

# Differentiation of highly pathogenic strains of human JC polyomavirus in neurological patients by next generation sequencing

Eeva Auvinen<sup>a,\*</sup>, Anni Honkimaa<sup>a</sup>, Pia Laine<sup>b</sup>, Sara Passerini<sup>c</sup>, Ugo Moens<sup>d</sup>, Valeria Pietropaolo<sup>c</sup>, Mika Saarela<sup>e</sup>, Leena Maunula<sup>f</sup>, Laura Mannonen<sup>a</sup>, Olli Tynninén<sup>g</sup>, Hannu Haapasalo<sup>h</sup>, Tuomas Rauramaa<sup>i</sup>, Petri Auvinen<sup>b</sup>, Hanna Liimatainen<sup>a</sup>

<sup>a</sup> Department of Virology, Helsinki University Hospital and University of Helsinki, Helsinki, Finland

<sup>b</sup> Institute of Biotechnology, DNA Sequencing and Genomics Laboratory, University of Helsinki, Helsinki, Finland

<sup>c</sup> Department of Public Health and Infectious Diseases, Sapienza University of Rome, Rome, Italy

<sup>d</sup> Institute of Medical Biology, UiT The Arctic University of Norway, Norway

<sup>e</sup> Department of Neurology, Helsinki University Hospital and University of Helsinki, Helsinki, Finland

<sup>f</sup> Department of Food Hygiene and Environmental Health, Faculty of Veterinary Medicine, University of Helsinki, Helsinki, Finland

<sup>g</sup> Department of Pathology, Helsinki University Hospital and University of Helsinki, Helsinki, Finland

<sup>h</sup> Department of Pathology, FIMLAB Laboratories Ltd and Tampere University, Tampere, Finland

<sup>i</sup> Department of Pathology, Kuopio University Hospital, Kuopio, Finland and 12. Unit of Pathology, Institute of Clinical Medicine, University of Eastern Finland, Kuopio, Finland

## ARTICLE INFO

### Keywords:

Diagnostics  
JC polyomavirus  
NGS  
PML  
Targeted sequencing

## ABSTRACT

**Background:** JC polyomavirus (JCPyV) persists asymptomatic in more than half of the human population. Immunocompromising conditions may cause reactivation and acquisition of neurotropic rearrangements in the viral genome, especially in the non-coding control region (NCCR). Such rearranged JCPyV strains are strongly associated with the development of progressive multifocal leukoencephalopathy (PML).

**Methods:** Using next-generation sequencing (NGS) and bioinformatics tools, the NCCR was characterized in cerebrospinal fluid (CSF;  $N = 21$ ) and brain tissue ( $N = 16$ ) samples from PML patients ( $N = 25$ ), urine specimens from systemic lupus erythematosus patients ( $N = 2$ ), brain tissue samples from control individuals ( $N = 2$ ) and waste-water samples ( $N = 5$ ). Quantitative PCR was run in parallel for diagnostic PML samples.

**Results:** Archetype NCCR (i.e. ABCDEF block structure) and archetype-like NCCR harboring minor mutations were detected in two CSF samples and in one CSF sample and in one tissue sample, respectively. Among samples from PML patients, rearranged NCCRs were found in 8 out of 21 CSF samples and in 14 out of 16 brain tissue samples. Complete or partial deletion of the C and D blocks was characteristic of most rearranged JCPyV strains. From ten CSF samples and one tissue sample NCCR could not be amplified.

**Conclusions:** Rearranged NCCRs are predominant in brain tissue and common in CSF from PML patients. Extremely sensitive detection and identification of neurotropic viral populations in CSF or brain tissue by NGS may contribute to early and accurate diagnosis, timely intervention and improved patient care.

## 1. Introduction

JC polyomavirus (JCPyV) is a ubiquitous human virus acquired by respiratory or fecal-oral route mostly in childhood. Seroprevalence among healthy adults is 60–80 % worldwide [1,2]. After primary infection lifelong persistence is established in the kidney or other

peripheral tissues [3]. Persistent infection in immunocompetent individuals is asymptomatic, but occasional activation of virus replication may lead to viraemia in 20–25 % of healthy individuals [2]. Persistent virus may be reactivated due to weakened immunological status, and enhanced replication of the virus may lead to the development of progressive multifocal leukoencephalopathy (PML) or rarer neurological conditions such as cerebellar granule cell neuronopathy,

Suggested Referees: Axel.zurHausen@mumc.nl, Flore.Rozenberg@aphp.fr

\* Corresponding author at: University of Helsinki and Helsinki University Hospital Diagnostic Center, Department of Virology, POB 21 (Haartmaninkatu 3), 00014 University of Helsinki, Helsinki, Finland.

E-mail address: [eeva.auvinen@helsinki.fi](mailto:eeva.auvinen@helsinki.fi) (E. Auvinen).

<https://doi.org/10.1016/j.jcv.2024.105652>

Received 10 December 2023; Received in revised form 1 February 2024;

Available online 12 February 2024

1386-6532/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

### Abbreviations

ASV,	amplicon sequence variant
CNS,	central nervous system
CSF,	cerebrospinal fluid
FFPE,	formalin-fixed paraffin-embedded
JCPyV,	JC polyomavirus
MRI,	magnetic resonance imaging
MS,	multiple sclerosis
NCCR,	non-coding control region
NGS,	next-generation sequencing
PML,	progressive multifocal leukoencephalopathy
rr-NCCR,	rearranged non-coding control region
SLE,	systemic lupus erythematosus
TFBS,	transcription factor binding site

encephalopathy, or meningitis [4,5]. PML may rarely emerge in patients suffering from systemic lupus erythematosus (SLE) [6]. Immunosuppression due to HIV and AIDS, or the use of immunomodulatory treatments for severe chronic diseases including multiple sclerosis (MS), rheumatoid arthritis or lymphoma are among the most frequently identified factors predisposing to PML [7]. However, PML may also develop due to transient immunosuppression [8] or without known predisposing factors [9]. JCPyV may in rare cases be the causative agent of nephropathy in kidney transplant recipients [10]. A role in some malignancies in the central nervous system (CNS) has also been suggested [11].

