIRUS INFECTION, A FIVE-LOCUS MODEL IN THE FERGANA VALLEY POPULATION

Djuraev M. G.
Azimova S. B.
Vasilevskiy E. A.
Dadabaev O. T.
Andijan State Medical Institute, Andijan, Uzbekistan

ABSTRACT

Objective: to develop and evaluate a cumulative genetic risk score based on five single nucleotide polymorphisms (SNPs) for predicting the manifest persistent form of cytomegalovirus infection (CMVI).

Materials and methods. Genotyping for TNF-α (G308A), IL-10 (G>A), TLR4 (A896G), eNOS (G894T), and SOD2 (Ala16Val) was performed in 100 CMVI patients (Group I, n=42, latent; Group II, n=58, manifest persistent) and 80 healthy controls using real-time PCR with TaqMan probes (CFX96, Bio-Rad). A cumulative genetic risk score (0–10 points) was calculated by assigning 0 points for each protective homozygous genotype, 1 for heterozygous, and 2 for risk homozygous per locus. ROC analysis was used to evaluate diagnostic performance.

Results. A high genetic score (≥6) was found in 65.5% of patients with the manifest form and only 16.7% with the latent form (χ^2=24.12; p<0.001; OR=9.64; 95% CI 3.58–25.9). At the threshold of ≥6, sensitivity was 65.5%, specificity 83.3%, and AUC=0.78 (95% CI 0.69–0.87). Combining the genetic score with viral load (>10^3 copies/mL) increased AUC to 0.85 (95% CI 0.77–0.93), specificity to 90.5%, and positive predictive value to 90.7%. No individual polymorphism achieved comparable diagnostic accuracy (individual AUC range: 0.61–0.68).

Conclusion. The five-locus cumulative genetic risk score provides clinically meaningful stratification of CMVI patients. The combined model integrating genetic and virological data achieves high diagnostic accuracy and may serve as a practical tool for personalized management.

Keywords: Cytomegalovirus infection, genetic risk score, single nucleotide polymorphisms, ROC analysis, personalized medicine, risk stratification.

Introduction

Cytomegalovirus infection (CMVI), caused by Cytomegalovirus humanbeta5 of the Orthoherpesviridae family, represents a major global health burden with seroprevalence rates ranging from 40% in developed countries to over 95% in developing regions, including Central Asia (3, 15). The clinical spectrum of CMVI extends from asymptomatic latency to severe multi-organ involvement in neonates, pregnant women, and immunocompromised patients (2, 12). Predicting which patients will progress from latent to manifest persistent infection remains a key challenge for clinicians, as traditional laboratory markers such as viral load and C-reactive protein reflect the current disease state rather than the inherent susceptibility of the host (6, 12).

Genetic factors modulating the immune response, oxidative stress defense, and endothelial function play a significant role in determining the clinical outcome of CMVI. Single nucleotide polymorphisms (SNPs) in genes encoding tumor necrosis factor alpha (TNF-α, rs1800629), interleukin-10 (IL-10, rs1800896), Toll-like receptor 4 (TLR4, rs4986790), endothelial nitric oxide synthase (eNOS, rs1799983), and mitochondrial superoxide dismutase (SOD2, rs4880) have been individually associated with susceptibility to various infectious diseases (1, 5, 7, 8, 14). However, the individual discriminative power of any single SNP for predicting CMVI progression has proven insufficient for clinical application, with reported AUC values typically below 0.70 (7, 8).

The polygenic nature of susceptibility to chronic viral infections suggests that a composite approach integrating multiple genetic variants may yield superior predictive accuracy. Cumulative genetic risk scores, which aggregate the effects of several SNPs into a single quantitative measure, have demonstrated clinical utility in cardiovascular disease, oncology, and pharmacogenomics (3, 6). However, this approach has not been applied to CMVI in Central Asian populations, where high seroprevalence and intense viral circulation create conditions in which host genetic factors may be particularly influential (3, 15).

The present study aimed to develop and evaluate a cumulative genetic risk score based on five SNPs (TNF-α, IL-10, TLR4, eNOS, SOD2) for predicting the manifest persistent form of CMVI in the Fergana Valley population of Uzbekistan, and to assess whether combining the genetic score with viral load data improves diagnostic performance.

