Viral metagenomics applied to blood donors and recipients at high risk for blood-borne infections

Virginie Sauvage, Syria Laperche, Justine Cheval, Erika Muth, Myriam Dubois, Laure Boizeau, Charles Hebert, Francois Lionnet, Jean-Jacques Lefrere, Marc Eloit

Viral metagenomics applied to blood donors and recipients at high risk for blood-borne infections

Virginie Sauvage[1], Syria Laperche[1], Justine Cheval[2], Erika Muth[2], Myriam Dubois[2], Laure Boizeau[1], Charles Hébert[2], François Lionnet[3], Jean-Jacques Lefrère[1,4], Marc Eloit[2,5,6]

[1]National Institute of Blood Transfusion (INTS), Department of Blood-borne agents, National Reference Center for Hepatitis B and C and HIV in Transfusion, Paris; [2]Pathoquest, Paris; [3]Internal Medicine, Assistance Publique-Hôpitaux de Paris, Tenon Hospital, Paris; [4]Paris-Descartes University, Paris; [5]Institut Pasteur, Laboratory of Pathogen Discovery, Biology of Infection Unit, Inserm U1117, Paris; [6]Alfort National Veterinary School, UMR 1161 Virology ENVA, INRA, ANSES, Maisons-Alfort, France

Background. Characterisation of human-associated viral communities is essential for epidemiological surveillance and to be able to anticipate new potential threats for blood transfusion safety. In high-resource countries, the risk of blood-borne agent transmission of well-known viruses (HBV, HCV, HIV and HTLV) is currently considered to be under control. However, other unknown or unsuspected viruses may be transmitted to recipients by blood-derived products. To investigate this, the virome of plasma from individuals at high risk for parenterally and sexually transmitted infections was analysed by high throughput sequencing (HTS).

Materials and methods. Purified nucleic acids from two pools of 50 samples from recipients of multiple transfusions, and three pools containing seven plasma samples from either HBV-, HCV- or HIV-infected blood donors, were submitted to HTS.

Results. Sequences from resident anelloviruses and HPgV were evidenced in all pools. HBV and HCV sequences were detected in pools containing 3.8×10^3 IU/mL of HBV-DNA and 1.7×10^5 IU/mL of HCV-RNA, respectively, whereas no HIV sequence was found in a pool of 150 copies/mL of HIV-RNA. This suggests a lack of sensitivity in HTS performance in detecting low levels of virus. In addition, this study identified other issues, including laboratory contaminants and the uncertainty of taxonomic assignment of short sequence. No sequence suggestive of a new viral species was identified.

Discussion. This study did not identify any new blood-borne virus in high-risk individuals. However, rare and/or viruses present at very low titre could have escaped our protocol. Our results demonstrate the positive contribution of HTS in the detection of viral sequences in blood donations.

Keywords: viral metagenomics, high throughput sequencing, blood-borne viruses, sensitivity, blood safety.

Introduction

Although the safety of blood transfusion has been dramatically improved, especially in high-resource countries, the emergence of infectious agents that may be transmitted by blood products remains a challenge for transfusion medicine. Several measures to prevent transmission of viruses by blood are in place, but because most of them are target specific, they are not efficient against uncharacterised infections. Only new application of at risk donor criteria and pathogen inactivation methods would be able to preclude those potential infections yet to be characterised. Pathogen inactivation is certainly a promising proactive method to prevent transmission of infectious agents due to its broad and non-specific spectrum of action. However, there is currently no universal method that could be applied to all blood components and be equally effective for all viral families. Some viruses, particularly when non-enveloped, are partially inactivated[1], leading to possible viral transmission with inactivated blood products, as recently reported with hepatitis E virus[2]. In addition, recent studies have confirmed the inadequate capacity of some viral reduction methods to inactivate very high viraemic concentration in plasma pools and plasma derivates[3,4]. As a result, improving the safety of blood products remains an ongoing challenge, not only through surveillance of well characterised viral pathogens (human immunodeficiency virus [HIV], hepatitis B virus [HBV], hepatitis C virus [HCV], human parvovirus B19 [B19V]) but also by implementing novel methodologies to identify unknown infectious agents that may affect blood transfusion safety.

