Article

Database and Statistical Analyses of Transcription Factor Binding Sites in the Non-Coding Control Region of JC Virus

Kazuo Nakamichi [1],* and Toshio Shimokawa [2]

[1] Department of Virology 1, National Institute of Infectious Diseases, Tokyo 162-8640, Japan
[2] Department of Medical Data Science, Graduate School of Medicine, Wakayama Medical University, Wakayama 641-8509, Japan; shimokaw@wakayama-med.ac.jp
*   Correspondence: nakamich@niid.go.jp

Abstract: JC virus (JCV), as an archetype, establishes a lifelong latent or persistent infection in many healthy individuals. In immunocompromised patients, prototype JCV with variable mutations in the non-coding control region (NCCR) causes progressive multifocal leukoencephalopathy (PML), a severe demyelinating disease. This study was conducted to create a database of NCCR sequences annotated with transcription factor binding sites (TFBSs) and statistically analyze the mutational pattern of the JCV NCCR. JCV NCCRs were extracted from >1000 sequences registered in GenBank, and TFBSs within each NCCR were identified by computer simulation, followed by examination of their prevalence, multiplicity, and location by statistical analyses. In the NCCRs of the prototype JCV, the limited types of TFBSs, which are mainly present in regions D through F of archetype JCV, were significantly reduced. By contrast, modeling count data revealed that several TFBSs located in regions C and E tended to overlap in the prototype NCCRs. Based on data from the BioGPS database, genes encoding transcription factors that bind to these TFBSs were expressed not only in the brain but also in the peripheral sites. The database and NCCR patterns obtained in this study could be a suitable platform for analyzing JCV mutations and pathogenicity.

Keywords: database; JC virus; non-coding control region; mutational pattern; transcription factor binding sites; statistical analysis

1. Introduction

The JC virus (JCV) belongs to the family Polyomaviridae and has a circular double-stranded DNA genome. JCV is widely distributed in the human population, with 60% to 80% of adults serologically positive for this virus [1–5]. The first infection with JCV occurs mainly in childhood, after which the virus establishes a persistent asymptomatic infection in the kidney and urinary tract [6,7]. Additionally, JCV persists or is latent in other sites, such as lymphoid tissue and bone marrow [1,8–11]. This particular JCV presents a stable genomic structure and sequence and is referred to as the archetype [12,13]. Archetype JCV is detected in the urine of 7% to 46% of immunocompetent individuals [4,14–27].

In immunocompromised patients or those treated with agents that affect cellular immunity, JCV can reactivate and cause progressive multifocal leukoencephalopathy (PML) due to lytic infection of oligodendrocytes and fatal demyelination in the brain [28–32]. JCV isolates from the brain and cerebrospinal fluid (CSF) of PML patients shows hypervariable mutations in the non-coding control region (NCCR; also referred to as the regulatory or transcription control regions) of the viral genome, and these variants are termed prototypes [8,28,33–36]. The nucleotide sequence of the NCCR includes transcription factor binding sites (TFBSs) and is responsible for the expression of viral early and late genes [8,28,37,38]. Rearrangement in the NCCRs of prototype JCV is thought to be generated by deletions and/or duplications in the archetypal sequences [12,39], which alter promoter activity [40,41].

In previous studies, we observed patient-specific mutation patterns in the NCCR sequences of prototype JCV derived from CSF specimens of PML cases [42,43] and similar to those reported in other investigations [10,34,36,40,44–55]. Furthermore, the mutation pattern of the most frequently detected JCV NCCRs in each patient changed during disease progression [42]. However, because of the limited number of JCV isolates obtained by a single research group and the highly complex variations, it was not feasible to distinguish apparent features in the NCCRs of prototype JCV by schematically comparing the sequences, except for the deletion occurring mainly in region D [42,43]. The nucleotide sequences of JCV NCCRs have been determined mainly in anthropological or epidemiological studies of archetype viruses and analyses of the pathogenic mechanism of prototype viruses by many researchers, including the authors [7,10,12,13,34–40,42–55]. To better understand the diversity and regularity of NCCR rearrangement, a platform for compiling these sequences is necessary.

Additionally, to analyze the NCCRs of polyomaviruses, including JCV, several research groups have used computer simulations to identify the types and locations of TFBSs [56–61]. Using this technique might enable the comprehensive estimation of TFBSs present in the NCCR sequences of a large variety of JCVs and analysis of these data to reveal mutation patterns. In this study, we created an annotation database of the NCCR sequences of JCV isolates registered in GenBank and examined the characteristics of TFBSs using statistical methods.

2. Materials and Methods

2.1. Acquisition of Sequence Data for JCV NCCRs

The overall workflow of this study is summarized in Figure 1. A total of 2337 sequences of JCV DNA and their metadata were downloaded from the nucleotide database of the National Center for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov/, accessed date: 1 October 2013) using search terms "JC", "polyomavirus", and "region." To avoid any error or bias caused by mixing data from Sanger sequencing and next-generation sequencing, DNA sequences from JCV registered over the course of 23 years (1990–2013) were used for the analysis. During this registration period, as far as we could ascertain, the DNA sequence of JCV was determined by Sanger sequencing. After importing the sequence data into the CLC Genomics Workbench software program version 7.0 (Qiagen, Aarhus, Denmark), alignments were performed using the genomic DNA sequence of the representative archetype JCV (CY strain; GenBank: AB038249.1) as a reference. The sequences were roughly aligned using 20 sequences each, and the alignment was repeated by manually checking each sequence. After this process, 1024 sequences of JCV NCCRs were obtained (Supplementary Dataset S1).
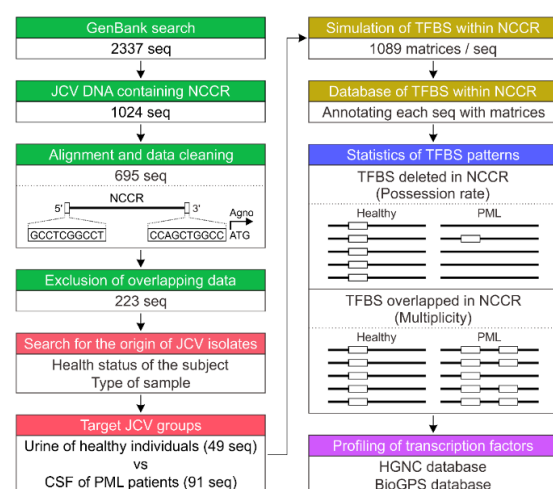

Figure 1. Overall workflow of the data processing and analysis of transcription factor binding sites (TFBSs) in the non-coding control region (NCCR) of JC virus (JCV). The nucleotide sequences (seq)

of JCV NCCRs in GenBank were extracted and aligned, and their origins were confirmed. TFBSs in the NCCRs of JCV isolates from the urine of healthy individuals and cerebrospinal fluid (CSF) of PML patients were identified using computer simulation. A database was created based on the NCCR sequences and TFBS annotations, and the patterns and metadata of TFBSs were examined using statistical analysis and public databases.

2.2. Data Cleaning and Extraction of JCV NCCR Sequences

The collected NCCR data were in a miscellaneous state and included partial fragments, sequences where some nucleotides had not been sequenced, or many identical sequences of NCCR clones derived from the same individuals. To obtain results with a higher degree of reliability for the computer simulation and statistical analysis of TFBSs, we cleaned the NCCR sequence data. Sequence data were extracted for a group of NCCR sequences with 5′ and 3′ nucleotide positions (1–10 and 258–267, respectively) located at both ends in the NCCR (regions A–F) of archetype JCV (CY strain) (Figure 1). Additionally, NCCR sequences containing uncertain characters other than ATGC were excluded, resulting in 695 sequences. The duplicated identical NCCR sequences were then removed by alignments, leaving 223 sequences. Review of the metadata and published literature registered in GenBank enabled examination of the types of specimens and health conditions of individuals in which each JCV sequence was detected. To accurately compare the TFBS patterns in the archetypal and prototype NCCRs, two groups of JCV NCCRs derived from the urine of healthy individuals and CSF of PML patients (49 and 91 sequences, respectively) were subjected to the following analyses as target sequences.