Materials and Methods

The study was conducted at medical institutions of the Fergana Valley (Zam-Zam Clinic, Andijan; Namangan Central Polyclinic; Fergana Regional Infectious Disease Hospital) during 2019–2025. A total of 100 patients with verified CMVI and 80 healthy controls matched by sex, age, and ethnicity were enrolled. Diagnosis was established by detection of anti-CMV IgM and/or rising IgG titers (ELISA, Cobas e411, Roche), CMV DNA quantification by real-time PCR (Rotor-Gene Q, Qiagen), and compatible clinical presentation.

Patients were classified into Group I (n=42) with latent or subclinical CMVI (viral load <1000 copies/mL, no significant complications) and Group II (n=58) with manifest persistent CMVI (viral load >1000 copies/mL, organ involvement, recurrent course). Groups were comparable by sex and age distribution ($p>0.05$). The multicenter design ensured population representativeness across the Fergana Valley.

Genomic DNA was extracted from venous blood (QIAamp DNA Blood Mini Kit, Qiagen). Allelic discrimination was performed by real-time PCR with TaqMan probes on CFX96 amplifier (Bio-Rad) for five polymorphisms: TNF-α G308A (rs1800629), IL-10 G>A (rs1800896), TLR4 A896G (rs4986790), eNOS G894T (rs1799983), and SOD2 Ala16Val (rs4880). Hardy-Weinberg equilibrium (HWE) was verified in the control group for all loci. The cumulative genetic risk score was calculated according to a three-tier system: 0 points for each protective homozygous genotype, 1 point for each heterozygous genotype, and 2 points for each risk homozygous genotype, yielding a total range of 0 to 10 points across five loci. Risk categories were defined as: low (0–2), intermediate (3–5), high (6–8), and very high (9–10). The combined model integrated the genetic score (≥6) with high viral load ($>10^3$ copies/mL).

Statistical analysis was conducted using SPSS 26.0 and R 4.2. Genotype frequencies were compared using Pearson χ^2 with Yates correction; Fisher exact test was applied when expected frequencies were <5. Odds ratios (OR) and relative risks (RR) with 95% confidence intervals (CI) were calculated. Diagnostic performance was assessed by ROC analysis with calculation of AUC, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). Statistical significance was set at $p<0.05$.

Results

HWE was confirmed in the control group for all five loci ($p>0.05$ for all), validating the genotyping accuracy and sample representativeness. Individual polymorphism analysis demonstrated significant associations of all five SNPs with the manifest persistent form, with risk allele frequencies consistently higher in Group II compared to controls. However, no single polymorphism achieved an AUC above 0.70 when evaluated individually, underscoring the need for a composite approach.

Distribution of cumulative genetic risk scores

The distribution of genetic risk scores differed profoundly between clinical groups (Table 1). In Group I, the majority of patients (83.3%) had low or intermediate scores (0–5), whereas in Group II, the distribution shifted markedly toward higher values, with 65.5% of patients scoring ≥6. This pattern indicates that the accumulation of unfavorable genetic variants across multiple pathogenetic pathways substantially increases the probability of manifest disease.

American Journal of Interdisciplinary Research and Development
ISSN Online: 2771-8948
Website: www.ajird.journalspark.org
Volume 50, March- 2026

Table 1. Distribution of cumulative genetic risk scores by clinical group

№	Risk category (score)	Group I (n=42)	Group II (n=58)	p	OR (95% CI)
1.	Low (0–2)	16 (38.1%)	4 (6.9%)	<0.001	0.12 (0.04–0.39)
2.	Intermediate (3–5)	19 (45.2%)	16 (27.6%)	0.07	0.46 (0.20–1.04)
3.	High (6–8)	6 (14.3%)	25 (43.1%)	0.002	4.56 (1.68–12.4)
4.	Very high (9–10)	1 (2.4%)	13 (22.4%)	0.004	11.9 (1.50–94.0)

Among patients with low genetic risk (0–2 points), 38.1% belonged to Group I and only 6.9% to Group II, yielding an OR of 0.12 (95% CI 0.04–0.39; p<0.001), which confirms the strong protective effect of a favorable genetic profile. Conversely, high scores (≥6) were present in 65.5% of Group II patients compared to 16.7% of Group I (χ^2=24.12; p<0.001; OR=9.64; 95% CI 3.58–25.9; RR=3.93; 95% CI 1.91–8.08). The very high category (9–10) was almost exclusively represented in Group II (22.4% vs 2.4%; OR=11.9; p=0.004), indicating that maximal genetic burden carries the most pronounced clinical consequences.