The use of high throughput sequencing (HTS) in viral metagenomics has demonstrated the power of this "without *a priori*" technology to identify novel viruses that are too divergent from known agents to be detected by conventional polymerase chain reaction (PCR) or microarray techniques. By overcoming conventional methods of viral identification, metagenomics, which gives access to all nucleic acids present in a given sample, can describe viral communities and their diversity in environmental[5], human[6-8] and animal samples[9,10]. This approach has also been demonstrated to be useful in infectious diseases of unknown aetiology. This was seen in, for example, human Merkel cell carcinoma for which a polyomavirus was identified as responsible[11], in a cluster of fatal post-transplant-associated diseases in patients grafted from the same infected donor due to a novel arenavirus[12], and for the thrombocytopenia and leukopenia syndrome in China associated with a new bunyavirus[13]. In addition, viral metagenomics is suitable for detecting adventitious viruses in live-attenuated vaccines[14] and raw materials used in the manufacture of biologicals[15,16]. Overall, metagenomics appears to be a potent tool for global viral surveillance, diagnostics and for the discovery of emerging agents[17,18]. However, despite the analytical strength of this approach, difficulties remain in computational analysis and the detection of very low levels of pathogens in the clinical sample and represent a significant challenge.

This study illustrates the contribution of viral metagenomics to provide in depth analysis of virus populations present in plasma from individuals at high risk for parenterally and sexually transmitted infections.

Materials and methods

Study population

All studied plasma samples were collected from individuals who gave written consents for the use of these samples in conformity with French ethical legislation.

Plasma samples included in this study were split into two groups. Group 1 was made up of 50 multiple transfusion patients suffering from thalassaemia major or sickle cell disease, prospectively followed at the Tenon Hospital, Paris, France. Study subjects had received more than 100 red blood cell units. There were 21 males and 29 females; median age at the beginning and the end of follow up was 21 years (range: 3-82) and 35 years (range: 18-86), respectively. All tested negative for HIV and human T leukaemia virus (HTLV-1) antibodies, 34 were positive for HCV antibodies, 37 had anti-HBs only due to vaccination, 12 showed a past infection pattern (anti-HBc ± anti-HBs) and one had no HBV markers. For each study subject, one sample collected at the beginning and one at the end of the medical follow up (from 1988 to 2012) were analysed to investigate blood-borne infections that could have occurred during follow up. A total of 100 samples were collected and these were divided into two pools of 50 samples each, from 25 individuals.

Group 2 included 21 blood donors who tested positive for at least one viral marker. As shown in Table I, 7 were chronically infected with HBV, 7 were HCV-positive and declared intravenous drug abuse at the post-donation interview, and 7 were HIV-positive men who had had undeclared sex with men. Only one donor was co-infected with HBV and HIV. The three mini pools of Group 2 were prepared with a mix of plasma of the 7 blood donors who tested positive for HBV (HBV pool accounting for an estimated HBV-DNA concentration of 3.8×10^3 IU/mL), HCV (HCV pool containing 1.7×10^5 IU/mL of HCV-RNA) or HIV (HIV pool with a HIV viral load estimated at 150 copies/mL).

Nucleic acid extraction and amplification

A volume of 150 µL of each of the five plasma pools was treated with a cocktail of DNases (2 h at 37 °C) in order to digest unprotected nucleic acids. Total nucleic acids were then extracted using a QIAamp cador Pathogen Mini kit (Qiagen, Courtaboeuf, France), and randomly amplified as previously described[15,19]. Particular precautions to avoid contamination of samples, nucleic acid extracts and amplification products by DNA molecules from environment were implemented throughout the process including ultra-violet light treatment of some kit reagents[20].

High throughput sequencing and sequence analysis

The HTS and bioinformatics analyses were performed as previously described[15]. Briefly, 1 µg of high molecular weight DNA resulting from isothermal amplification was fragmented into 200- to 350-nt fragments to which adapters were ligated. Adapters included a nucleotide tag allowing for multiplexing several samples per lane. Samples were multiplexed in two separate lanes: 1) plasma pools from recipients; and 2) HBV, HCV and HIV plasma pools. