



OPEN

Molecular characterization of JC virus in progressive multifocal leukoencephalopathy cases from India

Gnana Soundari Palani, Surbhi Telang, Vijayalakshmi Reddy, K. Dhanya, M. A. Ashwini & Reeta S. Mani[✉]

Progressive Multifocal Leukoencephalopathy (PML) is a rare, often fatal demyelinating disease of the central nervous system caused by reactivation of the John Cunningham virus (JC virus) in immunocompromised individuals. Despite an estimated 2.4 million people living with HIV in India, the reported incidence of PML remains lower than in Western countries, likely due to underdiagnosis, underreporting, and distinct host genetic and viral factors. The rising number of individuals on immunosuppressive therapies, including organ transplant recipients and those with autoimmune disorders, further emphasizes the need to study JC virus diversity in the Indian context. This study aimed to characterize the genetic diversity of JCV in India by sequencing the VP1 and non-coding control region (NCCR) from cerebrospinal fluid (n=30) of confirmed PML cases using Sanger sequencing. VP1 sequencing (n=23) revealed a predominance of genotypes 2 (subtypes 2D, 2A, 2B) and 3A. NCCR analysis (n=17) showed extensive rearrangements relative to the archetype form, with most sequences classified as Type II-R. Structural variations, including deletions, duplications, and insertions were common, particularly in blocks D, C, and F. Transcription factor binding sites (TFBS) were identified for TATA box, Tst-1, SP-1, p53, CEBPB, AP-1, NF-1, EGR-1, GF-1, CRE-TAR and NFkB. Additional TFBS were created due to rearrangements, often spanning two blocks. These findings underscore the genomic diversity of JCV in India and highlight the need for continued molecular surveillance to better understand its implications for high-risk populations.

Keywords Progressive multifocal leukoencephalopathy, JC virus, Genotype, Non-coding control region, Transcriptional factor binding sites

Progressive Multifocal Leukoencephalopathy (PML) is a rare, often fatal, demyelinating disease of the central nervous system (CNS) primarily caused by reactivation of the John Cunningham virus (JC virus). JCV is a ubiquitous virus belonging to family *Polyomaviridae*. The seroconversion levels for JCV reach ~60–80% by the age of 70 years in general population¹. Although the established routes of transmission remain unclear, the respiratory and urine-oral routes, particularly in children, are proposed as the major modes of infection². JCV remains latent in the kidneys and lymphoid tissues of immunocompetent individuals. However, in those who are immunocompromised, such as individuals with HIV/AIDS, certain cancers, or those receiving immunosuppressive therapy for organ transplants or autoimmune diseases, it can reactivate and cause PML. PML causes progressive neurological decline, including cognitive deficits, motor dysfunction, and visual disturbances. Currently, no specific antiviral therapy is approved for its treatment.

JCV is a circular, non-enveloped, double-stranded DNA virus with a genome of ~5,130 bp. It encodes six major proteins: the early proteins (large T and small T antigens) and the late capsid proteins (VP1, VP2, VP3) along with the agnoprotein. Early and late genes are transcribed in opposite directions, separated by the non-coding control region (NCCR), which contains the origin (ORI) and promoter/enhancer elements organized into blocks A–F³. JCV is classified into eight genotypes based on VP1 sequences, with distinct geographic patterns: Types 1 and 4 in Europe and the USA, Types 2 and 7 in Asia, Types 3 and 6 in Africa, and Type 8 in Papua New Guinea and the Western Pacific⁴.

Department of Neurovirology, National Institute of Mental Health and Neuro Sciences, Bangalore, India. [✉]email: drreeta@gmail.com

JCV exists in two forms based on the NCCR structure: the non-pathogenic, latent archetype, typically found in the urine of both healthy and infected individuals, and the rearranged, pathogenic form associated with PML⁵⁻⁷. The NCCR of archetype consists of conserved blocks including - A (36bp), B (23bp), C (55bp), D (66bp), E (18bp) and F (69bp)^{8,9}. The NCCR appears in a rearranged state during the neurotropic phase, typically found in the brain, cerebrospinal fluid (CSF), or blood of a PML patient. Rearrangements of the blocks in the NCCR lead to duplications and deletions of specific sequence elements¹⁰, which are thought to play a role in virus pathogenesis by modifying its cellular tropism.

In India, studies on PML and JCV are limited, with most focusing on HIV-infected patients. Reported PML incidence in this group ranges from 1.2% to 3.5%^{11,12}, lower than the ~5% seen in Western countries¹³. This discrepancy is attributed to underreporting, underdiagnosis, and factors such as host genetics and viral strain diversity¹⁴.

With an estimated 2.4 million people living with HIV in India¹⁵, along with a growing population of patients on immunosuppressive therapies, there is a clear need for large-scale studies on JCV diversity and molecular epidemiology in the Indian context.

This study aimed to identify JCV genotypes and characterize their NCCR in CSF samples from confirmed PML cases. The objectives were to identify JCV genotypes based on VP1 sequences, assess sequence diversity within the NCCR, and map transcription factor binding sites (TFBS) across the NCCR, using Sanger sequencing, a high-accuracy DNA sequencing method.

Results

Genotypes of JCV

VP1 was amplified in 23/30 CSF samples. Sequence analysis of the VP1 from these 23 samples identified Genotype 2 in 18 samples, with subtypes 2D (n=10), 2A (n=6), and 2B (n=1). Genotype 3 was detected in 5 CSF samples, all of which belonged to subtype 3A (Figure 1).

Genetic diversity of the NCCR

NCCR was amplified in 17/30 CSF samples. All the 17 sequences exhibited variations compared to the archetype sequence and varied among each other. Analysis revealed that 'A' block was conserved among all 17 samples where NCCR was amplified. Deletions were a predominant feature across majority of 17 sequences, although insertions and duplications were also observed (Figure 2).

Based on further analysis of insertion and repeat patterns, 3/17 CSF samples were classified as NCCR Type IIS, while 14/17 samples were classified as NCCR Type IIR. Among the three NCCR type IIS samples, one sample (NNV-JCV-NCCR-CSF01) had 18bp insertion, one (NNV-JCV-NCCR-CSF22) exhibited a complete deletion of block 'D', while other two samples (NNV-JCV-NCCR-CSF01 and 15) had partial deletions of block 'D' and were classified as 'Archetype like' (Figure 2A).

Among the 14 Type IIR samples, four (28.5%) samples (NNV-JCV-NCCR-CSF02, 11, 18 and 19) had a block structure of 'ABCDE-EF' with a 69-base insertion between the two 'E' blocks. The insert size is the largest observed among the samples in the present study, the size of other inserts ranging from 6bp to 20bp (Figure 2C). Additionally, 3/14 (NNV-JCV-NCCR-CSF08, 13 and 24) samples of Type IIR had complete deletion of block 'D' (Figure 2B, 2C) and one sample (NNV-JCV-NCCR-CSF07) had a complete deletion of block 'B' (Figure 2B). Tandem repeats of blocks in various combinations were found in ten samples from the study. Duplication of blocks B, C, D, E and F were noted in several samples, while duplication of block 'A' was observed in only one sample (NNV-JCV-NCCR-CSF07). Truncation of blocks were observed frequently in the duplicated regions. Nucleotide substitutions were also observed in block F (7 samples), block C (6 samples) and block D (1 sample).

Host transcription factor binding sites (TFBS) in NCCR

Type IIR NCCR sequences had a minimum of 33 to a maximum of 49 TFBS while type IIS sequences had 23 to 27 TFBS (Supplementary file S1). In the present study, block D had maximum number of TFBS (7 TFBS) followed by block C (6 TFBS) and block F (5 TFBS). Truncations of blocks led to loss of one or more TFBS in all the sequences in the study, while additional TFBS were created at the junctions of two blocks due to rearrangements in 11/17 samples. All the isolates had a single binding site for TATA in block A, notably an additional TATA box was found in only one sample (NNV-JCV-NCCR-CSF07) in the truncated block D (Figure 2B_(k)). TATA binding proteins (TBP) were found in four samples (Figure 2C) between the duplicated E block which had the insert of 69bp. Additional TFBS were also found on the nucleotide inserts (1bp to 69bp) observed in eight samples (Figure 2A, 2C).

Discussion

JCV, was the first among human polyomavirus to be isolated in 1971¹⁶ and is believed to have co-evolved with humans¹⁷. Sero-epidemiological studies indicate that over half of the global population is either transiently or latently infected with JCV¹⁸ although there is variability reported across different populations. Additionally, 10–30% of the asymptomatic individuals shed JCV in the urine as the virus is latent in kidneys¹⁹. Under immunosuppressive conditions such as HIV/AIDS, organ transplantation, or autoimmune therapies, JCV can reactivate and cause PML. It primarily targets oligodendrocytes in the brain, leading to lytic infection and white matter lesions characteristic of the disease. PML incidence is significantly higher in AIDS patients compared to those with other causes of immunosuppression¹⁹. In India, the reported incidence of PML is lower despite a large population of people living with HIV^{14,20}, with viral strain diversity being one of several factors proposed to explain this discrepancy. Most studies to date have been limited to case reports or have focused on JCV in immunocompetent individuals from specific regions²⁰. This underscores the need for comprehensive molecular

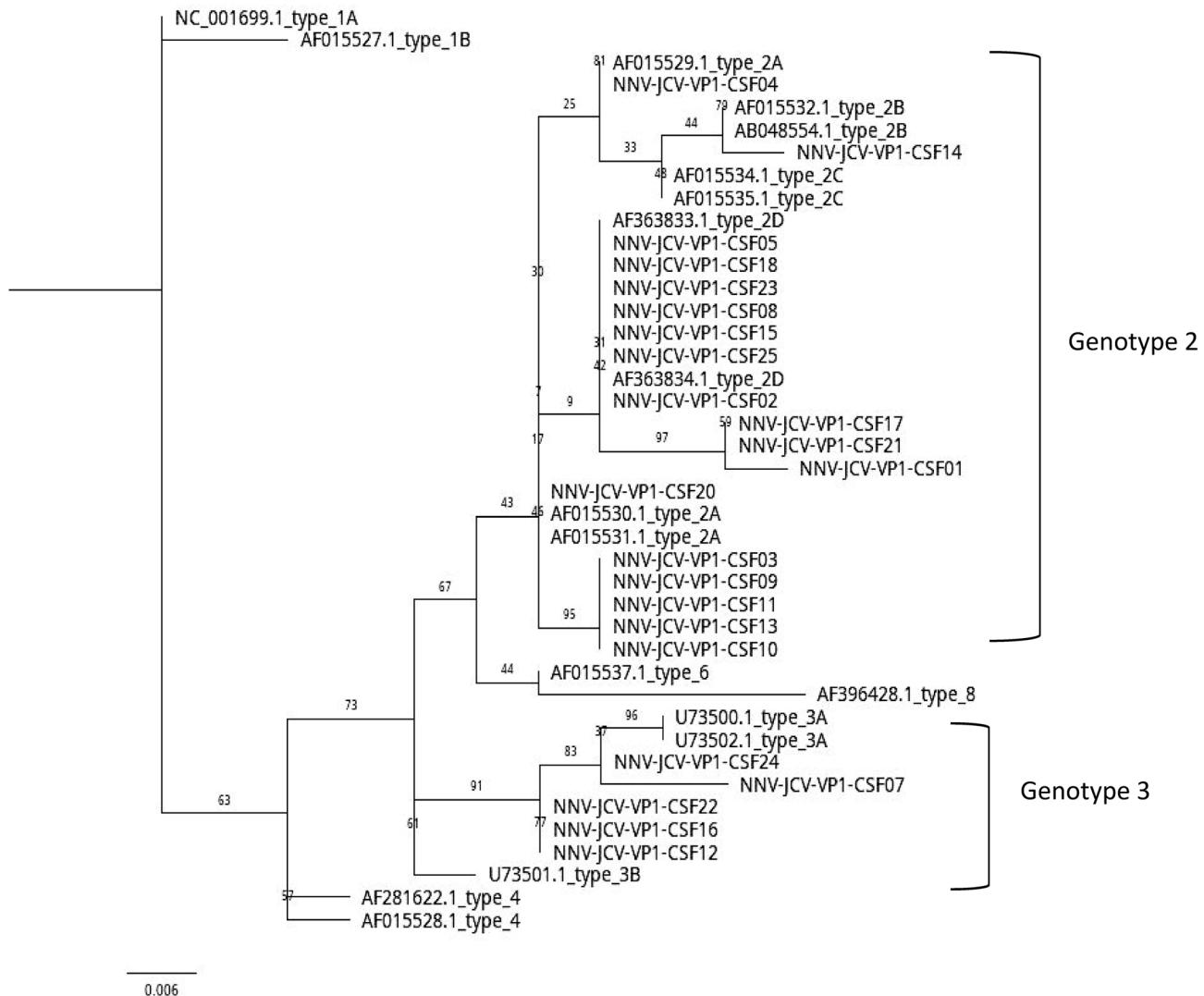
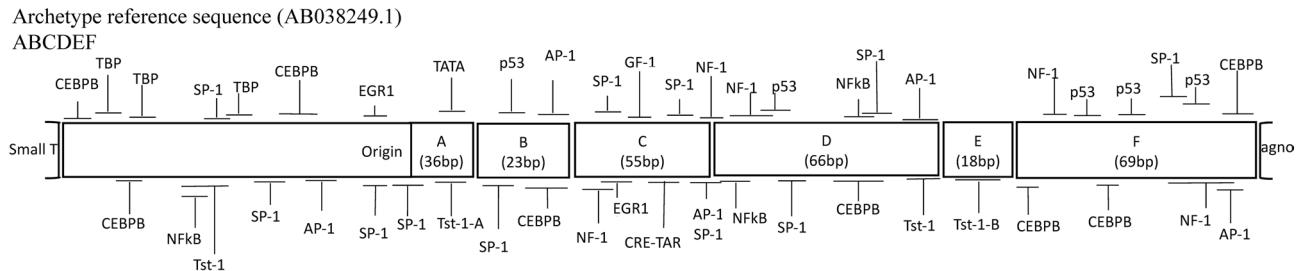