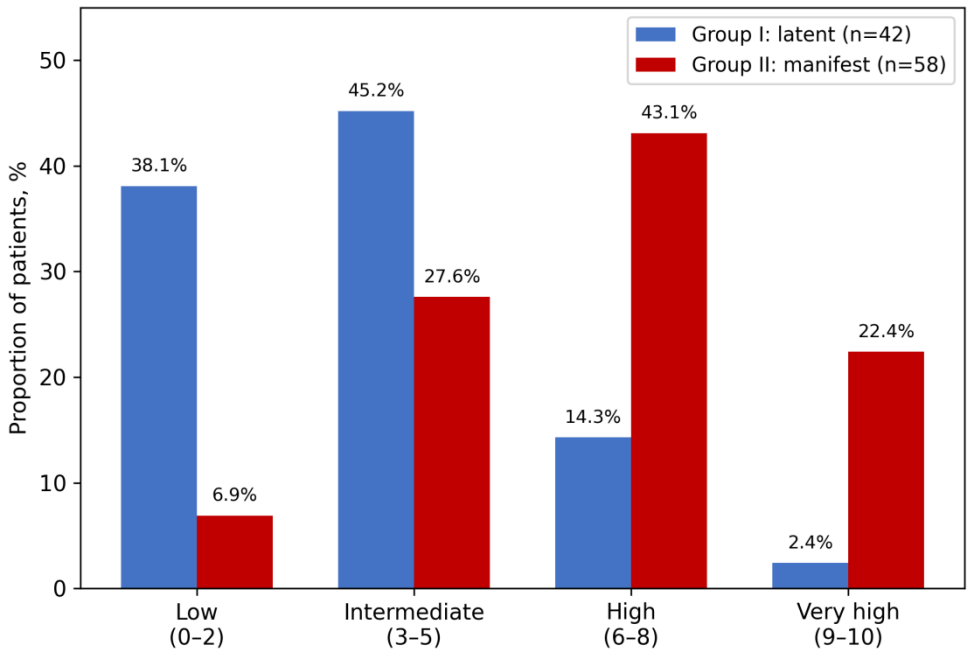

Fig. 2. Distribution of cumulative genetic risk scores by clinical group

The bar chart demonstrates the contrasting distribution patterns between groups. In Group I, the modal category is intermediate (45.2%), with a pronounced leftward skew. In Group II, the distribution is shifted rightward, with the modal category being high (43.1%).

This visual representation reinforces the quantitative findings and illustrates the practical utility of the score for patient stratification.

ROC analysis and diagnostic performance

At the threshold of ≥6, the genetic score demonstrated a sensitivity of 65.5% (95% CI 51.9–77.5), specificity of 83.3% (95% CI 68.6–93.0), PPV of 79.2%, NPV of 71.4%, and AUC of 0.78 (95% CI 0.69–0.87; p<0.001). When the genetic score was combined with a viral load criterion (>10^3 copies/mL), the diagnostic performance improved substantially: AUC increased to 0.85 (95% CI 0.77–0.93; p<0.001), sensitivity was 67.2%, specificity reached 90.5%, PPV rose to 90.7%, and NPV was 66.7% (Table 2).

Table 2. Diagnostic performance of the genetic score and combined model

№	Parameter	Genetic score ≥6	Combined model*
1.	AUC (95% CI)	0.78 (0.69–0.87)	0.85 (0.77–0.93)
2.	Sensitivity, %	65.5 (51.9–77.5)	67.2
3.	Specificity, %	83.3 (68.6–93.0)	90.5
4.	PPV, %	79.2	90.7
5.	NPV, %	71.4	66.7
6.	p-value	<0.001	<0.001

* Combined model: genetic score ≥6 + viral load >10^3 copies/mL

The superiority of the combined model over the genetic score alone is attributable to the complementary nature of the two components. The genetic score captures the stable, lifelong predisposition of the host, while viral load reflects the current state of viral replication. Integrating these static and dynamic parameters yields a more comprehensive risk assessment than either measure provides individually.