The genome of JCPyV consists of the early region encoding the small t and large T antigens, the late region encoding the capsid proteins VP1, VP2, VP3, and agnoprotein, and the non-coding control region (NCCR) harboring the origin of replication, regulatory elements and binding sites for host transcription factors (TFBS) such as nuclear transcription factor-1 (NF-1), activating protein (AP-1), Ets subfamily members Spi-1 and Spi-B, and specificity protein-1 (Sp1) [4]. The NCCR contains determinants for the efficiency of viral replication and cellular tropism [4]. The NCCR of archetype CY strain consists of a linear arrangement of conserved sequence blocks named A (36 bp), B (23 bp), C (55 bp), D (66 bp), E (18 bp) and F (69 bp) [12]. Archetype JCPyV is thought to reside in the environment, circulate in the population and cause primary infections. Rearranged variants are derived from archetype strains pre-existing in the body upon reactivation [13,14]. Rearranged (rr) or neurotropic strains contain point mutations, deletions and duplications in the NCCR, and often point mutations in VP1 [13,15]. Compared with archetype CY strain, neurotropic Mad-1 strain is characterized by deleted B and D and duplicated A, C and E boxes [14]. rr-NCCRs reveal modified patterns of TFBSs, suggesting that the rearrangements may modify cellular tropism and increase viral transcription and replication within the brain [14]. PML patients frequently have one or more individual JCPyV strains with highly variable NCCR in blood, brain, and cerebrospinal fluid (CSF) [16–18]. Major strains in urine seem to be archetype [5,10], although minor rearranged virus populations have recently been reported [14]. Whether the rearrangements emerge or are initiated in peripheral tissue, are required for brain entry, or emerge first in the brain, remains a topic of active research.

PML diagnosis is based on clinical symptoms, magnetic resonance imaging (MRI) of the brain, and detection of JCPyV DNA in CSF or brain tissue by PCR. Occasionally PML is diagnosed when clinical symptoms are already severe, delaying appropriate patient management. In this work we explore the possibilities of next generation sequencing (NGS) to detect and differentiate clinically relevant, neurotropic JCPyV sequence variants to improve diagnosis, risk assessment and patient care.

## 2. Patients and methods

### 2.1. Patients and samples

In this study we included altogether 37 CSF or brain tissue samples from 25 patients with confirmed PML diagnosis. One CSF sample was available from eleven and two from two patients. From six patients we had CSF and tissue, and from 10 patients only tissue sample. For comparison we included urine samples from two SLE patients. Brain tissue samples from two non-PML individuals who had a condition unrelated to JCPyV were included as controls. Finally, five batches of waste-water samples were collected on different days from two water purification plants in Southern Finland, to represent the burden from human excretion to the environment. Of all PML samples 27 were originally taken for diagnostic JCPyV qPCR (1–17, 19–28), one for BKPyV qPCR (18) and nine for histological diagnostics (29–37). Detailed NCCR description of samples 7–9, 20–27 and 29–31 has been published elsewhere [16]. All samples are described in Table 1. The use of patient samples in this study was approved by the Ethical Board of the Helsinki University Hospital.

### 2.2. DNA extraction

Nucleic acids were extracted from CSF, tissue, urine and waste-water samples as described in detail in Supplementary Data.

### 2.3. NCCR amplification and diagnostic quantitative PCR

For short-read sequencing, the NCCR region was amplified using nucleic acid templates extracted from CSF, brain tissue, urine or waste-water samples. The 50  $\mu$ l PCR reactions consisted of 6  $\mu$ l template, 0.2 mmol/l each dNTP, 0.5  $\mu$ mol/l each primer, 1x buffer, and 1 U Q5 Hot-start polymerase (New England Biolabs). The forward and reverse primer sequences containing Illumina overhangs were 5'-ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT GGC CTC CTG TAT ATA TAA AAA AAA-3' and 5'-GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC TTT ACG TGA CAG CTG GCG AAG-3', respectively. After initial denaturation of 30 s at 98 °C, amplification was done for 35 cycles of 10 s at 98 °C, 30 s at 68 °C and 10 s at 72 °C.

Quantitative PCR (qPCR) for diagnostic PML samples is described in Supplementary Data.

### 2.4. Sequencing and analysis of NCCR sequences

The NCCR region was amplified and sequenced using short-read NGS (Illumina Inc., San Diego, CA) in order to characterize the JCPyV strains in samples. Approaches for library preparation and sequencing are described in Supplementary Data. NCCR sequence clusters and block structures in individual samples were analyzed manually. In order to examine the NCCR sequence variants and their frequencies across samples we applied DADA2 or amplicon sequence variant (ASV) analysis of NCCR sequence reads. Details of sequencing and bioinformatics methods are described in Supplementary Data.

## 3. Results

We analyzed the JCPyV strains in 21 CSF and 16 brain tissue samples from 25 individuals with confirmed PML diagnosis (Table 1, samples 1–37). Most of the patients had received PML-pre-exposing medications such as rituximab, bendamustin or prednisolone to treat e.g. lymphoma or leukemia, but some of the patients had no background condition to established PML pre-exposing medication, as described previously for part of the patients [16]. Based on PCR amplification and short-read sequencing of the ca 260–340 bp long NCCR amplicons, 26 samples from PML patients were interpreted JCPyV positive and eleven, including ten CSF samples and one tissue sample, remained negative

**Table 1**

Samples from PML patients, SLE patients and control individuals; waste-water samples included in the study. CSF, cerebrospinal fluid. FFPE, formalin-fixed, paraffin-embedded tissue. NCCR, non-coding control region. rr-NCCR, rearranged NCCR. qPCR, diagnostic quantitative PCR. N/D, not done. N/A, not applicable. ASV, amplicon sequence variant. \*Positive, below quantitation level.