Figure 1. Phylogenetic analysis of VP1 region: MAFFT was used to align the sequence. Phylogenetic tree was build using iqtree with best fit model according to BIC with 1000 bootstrap replicates. Tree was visualized using figtree. Reference GenBank sequences: type 1A (NC_001699.1); type 1B (AF015527.1); type 2A (AF015529.1, AF015530.1, AF015531.1); type 2B (AF015532.1, AB048554.1); type 2C (AF015534.1, AF015535.1); type 2D (AF363833.1, AF363834.1); type 3A (U73500.1, U73502.1); type 3B (U73501.1); type 4 (AF281622.1, AF015528.1); type 6 (AF015537.1); type 8 (AF396428.1).

studies to understand JCV diversity in immunocompromised populations and its potential impact on disease outcomes relative to strain diversity.

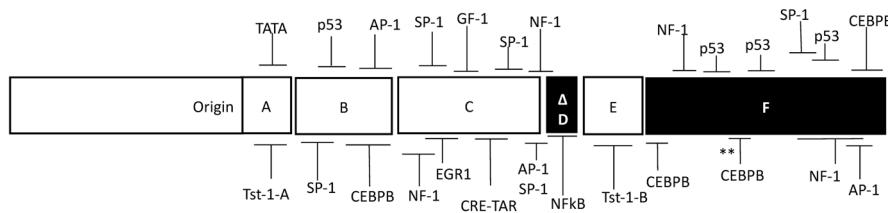
This study aimed to investigate genetic variation in JCV by analyzing CSF samples from confirmed PML cases collected between 2020 and 2023 (n=30), focusing on two key genomic regions: VP1 and NCCR. The VP1 region, part of the late gene segment, was sequenced to determine viral genotypes, as it serves as a reliable marker for classification. Genotyping is important for both epidemiological and pathological insights, as JCV types are linked to specific geographic populations²¹ and have been associated with PML risk in AIDS patients²². The NCCR, composed of blocks A-F, is highly variable and plays a key role in JCV replication and host-cell tropism. These blocks contain binding sites for host transcription factors that modulate viral gene expression and influence pathogenicity. JCV exists in two main forms based on NCCR structure: the archetype form, with a stable block sequence commonly found in urine, and the pathogenic form, characterized by rearrangements, deletions, or duplications. These alterations can modify transcription factor binding sites, enhancing viral replication and neuroinvasion, thereby contributing to PML development.

In this study, all 30 CSF samples tested positive for JCV DNA and were subjected to VP1 and NCCR amplification. The VP1 region was successfully amplified in 23 samples. Genotype analysis revealed that 18 samples (78%) belonged to genotype 2, and 5 (22%) to genotype 3. Among genotype 2 samples, 10 were subtype 2D, 6 were 2A, and 1 was 2B, with one sample unclassified. All genotype 3 samples were identified as subtype 3A. Genotype 2, prevalent in Asia, was the dominant genotype found in HIV-associated PML cases in India.^{4,23}



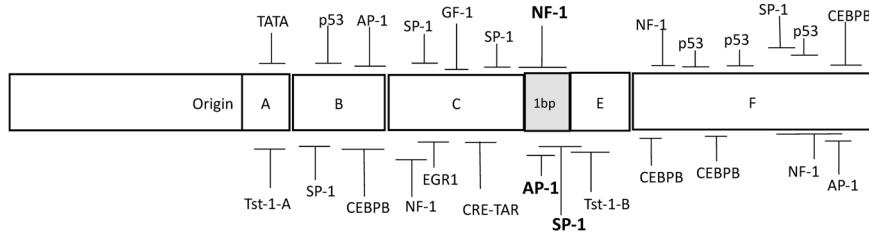
a) NNV-JCV-NCCR-CSF15 (318bp)

ABCDEF



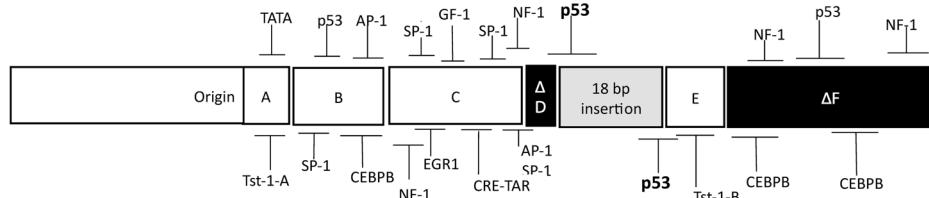
b) NNV-JCV-NCCR-CSF22 (317 bp)

ABC1bpEF



c) NNV-JCV-NCCR-CSF01 (314 bp)

ABCD18bpEF



While Agostini et al. 1997²⁴ reported a strong association of genotype 2B with PML in HIV patients, particularly in brain tissues, our findings align with Kannangai et al.²³, who detected only genotype 2D in CSF and brain tissue from both PML-positive and negative cases in India. Interestingly, subtype 2A, typically found in China and Japan, was detected in six samples, while the East Asian subtype 2B was found in only one sample²⁴. Subtype 3, considered an African type, was identified in three samples. This genotype and subtype diversity suggests regional variability, with genotype 2 likely representing the predominant circulating strain in India. Further studies are needed to determine whether specific subtypes are more strongly associated with PML in the Indian population.

In addition to genotyping, the NCCR was analyzed for sequence variation and transcription factor binding sites. NCCR was successfully amplified in 17 of 30 CSF samples. Compared to the archetype sequence, all showed extensive variability, with no two sequences identical. Most were classified as Type IIR (14/17), while 3 were Type IIS, consistent with previous reports identifying Type IIR as the dominant form in PML cases²⁵. Deletions were the most common alteration, though insertions and duplications were also observed.