Comparison with individual polymorphisms

To evaluate the added value of the composite approach, the diagnostic performance of each individual polymorphism was compared with the cumulative genetic score and the combined model (Table 3). None of the individual SNPs reached the clinically meaningful threshold of AUC > 0.75, confirming that a single-marker strategy is insufficient for reliable patient stratification. The best-performing individual SNP was IL-10 (AUC = 0.68), followed by SOD2 (AUC = 0.66), while TLR4 demonstrated the lowest individual discriminative power (AUC = 0.61) despite having one of the highest individual OR values, reflecting the low population frequency of the mutant allele.

American Journal of Interdisciplinary Research and Development
ISSN Online: 2771-8948
Website: www.ajird.journalspark.org
Volume 50, March- 2026

Table 3. Comparative diagnostic performance of individual SNPs, cumulative genetic score, and combined model

№	Predictor	AUC	Sens., %	Spec., %	PPV, %	NPV, %
1.	IL-10 (rs1800896)	0.68	56.9	72.5	–	–
2.	SOD2 (rs4880)	0.66	53.4	70.8	–	–
3.	TNF-α (rs1800629)	0.64	51.7	68.3	–	–
4.	eNOS (rs1799983)	0.63	48.3	66.7	–	–
5.	TLR4 (rs4986790)	0.61	34.5	87.5	–	–
6.	Genetic score ≥6	0.78	65.5	83.3	79.2	71.4
7.	Combined model*	0.85	67.2	90.5	90.7	66.7

* Combined model: genetic score ≥6 + viral load >10^3 copies/mL; Sens. = sensitivity; Spec. = specificity

As shown in Table 3, the composite genetic score provided an absolute AUC improvement of 0.10–0.17 compared to individual SNPs, while the combined model yielded a further increment of 0.07. The stepwise improvement from single SNP to composite score to combined model confirms the additive value of each level of integration. Notably, TLR4 exhibited the highest specificity among individual SNPs (87.5%) but the lowest sensitivity (34.5%), reflecting the low frequency but high impact of the mutant allele. This pattern is consistent with the role of TLR4 as a binary gatekeeper of innate immunity: when disrupted, the consequences are severe, but the population frequency of disruption is relatively low.

The correlation between the genetic score and viral load was moderate ($r=0.44$; $p<0.001$), indicating that these parameters capture partially overlapping but largely independent aspects of infection pathogenesis. This moderate correlation supports the biological rationale for combining them: the genetic score reflects the inherent capacity of the host to resist viral persistence, while viral load reflects the current balance between viral replication and host defense.

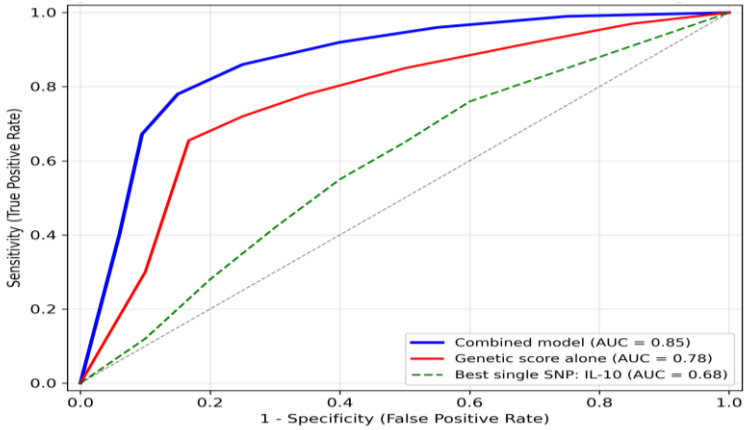

Fig. 1. ROC curves: genetic score alone, combined model, and best individual SNP (IL-10)

The ROC curves clearly illustrate the incremental gain in diagnostic accuracy from single SNP to composite score to combined model. The curve of the best individual SNP (IL-10, green dashed line) remains close to the diagonal, reflecting limited discriminative ability. The genetic score curve (red) demonstrates substantially better performance, and the combined model (blue) achieves the highest accuracy, with the curve approaching the upper left corner of the plot.