Sample number	Sample matrix	NCCR type	Result of diagnostic qPCR, copies/mL	References
1	CSF	rr-NCCR	12,600	
2	CSF	Negative	4700	
3	CSF	rr-NCCR	2700	
4	CSF	rr-NCCR	580	
5	CSF	rr-NCCR	290	
6	CSF	rr-NCCR	210	
7	CSF	rr-NCCR	Positive*	
8	CSF	Negative	Positive*	
9	CSF	Negative	Positive*	
10	CSF	Negative	<125 negative	
11	CSF	Negative	<125 negative	
12	CSF	Archetype	290	
13	CSF	Archetype	<125 negative	
14	CSF	Negative	156	
15	CSF	Negative	Positive*	
16	CSF	Negative	Positive*	
17	CSF	Negative	Positive*	
18	CSF	Negative	N/D	
19	CSF	rr-NCCR	937,900	Seppälä et al. 2017
20	CSF	rr-NCCR	Positive*	Seppälä et al. 2017
21	CSF	Archetype-like	18,365	Seppälä et al. 2017
22	fresh/ frozen tissue	rr-NCCR	Positive	
23	fresh/ frozen tissue	rr-NCCR	Positive	
24	fresh/ frozen tissue	rr-NCCR	Positive	
25	fresh/ frozen tissue	rr-NCCR	Positive	
26	fresh/ frozen tissue	rr-NCCR	Positive	
27	fresh/ frozen tissue	rr-NCCR	Positive	
28	fresh/ frozen tissue	Archetype-like	Positive	
29	FFPE	rr-NCCR	N/D	Honkima et al. 2023
30	FFPE	rr-NCCR	N/D	Honkima et al. 2023
31	FFPE	rr-NCCR	N/D	Honkima et al. 2023
32	FFPE	rr-NCCR	N/D	Honkima et al. 2023
33	FFPE	rr-NCCR	N/D	Honkima et al. 2023
34	FFPE	rr-NCCR	N/D	Honkima et al. 2023
35	FFPE	rr-NCCR	N/D	Honkima et al. 2023
36	FFPE	rr-NCCR	N/D	Honkima et al. 2023
37	FFPE	Negative	N/D	Honkima et al. 2023
38	FFPE control	rr-NCCR	N/D	Honkima et al. 2023
39	FFPE control	Negative	N/D	Honkima et al. 2023

**Table 1 (continued)**

Sample number	Sample matrix	NCCR type	Result of diagnostic qPCR, copies/mL	References
40	urine	Archetype and rr-NCCR	N/D	
41	urine	Archetype	N/D	
42	waste-water	Archetype	N/A	
43	waste-water	Archetype	N/A	
44	waste-water	Archetype	N/A	
45	waste-water	Archetype	N/A	
46	waste-water	Archetype	N/A	

(Table 1). Of the two control tissue samples, one was JCPyV positive (sample 38) and one was negative. Both SLE urine samples (samples 40 and 41) and all waste-water samples (samples 42–46) were JCPyV positive. Diagnostic JCPyV qPCR was run for 27 samples (Table 1). Both NCCR PCR and qPCR targeting large T antigen region amplified JCPyV DNA from 10 out of 20 CSF samples and from all seven tissue samples. Of the ten CSF samples remaining negative in NCCR PCR, seven (2, 8, 9, 14–17) were positive in qPCR. One CSF sample remaining negative in qPCR was amplified in NCCR PCR and further subjected to NGS (13). Two samples remained negative in both assays (10, 11).

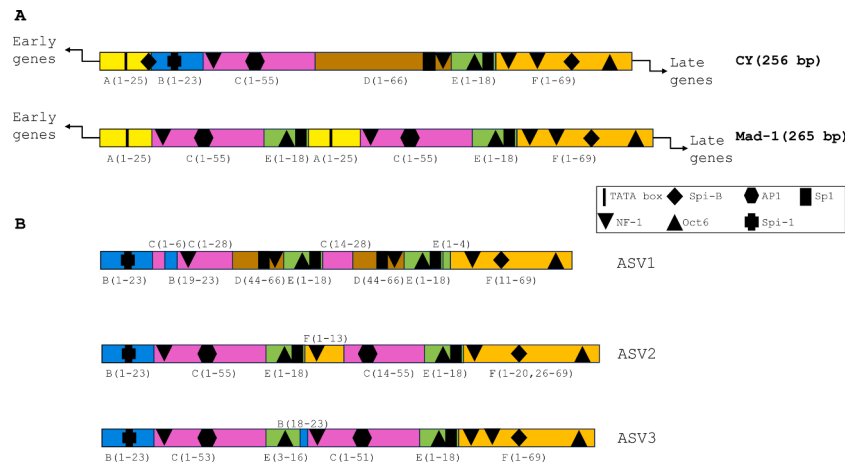
The NCCR region is the most variable region of the JCPyV genome [3] and it is used to characterize viral strains, as exemplified by the different NCCR block structures of archetype CY and rearranged Mad-1 strains (Fig. 1A). The presence of JCPyV strains with archetype, archetype-like and rearranged NCCR in the samples was established by short-read NGS (Table 1). Out of 16 brain tissue samples, rr-NCCR was found in 14 samples and archetype-like NCCR in one sample. In one sample NCCR could not be amplified. Out of 21 CSF samples, rearranged strains were found in eight, archetype-like in one (sample 21) and archetype in two samples (12, 13). From ten CSF samples JCPyV NCCR could not be amplified. Previously reported NCCR analyses using long-read sequencing of CSF samples 7–9 yielded similar results [16].

