

Short communication

Next-generation sequencing (NGS) in the identification of encephalitis-causing viruses: Unexpected detection of human herpesvirus 1 while searching for RNA pathogens



Karol Perlejewski^a, Marta Popiel^{a,b}, Tomasz Laskus^a, Shota Nakamura^c, Daisuke Motooka^c, Tomasz Stokowy^d, Dariusz Lipowski^e, Agnieszka Pollak^f, Urszula Lechowicz^f, Kamila Caraballo Cortés^a, Adam Stępień^g, Marek Radkowski^a, Iwona Bukowska-Ośko^{a,*}

^a Department of Immunopathology of Infectious and Parasitic Diseases, Warsaw Medical University, 3C Pawinskiego Street, 02-106 Warsaw, Poland

^b Postgraduate School of Molecular Medicine, 61 Żwirki i Wigury Street, 02-091 Warsaw, Poland

^c Department of Infection Metagenomics, Genome Information Research Center, Research Institute for Microbial Diseases, Osaka University 3-1 Yamadaoka, Suita-City, Osaka, Japan

^d Department of Clinical Science, University of Bergen, 5021 Bergen, Norway

^e Municipal Hospital for Infectious Diseases, 37 Wolska Street, 01-201 Warsaw, Poland

^f Department of Genetics, Institute of Physiology and Pathology of Hearing, Mochnickiego 10, 02-042 Warsaw, Poland

^g Department of Neurology, Military Institute of Medicine, 128 Szaserów Street, 04-141 Warsaw, Poland

ABSTRACT

Article history:

Received 18 May 2015

Received in revised form

24 September 2015

Accepted 24 September 2015

Available online 28 September 2015

Keywords:

Encephalitis

Next-generation sequencing

Unknown etiology

Background: Encephalitis is a severe neurological syndrome usually caused by viruses. Despite significant progress in diagnostic techniques, the causative agent remains unidentified in the majority of cases. The aim of the present study was to test an alternative approach for the detection of putative pathogens in encephalitis using next-generation sequencing (NGS).

Methods: RNA was extracted from cerebrospinal fluid (CSF) from a 60-year-old male patient with encephalitis and subjected to isothermal linear nucleic acid amplification (Ribo-SPIA, NuGen) followed by next-generation sequencing using MiSeq (Illumina) system and metagenomics data analysis.

Results: The sequencing run yielded 1,578,856 reads overall and 2579 reads matched human herpesvirus 1 (HHV-1) genome; the presence of this pathogen in CSF was confirmed by specific PCR. In subsequent experiments we found that the DNase I treatment, while lowering the background of host-derived sequences, lowered the number of detectable HHV-1 sequences by a factor of 4. Furthermore, we found that the routine extraction of total RNA by the Chomczynski method could be used for identification of both DNA and RNA pathogens in typical clinical settings, as it results in retention of a significant amount of DNA.

Conclusion: In summary, it seems that NGS preceded by nucleic acid amplification could supplement currently used diagnostic methods in encephalitis.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Abbreviations: NGS, next-generation sequencing; CSF, cerebrospinal fluid; HHV-1, human herpesvirus 1.

* Corresponding author.

E-mail addresses: karol.perlejewski@wum.edu.pl

(K. Perlejewski), mszabi@wp.pl (M. Popiel), t.laskus@yahoo.com

(T. Laskus), nshota@gen-info.osaka-u.ac.jp (S. Nakamura), daisukem@gen-info.osaka-u.ac.jp (D. Motooka), tomasz.stokowy@k2.uib.no

(T. Stokowy), dariusz.lipowski@wum.edu.pl (D. Lipowski), a.pollak@ifps.org.pl

(A. Pollak), u.lechowicz@ifps.org.pl (U. Lechowicz), kcaraballo@wum.edu.pl (K. Caraballo Cortés), astepien@wim.mil.pl

(A. Stępień), marek.radkowski@wum.edu.pl (M. Radkowski), ibukowska@wum.edu.pl (I. Bukowska-Ośko).

Encephalitis is a severe neurological syndrome, which may be caused by viruses as well as bacteria, fungi and parasites (Chaudhuri and Kennedy, 2002; Glaser et al., 2006; Solomon et al., 2007). Diagnostic procedures include serological testing for specific antibodies in blood and cerebrospinal fluid (CSF), detection of causative pathogens by molecular methods in CSF and blood, and occasionally even pathogen culture (Solomon et al., 2007). Correct diagnosis is facilitated by epidemiological data and clinical symptoms, routine CSF analysis and neuroimaging studies (Chaudhuri and Kennedy, 2002; Solomon et al., 2007). Surprisingly, in a large proportion

of cases the exact etiology of encephalitis remains unknown. For example, The Californian Encephalitis Project, which included 1570 patients, failed to identify cause of encephalitis in 63% of patients (Glaser et al., 2006), while in a French study the etiology of encephalitis remained unknown in 80% of cases (Mailles et al., 2007). However, identification of causative agents is crucial for the correct diagnosis and may be necessary for the timely introduction of specific treatment (Studahl et al., 2013).

The high rate of cases with unclear etiology is partly due to the limited sensitivity and specificity of currently used diagnostic assays, but also to the suboptimal time of sample collection, since viral pathogens may be detectable only transiently (Glaser et al., 2006; Studahl et al., 2013). Moreover, a large and still expanding number of pathogens reported to be capable of causing encephalitis makes specific diagnosis challenging due to the sheer scope of required testing (Kennedy, 2005). A recently developed next-generation sequencing (NGS) technique offers a potential breakthrough in the diagnostic process of encephalitis by eliminating the need for testing of individual pathogens. High-throughput sequencing has already established a firm foothold in the area of microbiology as it has been successfully employed for pathogen detection (including novel microorganisms), analysis of genetic evolution and diversity, as well as for the determination of drug resistance (Radford et al., 2012). The aim of the present study was to determine the value of NGS in detecting putative causative agent(s) in CSF in encephalitis.

For this purpose CSF from 60-year-old male with suspected encephalitis, admitted to the hospital with a 1-week history of fever, drowsiness and severe headaches was used. These symptoms were preceded by a mild cold. At the time of admission the patient was comatose with positive meningeal signs and his temperature was 38.5 °C. Analysis of CSF revealed 480 leukocytes/mm³ (90% lymphocytes), protein level of 3.21 g/l (normal value: 0.12–0.60 g/l), and glucose value of 2.32 mmol/l (normal value: 2.20–4.40 mmol/l). The peripheral blood leukocyte count was 6900/mm³ with 66.7% neutrophils and 22.2% lymphocytes. Serum and CSF samples collected at admission were negative for the presence of IgM and IgG antibodies to human herpesvirus (NovaLisa HSV 1/HSV 2/HSV 1+2 IgG/IgM ELISA; NovaTec Immundiagnostica GmbH, Dietzenbach, Germany). However, both anti-HHV-1 IgM and anti-HHV-1 IgG were detected in subsequent CSF samples collected at 1.5, 3 and 5 months. Magnetic resonance imaging (MRI) at admission showed hyperintensity in the right temporal lobe. Patient was treated with Acyclovir (Zovirax IV) for 15 days and fully recovered with no residual neurological sequelae. PCR for HSV-1/2 was not performed at admission, as it was not a part of routine diagnostics in Poland at that time, and our analysis was done retrospectively.

Total RNA was isolated from 1 ml of CSF collected at admission using the modified Chomczynski method (TRIZOL LS Reagent; Life Technologies, Carlsbad, USA) and split into two aliquots, one of which (sample SDNase) was treated with DNase I (DNA-free DNA Removal Kit; Invitrogen, Waltham, USA) at a concentration of 2 U/μl for 20 min in 37 °C in a final volume of 100 μl to reduce the amount of DNA present. The other aliquot remained DNase untreated (sample S1). Control samples included HCV-RNA-positive serum sample adjusted to contain 100 viral copies per reaction (sample C-HCV) and three HHV-1-negative CSF samples (samples C1–C3; Table 1).

The quantity of RNA extracted from CSF was small and was below the sensitivity limits of DeNovix DS-11 Spectrophotometer (DeNovix, Wilmington, USA; sensitivity ≥2 ng/μl) and Qubit Fluorimeter (Thermo Scientific, Waltham, USA; sensitivity ≥250 pg/μl).

Both CSF samples (S1 and SDNase) and all control samples underwent reverse transcription followed by single-primer isothermal amplification (Ribo-SPIA), using commercially available kit (Ovation RNA-Seq V2 system (NuGEN, San Carlos, USA). The

protocol closely followed manufacturer's recommendations. The resulting amplified ds cDNA was purified using Agencourt AMPure XP beads (Beckman Coulter, Pasadena, USA).

Libraries for NGS were prepared from 1 ng of ds cDNA using Nextera XT Kit (Illumina, San Diego, USA) following manufacturer's protocol. First, cDNA was fragmented using transposon-based method followed by addition of indexes by PCR. The resulting PCR products were purified with 1.8 volumes of AMPure XP beads (Beckman Coulter, Pasadena, USA). The quality and average length of sequence library for each sample was assessed using Bioanalyzer (Agilent Technologies, Santa Clara, USA) and either the DNA 1000 or DNA HS kit, depending on sample concentration. The indexed samples were pooled equimolarily and sequenced on Illumina MiSeq (150 nt, paired-end reads) in case of samples S1, SDNase, C-HCV and on Illumina HiSeq (101 nt, paired-end reads); in case of samples C1–C3.

Raw reads were trimmed by the following procedures: (i) removing adaptor sequences using cutadapt-1.2.1 (Martin, 2011), (ii) removing artifact sequences using fastx.artifacts.filter (http://hannonlab.cshl.edu/fastx_toolkit/index.html, 2015), and (iii) trimming bases with quality below Q20 (phred quality score) from 3' end of each read and removing reads shorter than 50 bp using fastq_quality_trimmer (http://hannonlab.cshl.edu/fastx_toolkit/index.html, 2015). Trimmed sequences were then mapped onto the human reference sequence (hg19) with the mapping software Stampy (Lunter and Goodson, 2011). The unmapped sequences were subjected to similarity search using the program blastn (Altschul et al., 1990) against the whole unfiltered NCBI-nt database with e-value cutoff of 1e–5. The taxonomic information of each sequence was assigned and the relative abundance of detected organisms was summarized by text mining of blastn output files using BioRuby (Goto et al., 2010) scripts.

The initial sequencing run included samples S1, SDNase, C-HCV and yielded 4,529,322 reads altogether. After trimming, there were 1,190,455 reads for sample S1 (no DNase treatment) and 1,030,284 reads for sample SDNase (treated with DNase I). Sequencing of Hepatitis C Virus (HCV) positive control (sample C-HCV) provided 1,184,251 reads (Table 1). Human sequences were the most abundant (43.913–99.506% of all sequences; Table 1). Viral sequences represented 0.07–0.246% of reads and the most frequent among them corresponded to HHV-1 (99.8% of all viral sequences in sample S1 and 86.6% in sample SDNase). In the control HCV RNA-positive sample 99.8% of viral reads were positively aligned to HCV genome (Table 2). Fig. 1 shows the coverage plot of detected HHV-1 and HCV sequences with respect to reference genomes. Importantly, the DNase I treatment, while lowering the background of host-derived sequences from 98.2% to 43.9%, at the same time lowered the number of detectable HHV-1 sequences by a factor of 4 (from 2579 to 629 sequences).

Control CSF samples C1–C3 were analyzed in a separate run yielding a total of 107,219,264 reads. After trimming, there were 37,516,857 reads for sample C1, 35,638,335 reads for sample C2 and 34,064,072 reads for sample C3. The metagenomic approach did not detect any HSV-1 sequences in any of these three samples. The results of NGS sequencing and the most frequently identified species/families are presented in Tables 1 and 2.

The presence of HHV-1 in CSF of the analyzed patient was confirmed by specific PCR (Fig. 2) performed as follows. Human herpesvirus 1 (HHV-1) DNA was isolated from cerebrospinal fluid (200 μl) using QIAamp DNA Mini Kit (Qiagen, Venlo, Netherlands), then eluted in 5 μl of water (Water Molecular Biology Reagen; Sigma-Aldrich, Saint Louis, USA) and the HHV specific PCR was performed using primers specific for both HHV-1 and HHV-2 (5'-TTGAAGCGTCGGCGCTA-3' and 5'-GCATCGTCGAGGTGGA-3') resulting in 148 bp amplification product (54836 nt–54983 nt; [GenBank:NC_001806.1]). Extracted

Table 1

Next-generation sequencing (NGS) of CSF from patient with HHV-1 encephalitis and from HHV-1-negative controls. Sequences were compared to the NCBI-nt database.

Age/gender	60/male		56/female	20/female	43/female	
Diagnosis	HHV-1 encephalitis		Encephalitis of unknown etiology	Encephalitis of unknown etiology	Multiple sclerosis	
Sample	S1	SDNase	C1	C2	C3	C-HCV
Total reads	1,578,856	1,427,356	37,516,857	35,638,335	34,064,072	1,523,110
Reads after trimming	1,190,455	1,030,284	375,168,57	35,638,335	34,064,072	1,184,251
Human	1,169,244 (98.218%)	452,428 (43.913%)	36,061,952 (96.122%)	35,374,962 (99.261%)	32,616,258 (95.750%)	1,178,397 (99.506%)
Viral	2583 (0.217%)	726 (0.070%)	52,226 (0.139%)	71,493 (0.201%)	140,890 (0.414%)	2916 (0.246%)
Bacterial	5112 (0.429%)	242,633 (23.550%)	830,811 (2.215%)	120,107 (0.337%)	668,362 (1.962%)	1142 (0.096%)
Fungal	251 (0.021%)	2591 (0.251%)	45,550 (0.121%)	3190 (0.009%)	78,112 (0.229%)	32 (0.003%)
Protozoan	22 (0.002%)	795 (0.077%)	6386 (0.017%)	1100 (0.003%)	16,298 (0.048%)	13 (0.001%)
Other ^a	12,018 (1.010%)	315,615 (30.634%)	428,866 (1.143%)	36,106 (0.104%)	468,183 (1.374%)	1015 (0.086%)
No match	1225 (0.103%)	15,496 (1.504%)	91,066 (0.243%)	31,377 (0.088%)	75,969 (0.223%)	736 (0.062%)

S1, no DNase I treatment; SDNase, treated with DNase I; C, HCV-RNA positive serum.