Consistent with previous studies, block A was conserved across all NCCR sequences, with duplication observed in only one sequence (NNV-JCV-NCCR-CSF07) in a truncated form. The TATA box was also conserved and present as a single copy in all complete block A sequences. Notably, block E was duplicated in four sequences, each containing a unique 69-base insert between the duplicated E blocks, a feature not previously reported. BLAST analysis of this insert showed similarity to the JCV Large T antigen and contained seven transcription factor binding sites including two each for TBP and p53, and one each for CEBPB, AP-1, and SP-1. A significant finding in this study was the detection of NCCR Type IIS or archetype-like strains, in three CSF samples (NNV-JCV-NCCR-CSF 01, 15 and 22). All showed deletions in block D, either complete or partial.

Figure 2. DNA sequence blocks of NCCR rearrangement obtained in this study in comparison to archetype reference sequence: Upper case letters A, B, C, D, E, F indicate sequence blocks. Transcriptional binding sites are labelled on each block. Clear blocks indicate sequence blocks similar to archetype. Black colored blocks with Δ indicates truncated regions with the loss of one or more transcriptional binding sites. Black colored blocks without Δ indicates mutations or one to three nucleotide deletion which does not affect the transcriptional binding sites. Grey colored blocks indicate insertions with their base pairs. Additional transcriptional binding formed at the rearranged junctions are highlighted in bold. TATA (TATA box), TBP (TATA binding proteins), Oct6/Tst-1A, Tst-1B (POU domain protein or Octamer binding protein 6), SP-1 (Specificity protein 1), p53 (cellular tumor antigen p53), CEBPB (CCAAT/enhancer binding protein beta), AP-1 (Activating protein 1), NF-1 (Nuclear factor 1), EGR-1 (Early growth response protein-1), GF-1 (Glial factor 1), CRE-TAR (cyclic AMP response element, Transactivating response element), NFkB (Nuclear factor NF-kappa-B subunit). ** Indicates 2nt deletion and/or 1nt substitution observed in this study and retains all the binding sites in F block. * Indicates 1nt substitution or deletion observed in this study and results in the loss of binding sites for AP-1, SP-1 and/or NF-1. (A) **Type IIS archetype like sequence:** (a) NNV-JCV-NCCR-CSF15 with truncated D; (b) NNV-JCV-NCCR-CSF22 with complete deletion of block D and 1 bp insertion between C and F retains transcriptional binding site for NF-1, AP-1 and SP-1; (c) NNV-JCV-NCCR-CSF01 with truncated D and F, 18bp insertion between truncated D and E with two copies of p53. (B) **Type IIR NCCR rearrangement with deletions:** Deletions and repeats of different blocks. Rearrangements are frequently targeted in the junction of blocks-p53 between F/C (e); NF-1 between E/C (f); p53 between C/D (h); NF-1, SP-1 and/or AP-1 between C/E (i,j); p53 between F/B (i); AP-1 between A/C (k); Truncated D with second copy of NF-1 and TATA box (k). Complete loss of D block (i,j) and complete loss of B block (k). (C) **Type IIR NCCR rearrangement with deletions and insertions:** (l) 69bp insertion between two E blocks with two copies of TBP, two copies of p53 and single copy of Tst-1, CEBPB, AP-1 and SP-1; (m) SP-1 in the junction of E/F, 18bp insertion between F/D with SP-1, AP-1 and NF-1; (n) Complete loss of D block, 6bp insertion between C/F with SP-1 and p53, 20bp insertion between F/E with SP-1, AP-1 and NF-1.

While such changes are typically observed in urine samples from healthy individuals²⁶, previous studies have also reported archetype-like strains in CSF of PML patients with prolonged survival, suggesting their potential as prognostic markers for better outcomes²⁷.

The widely accepted hypothesis is that deletions in the regions closer to the late genes represent a loss of function, by removing a suppressing control transcriptional sequence, whereas duplications in the left end of the NCCR represent a gain of function, increasing activating control sequences that enhance viral replication and gene transcription. It is of interest to note that in the present study, we too observed higher frequency of rearrangements (deletions/duplications) in the NCCR closer to the late genes²⁸.

We also examined the presence of transcription factor binding sites (TFBS) across the NCCR blocks A-F, which are essential for the JCV life cycle including TATA box, SP-1, NF-1, CRE-TAR, GF-1, AP-1, Tst-1, p53, CEBPB and EGR-1. These TFBS are essential for viral replication, transcription, and reactivation, and the presence of these binding sites may influence the ability of the virus to persist and reactivate, especially in immuno-compromised individuals.

In this study, blocks ABC proximal to origin of replication are present in 16/17 samples with the binding sites for Tst-1, TATA (block A), SP-1, p53, CEBPB, AP-1 (block B) and NF-1, SP-1, EGR-1, GF-1, CRE-TAR, AP-1 (block C). Though block D had maximum number of binding sites, partial or complete deletions in most of the samples leads to loss of one or more TFBS (NFkB, NF-1, p53, SP-1, Tst-1, AP-1, CEBPB). All the samples had an intact block E with the binding site for Tst-1; While block F was present in all the samples with TFBS for CEBPB, NF-1, p53, SP-1, AP-1, except one where the truncated F resulted in the loss of all TFBS. Furthermore, rearrangements led to the emergence of additional TFBS spanning two blocks (C/E, D/E, F/C, E/C, C/D, F/B, A/C, E/F, E/F, C/F, F/E) in ten samples.

Among these factors, NF-1, SP-1, TAR, Tst-1 have been well studied for their role in JCV gene expression, regulation and replication. Multiple studies have shown that NF-1 binding sites in the NCCR of JCV genome are associated with JCV replication in glial cells^{29,30}. It has been demonstrated that the NF-1 site increases the replication of JCV DNA in transient transfection assay³¹. In the present study, rearrangement of blocks C, D and F resulted in the multiple binding sites for NF-1. In contrast, the archetype form of JCV does not have repeats, leading to reduction in the number of binding sites for NF-1 family of TFBS, which are critical for the activation of viral transcription in brain tissue and lymphoid cells⁴.

SP-1, a zinc finger-containing transcription factor is retained in all the samples and has binding sites in block B, C, D and F. Research has shown that the SP-1 binding site, often referred to as the GA box, is conserved in many promoters of glial cell-specific genes, suggesting that SP-1 may play a role in activating the JCV promoter in glial cells^{32,33}. SP-1 has dual function in JCV replication, it activates the JCV early promoter in both glial and non-glial cells and it also facilitates TAg-mediated transactivation of viral early genes^{33,34}. Previous studies have also shown that both SP-1 binding sites in the A-F blocks are typically intact in non-AIDS PML isolates but often deleted in AIDS-related cases, suggesting a role for SP-1 in PML among non-HIV patients³⁵. However, in our study, both SP-1 sites were present in all the 14 sequences from HIV-infected patients. This variation may reflect differences in JCV strains or other factors influencing PML in the Indian population, indicating a more complex role for SP-1 in PML pathogenesis. These findings underscore the need for further research on SP-1's role across JCV variants and patient groups.

All the NCCR sequenced in this study had a binding site for GF-1 and TAR. GF-1 is a key transcription factor believed to interact with specific binding sites in the JCV promoter region, enhancing or regulating viral

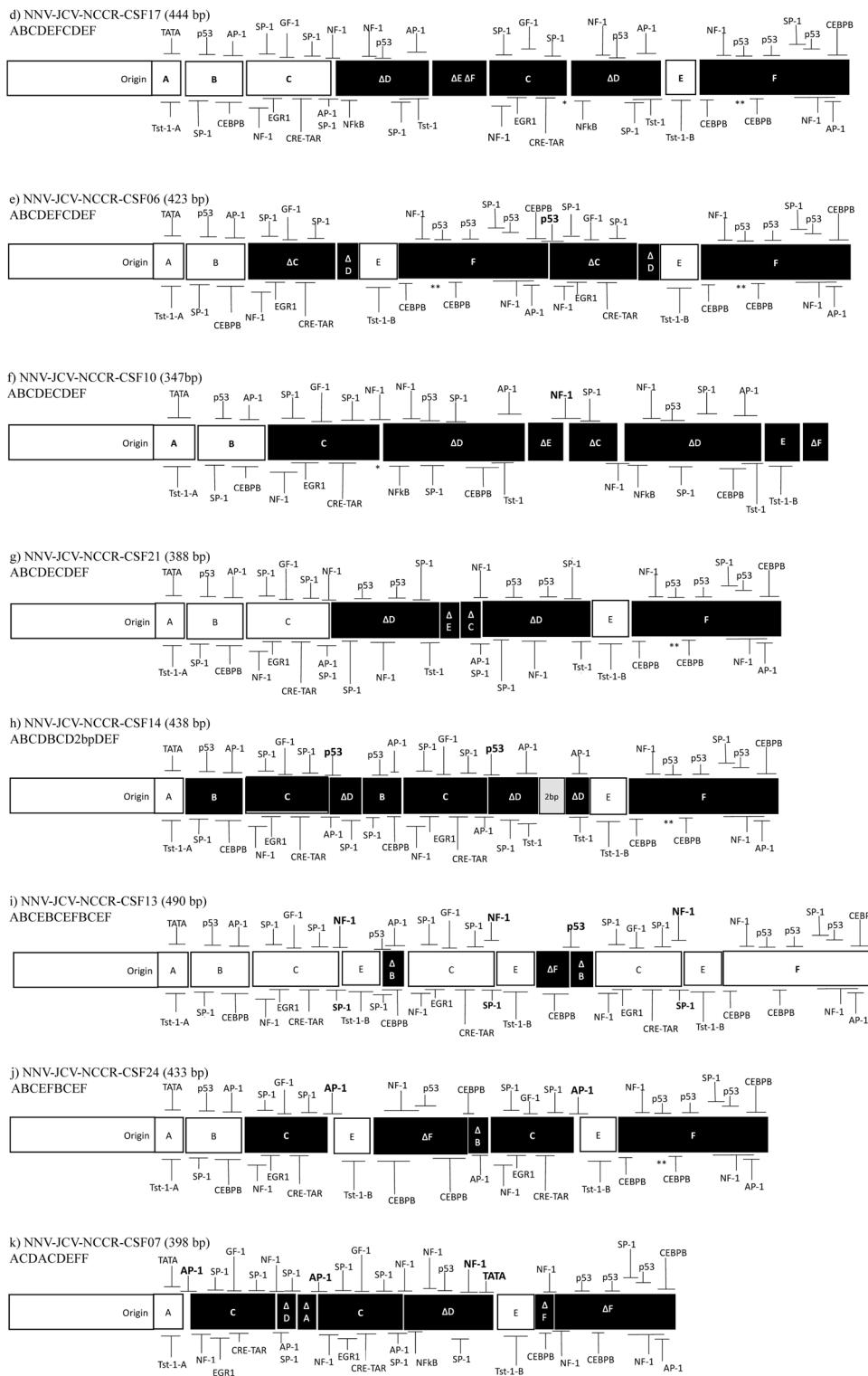
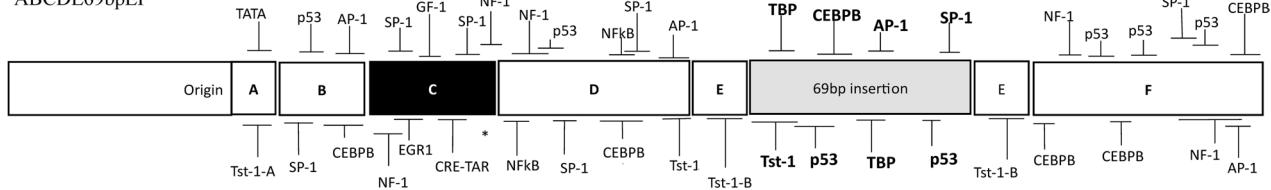


Fig. 2. (continued)

replication in glial cells such as oligodendrocytes and astrocytes, the primary targets of JCV. This suggests that GF-1 may play a role in modulating JCV replication and persistence within these cells^{35,36}. The Transactivation Response Element (TAR) is another important transcription factor known to modulate JCV gene expression. The HIV-1 Tat protein is known to bind to the TAR site and stimulate the activation of the JCV late promoter, thereby facilitating viral replication. This interaction may explain the higher incidence of PML in HIV-infected individuals^{37,38}.

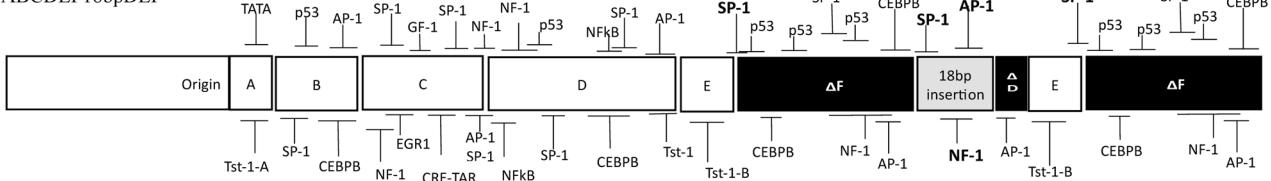
i) NNV-JCV-NCCR-CSF02 (435 bp), CSF11
(469BP), CSF18 (441bp), CSF19 (440bp)

ABCDE69bpEF



m) NNV-JCV-NCCR-CSF23 (433 bp)

ABCDEF18bpDEF



n) NNV-JCV-NCCR-CSF08 (424 bp)

ABC6bpFBC6bpF20bpEF

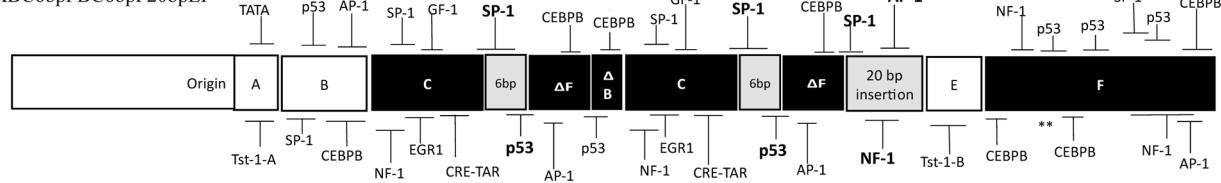


Fig. 2. (continued)

Tst-1 binding sites were present in blocks A, D and E. Four sequences had duplicated block E resulting in two binding sites for Tst-1 except one (NNV-JCV-CSF13) which had three. Tst-1 binding site in block A was conserved in all the sequences in comparison to block D. Tst-1 has been shown to stimulate viral gene expression and may contribute to the glial specificity of JCV by promoting its replication in glial cells³⁹.

In addition to the above transcription factors, binding sites for transcription factor AP-1 were identified in block C of seven sequences, whereas multiple AP-1 sites were formed in the newly created junctions. AP-1 is known to regulate JCV transcription in glial cells⁴⁰. EGR-1 binding sites were found in all sequences, consistent with its role in activating the JCV late promoter⁴¹. NF- κ B sites were detected in nine sequences and CEBPB in all, both factors known to influence JCV latency and reactivation⁴². Additionally, TP53 binding sites were present in all the sequences including in the newly created junctions suggesting a potential role for p53 in JCV transcriptional regulation and PML pathogenesis^{43,44}.

While this study offers valuable insights into the molecular characteristics of JCV from PML cases in India, it has a few limitations. Exploring the correlation between viral genotypes, NCCR rearrangements, and clinical outcomes could further enhance understanding of how specific JCV variants influence PML progression. The genomic analysis in this study was restricted to the VP1 and NCCR regions of JCV. Further, VP1 sequencing did not include analysis of C-terminal region which has been implicated in the development of JC virus granule cell neuronopathy (JCV GCN), a lesser known but clinically relevant manifestation of JCV infection. Investigating other JCV genomic regions, along with host genetic factors and immune responses, may offer a more comprehensive view of the virus-host interactions that influence disease outcomes. Future research incorporating these aspects and larger, longitudinal cohorts are essential to fully understand JCV diversity and its clinical implications in the Indian population.

Conclusion

The genetic variability observed in the VP1 and NCCR regions underscores the need for continuous surveillance and molecular epidemiological studies to track circulating JCV strains, especially given the growing high-risk populations in India. Moreover, identifying specific transcription factor binding sites within the NCCR of Indian JCV isolates may offer promising targets for developing antiviral therapies to prevent JCV reactivation.

Methods

Clinical samples

A total of 30 archived de-identified CSF samples from PML cases, which tested positive for JCV DNA by commercial real-time PCR, and received during 2020-2023 were used in this study. The samples included 28 patients with HIV, 1 each from a renal transplant recipient and a patient with chronic lymphocytic leukemia.

The study protocol was reviewed and approved by the Institutional Ethics Committee of NIMHANS (Letter No. NIMHANS/50th IEC (BS & NS DIV.)/2024 dated 26-10-2024). The requirement for informed consent was waived by the ethics committee, as this retrospective study involved no direct contact with participants and used anonymized leftover clinical samples collected during routine diagnostic procedures. All the experiments were carried out in accordance with relevant guidelines and regulations.

DNA extraction

Viral DNA was extracted from the CSF samples using QIAamp DNA blood mini kit (QIAGEN, cat.no 511904) according to the manufacturer's instructions. The eluted DNA was stored at -80°C until further testing.

Amplification of VP1 and NCCR region

The initial step of the PCR involved linearization of the JCV circular genome followed by the whole genome amplification. The resulting PCR product was then used as a template for the amplification of VP1 and NCCR in separate PCR reactions. The linearization of the genome through cleavage at BamHI restriction sites and whole genome amplification was achieved using protocol described earlier by Agostini et al., 1995⁴⁵. Briefly, 5µl of extracted DNA was digested with 2U of BamHI in a total volume of 15µl for 1 hour at 37°C, followed by heat inactivation at 65°C for 10 minutes. Whole genome PCR amplification was performed using JCV-specific primers overlapping the BamHI site: BAM-1 (5'-GGGATCCTGTGTTTCATCATCACTGGC-3') and BAM-2 (5'- AGGATCCAACTCTACCCCCACC-3'). Hot Start PCR was used with the following cycling conditions: denaturation at 94°C for 1 minute, followed by 40 cycles of 94°C for 40 seconds, 64°C for 1 minute, and 72°C for 6 minutes, with a final extension at 72°C for 10 minutes. The resulting PCR product was diluted 1:200 and used for subsequent amplification of VP1 and NCCR.

The VP1 gene (~400bp) was amplified using in-house primers, forward primer 5'-CCCAATCTAAATGAGG ATCTAACCTGT-3' (nt 1733-1759); reverse primer 5'-TTTCTCCTCCTGTTAGTGTCCC-3' (nt 2132-2110) in a PCR reaction with 50 cycles. The NCCR (~350-550bp) was amplified via nested PCR approach using outer primers A1: 5'-TCCATGGATTCCCTCCCTATTCAAGCACTTTGT-3' (nt 4979-5009), A2: 5'-TTACTTACCTAT GTAGCTTT-3' (nt 500-481) and inner primers B1: 5'-GCAAAAAGGGAAAAACAAAGGG-3' (nt 5041-5062) and B3: 5'-CAGAAGCCTTACGTGACAGCTGG-3' (nt 310-288)⁴⁶. The PCR master mix had 1X PCR buffer, forward and reverse primers at 10pmol concentration. The first round of PCR for NCCR had 50 cycles, while the second round had 40 cycles. The cycling conditions for both VP1 and NCCR were as follows, denaturation at 95°C for 1 minute, followed by amplification at 95°C for 30 seconds, 60°C for 45 seconds, and 72°C for 1 minute, with a final extension at 72°C for 5 minutes. The presence of PCR products was confirmed by agarose gel electrophoresis. The PCR products were purified by gel extraction method and Sanger's sequencing (di-deoxy nucleotide termination-based method) was performed to determine the nucleotide sequences of the VP1 and NCCR (Sakhala Enterprises, Bangalore).