No association of viral load (qPCR) or features of the NCCR with severity of the disease could be established.

A subset of samples represents two or three consecutive samples from altogether eight patients. From CSF sample pairs of three patients, NCCR could not be amplified from the earlier CSF sample, whereas the later sample (collected 38, 64 and 23 days later, respectively) contained rr-NCCR. CSF samples of five patients remained negative while tissue samples collected 5–91 days later contained rr-NCCR. In one case archetype NCCR was found in the CSF while tissue samples collected 31 days later harbored rr-NCCR. Comparison of the viral strains among those samples is presented separately in Supplementary Data.

We also analyzed two control brain tissue samples with no morphological abnormalities from non-PML individuals who died suddenly. In one control from a 36-year-old male with history of epilepsy, rearranged JCPyV was identified (sample 38). The other control from a 50-year-old male with history of atrial fibrillation and a single seizure episode was JCPyV negative (sample 39). For comparison we also analyzed urine samples from two SLE patients (Table 1). The urine sample from one SLE patient harbored both rr-NCCR and archetype NCCR, rr-NCCR being in majority (sample 40), while the other had only archetype NCCR (sample 41). All waste-water samples (42–46) contained archetype JCPyV as analyzed by long-read sequencing.

Cluster weight analysis of NCCR sequences from individual samples revealed one major NCCR sequence cluster in most samples. Major clusters in each sample were manually analyzed to reveal block structure (Table 2). All main clusters start with intact or minutely mutated A and with intact B, suggesting that they are essential in viral replication.



**Fig. 1.** NCCR structure of JCPyV archetype and rearranged variants. (A) NCCR structure of archetype CY strain with linear arrangement of A, B, C, D, E and F blocks. The number of base pairs in each block is shown in parenthesis. NCCR structure of rearranged Mad-1 strain showing duplications of A, C and E blocks and deletion of B and D blocks. Binding sites for transcription factors TATA box binding protein, Sp1, Spi-B, AP1, NF-1, Oct6 and Spi-1 are shown. (B) Rearranged NCCR organization of the three most common amplicon sequence variants ASV1, ASV2, and ASV3 among our PML patient data (Supplementary Data), together with the number of nucleotides in each block and putative transcription factor binding sites indicated. Of note, the A block was not included in the ASV analysis.

**Table 2**

Block structure of the rearranged NCCR regions identified among the samples. Letters A to F depict NCCR sequence blocks. Block features are marked as follows: A, complete block A; (A) partial block, minimum 5 nucleotides; A\*, A block harboring one single-nucleotide polymorphism (SNP); A\*\*, A block harboring two SNPs; A\*\*\*, A with a larger mutation; (A)\*, most part of A block with one SNP; (A)\*\*, most part of A with two SNPs.

Sample number	Block structure	A	B	C	D	E	F	(B)	(C)	(D)	(E)	(F)					
1	A-B-(C)-(D)-E-(B)-(C)-(D)-E-F*	A	B	(C)	(D)	E		(B)	(C)	(D)	E	F*					
3	A-B-C-(D)-(C)-(D)-E-F*	A	B	C	(D)				(C)	(D)	E	F*					
4	A-B-C-(D)-(C)-(E)-(F)	A	B	C	(D)				(C)		(E)	(F)					
5	A-B-(C)-(B)-(C)-(D)-E-(C)*-(D)-(E)-F*	A	B	(C)				(B)	(C)	(D)	E	(C)	(D)	(E)	F*		
6	A-B-(C)*-E-(F)*-(A)-B-(C)*-E-F*	A	B	(C)*		E	(F)*	(A)	B	(C)*		E	F*				
7	A-B-C*(D)-E-(B)-C*(D)-E-F*	A	B	C*	(D)	E		(B)	C*	(D)	E	F*					
19/1	A-B-C-E-F	A	B	C							E	F					
19/2	A-B-C-E-F-(C)-E-F	A	B	C		E	F		(C)		E	F					
19/3	A-B-C-E-A-B-C-E-F	A	B	C		E	A	B	C		E	F					
20	A-B-C-D-E-(F)*	A	B	(C)	D						E	(F)*					
21	A-B-(C)-E-(B)-(C)-E-(F)*	A	B	(C)		E		(B)	(C)		E	(F)*					
22	A-B-C*-E-(F)-(C)-E-F***	A	B	C*		E	(F)		(C)		E	F***					
23	A-B-(C)-(D)-E-(B)**-(C)-(D)-E-F**	A	B	(C)	(D)	E		(B)	(C)	(D)	E	F**					
24	A-B-C-E-(B)-(C)-E-F*	A	B	C		E		(B)	(C)		E	F*					
25	A-B-(C)-(B)-(C)-(D)-E-(C)*-(D)-(E)-F*	A	B	(C)				(B)	(C)	(D)	E	(C)	(D)	(E)	F*		
26	A-B-C-(D)-E-(F)-(B)-C-(D)-E-F	A	B	C	(D)	E	(F)	(B)	C	(D)	E	F					
27	A-B-C*(D)-(D)-(C)*-D-E-(F)	A	B	C*	(D)					(D)	(C)	D	E	(F)			
28	A-B*(C)**-D-E-(F)*	A	B*	(C)**	D						E	(F)*					
29	A-B-C-(E)-(C)-(E)-(F)-(C)-(E)-(C)-(E)-F	A	B	C		(E)			(C)		(E)	(F)	(C)	(E)	(C)	(E)	F
30	A-B-C*-E-(F)-(C)-E-F***	A	B	C*		E	(F)			(C)	E	F***					
31/1	A-B-(C)-(D)-(E)-(B)-(C)-(D)-E-(F)	A	B	(C)	(D)	(E)		(B)	(C)	(D)	E	(F)					
31/2	A-B-(C)-(D)-E-(B)-(C)-(D)-E-F*	A	B	(C)	(D)	E		(B)	(C)	(D)	E	F*					
32	A-B-(C)-(B)-(C)-(D)-E-(C)-(D)-(E)-F	A	B	(C)				(B)	(C)	(D)	E	(C)	(D)	(E)	F		
33	A*-B-(C)-(E)-(B)-(C)-E-F	A*	B	(C)		(E)		(B)	(C)		E	F					
34	A*-B-(C)-(D)-E-(F)-(C)-F*	A*	B	(C)	(D)	E	(F)		(C)		E	F*					
35	A-B-(C)-(E)-(B)-(C)-E-F	A	B	(C)		(E)		(B)	(C)		E	F					
36	A-B-(C)-(E)-(F)-(C)-(E)-F	A	B	(C)		(E)	(F)		(C)		(E)	F					
38	A-B-(C)-(B)-(C)-(D)-E-(C)*-(D)-(E)-F*	A	B	(C)				(B)	(C)	(D)	E	(C)	(D)	(E)	F*		
40	A-B-C-D-(D)-E-F*	A	B	C	D					(D)	E	F*					