Samples C1–C3 were negative for HHV-1 by PCR.

^a Sequences related to plants, plant viruses, synthetic constructs.

Table 2

The most frequently identified species/families in cerebrospinal fluid from a patient with HHV-1 encephalitis and controls.

	Viruses	Bacteria ^a	Fungi ^a	Protozoa
S1	Human herpesvirus 1 (2579) Influenza C virus (3) Hepatitis C virus (1)	Moraxellaceae (871) Micrococcaceae (554) Corynebacteriaceae (321) Propionibacteriaceae (204) Gordoniaceae (206)	<i>Pyrenophora teres</i> (151) <i>Pyrenophora tritici-repentis</i> (29) <i>Malassezia sympodialis</i> (14) <i>Malassezia globosa</i> (10) <i>Debaryomyces hansenii</i> (7)	–
SDNase	Human herpesvirus 1 (629) Influenza C virus (55) Hepatitis C virus (39)	Moraxellaceae (37,130) Staphylococcaceae (35,753) Micrococcaceae (48,011) Corynebacteriaceae (19,993) Neisseriaceae (8240)	Uncultured fungus (519) Uncultured Cryptococcus (206) Uncultured Rhodotorula (180) <i>Malassezia sympodialis</i> (142) <i>Malassezia globosa</i> (147)	<i>Albugo laibachii</i> (441) <i>Besnoitia besnoiti</i> (89) <i>Trypanoplasma borreli</i> (83) <i>Ichthyophthirius multifiliis</i> (75)
C1	–	Corynebacteriaceae (2,75,825) Staphylococcaceae (93,611) Micrococcaceae (73,191) Moraxellaceae (51,378) Sphingobacteriaceae (29,855)	Uncultured fungus (10,062) <i>Pseudogymnoascus destructans</i> (1199) <i>Melampsora larici-populina</i> (995) <i>Leucosporidium</i> sp. AY30 (787) Uncultured soil fungus (786)	<i>Nannochloropsis oceanica</i> (1549) <i>Albugo laibachii</i> (401) Amphifilidae sp. H-1 (373) <i>Babesia divergens</i> (371) <i>Exocolpoda augustini</i> (324)
C2	–	Burkholderiaceae (40,597) Sphingobacteriaceae (30,089) Streptococcaceae (5702) Bradyrhizobiaceae (4649) Enterobacteriaceae (4546)	Uncultured fungus (551) <i>Geotrichum klebahnii</i> (250) <i>Malassezia globosa</i> (225) <i>Dipodascus tetrasporaeus</i> (132) <i>Mortierella alpina</i> (82)	<i>Nannochloropsis oceanica</i> (3204) <i>Plasmodium ovale</i> (1201) <i>Albugo laibachii</i> (964) <i>Babesia divergens</i> (549) <i>Plasmodium inui</i> (312)
C3	–	Moraxellaceae (94,524) Enterobacteriaceae (67,339) Corynebacteriaceae (59,747) Micrococcaceae (48,846) Bradyrhizobiaceae (47,559)	Uncultured fungus (12,841) <i>Triplosporium cycadicola</i> (4346) <i>Mollisina uncinata</i> (3997) <i>Pseudogymnoascus destructans</i> (2790) <i>Melampsora larici-populina</i> (2465)	<i>Nannochloropsis oceanica</i> (2468) <i>Albugo laibachii</i> (2274) <i>Plasmodium inui</i> (2075) <i>Eunotia</i> sp. ECR-2014 (1780) <i>Plasmodium vincke</i> (588)
C-HCV	Hepatitis C virus (2905) Influenza C virus (3)	Moraxellaceae (197) Micrococcaceae (166) Propionibacteriaceae (67) Streptococcaceae (66) Comamonadaceae (62)	<i>Debaryomyces hansenii</i> (19) <i>Staninwardia suttonii</i> (2) <i>Malassezia sympodialis</i> (2) <i>Leptosphaeria maculans</i> (2) <i>Entyloma calendulae</i> (2)	–

S1, no DNase I treatment; SDNase, treated with DNase I; C-HCV, HCV-RNA positive serum; C1 and C2, CSF of patients with encephalitis of unknown etiology; C3, CSF of patient with multiple sclerosis.

Samples C1–C3 were negative for HHV-1 by PCR.

The numbers of sequences representing each species/family are shown in brackets.

^a 5 most numerous families.