Analysis of VP1 region and genotyping

The VP1 sequences obtained from the CSF samples were aligned with the representative sequences from the known JCV genotypes (1-8) and subtypes available in the GenBank database using MAFFT (v7.526). The aligned sequences were then used to construct the phylogenetic tree using iqtree2 (v2.0.7) with the best fit model according to Bayesian information criterion (BIC) with a bootstrap of 1000 replicates⁴⁷⁻⁴⁹ and the resulting tree was visualized using Figtree (v1.4.4).

Analysis of NCCR

For the analysis of the NCCR, the sequences obtained were mapped against the archetype NCCR reference sequence (AB038249, CY) using Bio-edit software (v7.7). The origin (ORI) and agnoprotein regions were identified at either end of the sequence, while the ~300-500bp region in between was manually inspected and divided into segments labelled A-F sequentially, based on the comparison with the archetype sequence. The NCCR sequenced from the study samples were analysed for patterns of duplications, deletions and insertion in the regions from A-F (Figure 2). NCCR was also classified further as described previously²⁶. NCCR were classified into four groups which include types - IS, IR, IIS and IIR based on the pattern of insertions and repeats. Type I NCCRs do not contain B and D block whereas Type II NCCRs have an insertion that includes at least a part of the sequence from either sections B or D. The "S" (singular) subtypes do not have a repeat, while the "R" (repeat) subtypes do. Type IIS NCCRs are also referred to as archetype or archetype-like when deletions are present.

Analysis of the host transcription factor binding sites in NCCR

Transcriptional binding sites in the sequences were analysed using TFbind tool (<https://tfbind.hgc.jp>)⁵⁰. TFbind employs motifs from the TRANSFAC database R3.2. The location of binding sites against each block were analysed manually and compared with the archetype sequence to determine the frequency of TBs (transcriptional binding sites).

Data availability

The accession numbers for the sequences generated in this study can be found in Supplementary file S2. Accession numbers for the sequences in this study in GenBank are as follows: VP1 region-PV685521, PV691829-PV691850; NCCR-PV685520, PV730286-PV730289, PV740670-PV740681.

Received: 18 June 2025; Accepted: 31 October 2025

Published online: 01 December 2025

References

- Cortese, I., Reich, D. S. & Nath, A. Progressive multifocal leukoencephalopathy and the spectrum of JC virus-related disease. *Nat. Rev. Neurol.* **17**, 37–51 (2021).
- Sarıbaş, A. S., Özdemir, A., Lam, C. & Safak, M. JC virus-induced progressive multifocal leukoencephalopathy. *Futur. Virol.* **5**, 313–323 (2010).
- White, M. K., Safak, M. & Khalili, K. Regulation of gene expression in primate polyomaviruses. *J. Virol.* **83**, 10846–10856 (2009).
- Ferenczy, M. W. et al. Molecular biology, epidemiology, and pathogenesis of progressive multifocal leukoencephalopathy, the JC virus-induced demyelinating disease of the human brain. *Clin. Microbiol. Rev.* **25**, 471–506 (2012).
- Yogo, Y. et al. Isolation of a possible archetypal JC virus DNA sequence from nonimmunocompromised individuals. *J. Virol.* **64**, 3139–3143 (1990).
- Bellizzi, A. et al. New insights on human polyomavirus JC and pathogenesis of progressive multifocal leukoencephalopathy. *Clin. Dev. Immunol.* **2013**, 839719 (2013).
- S, S. et al. Progressive multifocal leukoencephalopathy-epidemiology, immune response, clinical differences treatment. *Epidemiol. Mikrobiol. Immunol.* **68**, 24–31 (2019).
- Elsner, C. & Dörries, K. Human polyomavirus JC central region variants in persistently infected CNS and kidney tissue. *J. Gen. Virol.* **79**, 789–799 (1998).
- Ault, G. S. & Stoner, G. L. Human polyomavirus JC promoter/enhancer rearrangement patterns from progressive multifocal leukoencephalopathy brain are unique derivatives of a single archetypal structure. *J. Gen. Virol.* **74**, 1499–1507 (1993).
- Yogo, Y. et al. JC Virus regulatory region rearrangements in the brain of a long surviving patient with progressive multifocal leukoencephalopathy. *Neurol. Neurosurg. Psychiatr.* **71**(3), 397–400 (2001).
- Sharma, S. K. et al. Progressive multifocal leucoencephalopathy in HIV/AIDS: Observational study from a tertiary care centre in Northern India. *Ind. J. Med. Res.* **138**(1), 72–77 (2013).
- Shah, V., Toshniwal, H., & Shevkani, M. Clinical profile and outcome of progressive multifocal leukoencephalopathy in HIV infected Indian patients - PubMed. *J. Assoc. Phys. India.* **65**(3) 40–44 (2017).
- Berger, J. R. & Major, E. O. Progressive multifocal leukoencephalopathy. *Semin. Neurol.* **19**(2), 193–200 (1999).
- Shankar, S. K. et al. Low prevalence of progressive multifocal leukoencephalopathy in India and Africa: Is there a biological explanation?. *J. NeuroVirol.* **9**, 59–67 (2003).
- India HIV Estimates 2021 _Fact Sheets _Final_Shared_24_08_2022_0 (www.aidsdatahub.org) (2021).
- Padgett, B. L., Zurhein, G. M., Walker, D. L., Eckroade, R. J. & Dessel, B. H. Cultivation of papova-like virus from human brain with progressive multifocal leucoencephalopathy. *Lancet* **297**, 1257–1260 (1971).
- Agostini, H. T. et al. Genotypes of JC virus in East, Central and Southwest Europe. *J. Gen. Virol.* **82**, 1221–1331 (2001).
- Padgett, B. L. & Walker, D. L. Prevalence of antibodies in human sera against JC Virus, an isolate from a case of progressive multifocal leukoencephalopathy. *J. Infect. Dis.* **127**(4), 467–470 (1973).
- Major, E. O. Progressive multifocal leukoencephalopathy in patients on immunomodulatory therapies. *Annu. Rev. Med.* **61**, 35–47 (2010).
- Choudhary, S., Parashar, M., Parashar, N. & Ratre, S. AIDS-related progressive multifocal leukoencephalopathy-really rare in India: A case report and review of literature. *Ind. J. Sex. Transm. Dis and AIDS* **39**, 55–58 (2018).
- Shackelton, L. A., Rambaut, A., Pybus, O. G. & Holmes, E. C. JC virus evolution and its association with human populations. *J. Virol.* **80**, 9928–9933 (2006).
- Dubois, V. et al. JC virus genotypes in France: Molecular epidemiology and potential significance for progressive multifocal leukoencephalopathy. *J. Infect. Dis.* **183**(2), 213–217 (2001).
- Kannangai, R. et al. Association of neurotropic viruses in HIV-infected individuals who died of secondary complications of tuberculosis, cryptococcosis, or toxoplasmosis in south India. *J. Clin. Microbiol.* **51**, 1022–1025 (2013).
- Rgen, H. et al. JC virus (JCV) Genotypes in brain tissue from patients with progressive multifocal leukoencephalopathy (PML) and in urine from controls without PML: Increased frequency of JCV Type 2 in PML. *J. Infect. Dis.* **176**(1), 1–8 (1997).
- L'Honneur, A. S. et al. Exploring the role of nCCR variation on jc polyomavirus expression from dual reporter minicircles. *PLoS ONE* **13**, e0199171 (2018).
- Jensen, P. N. & Major, E. O. Basic science and immunobiology report a classification scheme for human polyomavirus JCV variants based on the nucleotide sequence of the noncoding regulatory region. *J. Neurovirol.* **7**, 280–287 (2001).
- Ferrante, P. et al. Analysis of JC virus genotype distribution and transcriptional control region rearrangements in human immunodeficiency virus-positive progressive multifocal leukoencephalopathy patients with and without highly active antiretroviral treatment. *J. Neurovirol.* **9**, 42–46 (2003).
- Ciardi, M. R. et al. JCPyV NCCR analysis in PML patients with different risk factors: Exploring common rearrangements as essential changes for neuropathogenesis. *Virol. J.* **17**, 23 (2020).
- Amemiya, K., Traub, R., Durham, L. & Major, E. O. Interaction of a nuclear factor-1-like protein with the regulatory region of the human polyomavirus JC virus. *J. Biol. Chem.* **264**(12), 7025–7032 (1989).
- Tamura, T.-A., Inoue, T., Nagata, K. & Mikoshiba, K. Enhancer of human polyoma JC virus contains nuclear factor I- binding sequences; analysis using mouse brain nuclear extracts. *Biochem. Biophys. Res. Commun.* **157**(2), 419–425 (1988).
- Sock, E., Wegner, M. & Grummt, F. DNA replication of human polyomavirus JC Is stimulated by NF-I in Vivo. *Virology* **182**, 298–308 (1991).
- Henson, J., Saffer, J. & Furneaux, H. The transcription factor Sp1 binds to the JC virus promoter and is selectively expressed in glial cells in human brain. *Ann. Neurol.* **32**, 72–77 (1992).
- Henson, J. W. Regulation of the glial-specific JC virus early promoter by the transcription factor Sp1. *J. Biol. Chem.* **269**, 1046–1050 (1994).
- Kim, H.-S., Goncalves, N. M. & Henson, J. W. Glial cell-specific regulation of the JC virus early promoter by large T antigen. *J. Virol.* **74**(2), 755–763 (2000).
- Mischitelli, M. et al. Investigation on the role of cell transcriptional factor Sp1 and HIV-1 TAT protein in PML onset or development. *J. Cell. Physiol.* **204**, 913–918 (2005).
- Kerr, D. & Khalili, K. A recombinant cDNA derived from human brain encodes a DNA binding protein that stimulates transcription of the human neurotropic virus JCV. *J. Biol. Chem.* **266**, 15876–15881 (1991).
- Peter, R., Michel, U., Kehrl, & H. J. Regulation of JC virus expression in B lymphocytes. *J. Virol.* **68**(1), 217–222 (1994).

38. Chowdhury, M., Taylor, J. P., Chang, C.-F., Rappaport, J. & Khalilil, K. Evidence that a sequence similar to TAR is important for induction of the JC virus late promoter by human immunodeficiency virus type 1 Tat. *J Virol.* **66**(12), 7355–7361 (1992).
39. Wegner, M., Drolet, D. W. & Rosenfeld, M. G. Regulation of JC virus by the POU-domain transcription factor Tst-1: Implications for progressive multifocal leukoencephalopathy. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 4743–4747 (1993).
40. Sadowska, B., Barrucco, R., Khalili, K. & Safak, M. Regulation of human polyomavirus JC virus gene transcription by AP-1 in glial cells. *J. Virol.* **77**, 665–672 (2003).
41. Romagnoli, L. et al. Early growth response-1 protein is induced by JC virus infection and binds and regulates the JC virus promoter. *Virology* **375**, 331–341 (2008).
42. Romagnoli, L. et al. Modulation of JC virus transcription by C/EBP β . *Virus Res.* **146**, 97–106 (2009).
43. Honkima, A. et al. Exploring JC polyomavirus sequences and human gene expression in brain tissue of patients with progressive multifocal leukoencephalopathy. *J. Infect. Dis.* **230**(3), e732–e736 (2024).
44. Kim, H.-S. & Woo, M.-S. Transcriptional regulation of the glial cell-specific JC virus by P53. *Arch. Pharm. Res.* **25**, 208–213 (2002).
45. Agostini, H. T. & Stoner, G. L. Amplification of the complete polyomavirus JC genome from brain, cerebrospinal fluid and urine using Pre-PCR restriction enzyme digestion. *J. Neuropathol. Exp. Neurol.* **54**, 316–320 (1995).
46. Sugimoto, C. et al. Amplification of JC virus regulatory DNA sequences from cerebrospinal fluid: Diagnostic value for progressive multifocal leukoencephalopathy. *Arch. Virol.* **143**, 249–262 (1998).
47. Minh, B. Q. et al. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
48. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Haeseler, A. V. & Jermiin, L. S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Method.* **14**, 587–589 (2017).
49. Hoang, D. T., Chernomor, O., Haeseler, A. V., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).
50. Tsunoda, T. & Takagi, T. Estimating transcription factor bindability on DNA. *Bioinformatics* **15**(7–8), 622–630 (1999).

Author contributions

RSM and GS conceived and designed the study. GS, ST, and DK performed all the experiments. GS and ST performed the Data analysis and interpretation. GS, VR and MAA contributed to drafting the initial manuscript. RSM supervised overall work, reviewed and finalised the manuscript. All authors have read and approved the manuscript.

Funding

This study was supported by funds from the Department of Health Research-Indian Council of Medical Research (DHR-ICMR) to the Viral Research and Diagnostic Laboratory (VRDL) at NIMHANS (Dr. Reeta S Mani, Principal Investigator). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-26996-w>.

Correspondence and requests for materials should be addressed to R.S.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025