2.3. Computer Simulation of TFBSs in NCCR Sequences

TFBSs within the NCCR sequences were analyzed by computer simulation using MatInspector software (Genomatix, Munich, Germany) with Matrix Family Library version 9.2 (Genomatix). This program identifies TFBSs in nucleotide sequences using a large library of weight matrices, annotates the corresponding sites with the matrices, and presents simple metadata of TFBSs [62,63]. The matrix library used in this study included 1072 and 17 weight matrices for vertebrate TFBSs and general core promoter elements, respectively. The search parameters were configured using software defaults. The sequences processed by MatInspector were imported into the CLC Genomics Workbench as FASTA format files. Because these numerous annotations were difficult to handle owing to a combination of matrix names and their family types, each of them was manually corrected to the name of the individual matrix.

2.4. Statistical Analyses of TFBS Patterns

The proportions of NCCRs of JCV isolates that possessed the respective TFBSs (referred to here as "possession rates") between groups were statistically compared using Fisher's exact test, and the Benjamini–Hochberg method was used to adjust for multiple comparisons. Statistical significance was considered at a false discovery rate (FDR)-adjusted Q-value < 0.05 [64]. The multiplicity of each TFBS matrix in the NCCR of JCV isolates from healthy individuals and PML patients was analyzed using Poisson regression analysis. All analyses were conducted using R version 3.6.1 (R Foundation for Statistical Computing, Vienna, Austria). Statistical significance was set at $P < 0.05$.

2.5. The Gene Ontology and Expression Profiles of Transcription Factors

Transcription factors that bind to each TFBS are briefly indicated in the metadata of TFBSs presented by MatInspector. Based on this information, the ontology of the genes encoding transcription factors was searched using the Human Genome Organization Gene Nomenclature Committee (HGNC) database (https://www.genenames.org/, accessed date: 18 June 2021), and the HGNC identifiers (IDs) of these genes and their currently approved symbols and names were confirmed. The gene-expression profiles of transcription factors in human tissues or cells were retrieved by accessing the BioGPS database (http://biogps.org/, accessed date: 18 June 2021) [65] via the symbol reports of the HGNC database. The BioGPS plug-in

"Gene expression/activity chart" and Affymetrix microarray dataset "GeneAtlas U133A, and gcrma" [66] were used in these data searches.

3. Results

3.1. Creation of the TFBS Database for JCV NCCR Sequences

For the computational and statistical analyses of TFBSs according to the NCCR sequences of JCV isolates, the nucleotide sequences deposited in GenBank over a 20-year period were aligned and selected. Of the downloaded data, only ~10% (234 of 2337 sequences) comprised the full-length JCV genome, with the remainder representing partial fragments. Because a substantial number of these entries did not contain NCCRs and could not be automatically extracted as target sequences using the software, we extracted NCCRs manually by repeating the small-scale alignment using the archetype JCV genome as a reference. To simulate TFBS more accurately in the NCCRs of various JCV isolates, the target sequences were narrowed down using 10 bases at both ends of the CY strains as landmark sequences. For the 695 NCCRs that remained after data cleaning, duplicates of precisely identical sequences were removed, and ~32% (223 sequences) remained. Most of these identical sequences were deposited during the massive sequence analysis conducted by Reid et al. [46]. Examination of the origin of the 223 extracted sequences revealed that most NCCRs were detected in the urine of healthy individuals and the CSF of PML patients (Supplementary Dataset S2). Therefore, TFBSs in the NCCRs of both groups were identified using MatInspector, and a database of TFBSs within JCV NCCR sequences was created. This database comprises a set of nucleotide sequences of each JCV isolate along with a massive number of TFBS matrices added as annotations. Moreover, the database allows the positions and sequences of any TFBS to be visualized and tabulated using standard genetic analysis software. For example, Figure 2 shows the position of the TFBSs on the NCCR of a well-known archetype JCV (CY strain). In the NCCR of the CY strain, 54 and 43 TFBS matrices were detected in the 5' and 3' nucleotide positions (1–267; forward) and its complementary strand (reverse), respectively. TFBS matrices were especially visible in regions A and regions D through F, but they were also found in other regions.
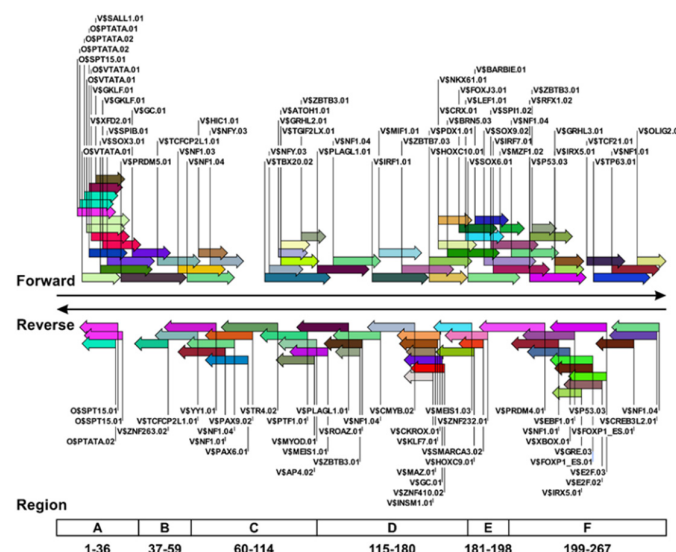

Figure 2. Representative example showing the type and location of TFBSs within the JCV NCCR. TFBS matrices in the NCCR sequences of 140 JCV isolates were identified using MatInspector. The data were imported into the CLC Genomics Workbench as FASTA format files, and the types and locations of TFBS matrices in each JCV isolate were annotated to NCCR sequences. Data represent all TFBS matrices identified only for the CY strain, a representative archetype JCV. The TFBS matrices (right, labeled "Forward") were identified in the 5' and 3' nucleotide positions (1–267), whereas those to the left (labeled "Reverse") were present in the complementary strand. The horizontal boxes (bottom) indicate the location and base number of regions A through F within the NCCR of the CY strain of JCV.

3.2. Overall View of TFBS Patterns in JCV NCCR

The patterns of TFBSs in the JCV NCCRs were observed by exporting the data from the created database. The NCCRs of 49 JCV isolates derived from the urine of healthy individuals were archetypal sequences, including those with very small insertions or deletions that could be regarded as genetic polymorphisms. The average numbers of TFBS matrices in these JCV isolates were 52.6 and 40.0 for the forward and reverse strands, respectively. The NCCRs of 91 JCV isolates detected in the CSF of PML patients showed prototypal sequences with averages of 57.1 and 38.0 TFBS matrices in the forward and reverse strands, respectively. There was no statistically significant difference in the mean total number of TFBS matrices within either strand of JCV NCCRs from the urine of healthy subjects and CSF of PML patients. These results indicated that the NCCR of JCV produces complex reconstructions in patients with PML, but that the total number of TFBSs within each NCCR is not significantly altered. Moreover, 464 types of TFBS matrices were detected in all NCCR sequences, among which there were sequences of shallow frequency and suspected to be nonspecific or biologically insignificant. Therefore, the TFBS matrices that were highly conserved in the archetypes were sorted and statistically examined to determine how they changed in the prototypes. Histogram analysis of the prevalence of each TFBS matrix in the NCCR of archetype JCV isolates showed that >95% of them had 34 and 18 matrices in the forward and reverse strands, respectively (Supplementary Figure S1). When the database was used to select and visualize each TFBS matrix in the NCCR alignments of the JCV isolates, we noticed that some matrices tended to be missing or overlapped in the prototype JCVs (Figure 1). Consequently, the following statistical analyses were conducted to analyze the patterns of TFBSs.

3.3. TFBSs Frequently Lost in the NCCR Sequences of Prototype JCV

Table 1 shows the TFBS matrices that showed significantly different possession rates between archetype JCVs from the urine of healthy individuals and prototype viruses from the CSF of PML patients. The NCCR sequences of prototype JCVs derived from the CSF of PML patients showed low percentages of the possession of 13 and nine matrices in the forward and reverse strands, respectively, and 15 of 22 TFBS matrices on both strands were absent in >50% of the prototypal NCCRs. Visualization of the location of these TFBSs in the NCCR of archetype JCV (CY strain) using the created annotation database revealed that the TFBS group, the possession rate of which decreased in the prototype JCVs, was mainly located in regions D through F of the NCCR of the archetype virus (Figure 3). Although the TFBS matrices identified by MatInspector are accompanied by metadata, such as the names and expression patterns of transcription factors that bind to the TFBSs, we found that this information included past designations and obscure expression sites. Therefore, the gene ontology and expression sites of transcription factors capable of binding to TFBS matrices were confirmed using the HGNC and BioGPS databases. Based on data retrieval from BioGPS using the Affymetrix microarray dataset, transcription factors that bind to TFBSs often lost in the NCCR sequences of the prototype JCVs were suggested to be expressed in various human tissues (Table 2). Additionally, some of these transcription factors were suggested to be highly expressed in sites of persistent or latent JCV infection, such as the kidney, bone marrow (CD34+ cells), and lymph nodes (Table 2). These data indicated that TFBSs, which are often lost in the NCCRs of prototype JCVs, are mainly located in regions D through F, and that the transcription factors that bind to them are expressed at a variety of peripheral sites.

Table 1. Possession rates of TFBSs within the NCCRs of JCVs in the urine of healthy individuals and CSF of PML patients.

Matrix Name	DNA Strand [a]	Urine of Healthy Individuals (n = 49)		CSF of PML Patients (n = 91)		*P*-value [d]
		JCV Isolates with Matrix [b]	Possession Rate (%) [c]	JCV Isolates with Matrix	Possession Rate (%)	
V$MIF1.01	FWD	47	95.9	19	20.9	<0.001
V$IRF1.01	FWD	47	95.9	20	22.0	<0.001
V$HOXC10.01	FWD	48	98.0	27	29.7	<0.001
V$PLAGL1.01	FWD	48	98.0	30	33.0	<0.001
V$PDX1.01	FWD	48	98.0	33	36.3	<0.001
V$CRX.01	FWD	48	98.0	34	37.4	<0.001
V$NKX61.01	FWD	48	98.0	34	37.4	<0.001
V$BRN5.03	FWD	48	98.0	36	39.6	<0.001
V$ZBTB3.01	FWD	47	95.9	59	64.8	<0.001
V$FOXJ3.01	FWD	48	98.0	62	68.1	<0.001
V$SPI1.02	FWD	49	100	80	87.9	0.008
V$IRF7.01	FWD	49	100	81	89.0	0.015
V$MZF1.02	FWD	49	100	81	89.0	0.015
V$CMYB.02	REV	47	95.9	23	25.3	<0.001
V$ROAZ.01	REV	48	98.0	28	30.8	<0.001
V$PLAGL1.01	REV	48	98.0	30	33.0	<0.001
V$ZNF232.01	REV	48	98.0	32	35.2	<0.001
V$ZBTB3.01	REV	47	95.9	32	35.2	<0.001
V$HOXC9.01	REV	48	98.0	33	36.3	<0.001
V$MEIS1.03	REV	48	98.0	33	36.3	<0.001
V$SMARCA3.02	REV	48	98.0	55	60.4	<0.001
V$PRDM4.01	REV	49	100	80	87.9	0.008

Abbreviations: CSF, cerebrospinal fluid; FWD, forward; JCV, JC virus; NCCR, non-coding control region; PML, progressive multifocal leukoencephalopathy; REV, reverse; TFBS, transcription factor binding site. [a] The direction of the DNA strand is mentioned in the Figure 2 legend. [b] The number of JCV isolates with the respective matrices for TFBSs in the NCCR. [c] The proportion of JCV isolates that possessed the respective matrices in each group. [d] The possession rates of 52 matrices between the two groups were analyzed using Fisher's exact test and the Benjamini–Hochberg method. Matrices with statistically significant differences are shown.
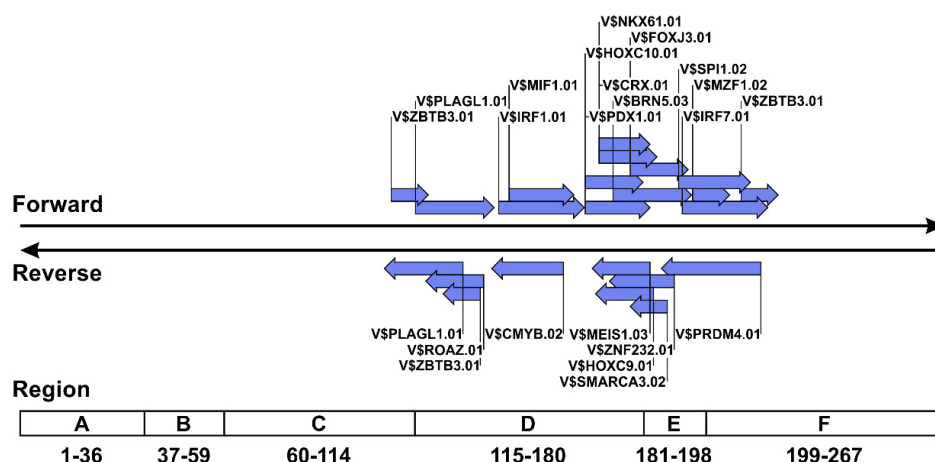

Figure 3. Types and locations of TFBSs lost in rearranged NCCR sequences of prototype JCV. TFBS matrices showing statistically significant decreases in possession rate in the NCCRs of prototype JCV isolates and shown in the sequence of the archetype virus (CY strain). The method of illustrating the TFBS matrices is the same as that shown in Figure 1.

Table 2. Profiles of TFBSs that are reduced in the NCCRs of JCVs in the CSF of PML patients.

| Matrix Name | DNA Strand [b] | Transcription Factor [a] | | | |
		HGNC ID	Symbol	Full Name	Gene Expression (>3 × Median) [c]
V$MIF1.01 [d]	FWD	4921/9982	HIVEP2/ RFX1	HIVEP zinc finger 2/ regulatory factor X1	Brain (cerebrum), Blood (T cells)/ NA (Ubiquitous)
V$IRF1.01	FWD	6116	IRF1	Interferon regulatory factor 1	Blood (Leukocytes), Bone marrow (CD34+ cells), Colon, Heart, Lung, Lymph node, Placenta, Small intestine, Thymus
V$HOXC10.01	FWD	5122	HOXC10	Homeobox C10	Kidney
V$PLAGL1.01	FWD REV	9046	PLAGL1	PLAG1-like zinc finger 1	Adrenal gland, Bone marrow (CD34+ cells), Colon, Pituitary gland, Placenta, Prostate, Retina, Small intestine, Smooth muscle, Uterus
V$PDX1.01	FWD	6107	PDX1	Pancreatic and duodenal homeobox 1	NA (Ubiquitous)
V$CRX.01	FWD	2383	CRX	Cone-rod homeobox	Pineal gland, Retina
V$NKX61.01	FWD	7839	NKX6-1	NK6 homeobox 1	NA (Ubiquitous)
V$BRN5.03	FWD	9224	POU6F1	POU class 6 homeobox 1	NA (Ubiquitous)
V$ZBTB3.01	FWD REV	22918	ZBTB3	Zinc finger and BTB domain containing 3	NA (Ubiquitous)
V$FOXJ3.01	FWD	29178	FOXJ3	Forkhead box J3	NA (Ubiquitous)
V$SPI1.02	FWD	11241	SPI1	Spi-1 proto-oncogene	Blood (Monocytes), Lung
V$IRF7.01	FWD	6122	IRF7	Interferon regulatory factor 7	Blood (Leukocytes), Bone marrow (CD34+ cells), Heart, Lung, Lymph node, Thymus, Tonsil
V$MZF1.02	FWD	13108	MZF1	Myeloid zinc finger 1	Blood (leukocytes), Blood vessel (endothelial cells), Bone marrow (CD34+ cells), Pineal gland, Prostate, Thyroid gland
V$CMYB.02	REV	7545	MYB	MYB proto-oncogene, transcription factor	Blood vessel (endothelial cells), Bone marrow (CD34+ cells), Thymus
V$ROAZ.01	REV	16762	ZNF423	Zinc finger protein 423	Brain (whole), Pineal gland, Retina, Small intestine, Uterus
V$ZNF232.01	REV	13026	ZNF232	Zinc finger protein 232	NA (Ubiquitous)
V$HOXC9.01	REV	5130	HOXC9	Homeobox C9	NA (Ubiquitous)
V$MEIS1.03	REV	7000	MEIS1	Meis homeobox 1	Adrenal gland, Bone marrow (CD34+ cells), Brain (cerebellum), Colon, Ovary, Salivary gland, Small intestine, Smooth muscle, Trachea, Uterus
V$SMARCA3.02	REV	11099	HLTF	Helicase like transcription factor	Blood (T cells and NK cells), Blood vessel (endothelial cells), Bone marrow (CD34+ cells), Pineal gland, Pituitary gland, Thyroid gland
V$PRDM4.01	REV	9348	PRDM4	PR/SET domain 4	Blood (B cells), Bone marrow (CD34+ cells), Pineal gland