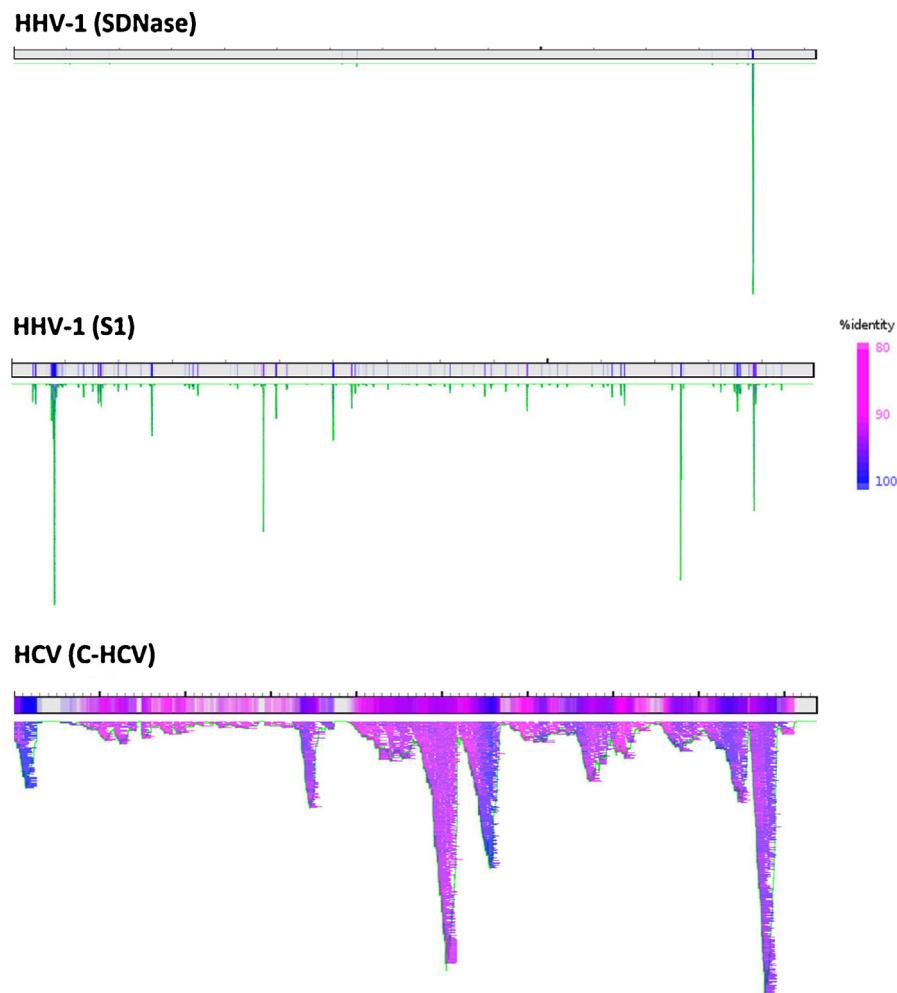


Fig. 1. Coverage plot of HHV-1 and HCV genomes. Sequencing reads from samples S1 and SDNase were mapped to the reference genome of HHV-1 (GenBank accession number AB618031) and control sample C-HCV was mapped to HCV genome (GenBank accession number EF407475).

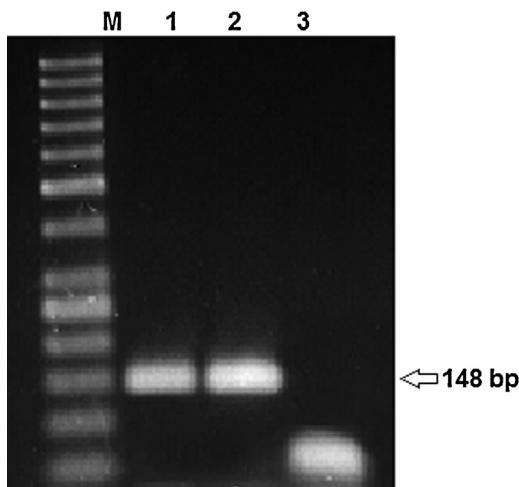


Fig. 2. PCR detection of human herpesvirus type 1 (HHV-1). Lane: 1, CSF sample drawn at admission; 2, positive control (400 target copies of HHV-1 DNA from HHV-1 infected Vero cells 3, negative control consisting of water; M, GeneRuler 50 bp DNA Ladder (Feremitas, Thermo Scientific, Waltham, USA; 50 bp). Expected product size was 148 bp.

DNA (38 μ l) was mixed with 5 μ l of 10 \times buffer (Life Technologies, Carlsbad, USA), 2 μ l of PCR primers (10 μ M), 2 μ l of 50 \times dNTP (10 mM), 5 μ l of 25 mM MgCl₂ (Life Technologies, Carlsbad, USA) and 2.5 U of Taq Polymerase (Life Technologies, Carlsbad, USA) in

final volume of 50 μ l. After the initial 5 min denaturation in 94 °C, 50 cycles of 94 °C for 30 s and 68 °C for 30 s were run using thermocycler 9700 Applied Biosystem (Applied Biosystem, Life Technologies, Carlsbad, USA). PCR products were analyzed by electrophoresis in 2% agarose. Detection limit of our PCR was \geq 44 viral copies per reaction. PCR sensitivity was determined using serial dilutions of DNA extracted from HHV-1 infected Vero cell culture (Virus Isolation and Reference Service Laboratory, New York State Health Department).