Discussion

The present study demonstrates that a cumulative genetic risk score integrating five SNPs across distinct pathogenetic pathways provides clinically meaningful stratification of CMVI patients. The rationale for combining these specific polymorphisms is grounded in their non-overlapping biological functions: TLR4 mediates innate immune recognition, TNF-α and IL-10 regulate the pro/anti-inflammatory balance of adaptive immunity, SOD2 controls mitochondrial antioxidant defense, and eNOS maintains endothelial function and NO-dependent antiviral activity (1, 5, 7, 8, 14). By capturing multiple levels of host defense simultaneously, the composite score overcomes the inherent limitation of single-marker approaches.

The AUC of 0.78 for the genetic score alone represents a substantial improvement over individual SNPs (AUC range 0.61–0.68) and exceeds the threshold of 0.75 generally considered acceptable for clinical screening tools (6). The combined model, incorporating viral load, further improves accuracy to AUC = 0.85 with a specificity of 90.5%, which is particularly valuable for ruling in high-risk patients who require intensive monitoring and early antiviral therapy. Griffiths P. and Reeves M. (2021) emphasized the importance of early identification of high-risk patients for optimizing CMVI management outcomes (6), and our findings provide a practical tool for achieving this goal.

A key advantage of the genetic score is its stability: unlike viral load, CRP, or transaminase levels, the genetic profile is determined once from a standard blood sample and remains valid throughout the patient's lifetime (3, 6). This property makes it particularly suitable for long-term prognostication and for identifying patients who require enhanced surveillance even during periods of apparent clinical remission. Goodrum F. et al. (2021) highlighted the unpredictable nature of CMV reactivation and the need for host-based prognostic markers that are independent of the current disease state (4), which aligns precisely with the characteristics of our genetic score.

The finding that 81.8% of children with congenital CMVI in Group II had genetic scores of ≥6 raises the intriguing possibility that maternal genetic predisposition may influence vertical transmission risk. Boppana S.B. et al. (2021) demonstrated the devastating consequences of congenital CMVI (2), and Lanzieri T.M. et al. (2017) called for improved screening strategies (9). If maternal genotyping proves predictive of vertical transmission risk, this could open new avenues for preconception screening and targeted prevention of congenital CMVI.

The population-specific relevance of our findings should be emphasized. The Fergana Valley, with its high population density, extensive household contact patterns, and estimated CMV seroprevalence of 85–95%, represents a setting where virtually all individuals are exposed to the virus (3, 15). In such an environment, the clinical outcome is determined not by whether infection occurs, but by how the host responds to it. Genetic profiling therefore becomes a particularly powerful prognostic tool, as it addresses the key determinant of disease progression in a population with near-universal exposure.

Several limitations should be acknowledged. The sample size of 100 patients, while adequate for detecting the observed effect sizes, limits the statistical power for subgroup analyses and for evaluating rare genotype combinations. The scoring system assigns equal weight to all five loci, whereas the actual contribution of each polymorphism may differ. Future studies employing weighted scoring based on regression coefficients or machine learning algorithms may achieve higher accuracy. Multicenter validation across other Central Asian populations and diverse ethnic groups is needed before widespread clinical implementation. Expansion of the panel to include interferon response genes (IFN-γ, IFNL3), HLA-G, and KIR receptors may further enhance predictive power (10, 11).

From an economic standpoint, the proposed five-SNP panel is a one-time investment that yields lifelong prognostic information. The cost of genotyping is comparable to a single biochemical blood test and is considerably less than a single MRI scan, while the clinical value of the result persists indefinitely (3). In a healthcare system with limited resources, such as Uzbekistan, this cost-effectiveness profile is particularly attractive. The preliminary implementation experience at Zam-Zam Clinic and Namangan Central Polyclinic (2024–2025) has demonstrated the feasibility of integrating genotyping into routine clinical workflows, with 68 patients genotyped and 24 (35.3%) classified as high genetic risk. Among these high-risk patients, intensified monitoring enabled timely initiation of antiviral therapy before the development of severe complications.