Abbreviations: CD, cluster of differentiation; FWD, forward; HGNC, Human Genome Organization Gene Nomenclature Committee; ID, identification; JCV, JC virus; NA, not applicable; NCCR, non-coding control region; PML, progressive multifocal leukoencephalopathy; REV, reverse; TFBS, transcription factor binding site. [a] The gene ontology of transcription factors predicted to bind each sequence was confirmed using the HGNC database in accordance with the metadata of the matrices. [b] The direction of the DNA strands is mentioned in the Figure 2 legend. [c] Gene-expression profiles of transcription factors in human tissues and blood were obtained using BioGPS microarray data, and the sites with 3-fold higher expression levels relative to the median are indicated. [d] This matrix is defined as the sequence targeted by the HIVEP2–RFX1 complex.

3.4. TFBSs Likely to Multiply in the NCCR Sequences of Prototype JCVs

We performed a final set of analyses to examine TFBSs that tend to overlap in the rearranged NCCRs of prototype JCVs using the Poisson distribution for modeling count data. Only seven of 52 target TFBS matrices showed a significantly higher multiplicity

within the NCCRs of prototype JCVs from PML patients as compared with those of archetype JCVs from healthy individuals (Table 3). In the archetype JCVs from the urine of healthy individuals, each of these TFBS matrices were present in the NCCRs, except for matrix V$NFY.03, which was presented in duplicate. However, prototype JCVs detected in the CSF of PML patients had a higher number of TFBSs in the NCCRs than observed in archetype viruses. When TFBS matrices, often repeated in the NCCRs of prototype JCVs, were depicted in the NCCR of archetype JCV, these matrices were mainly distributed in regions C and E (Figure 4). Table 4 shows the gene ontology and expression profiles of transcription factors capable of binding to TFBS matrices with increased multiplicity in the rearranged NCCR of prototype JCVs. Genes encoding these transcription factors, except for lymphoid enhancer binding factor 1 and nuclear factor I (NFI) C, are expressed in various human tissues, including the brain. The gene-expression profile of SRY-box transcription factor 6 was not included in the Affymetrix microarray dataset used in this study but is reportedly ubiquitously expressed [67]. These data suggested that the limited number of TFBSs in regions C and E tend to overlap during the rearrangement of the NCCR sequences of prototype JCVs from a statistical standpoint, and that transcription factors capable of binding to these TFBSs are mostly ubiquitously expressed.

Table 3. Multiplicity of TFBSs within the NCCRs of JCVs in the urine of healthy individuals and CSF of PML patients.

Matrix Name	DNA Strand [b]	Poisson Mean (95% CI) [a]				*P*-value [c]
		Urine of Healthy Individuals (*n* = 49)		CSF of PML Patients (*n* = 91)		
V$HIC1.01	FWD	1.04	[0.75, 1.33]	1.89	[1.59, 2.18]	<0.001
V$NF1.03	FWD	1.06	[0.77, 1.35]	1.93	[1.63, 2.23]	<0.001
V$NFY.03	FWD	2.00	[1.60, 2.40]	3.19	[2.81, 3.56]	<0.001
V$SOX6.01	FWD	1.02	[0.73, 1.31]	1.66	[1.39, 1.94]	0.001
V$LEF1.01	FWD	1.02	[0.73, 1.31]	1.63	[1.36, 1.90]	0.002
V$PAX9.02	REV	1.04	[0.75, 1.34]	1.89	[1.59, 2.18]	<0.001
V$PAX6.01	REV	1.04	[0.75, 1.34]	1.88	[1.58, 2.17]	<0.001

Abbreviations: CI, confidence interval; CSF, cerebrospinal fluid; FWD, forward; JCV, JC virus; NCCR, non-coding control region; PML, progressive multifocal leukoencephalopathy; REV, reverse; TFBS, transcription factor binding site. [a] The number of each matrix within the NCCRs of JCV isolates (multiplicity) was predicted by using the Poisson distribution for modeling count data. [b] The direction of the DNA strands is mentioned in the Figure 2 legend. [c] *P*-values were adjusted for multiple testing using the Benjamini–Hochberg method, and the matrices with statistically significant differences are shown.
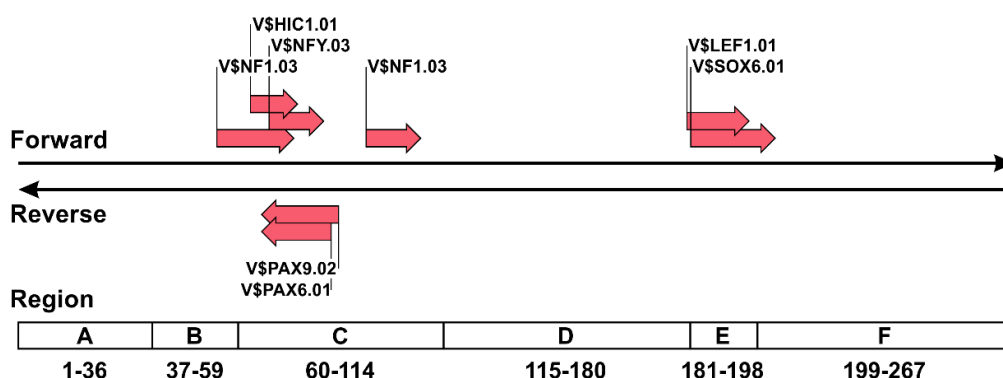

Figure 4. Types and locations of TFBSs tend to overlap in rearranged NCCR sequences of prototype JCV. TFBS matrices with statistically higher multiplicity in the NCCR of the prototype JCV as compared with those of the archetype JCV were determined by Poisson distribution for modeling count data. The locations of these matrices are depicted in the NCCR sequence of archetype JCV (CY strain).

Table 4. Profiles of TFBSs likely to multiply in the NCCRs of JCVs in the CSF of PML patients.

Matrix Name	DNA Strand [b]	Transcription Factor [a]			
		HGNC ID	Symbol	Full Name	Gene Expression (>3 × Median) [c]
V\$HIC1.01	FWD	4909	HIC1	HIC ZBTB transcriptional repressor 1	NA (Ubiquitous)
V\$NF1.03	FWD	7784	NFIA	Nuclear factor I A	NA (Ubiquitous)
V\$NF1.03	FWD	7785	NFIB	Nuclear factor I B	Brain (cerebrum, cerebellum, olfactory bulb), Colon, Ovary, Pancreatic islet, Prostate, Retina, Salivary gland, Skin, Small intestine, Smooth muscle, Tongue, Trachea, Uterus
V\$NF1.03	FWD	7786	NFIC	Nuclear factor I C	Skeletal muscle
V\$NF1.03	FWD	7788	NFIX	Nuclear factor I X	NA (Ubiquitous)
V\$NFY.03	FWD	7804	NFYA	Nuclear transcription factor Y subunit alpha	NA (Ubiquitous)
V\$SOX6.01	FWD	16421	SOX6	SRY-box transcription factor 6	NA [d]
V\$LEF1.01	FWD	6551	LEF1	Lymphoid enhancer binding factor 1	Blood (T cells), Thymus
V\$PAX9.02	REV	8623	PAX9	Paired box 9	NA (Ubiquitous)
V\$PAX6.01	REV	8620	PAX6	Paired box 6	Brain (cerebrum, cerebellum), Pancreatic islet, Pineal gland, Retina, Skeletal muscle

Abbreviations: CSF, cerebrospinal fluid; FWD, forward; HGNC, Human Genome Organization (HUGO) Gene Nomenclature Committee; ID, identification; JCV, JC virus; NA, not applicable; NCCR, non-coding control region; PML, progressive multifocal leukoencephalopathy; REV, reverse; TFBS, transcription factor binding site. [a] The gene ontology of transcription factors predicted to bind to each sequence was confirmed using the HGNC database according the metadata of the matrices. [b] The direction of the DNA strands is mentioned in the Figure 2 legend. [c] Gene-expression profiles of transcription factors in human tissues and blood were obtained using BioGPS microarray data, and the sites with 3-fold higher expression levels relative to the median are indicated. [d] The gene-expression profile for SOX6 was not included in the dataset.