In the present study we unexpectedly detected HHV-1, which is a DNA virus, while analyzing CSF sample subjected to RNA extraction by the most widely used Chomczynski method followed by RNA amplification. This finding could have been the result of viral mRNA detection, or alternatively, a substantial quantity of DNA could have been retained after RNA extraction, thus allowing for the detection of both RNA and DNA sequences. To verify the latter possibility, serial 10-fold dilutions of plasmid template (pCR2.1-TOPO, Invitrogen, Waltham, USA) were spiked into CSF sample from an uninfected patient and subjected to parallel RNA and DNA extractions (TRIZOL LS Reagent, Life Technologies and QIAamp DNA Mini Kit, Qiagen, Venlo, Netherlands). Genetic material extracted from each sample was subjected to real-time PCR (10 min denaturation in 95 °C followed by 55 cycles of 95 °C for 10 s and 50 °C for 15 s) using LightCycler FastStart DNA Master SYBR Green I (Roche Molecular Diagnostics, Basel, Switzerland) and primers (5'-GTAAACGACGGCCAG-3'; 5'-CAGGAAACAGCTATGAC-3') resulting in 202 bp amplification product. As seen in Fig. 3, RNA extraction retained a significant amount of DNA which was sufficient for

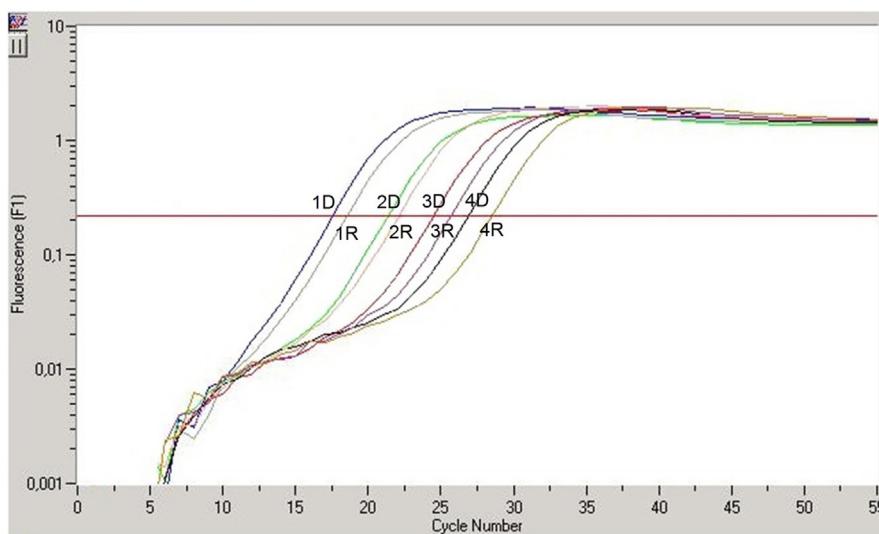


Fig. 3. Real-time PCR detection of serial 10-fold dilutions of plasmid template (pCR2.1-TOPO; Invitrogen, Waltham, USA) spiked into CSF from an uninfected patient. Samples were either extracted for DNA (D), or RNA (R). Samples 1, 2, 3, 4 contained 10^3 , 10^4 , 10^5 and 10^6 plasmid copies per reaction, respectively.

real-time PCR detection of $\geq 10^3$ plasmid copies. Surprisingly, for a wide range of plasmid concentrations tested (10^3 , 10^4 , 10^5 and 10^6 per reaction) the RNA extraction resulted in only 1.6–3.1 fold sensitivity decrease when compared to DNA extraction.

Determining the etiology of encephalitis remains challenging in clinical practice, partly due to the sheer number of potential pathogens. So far over 100 viruses were reported to be capable of causing encephalitis, and this number is still growing (Debiasi and Tyler, 2004). The results of some recent studies indicate that high-throughput sequencing could be successfully used to identify causative pathogens in patients with central nervous system infections. For example, Chan et al. (2014) identified measles virus and HHV-1 in frozen brain tissue samples using Illumina HiSeq 2000, whereas Tan le et al. (2013) used FLX genome sequencer (454 Life Science, Roche) and Wilson et al. (2014) used Illumina MiSeq to detect Cyclovirus and Leptospira sanatarosa in CSF, respectively. Moreover, NGS based approaches enable identification of all agents in a single run, regardless whether their sequence is known, or not (Big et al., 2009; Dupuis et al., 2011; Glaser et al., 2006; Quan et al., 2010).