The clinical implications of our findings extend beyond individual patient management. At the population level, genetic profiling could inform targeted screening programs for pregnant women, neonates, and transplant recipients, allowing healthcare systems to concentrate resources on individuals most likely to benefit from intervention (2, 9, 12). The identification of high-risk genotypes before pregnancy could facilitate preconception counseling and early antiviral prophylaxis in women with unfavorable genetic profiles, potentially reducing the burden of congenital CMVI in the region (2, 9). These population-level applications represent an important direction for future research and policy development.

Conclusions

1. The five-locus cumulative genetic risk score (TNF-α, IL-10, TLR4, eNOS, SOD2; range 0–10) enables effective stratification of CMVI patients: a score of ≥6 is associated with 9.64-fold increased odds of manifest persistent infection (95% CI 3.58–25.9; $p<0.001$).

2. The genetic score alone provides a sensitivity of 65.5%, specificity of 83.3%, and AUC of 0.78, substantially exceeding the diagnostic performance of any individual polymorphism (AUC range 0.61–0.68).
3. Combining the genetic score (≥6) with viral load (>10^3 copies/mL) increases AUC to 0.85, specificity to 90.5%, and PPV to 90.7%, offering a robust tool for identifying patients requiring intensive monitoring and early antiviral intervention.
4. The genetic score is a stable, one-time measurement independent of the patient's current clinical state, making it uniquely suited for lifelong risk assessment and personalized management of CMVI.

References

1. Biolatti M., Dell'Oste V., Pautasso S. et al. Human CMV tegument protein pp65 modulates innate immune response // J. Virol. – 2018. – Vol.92(6). – P.87–94.
2. Boppana S.B., Ross S.A., Fowler K.B. Congenital cytomegalovirus infection: clinical outcome // Clin. Infect. Dis. – 2021. – Vol.57(Suppl 4). – P.S178–S181.
3. Cannon M.J., Schmid D.S., Hyde T.B. Review of cytomegalovirus seroprevalence // Rev. Med. Virol. – 2010. – Vol.20. – P.202–213.
4. Goodrum F., Caviness K., Zagallo P. Human cytomegalovirus persistence // Cell. Microbiol. – 2021. – Vol.14(5). – P.644–655.
5. Förstermann U., Sessa W.C. Nitric oxide synthases: regulation and function // Eur. Heart J. – 2017. – Vol.33. – P.829–837.
6. Griffiths P., Reeves M. Pathogenesis of human cytomegalovirus // Nat. Rev. Microbiol. – 2021. – Vol.19. – P.759–773.
7. Jabłońska A., Kocięcka B., Adrych K. TLR4 polymorphisms and viral infections // Immunobiology. – 2022. – Vol.227(1). – Art.152156.
8. Wilson A.G., Symons J.A. et al. Polymorphism in human TNF-alpha promoter // Proc. Natl. Acad. Sci. USA. – 2018. – Vol.94. – P.3195–3199.
9. Lanzieri T.M., Dollard S.C. et al. Systematic review of congenital CMV // J. Clin. Virol. – 2017. – Vol.68. – P.6–14.
10. Fishman J.A. Infection in organ-transplant recipients // N. Engl. J. Med. – 2017. – Vol.338. – P.1741–1751.
11. Slobedman B., Stern J.L. et al. Impact of CMV latent infection on myeloid progenitor cells // J. Virol. – 2019. – Vol.78(8). – P.4458–4467.
12. Razonable R.R., Humar A. Cytomegalovirus in solid organ transplantation // Am. J. Transplant. – 2020. – Vol.13(Suppl 4). – P.93–106.
13. Zelini P., Fornara C., Comolli G. Oxidative stress in pathogenesis of viral infections // J. Clin. Virol. – 2021. – Vol.94. – Art.104884.
14. Wang H., Li T., Zhang X. Global epidemiology of cytomegalovirus // J. Infect. Dis. – 2020. – Vol.216(Suppl 5). – P.S429–S435.
15. Zuhair M., Smit G.S.A., Wallis G. Systematic review of global CMV seroprevalence // BMC Public Health. – 2022. – Vol.22. – P.1659.