4. Discussion

In this study, the NCCR sequences of many JCV isolates were sorted and aligned, and their origins were checked individually. Additionally, we identified TFBSs by computer simulation and added these to the NCCR sequences as annotations. The resulting database can be used to visually display the location of TFBSs within the NCCRs of both archetype and prototype JCVs or extract a tabulated list of NCCRs that possess particular TFBSs. Although the TFBSs highly conserved in the archetype NCCRs were analyzed by data science or statistics in this work, the actual sequence list contained all identified TFBSs. This database could be used to compare not only the differences between archetype and prototype JCVs, but also the NCCR sequences of each virus type. For example, it may be applied to the analysis of TFBS among genotypes of archetype JCVs or to the statistical analysis of NCCR sequences of prototype JCVs among groups of PML patients divided by factors such as underlying disease and prognosis. Furthermore, the FASTA file dataset with TFBS annotations can be utilized by standard genetic analysis software without requiring advanced database-management skills, making it versatile for research on NCCRs.

We used the database created in this study to analyze features and trends in the patterns of TFBSs within the NCCR of the prototype JCV. A notable finding was that a limited number of TFBSs conserved in the NCCR sequences of archetype JCVs was often lost or overlapped in prototype viruses. Previous studies reported that the balance of promoter activities of early and late genes differs in the NCCRs of archetype and prototype JCVs [40,41]. It is likely that changes in TFBS patterns are associated with alterations in promoter

activity. Interestingly, although we identified TFBS matrices with statistically distinct possession rates or multiplicities in the NCCRs of prototype JCVs as compared with the archetype viruses, these matrices were neither lost nor duplicated in some prototype JCVs. These observations imply that there is no universal pattern for TFBS deletion or multiplication during NCCR rearrangement, and that a vast number of TFBS combinations are being generated.

The TFBSs with lower possession rates in prototype JCVs were mainly distributed in regions D through F of the NCCR sequences of archetype viruses. There are numerous reports indicating that NCCRs of prototype JCVs frequently lack region D [68], which is consistent with the present findings. Additionally, we showed that regions E and F also tended to be deleted according to statistical analysis of the TFBS patterns in a large number of prototype JCVs. Moreover, analysis of data from the BioGPS database indicated that the transcription factors that bind to these TFBSs were expressed in diverse peripheral tissues, including sites of persistent or latent JCV infection. Notably, several transcription factors that bind to these TFBSs with reduced possession rates are highly expressed in CD34+ cells, which have been implicated in PML pathogenesis as sites of latent or persistent JCV infection [29]. By contrast, we did not observe specific and high-level expression of transcription factors, which bind to TFBSs with increased multiplicity, in CD34+ cells. Thus, several TFBSs lost upon NCCR rearrangement might control the promoter activity of archetype JCV in CD34+ cells during latent or persistent infection and are not required for lytic infection of prototype JCV in the brain.

The number of TFBSs prone to overlap was small as compared with those likely to be lost in prototype NCCRs. This feature can be attributed to the fact that it is not unusual to find prototype JCVs without duplicated sequences in NCCRs, and that the pattern of rearrangement is highly variable, causing few TFBSs to overlap in common. We found that the NCCR sequences in regions C and E of archetype JCV tended to be duplicated in the prototype viruses based on results from the count data model using TFBSs as landmarks. It is difficult to reveal the mutational patterns of rearranged NCCRs by simply comparing the alignment of nucleotide sequences; therefore, this result could provide valuable insights into the mutational trends of NCCRs. Additionally, examination of the gene-expression profiles of transcription factors capable of binding to the TFBSs (which are likely to overlap in the prototype JCV) revealed their expression in various tissues. Although these observations suggest the possibility that overlapping TFBSs in rearranged NCCRs facilitate JCV proliferation, this duplication event might not be involved in the tropism of this virus in the brain.

We expect that the database created in this study will serve as a convenient roadmap for future studies of NCCR functions and especially analyses of the activities of early and late promoters of JCV. It would be interesting to compare the results obtained from the TFBS database with the transcriptional activity of JCV promoters in vitro. For example, the TFBS matrix V$NF1.03, which tends to duplicate in the NCCRs of prototype JCV, is targeted by transcription factors belonging to the NFI family, which are reportedly involved in JCV gene expression and proliferation [61,69]. Another interesting example is that the TFBS matrix V$SPIB.01 for the Spi-B transcription factor, which reportedly plays an important role in JCV gene expression [61,70,71], is highly conserved in both archetype and prototype JCVs (possession rates: 100% and 96.7%, respectively) in the created database. Notably, the possession rate and multiplicity of V$SPIB.01 were not statistically significant in the archetype and prototype JCVs ($P = 0.552$ and $P = 0.899$, respectively), suggesting that TFBSs that are not lost or duplicated upon NCCR rearrangement play roles in JCV replication. As summarized in Supplementary Dataset S3, there are other TFBSs that are highly conserved in the NCCRs of both archetype and prototype JCVs. Among these TFBSs, it would be interesting to investigate the function of the TFBS matrix V$OLIG2.01, which is located at the downstream end of the forward strand of the NCCR. This TFBS includes a binding sequence for oligodendrocyte transcription factor 2 (OLIG2), which is expressed in oligodendrocytes and required for myelination [72]. It is speculated that OLIG2 might be involved in the glia-specific gene expression of JCV.

Furthermore, it would be interesting to analyze the promoter machinery of NCCRs in more detail by deleting these TFBSs individually or in combination based on their sequences and locations. MatInspector is available to perform TFBS simulations and use metadata based on a paid subscription. Under the terms of the license agreement of MatInspector, disclosing more than 10 DNA sequences of TFBSs in publications is legally restricted. However, in this study, Precigen Bioinformatics Germany GmbH, which does business as Genomatix and is the provider of this product, has kindly allowed us to disclose the actual DNA sequences presented in the Figures and Tables. Thus, we listed the DNA sequences of TFBSs within NCCR of representative CY strain of JCV and attached a digital file with the names and HGNC IDs of their corresponding transcription factors (Supplementary Dataset S4). In these TFBSs, the core sequences are common in the archetype and prototype JCVs. However, in the rearranged NCCRs of prototype JCVs, there are occasionally differences in DNA sequences other than the core of the TFBSs, which are defined as one of the same TFBS matrices. In such cases, it may be possible to find the TFBS by alignment with the annotated sequence list of NCCRs from many prototype JCVs, although a simple sequence search may not yield any hits. The dataset established in the present study will be made available for non-commercial purposes upon reasonable request. Additionally, for detailed analysis of TFBSs in the NCCR sequences, we recommend the use of MatInspector.

5. Conclusions

In conclusion, we described the creation and application of a database containing the NCCR sequences of archetype and prototype JCVs, as well as annotations of TFBSs within these NCCRs. Furthermore, statistical analyses clarified the observed alterations in TFBS patterns during NCCR rearrangements. We believe that the generated database and subsequent insights obtained in this study will contribute to further advances in the analysis of NCCR function and JCV pathogenicity.