Our approach, which employed NuGen Ribo-SPIA followed by NGS and metagenomic data analysis, led to the identification of HHV-1 in CSF from a patient with acute encephalitis at a time when specific antibodies could not yet be detected. Subsequently, the etiology of encephalitis was confirmed by specific PCR as well as HHV-1 seroconversion in the follow up CSF samples.

Host-derived genetic material is commonly present in clinical samples and presents a major challenge for NGS analysis (Lim et al., 2013). A commonly used method to reduce human genetic background is DNase I treatment of the samples (Allander et al., 2001). While we found that this approach would result in more than two fold decrease in the number of host-derived sequences, it significantly (30–50 fold) increased the number of bacterial and “other” sequences probably because of diminished competition from human DNA. Not unexpectedly, the use of DNase I resulted also in the reduction of HHV-1 sequences.

Interestingly, samples provided a significant number of bacterial reads and sequences classified as other (plants, plant viruses). This problem was previously reported for other NGS studies (Bzhalava et al., 2012; Salter et al., 2014). Bacterial contamination is often found in existing human-derived RNA sequence datasets and will remain an important issue for future NGS diagnostic application (Laurence et al., 2014; Strong et al., 2014). The presence of HCV

		Number of plasmid copies	Ct value	Fold decrease
1D	DNA	10^6	17.54	
1R	RNA	10^6	18.51	1.96
2D	DNA	10^5	21.35	
2R	RNA	10^5	22.03	1.60
3D	DNA	10^4	24.51	
3R	RNA	10^4	25.71	2.23
4D	DNA	10^3	26.90	
4R	RNA	10^3	28.55	3.14

sequences in analyzed CSF samples was most likely the result of index mis-assign in Illumina sequencing (Kircher et al., 2012) and the second independent run with three HHV-1 negative control CSF samples revealed neither HCV nor HHV-1 sequences. Intriguing, samples from the first run contained a small number of influenza C sequences, which raises the question of their origin. However, the analysis was conducted on the peak of flu season (end of February) and it is currently almost impossible to completely eliminate environmental contaminants during library preparation process and mis-assigned reads from other samples.

Surprisingly, we detected HHV-1 genetic material in CSF which was extracted for RNA and similar occurrence has been reported by Chan et al. (2014). Modified acid guanidinium thiocyanate-phenol-chloroform RNA extraction is known to generate genomic DNA contaminants (Siebert and Chenchik, 1993) and in our study significant DNA presence after RNA extraction was confirmed by quantitative analysis of plasmid template.

In conclusions, it seems that NGS, when necessary preceded by unspecific nucleic acid amplification, could supplement currently used diagnostic methods for the determination of encephalitis etiology. However, due to common index mis-assignment and contamination with reagents-derived nucleic acids, the presence of any identified pathogen should be verified by PCR or culture. We also found that extraction of total RNA could be used for the identification of both DNA and RNA pathogens in a single run.

Acknowledgment

This study was supported by Foundation for Polish Science, grant No. POMOST/2013-7/2 and Polish National Science Center (NCN), grant No. N/N401/646940.

References

- Allander, T., Emerson, S.U., Engle, R.E., Purcell, R.H., Bukh, J., 2001. A virus discovery method incorporating DNase treatment and its application to the identification of two bovine parvovirus species. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 11609–11614.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.
- Big, C., Reineck, L.A., Aronoff, D.M., 2009. Viral infections of the central nervous system: a case-based review. *Clin. Med. Res.* **7**, 142–146.
- Bzhalava, D., Ekstrom, J., Lysholm, F., Hultin, E., Faust, H., Persson, B., Lehtinen, M., de Villiers, E.M., Dillner, J., 2012. Phylogenetically diverse TT virus viremia among pregnant women. *Virology* **432**, 427–434.