All clusters except one end with complete or partial E and F blocks. The middle region is extremely variable. Concerning rr-NCCR, the C and D blocks are most informative, as almost all clusters have only partial C and D blocks, or the D block is deleted completely. This points to their possible inhibitory function and suggests that enhanced viral replication may be connected to their rearrangements.

For alternative analysis of the NCCR region across samples, DADA2 or ASV analysis was carried out on a set of samples including 31 samples from PML patients (Table 3, Supplementary Table 2). The analysis of short-read sequence reads between 5' and 3' primers used in NCCR PCR revealed altogether 79 sequence variants representing one unique sequence each. The most common deletions and rearrangements affected blocks D and C, whereas B, E and F blocks were often intact or minutely mutated. The A block was not included in ASV sequences. The results of the ASV analysis are described in more detail in Supplementary Data. NCCR block structures of the three most common ASVs (ASV1–3) are shown in Fig. 1B to exemplify the diversity of the NCCR region.

#### 4. Discussion

Reactivation of dormant JCPyV in our body under immunosuppression or immunomodulation may lead to the emergence of PML or other fatal neurological conditions [4,5]. Timely diagnosis of these conditions is challenging, as clinical symptoms may overlap with other neurological diseases such as MS, and the performance of current laboratory tools to detect JCPyV may be suboptimal. In this work we have explored the possibilities to improve diagnostics and risk assessment by engaging NGS.

Our analysis of the JCPyV NCCR regions in brain tissue and CSF of PML patients reinforced previous findings on the frequent identification of predominantly neurotropic strains with PML [5,16,18]. Majority of the deletions and duplications occurred in C, D, E and F blocks, with D block deletion being most frequent. Deletions in D block have frequently been reported among PML patients and in the prototype Mad-1 strain as well [14,19,20]. In a large meta-analysis of 989 NCCR sequences including 32 sequences from brain samples and of those 26 from PML patients it was shown that the number of TFBSs is mostly affected in blocks C, D and F [20]. Deletion of the D block leads to a reduced number of binding sites for SP1, NFI, CEBP $\beta$  or MEIS1, putatively enhancing virus replication and contributing to the development of PML [20,21]. Exploring the presence or emergence of rearranged strains might be done by inspecting the C and D blocks. This could even be done by Sanger sequencing, but the better sensitivity of NGS in identifying multiple strains outperforms Sanger in revealing emerging minor rearranged populations.

However, archetype NCCR alone was found in two CSF samples from PML patients, suggesting that rearranged JCPyV may not be required for PML pathogenesis. Additionally, archetype-like strains were found in one CSF and one tissue. Archetype JCPyV in samples from the CNS has previously been considered a rare phenomenon [18,22]. Nevertheless, the expanding use and rapid development of different NGS technologies and bioinformatics tools may enable more detailed characterization of viral populations. Finding minor rearranged viral populations among major archetype virus may suggest enhanced risk of PML.

DADA2 amplicon sequencing analysis infers sequence variants (ASVs) from paired end short-read data with single nucleotide resolution. DADA2 analysis has been frequently used for 16S analysis of bacterial strains [23]. As opposed to per-sample NCCR block analysis, DADA2 analyzes the composition of NCCR sequences in a whole dataset and gives an overview of the possible rearrangements within a certain sequence. In our case DADA2 analysis identified altogether 79 variants, which is a fairly low number. Together with the small number of variants in any given sample suggests that the generation of rearrangements is not random and each ASV represents a true sequence variant. Similar to the analysis of individual samples, DADA2 revealed that the D and C

blocks are most frequently affected, and any sequence rearrangements within those blocks is indicative of a rearranged strain.

Multiple samples from a subset of patients showed that CSF and tissue may harbor different JCPyV strains. More often, CSF may remain negative while tissue is JCPyV positive. CSF may remain negative in early phases of PML, or if the PML lesion is located deep in brain parenchyma, where little virus is secreted into CSF. In one case CSF contained archetype and tissue collected 31 days later harbored rearranged virus. This can be explained by the time interval, as the emergence of rearranged strains as rapidly as in 19 days has been reported [14]. The role of technical issues such as suboptimal PCR sensitivity, unequal amplification of virus strains prior to NGS or cutoffs applied in bioinformatics tools cannot be fully excluded.

The fairly high negative rate of NCCR amplification among CSF samples from confirmed PML patients altogether suggests that diagnostics should not solely rely on NCCR amplification followed by NGS, at least unless assay sensitivity can be improved, for example by concentrating the CSF samples. The diagnostic qPCR assay shows better sensitivity than the current assay for NCCR amplification.