References

1. Cortese, I.; Reich, D.S.; Nath, A. Progressive multifocal leukoencephalopathy and the spectrum of JC virus-related disease. *Nat. Rev. Neurol.* **2021**, *17*, 37–51, doi:10.1038/s41582-020-00427-y.
2. Knowles, W.A.; Pipkin, P.; Andrews, N.; Vyse, A.; Minor, P.; Brown, D.W.; Miller, E. Population-based study of antibody to the human polyomaviruses BKV and JCV and the simian polyomavirus SV40. *J. Med. Virol.* **2003**, *71*, 115–123, doi:10.1002/jmv.10450.
3. Kean, J.M.; Rao, S.; Wang, M.; Garcea, R.L. Seroepidemiology of Human Polyomaviruses. *PLOS Pathog.* **2009**, *5*, e1000363, doi:10.1371/journal.ppat.1000363.
4. Egli, A.; Infanti, L.; Dumoulin, A.; Buser, A.; Samaridis, J.; Stebler, C.; Gosert, R.; Hirsch, H.H. Prevalence of Polyomavirus BK and JC Infection and Replication in 400 Healthy Blood Donors. *J. Infect. Dis.* **2009**, *199*, 837–846, doi:10.1086/597126.
5. Kornek, B.; Huppke, P.; Reindl, M.; Rostasy, K.; Berger, T.; Hennes, E.M. Age-Dependent Seroprevalence of JCV Antibody in Children. *Neuropediatrics* **2015**, *47*, 112–114, doi:10.1055/s-0035-1565272.
6. Dörries, K. Molecular biology and pathogenesis of human polyomavirus infections. *Dev. Boil. Stand.* **1998**, *94*, 71–79.
7. Zheng, H.-Y.; Kitamura, T.; Takasaka, T.; Chen, Q.; Yogo, Y. Unambiguous identification of JC polyomavirus strains transmitted from parents to children. *Arch. Virol.* **2003**, *149*, 261–273, doi:10.1007/s00705-003-0214-6.
8. Ferenczy, M.W.; Marshall, L.J.; Nelson, C.; Atwood, W.J.; Nath, A.; Khalili, K.; Major, E.O. Molecular Biology, Epidemiology, and Pathogenesis of Progressive Multifocal Leukoencephalopathy, the JC Virus-Induced Demyelinating Disease of the Human Brain. *Clin. Microbiol. Rev.* **2012**, *25*, 471–506, doi:10.1128/cmr.05031-11.
9. Monaco, M.C.; Atwood, W.J.; Gravell, M.; Tornatore, C.S.; Major, E.O. JC virus infection of hematopoietic progenitor cells, primary B lymphocytes, and tonsillar stromal cells: Implications for viral latency. *J. Virol.* **1996**, *70*, 7004–7012, doi:10.1128/jvi.70.10.7004-7012.1996.
10. Marzocchetti, A.; Wuthrich, C.; Tan, C.S.; Tompkins, T.; Bernal-Cano, F.; Bhargava, P.; Ropper, A.H.; Koralnik, I.J. Rearrangement of the JC virus regulatory region sequence in the bone marrow of a patient with rheumatoid arthritis and progressive multifocal leukoencephalopathy. *J. NeuroVirol.* **2008**, *14*, 455–458, doi:10.1080/13550280802356837.
11. Tan, C.S.; Dezube, B.J.; Bhargava, P.; Autissier, P.; Wüthrich, C.; Miller, J.; Koralnik, I.J. Detection of JC Virus DNA and Proteins in the Bone Marrow of HIV-Positive and HIV-Negative Patients: Implications for Viral Latency and Neurotropic Transformation. *J. Infect. Dis.* **2009**, *199*, 881–888, doi:10.1086/597117.
12. 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.* **1993**, *74*, 1499–1507, doi:10.1099/0022-1317-74-8-1499.
13. Yogo, Y.; Zhong, S.; Shibuya, A.; Kitamura, T.; Homma, Y. Transcriptional control region rearrangements associated with the evolution of JC polyomavirus. *Virology* **2008**, *380*, 118–123, doi:10.1016/j.virol.2008.07.016.
14. Verbeeck, J.; Van Assche, G.; Ryding, J.; Wollants, E.; Rans, K.; Vermeire, S.; Pourkarim, M.R.; Noman, M.; Dillner, J.; Van Ranst, M.; et al. JC viral loads in patients with Crohn's disease treated with immunosuppression: Can we screen for elevated risk of progressive multifocal leukoencephalopathy? *Gut* **2008**, *57*, 1393–1397, doi:10.1136/gut.2007.145698.
15. Randhawa, P.; Uhrmacher, J.; Pasculle, W.; Vats, A.; Shapiro, R.; Eghtsead, B.; Weck, K. A comparative study of BK and JC virus infections in organ transplant recipients. *J. Med. Virol.* **2005**, *77*, 238–243, doi:10.1002/jmv.20442.
16. Agostini, H.T.; Ryschkewitsch, C.F.; Stoner, G.L. Genotype profile of human polyomavirus JC excreted in urine of immunocompetent individuals. *J. Clin. Microbiol.* **1996**, *34*, 159–164, doi:10.1128/jcm.34.1.159-164.1996.
17. Atyabi, S.R.; Bouzari, M.; Kardi, M.T. John cunningham (JC) virus genotypes in kidney transplant recipients, rheumatoid arthritis patients and healthy individuals in Isfahan, Iran. *J. Med. Virol.* **2017**, *89*, 337–344, doi:10.1002/jmv.24626.
18. Karalic, D.; Lazarevic, I.; Banko, A.; Cupic, M.; Jevtović, Đ.; Jovanovic, T. Analysis of variability of urinary excreted JC virus strains in patients infected with HIV and healthy donors. *J. NeuroVirology* **2018**, *24*, 305–313, doi:10.1007/s13365-017-0608-y.
19. Markowitz, R.-B.; Thompson, H.C.; Mueller, J.F.; Cohen, J.A.; Dynan, W. Incidence of BK Virus and JC Virus Viruria in Human Immunodeficiency Virus-Infected and -Uninfected Subjects. *J. Infect. Dis.* **1993**, *167*, 13–20, doi:10.1093/infdis/167.1.13.
20. Stoner, G.L.; Agostini, H.T.; Ryschkewitsch, C.F.; Komoly, S. JC virus excreted by multiple sclerosis patients and paired controls from Hungary. *Mult. Scler. J.* **1998**, *4*, 45–48, doi:10.1177/135245859800400201.
21. Sundsfjord, A.; Osei, A.; Rosenqvist, H.; Van Ghelue, M.; Silsand, Y.; Haga, H.; Rekvig, O.P.; Moens, U. BK and JC Viruses in Patients with Systemic Lupus Erythematosus: Prevalent and Persistent BK Viruria, Sequence Stability of the Viral Regulatory Regions, and Nondetectable Viremia. *J. Infect. Dis.* **1999**, *180*, 1–9, doi:10.1086/314830.
22. Ferrante, M.M.P.; Mediati, M.; Caldarelli-Stefano, R.; Losciale, L.; Mancuso, R.; Cagni, A.E.; Maserati, R.; Ferrante, P. Increased frequency of JC virus type 2 and of dual infection with JC virus type 1 and 2 in Italian progressive multifocal leukoencephalopathy patients. *J. NeuroVirol.* **2001**, *7*, 35–42, doi:10.1080/135502801300069638.
23. Pagani, E.; Delbue, S.; Mancuso, R.; Borghi, E.; Tarantini, L.; Ferrante, P. Molecular Analysis of JC Virus Genotypes Circulating Among the Italian Healthy Population. *J. NeuroVirol.* **2003**, *9*, 559–566, doi:10.1080/13550280390241269.
24. Rossi, A.; Delbue, S.; Mazziotti, R.; Valli, M.; Borghi, E.; Mancuso, R.; Calvo, M.G.; Ferrante, P. Presence, quantitation and characterization of JC virus in the urine of Italian immunocompetent subjects. *J. Med. Virol.* **2007**, *79*, 408–412, doi:10.1002/jmv.20829.
25. Comar, M.; Delbue, S.; Lepore, L.; Martelossi, S.; Radillo, O.; Ronfani, L.; D'Agaro, P.; Ferrante, P. Latent viral infections in young patients with inflammatory diseases treated with biological agents: Prevalence of JC virus genotype 2. *J. Med. Virol.* **2013**, *85*, 716–722, doi:10.1002/jmv.23525.