- Chan, B.K., Wilson, T., Fischer, K.F., Kriesel, J.D., 2014. Deep sequencing to identify the causes of viral encephalitis. *PLOS ONE* 9, e93993.
- Chaudhuri, A., Kennedy, P.G., 2002. Diagnosis and treatment of viral encephalitis. *Postgrad. Med. J.* 78, 575–583.
- Debiasi, R.L., Tyler, K.L., 2004. Molecular methods for diagnosis of viral encephalitis. *Clin. Microbiol. Rev.* 17, 903–925, table of contents.
- Dupuis, M., Hull, R., Wang, H., Nattanmai, S., Glasheen, B., Fusco, H., Dziguia, L., Markey, K., Tavakoli, N.P., 2011. Molecular detection of viral causes of encephalitis and meningitis in New York State. *J. Med. Virol.* 83, 2172–2181.
- Glaser, C.A., Honarmand, S., Anderson, L.J., Schnurr, D.P., Forghani, B., Cossen, C.K., Schuster, F.L., Christie, L.J., Tureen, J.H., 2006. Beyond viruses: clinical profiles and etiologies associated with encephalitis. *Clin. Infect. Dis.* 43, 1565–1577.
- Goto, N., Prins, P., Nakao, M., Bonnal, R., Aerts, J., Katayama, T., 2010. BioRuby: bioinformatics software for the Ruby programming language. *Bioinformatics* 26, 2617–2619, http://hannonlab.cshl.edu/fastx_toolkit/index.html, 2015, FASTX-Toolkit.
- Kennedy, P.G., 2005. Viral encephalitis. *J. Neurol.* 252, 268–272.
- Kircher, M., Sawyer, S., Meyer, M., 2012. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* 40, e3.
- Laurence, M., Hatzis, C., Brash, D.E., 2014. Common contaminants in next-generation sequencing that hinder discovery of low-abundance microbes. *PLOS ONE* 9, e97876.
- Lim, Y.W., Schmieder, R., Haynes, M., Willner, D., Furlan, M., Youle, M., Abbott, K., Edwards, R., Evangelista, J., Conrad, D., Rohwer, F., 2013. Metagenomics and metatranscriptomics: windows on CF-associated viral and microbial communities. *J. Cyst. Fibros.* 12, 154–164.
- Lunter, G., Goodson, M., 2011. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res.* 21, 936–939.
- Mailles, A., Vaillant, V., Stahl, J.P., 2007. Infectious encephalitis in France from 2000 to 2002: the hospital database is a valuable but limited source of information for epidemiological studies. *Med. Mal. Infect.* 37, 95–102.
- Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J. (North America)*, 17.
- Quan, P.L., Wagner, T.A., Briese, T., Torgerson, T.R., Hornig, M., Tashmukhamedova, A., Firth, C., Palacios, G., Baisre-De-Leon, A., Paddock, C.D., Hutchison, S.K., Egholm, M., Zaki, S.R., Goldman, J.E., Ochs, H.D., Lipkin, W.I., 2010. Astrovirus encephalitis in boy with X-linked agammaglobulinemia. *Emerg. Infect. Dis.* 16, 918–925.
- Radford, A.D., Chapman, D., Dixon, L., Chantrey, J., Darby, A.C., Hall, N., 2012. Application of next-generation sequencing technologies in virology. *J. Gen. Virol.* 93, 1853–1868.
- Salter, S.J., Cox, M.J., Turek, E.M., Calus, S.T., Cookson, W.O., Moffatt, M.F., Turner, P., Parkhill, J., Loman, N.J., Walker, A.W., 2014. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* 12, 87.
- Siebert, P.D., Chenchik, A., 1993. Modified acid guanidinium thiocyanate-phenol-chloroform RNA extraction method which greatly reduces DNA contamination. *Nucleic Acids Res.* 21, 2019–2020.
- Solomon, T., Hart, I.J., Beeching, N.J., 2007. Viral encephalitis: a clinician's guide. *Pract. Neurol.* 7, 288–305.
- Strong, M.J., Xu, G., Morici, L., Splinter Bon-Durant, S., Baddoo, M., Lin, Z., Fewell, C., Taylor, C.M., Flemington, E.K., 2014. Microbial contamination in next generation sequencing: implications for sequence-based analysis of clinical samples. *PLoS Pathog.* 10, e1004437.
- Studahl, M., Lindquist, L., Eriksson, B.M., Gunther, G., Bengner, M., Franzen-Rohl, E., Fohlman, J., Bergstrom, T., Aurelius, E., 2013. Acute viral infections of the central nervous system in immunocompetent adults: diagnosis and management. *Drugs* 73, 131–158.
- Tan le, V., van Doorn, H.R., Nghia, H.D., Chau, T.T., Tu le, T.P., de Vries, M., Canuti, M., Dejjs, M., Jebbink, M.F., Baker, S., Bryant, J.E., Tham, N.T., Boni, N.T.B.K.M.F., Loi, T.Q., Phuong le, T., Verhoeven, J.T., Crusat, M., Jeeninga, R.E., Schultsz, C., Chau, N.V., Hien, T.T., van der Hoek, L., Farrar, J., de Jong, M.D., 2013. Identification of a new cyclovirus in cerebrospinal fluid of patients with acute central nervous system infections. *mBio* 4, e00231–e313.
- Wilson, M.R., Naccache, S.N., Samayoa, E., Biagtan, M., Bashir, H., Yu, G., Salamat, S.M., Somasekar, S., Federman, S., Miller, S., Sokolic, R., Garabedian, E., Candotti, F., Buckley, R.H., Reed, K.D., Meyer, T.L., Seroogy, C.M., Galloway, R., Henderson, S.L., Gern, J.E., DeRisi, J.L., Chiu, C.Y., 2014. Actionable diagnosis of neuroleptospirosis by next-generation sequencing. *N. Engl. J. Med.* 370, 2408–2417.