In agreement with previous literature, all waste-water samples were JCPyV positive and contained solely archetype virus [24,25], reinforcing that archetype virus resides in the environment and circulates in the human population.

Occasional JCPyV DNA findings in the brain of healthy individuals suggests that persistence of JCPyV in the brain cannot be excluded [26–28]. In the present study we also established the presence of rearranged JCPyV in the brain tissue of a non-PML individual. Low JCPyV DNA copy numbers have been detected in CSF of individuals who did not have PML at the time of sampling and did not develop PML in follow-up; viral DNA was reported to originate from virus-carrying blood lymphocytes circulating through the CSF instead of viral persistence in the CNS [29]. Viral DNA may thus be incidentally found in the CNS of individuals without JCPyV associated disease.

We found JCPyV with rr- and archetype NCCR in urine samples SLE of patients. SLE medication has been reported to predispose to PML development, and PML may even be more frequent among patients suffering from SLE than other rheumatic diseases [6,30]. Rearranged JCPyV strains have also been reported in the urine of PML patients due to enhanced viral replication [14].

In conclusion, sensitive PCR from CSF or brain tissue, preferably together with NCCR characterization using NGS, are recommended for detection and differentiation of highly pathogenic JCPyV in neurological patients. The predominance of rearranged viral strains in the CNS supports the use of CSF. Utmost sensitivity of diagnostic PCR or NCCR PCR prior to NGS would be required, and it may be enhanced by concentrating CSF in manifold. Rearranged strains are best defined by deletions or other rearrangements in the D and C blocks. Minor rearranged or archetype-like strains mixed with a major archetype viral population may suggest enhanced risk of emerging disease.

Based on our results we believe that rearranged JCPyV strains in the CNS are associated with and enhance the risk of PML, although archetype strains can be present in PML patients and rearranged strains can be found in individuals without neurological conditions. Identification of emerging rearranged strains may be an indication of enhanced PML risk due to putatively accelerated virus replication in the brain. For this purpose the superb resolution of NGS as compared to traditional Sanger sequencing in differentiating even minor viral populations is indispensable. However, high viral loads are not always present in PML patients, and emerging rearranged viral strains may present low proportions, stressing the need for highly sensitive amplification methods. Early diagnosis of this devastating disease is paramount to improve patient management and prognosis.

#### Funding

This work was supported by a grant from Maire Taponen Foundation

**Table 3**  
 Block structures of the 79 amplicon sequence variants (ASVs). After removal of NCCR primer sequences the structures begin with block B. The amount of sequencing reads containing the ASV sequence is given in the column most to the right. Block features are marked as follows: A, complete block A; (A) partial block, minimum 5 nucleotides; A\*, A block harboring one single-nucleotide polymorphism (SNP); A\*\*, A block harboring two SNPs; A\*\*\*, A with a larger mutation; (A)\*, most part of A block with one SNP; (A)\*\*\*, most part of A with two SNPs.