26. Melo, F.A.F.; Bezerra, A.C.F.; Santana, B.B.; Ishak, M.O.G.; Ishak, R.; Vallinoto, I.M.V.C.; Vallinoto, A.C.R. JC polyomavirus infection in candidates for kidney transplantation living in the Brazilian Amazon Region. *Memórias Inst. Oswaldo Cruz* **2013**, *108*, 145–149, doi:10.1590/0074-0276108022013003.

27. Zanotta, N.; Delbue, S.; Rossi, T.; Pelos, G.; D'Agaro, P.; Monasta, L.; Ferrante, P.; Comar, M. Molecular epidemiology of JCV genotypes in patients and healthy subjects from Northern Italy. *J. Med. Virol.* **2013**, *85*, 1286–1292, doi:10.1002/jmv.23585.

28. White, M.K.; Khalili, K. Pathogenesis of Progressive Multifocal Leukoencephalopathy—Revisited. *J. Infect. Dis.* **2011**, *203*, 578–586, doi:10.1093/infdis/jiq097.

29. Major, E.O.; Yousry, T.A.; Clifford, D.B. Pathogenesis of progressive multifocal leukoencephalopathy and risks associated with treatments for multiple sclerosis: A decade of lessons learned. *Lancet Neurol.* **2018**, *17*, 467–480, doi:10.1016/s1474-4422(18)30040-1.

30. Kleinschmidt-DeMasters, B.; Tyler, K. Progressive Multifocal Leukoencephalopathy Complicating Treatment with Natalizumab and Interferon Beta-1a for Multiple Sclerosis. *N. Engl. J. Med.* **2005**, *353*, 369–374, doi:10.1056/nejmoa051782.

31. Langer-Gould, A.; Atlas, S.W.; Green, A.J.; Bollen, A.W.; Pelletier, D. Progressive Multifocal Leukoencephalopathy in a Patient Treated with Natalizumab. *N. Engl. J. Med.* **2005**, *353*, 375–381, doi:10.1056/nejmoa051847.

32. Van Assche, G.; Van Ranst, M.; Sciot, R.; Dubois, B.; Vermeire, S.; Noman, M.; Verbeeck, J.; Geboes, K.; Robberecht, W.; Rutgeerts, P. Progressive Multifocal Leukoencephalopathy after Natalizumab Therapy for Crohn's Disease. *N. Engl. J. Med.* **2005**, *353*, 362–368, doi:10.1056/nejmoa051586.

33. McIlroy, D.; Halary, F.; Bressollette-Bodin, C. Intra-patient viral evolution in polyomavirus-related diseases. *Philos. Trans. R. Soc. B Biol. Sci.* **2019**, *374*, 20180301, doi:10.1098/rstb.2018.0301.

34. Seppälä, H.; Virtanen, E.; Saarela, M.; Laine, P.; Paulín, L.; Mannonen, L.; Auvinen, P.; Auvinen, E. Single-Molecule Sequencing Revealing the Presence of Distinct JC Polyomavirus Populations in Patients with Progressive Multifocal Leukoencephalopathy. *J. Infect. Dis.* **2016**, *215*, jiw399–895, doi:10.1093/infdis/jiw399.

35. Ciardi, M.R.; Zingaropoli, M.A.; Iannetta, M.; Prezioso, C.; Perri, V.; Pasculli, P.; Lichtner, M.; D'Ettorre, G.; Altieri, M.; Conte, A.; et al. JCPyV NCCR analysis in PML patients with different risk factors: Exploring common rearrangements as essential changes for neuropathogenesis. *Virol. J.* **2020**, *17*, 1–8, doi:10.1186/s12985-020-1295-5.

36. Prezioso, C.; Zingaropoli, M.A.; Iannetta, M.; Rodio, D.M.; Altieri, M.; Conte, A.; Vullo, V.; Ciardi, M.R.; Palamara, A.T.; Pietropaolo, V. Which is the best PML risk stratification strategy in natalizumab-treated patients affected by multiple sclerosis? *Mult. Scler. Relat. Disord.* **2020**, *41*, 102008, doi:10.1016/j.msard.2020.102008.

37. Frisque, R.J.; Bream, G.L.; Cannella, M.T. Human polyomavirus JC virus genome. *J. Virol.* **1984**, *51*, 458–469, doi:10.1128/jvi.51.2.458-469.1984.

38. White, M.K.; Safak, M.; Khalili, K. Regulation of Gene Expression in Primate Polyomaviruses. *J. Virol.* **2009**, *83*, 10846–10856, doi:10.1128/jvi.00542-09.

39. Yogo, Y.; Sugimoto, C. The Archetype Concept and Regulatory Region Rearrangement. In *Human Polyomaviruses*; Khalili, K., Stoner, G.L., Eds.; John Wiley & Sons, Ltd: New York, NY, USA, 2001; pp. 127–148.

40. Gosert, R.; Kardas, P.; Major, E.O.; Hirsch, H.H. Rearranged JC Virus Noncoding Control Regions Found in Progressive Multifocal Leukoencephalopathy Patient Samples Increase Virus Early Gene Expression and Replication Rate. *J. Virol.* **2010**, *84*, 10448–10456, doi:10.1128/jvi.00614-10.

41. L'Honneur, A.-S.; Leh, H.; Laurent-Tchenio, F.; Hazan, U.; Rozenberg, F.; Bury-Moné, S. Exploring the role of NCCR variation on JC polyomavirus expression from dual reporter minicircles. *PLoS ONE* **2018**, *13*, e0199171, doi:10.1371/journal.pone.0199171.

42. Nakamichi, K.; Kishida, S.; Tanaka, K.; Suganuma, A.; Sano, Y.; Sano, H.; Kanda, T.; Maeda, N.; Kira, J.-I.; Itoh, A.; et al. Sequential changes in the non-coding control region sequences of JC polyomaviruses from the cerebrospinal fluid of patients with progressive multifocal leukoencephalopathy. *Arch. Virol.* **2012**, *158*, 639–650, doi:10.1007/s00705-012-1532-3.

43. Nakamichi, K.; Tajima, S.; Lim, C.-K.; Saijo, M. High-resolution melting analysis for mutation scanning in the non-coding control region of JC polyomavirus from patients with progressive multifocal leukoencephalopathy. *Arch. Virol.* **2014**, *159*, 1687–1696, doi:10.1007/s00705-014-1988-4.

44. Tan, C.S.; Ellis, L.C.; Wüthrich, C.; Ngo, L.; Broge, T.A.; Saint-Aubyn, J.; Miller, J.S.; Koralnik, I.J. JC Virus Latency in the Brain and Extraneural Organs of Patients with and without Progressive Multifocal Leukoencephalopathy. *J. Virol.* **2010**, *84*, 9200–9209, doi:10.1128/jvi.00609-10.

45. Agostini, H.T.; Ryschkewitsch, C.F.; Singer, E.J.; Stoner, G.L. JC virus regulatory region rearrangements and genotypes in progressive multifocal leukoencephalopathy: Two independent aspects of virus variation. *J. Gen. Virol.* **1997**, *78*, 659–664, doi:10.1099/0022-1317-78-3-659.

46. Reid, C.E.; Li, H.; Sur, G.; Carmillo, P.; Bushnell, S.; Tizard, R.; McAuliffe, M.; Tonkin, C.; Simon, K.; Goelz, S.; et al. Sequencing and Analysis of JC Virus DNA From Natalizumab-Treated PML Patients. *J. Infect. Dis.* **2011**, *204*, 237–244, doi:10.1093/infdis/jir256.

47. Roux, D.; Bouldouyre, M.-A.; Mercier-Delarue, S.; Seilhean, D.; Zagdanski, A.-M.; Delaugerre, C.; Simon, F.; Molina, J.-M.; LeGoff, J. JC Virus Variant Associated with Cerebellar Atrophy in a Patient with AIDS. *J. Clin. Microbiol.* **2011**, *49*, 2196–2199, doi:10.1128/jcm.02057-10.

48. Delbue, S.; Elia, F.; Carloni, C.; Tavazzi, E.; Marchioni, E.; Carluccio, S.; Signorini, L.; Novati, S.; Maserati, R.; Ferrante, P. JC virus load in cerebrospinal fluid and transcriptional control region rearrangements may predict the clinical course of progressive multifocal leukoencephalopathy. *J. Cell. Physiol.* **2012**, *227*, 3511–3517, doi:10.1002/jcp.24051.

49. Ciappi, S.; Azzi, A.; De Santis, R.; Leoncini, F.; Sterrantino, G.; Mazzotta, F.; Mecocci, L. Archetypal and rearranged sequences of human polyomavirus JC transcription control region in peripheral blood leukocytes and in cerebrospinal fluid. *J. Gen. Virol.* **1999**, *80*, 1017–1023, doi:10.1099/0022-1317-80-4-1017.

50. Iida, T.; Kitamura, T.; Guo, J.; Taguchi, F.; Aso, Y.; Nagashima, K.; Yogo, Y. Origin of JC polyomavirus variants associated with progressive multifocal leukoencephalopathy. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 5062–5065, doi:10.1073/pnas.90.11.5062.

51. Sugimoto, C.; Ito, D.; Tanaka, K.; Matsuda, H.; Saito, H.; Sakai, H.; Fujihara, K.; Itoyama, Y.; Yamada, T.; Kira, J.; et al. Amplification of JC virus regulatory DNA sequences from cerebrospinal fluid: Diagnostic value for progressive multifocal leukoencephalopathy. *Arch. Virol.* **1998**, *143*, 249–262, doi:10.1007/s007050050284.

52. Yasuda, Y.; Yabe, H.; Inoue, H.; Shimizu, T.; Yabe, M.; Yogo, Y.; Kato, S. Comparison of PCR-amplified JC virus control region sequences from multiple brain regions in PML. *Neurology* **2003**, *61*, 1617–1619, doi:10.1212/01.wnl.0000096147.47128.7d.

53. Wharton, K.A.; Quigley, C.; Themeles, M.; Dunstan, R.W.; Doyle, K.; Cahir-McFarland, E.; Wei, J.; Buko, A.; Reid, C.E.; Sun, C.; et al. JC Polyomavirus Abundance and Distribution in Progressive Multifocal Leukoencephalopathy (PML) Brain Tissue Implicates Myelin Sheath in Intracerebral Dissemination of Infection. *PLoS ONE* **2016**, *11*, e0155897, doi:10.1371/journal.pone.0155897.

54. Takahashi, K.; Sekizuka, T.; Fukumoto, H.; Nakamichi, K.; Suzuki, T.; Sato, Y.; Hasegawa, H.; Kuroda, M.; Katano, H. Deep-Sequence Identification and Role in Virus Replication of a JC Virus Quasispecies in Patients with Progressive Multifocal Leukoencephalopathy. *J. Virol.* **2017**, *91*, 01335–16, doi:10.1128/jvi.01335-16.

55. Van Loy, T.; Thys, K.; Ryschkewitsch, C.; Lagatie, O.; Monaco, M.C.; Major, E.O.; Tritsmans, L.; Stuyver, L.J. JC Virus Quasispecies Analysis Reveals a Complex Viral Population Underlying Progressive Multifocal Leukoencephalopathy and Supports Viral Dissemination via the Hematogenous Route. *J. Virol.* **2015**, *89*, 1340–1347, doi:10.1128/jvi.02565-14.

56. Liang, B.; Tikhanovich, I.; Nasheuer, H.P.; Folk, W.R. Stimulation of BK Virus DNA Replication by NFI Family Transcription Factors. *J. Virol.* **2011**, *86*, 3264–3275, doi:10.1128/jvi.06369-11.

57. Chattaraj, S.; Bhattacharjee, S. Molecular Analysis of JC Polyomavirus Genotypes Circulating among Tribal Populations of North-Eastern West Bengal, India. *Pol. J. Microbiol.* **2014**, *63*, 191–201, doi:10.33073/pjm-2014-025.

58. Bethge, T.; Ajuh, E.; Hirsch, H.H. Imperfect Symmetry of Sp1 and Core Promoter Sequences Regulates Early and Late Virus Gene Expression of the Bidirectional BK Polyomavirus Noncoding Control Region. *J. Virol.* **2016**, *90*, 10083–10101, doi:10.1128/jvi.01008-16.

59. Ajuh, E.T.; Wu, Z.; Kraus, E.; Weissbach, F.H.; Bethge, T.; Gosert, R.; Fischer, N.; Hirsch, H.H. Novel Human Polyomavirus Noncoding Control Regions Differ in Bidirectional Gene Expression according to Host Cell, Large T-Antigen Expression, and Clinically Occurring Rearrangements. *J. Virol.* **2018**, *92*, e02231-17, doi:10.1128/jvi.02231-17.

60. Reoma, L.B.; Trindade, C.J.; Monaco, M.C.; Solis, J.; Montojo, M.G.; Vu, P.; Johnson, K.; Beck, E.; Nair, G.; Khan, O.I.; et al. Fatal encephalopathy with wild-type JC virus and ruxolitinib therapy. *Ann. Neurol.* **2019**, *86*, 878–884, doi:10.1002/ana.25608.

61. Ferenczy, M.W.; Johnson, K.R.; Marshall, L.J.; Monaco, M.C.; Major, E.O. Differentiation of Human Fetal Multipotential Neural Progenitor Cells to Astrocytes Reveals Susceptibility Factors for JC Virus. *J. Virol.* **2013**, *87*, 6221–6231, doi:10.1128/jvi.00396-13.

62. Cartharius, K.; Frech, K.; Grote, K.; Klocke, B.; Haltmeier, M.; Klingenhoff, A.; Frisch, M.; Bayerlein, M.; Werner, T. MatInspector and beyond: Promoter analysis based on transcription factor binding sites. *Bioinformatics* **2005**, *21*, 2933–2942, doi:10.1093/bioinformatics/bti473.

63. Quandt, K.; Frech, K.; Karas, H.; Wingender, E.; Werner, T. MatInd and MatInspector: New fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Res.* **1995**, *23*, 4878–4884, doi:10.1093/nar/23.23.4878.

64. Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **1995**, *57*, 289–300, doi:10.1111/j.2517-6161.1995.tb02031.x.

65. Wu, C.; Jin, X.; Tsueng, G.; Afrasiabi, C.; Su, A.I. BioGPS: Building your own mash-up of gene annotations and expression profiles. *Nucleic Acids Res.* **2016**, *44*, D313–D316, doi:10.1093/nar/gkv1104.

66. Su, A.I.; Wiltshire, T.; Batalov, S.; Lapp, H.; Ching, K.A.; Block, D.; Zhang, J.; Soden, R.; Hayakawa, M.; Kreiman, G.; et al. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 6062–6067, doi:10.1073/pnas.0400782101.

67. Hagiwara, N. Sox6, jack of all trades: A versatile regulatory protein in vertebrate development. *Dev. Dyn.* **2011**, *240*, 1311–1321, doi:10.1002/dvdy.22639.

68. Agostini, H.; Ryschkewitsch, C.; Stoner, G. Rearrangements of archetypal regulatory regions in JC virus genomes from urine. *Res. Virol.* **1998**, *149*, 163–170, doi:10.1016/s0923-2516(98)80034-4.

69. Ravichandran, V.; Major, E.O. DNA-binding transcription factor NF-1A negatively regulates JC virus multiplication. *J. Gen. Virol.* **2008**, *89*, 1396–1401, doi:10.1099/vir.0.2008/000059-0.

70. Marshall, L.J.; Moore, L.D.; Mirsky, M.M.; Major, E.O. JC virus promoter/enhancers contain TATA box-associated Spi-B-binding sites that support early viral gene expression in primary astrocytes. *J. Gen. Virol.* **2012**, *93*, 651–661, doi:10.1099/vir.0.035832-0.

71. Marshall, L.J.; Ferenczy, M.W.; Daley, E.; Jensen, P.N.; Ryschkewitsch, C.F.; Major, E.O. Lymphocyte Gene Expression and JC Virus Noncoding Control Region Sequences Are Linked with the Risk of Progressive Multifocal Leukoencephalopathy. *J. Virol.* **2014**, *88*, 5177–5183, doi:10.1128/jvi.03221-13.

72. Long, K.L.P.; Breton, J.M.; Barraza, M.K.; Perloff, O.S.; Kaufer, D. Hormonal Regulation of Oligodendrogenesis I: Effects Across the Lifespan. *Biomolecules* **2021**, *11*, 283, doi:10.3390/biom11020283.