ASV number															Amount of ASV			
ASV1	B-C-B-C-D-E-C-D-E-F	B	(C)					(B)	(C)	(D)	E		(C)	(D)	E	(F)	322,369	
ASV2	B-C-E-F-C-E-F	B	C*		E	(F)			(C)*							E	(F)**	178,923
ASV3	B-C-E-B-C-E-F	B	(C)		(E)			(B)	(C)							E	F*	165,779
ASV4	B-C-D-E-B-C-D-E-F	B	(C)	(D)*	E			(B)*	(C)	(D)*						E	F	97,986
ASV5	B-C-D-E-F	B	C		D											E	F*	95,005
ASV6	B-C-E-B-C-E-F	B	C		E			(B)	C							E	F*	92,254
ASV7	B-C-D-E-B-C-D-E-F	B	C	(D)	E			(B)	C	(D)						E	F*	71,303
ASV8	B-C-D-C-D-E-F	B	C*	(D)					(C)*	D						E	F	67,083
ASV9	B-C-D-E-F-B-C-D-E-F	B	C	(D)	E	(F)		(B)	C	(D)						E	F*	64,222
ASV10	B-C-D-E-F-C-F	B	(C)	(D)	E	(F)			(C)								F*	58,670
ASV11	B-C-D-E-C-D-E-F-C-D-E-C-D-E-F	B	C	(D)	(E)				(C)	(D)	(E)	(F)		(C)	(D)	(E)	(F)	56,298
ASV12	B-C-E-F-C-E-F	B	(C)		(E)	(F)*			(C)	(C)						(E)	F*	33,879
ASV13	B-C-D-C-D-D-E-F	B	C	(D)					(C)	(D)						E	F*	11,544
ASV14	B-C-D-E-B-C-D-E-F	B	C	(D)	E			(B)	(C)	(D)						E	F*	7028
ASV15	B-C-D-E-F	B	(C)	D												E	F**	5703
ASV16	B-C-D-E-B-C-D-E-F	B	(C)	(D)	E			(B)	(C)	(D)						E	(F)	4022
ASV17	B-C-E-F-A-F-B-C-E-F	B	(C)		E	(F)*	(A)					(F)	B	(C)		E	F*	3418
ASV18	B-C-E-F-B-C-E-F	B	C*		E	(F)*		B	C*							E	F**	2440
ASV19	B-C-D-E-F-B-C-D-E-F	B	C	(D)	E	(F)*		(B)	C	(D)						E	(F)	1216
ASV20	B-C-B-C-D-E-C-D-E-F	B	(C)					(B)	(C)	(D)	E			(C)	(D)		(F)*	1048
ASV21	B-C-E-F-E-F	B	(C)		(E)	(F)										(E)	F*	653
ASV22	B-C-D-E-B-C-D-E-F	B	(C)	(D)	E			(B)*	(C)	D*						E	F	574
ASV23	B-C-D-E-B-C-D-E-F	B	(C)	(D)	E			(B)*	(C)	(D)						E	(F)	550
ASV24	B-C-E-E-B-C-E-E-F	B	(C)		(E)						E	(B)		(C)		(E)	F	546
ASV25	B-C-C-D-D-E-F	B*	(C)						(C)	D						E	F*	410
ASV26	B-C-D-E-F-B-C-D-E-F	B	(C)	(D)	E	(F)		(B)	(C)	(D)						E	(F)	368
ASV27	B-C-E-B-C-E-F	B	C		E			(B)	C							E	(F)	328
ASV28	B-C-D-E-F	B	C	D												E	F	190
ASV29	B-C-E-F-C-E-F	B	(C)		E	(F)			(C)*							E	(F)**	178
ASV30	B-C-D-E-C-D-E-F-C-D-E-C-D-E-F	B	C	(D)	(E)				(C)	(D)	(E)	(F)		(C)	(D)	(E)	(F)	154
ASV31	B-C-D-E-B-C-D-D-E-F	B	(C)	(D)	E			(B)	(C)*	(D)						E	F*	138
ASV32	C-D-E-B-C-D-D-E-F		(C)	(D)	E			(B)	(C)	(D)						E	F**	128
ASV33	B-C-D-C-D-F	B	C*	(D)					(C)*	(D)							(F)	106
ASV34	B-C-B-C-D-E-C-D-E-C-D-E-F	B	(C)					(B)	(C)	(D)	E			(C)	(D)	E	(F)	100
ASV35	B-C-E-B-C-E-F	B	(C)		(E)			(B)	(C)							E	F*	100
ASV36	B-C-D-E-C-D-E-F-C-D-E-C-D-E-F	B	C	(D)	(E)				(C)	(D)	(E)	(F)		(C)	(D)	(E)	(F)	90
ASV37	B-C-D-E-F	(B)	(C)	D												E	F	86
ASV38	B-C-E-F-C-E-F	B	C		E	(F)			(C)							E	(F)	76
ASV39	B-C-D-E-F	(B)	(C)**	D												E	(F)*	74
ASV40	B-C-E-F-C-E-F	B	(C)		(E)	(F)***			(C)							(E)	F*	72
ASV41	B-C-B-C-D-E-C-D-E-F	B	(C)					(B)	(C)	(D)	E			(C)*	(D)	(E)	F*	71
ASV42	B-C-E-B-C-E-F	B	(C)		E			(B)	C							E	F*	71
ASV43	B-C-D-C-D-E-F	B	C*	(D)					(C)	(D)						E	F	40
ASV44	B-C-D-E-F	B	C	D												E	F*	39
ASV45	B-C-D-C-D-E-F	B	(C)	(D)					C*	D						E	(F)	30
ASV46	B-C-D-C-D-E-F	B	C	(D)					(C)	(D)						E	F*	30
ASV47	B-C-E-F-C-D-E-F	B	(C)		(E)	(F)			(C)	(D)						(E)	F	26
ASV48	B-C-D-E-C-D-E-C-D-E-C-D-E-F	B	C	(D)	(E)				(C)	(D)	(E)			(C)	(D)	(E)	F	26
ASV49	B-C-D-E-F	(B)	C	D												E	F	26
ASV50	B-C-D-E-C-D-E-F-C-F	B	C	(D)	(E)				(C)	(D)	(E)	(F)		(C)			F*	25
ASV51	B-C-D-C-E-F	B	C	(D)					(C)							(E)	(F)	25
ASV52	B-C-D-E-F	B	C	D												E	F**	19

(continued on next page)



- [7] M. Kartau, E. Auvinen, A. Verkkoniemi-Ahola, L. Mannonen, I. Helanterä, V. J. Anttila, JC polyomavirus DNA detection in clinical practice, *J. Clin. Virol.* 146 (2022) 105051, <https://doi.org/10.1016/j.jcv.2021.105051>.
- [8] Y.Y. Chang, M.Y. Lan, C.H. Peng, H.S. Wu, D. Chang, J.S. Liu, Progressive multifocal leukoencephalopathy in an immunocompetent Taiwanese patient, *J. Formos. Med. Assoc.* 106 (2 Suppl) (2007) S60–S64.
- [9] A. Honkima, J. Suppala, O. Tynnen, M. Saarela, H. Liimatainen, P. Laine, P. Auvinen, E. Auvinen, JC polyomavirus modifies the expression of human microRNAs in PML brain, *J. Infect. Dis.* 228 (2023) 829–833.
- [10] H. Seppälä, I. Helanterä, P. Laine, I. Lautenschlager, L. Paulin, T. Jahnukainen, P. Auvinen, E. Auvinen, Archetype JC polyomavirus prevails in a rare case of JC polyomavirus nephropathy and in stable renal transplant recipients with JC polyomavirus viraemia, *J. Infect. Dis.* 216 (2017) 981–989.
- [11] N. Ahye, A. Bellizzi, D. May, H.S. Wollebo, The role of the JC virus in central nervous system tumorigenesis, *Int. J. Mol. Sci.* 21 (17) (2020) 6236, <https://doi.org/10.3390/ijms21176236>.
- [12] Y. Yogo, T. Kitamura, C. Sugimoto, T. Ueki, Y. Aso, K. Hara, F. Taguchi, Isolation of a possible archetypal JC virus DNA sequence from nonimmunocompromised individuals, *J. Virol.* 64 (1990) 3139–3143.
- [13] G.S. Ault, G.L. Stoner, Human polyomavirus JC promoter/enhancer rearrangement patterns from progressive multifocal leukoencephalopathy brain are unique derivatives of a single archetypal structure, *J. Gen. Virol.* 74 (1993) 1499–1507.
- [14] A.S. L'Honneur, J. Pipoli Da Fonseca, T. Cokelaer, F. Rozenberg, JC polyomavirus whole genome sequencing at the single-molecule level reveals emerging neurotropic populations in progressive multifocal leukoencephalopathy, *J. Infect. Dis.* 226 (2022) 1151–1161.
- [15] C.E. Reid, H. Li, G. Sur, P. Carmillo, S. Bushnell, R. Tizard, M. McAuliffe, C. Tonkin, K. Simon, S. Goelz, P. Cinque, L. Gorelik, J.P. Carulli, Sequencing and analysis of JC virus DNA from natalizumab-treated PML patients, *J. Infect. Dis.* 204 (2011) 237–244.
- [16] H. Seppälä, E. Virtanen, M. Saarela, P. Laine, L. Paulin, L. Mannonen, P. Auvinen, E. Auvinen, Single-molecule sequencing reveals the presence of distinct JC polyomavirus populations in patients with progressive multifocal leukoencephalopathy, *J. Infect. Dis.* 215 (2017) 889–895.
- [17] C.S. Tan, L.C. Ellis, C. Wüthrich, L. Ngo, T.A. Broge Jr, J. Saint-Aubyn, J.S. Miller, I. J. Koralnik, JC virus latency in the rain and extraneural organs of patients with and without progressive multifocal leukoencephalopathy, *J. Virol.* 84 (2010) 9200–9209.
- [18] T. Van Loy, K. Thys, C. Ryschkewitsch, O. Lagatie, M.C. Monaco, E.O. Major, L. Tritsmans, L.J. Stuyver, JC virus quasispecies analysis reveals a complex viral population underlying progressive multifocal leukoencephalopathy and supports viral dissemination via the hematogenous route, *J. Virol.* 89 (2015) 1340–1347.
- [19] M.R. Ciardi, M.A. Zingaropoli, M. Iannetta, C. Prezioso, V. Perri, P. Pasculli, M. Lichtner, G. d'Etto, M. Altieri, A. Conte, V. Pietropaolo, C.M. Mastroianni, V. Vullo, JCPyV NCCR analysis in PML patients with different risk factors: exploring common rearrangements as essential changes for neuropathogenesis, *Virol. J.* 17 (2020) 23, <https://doi.org/10.1186/s12985-020-1295-5>.
- [20] M.P. Wilczek, A.M.C. Pike, S.E. Craig, M.S. Maginnis, B.L. King, Rearrangement in the hypervariable region of JC polyomavirus genomes isolated from patient samples and impact on transcription factor-binding sites and disease outcomes, *Int. J. Mol. Sci.* 2022 (23) (2022) 5699, <https://doi.org/10.3390/ijms23105699>.
- [21] J. Kim, S. Woolridge, R. Biffi, E. Borghi, A. Lassak, P. Ferrante, S. Amini, K. Khalili, M. Safak, Members of the AP-1 family, c-Jun and c-Fos, functionally interact with JC virus early regulatory protein large T antigen, *J. Virol.* 77 (2003) 5241–5252.
- [22] M.I. Domínguez-Mozo, M. García-Montojo, A. Atías-Leal, Á. García-Martínez, J. L. Santiago, I. Casanova, V. Galán, R. Arroyo, M. Fernández-Arquero, R. Alvarez-Lafuente, Monitoring the John Cunningham virus throughout natalizumab treatment in multiple sclerosis patients, *Eur. J. Neurol.* 23 (2016) 182–189.
- [23] B.J. Callahan, P.J. McMurdie, M.J. Rosen, A.W. Han, A.J.A. Johnson, S.P. Holmes, DADA2: high-resolution sample inference from Illumina amplicon data, *Nat. Methods* 13 (2016) 581–583.
- [24] S. Bofill-Mas, S. Pina, R. Girones, Documenting the epidemiologic patterns of polyomaviruses in human populations by studying their presence in urban sewage, *Appl. Environ. Microbiol.* 66 (2000) 238–245.
- [25] M. Iaconelli, S. Petricca, S.D. Libera, P. Di Bonito, G. La Rosa, First detection of human papillomaviruses and human polyomaviruses in river waters in Italy, *Food Environ. Virol.* 7 (2015) 309–315.
- [26] C. Elsner, K. Dörries, Evidence of human polyomavirus BK and JC infection in normal brain tissue, *Virology* 191 (1992) 72–80.
- [27] L. Vago, P. Cinque, E. Sala, M. Nebuloni, R. Caldarelli, S. Racca, P. Ferrante, G. Trabottoni, G. Costanzi, JCV-DNA and BKV-DNA in the CNS tissue and CSF of AIDS patients and normal subjects. Study of 41 cases and review of the literature, *J. Acquir. Immune Defic. Syndr. Hum. Retrovirol.* 12 (1996) 139–146.
- [28] F.A. White 3rd, M. Ishaq, G.L. Stoner, R.J. Frisque, JC virus DNA is present in many human brain samples from patients without progressive multifocal leukoencephalopathy, *J. Virol.* 66 (1992) 5726–5734.
- [29] E. Iacobaeus, C. Ryschkewitsch, M. Gravel, M. Khademi, E. Wallstrom, T. Olsson, L. Brundin, E.O. Major, Analysis of cerebrospinal fluid and cerebrospinal fluid cells from patients with multiple sclerosis for detection of JC virus DNA, *Mult. Scler.* 15 (2009) 28–35.
- [30] E.S. Molloy, L.H. Calabrese, Progressive multifocal leukoencephalopathy. A national estimate of frequency in systemic lupus erythematosus and other rheumatic diseases, *Arthritis Rheum.* 60 (2009) 3761–3765.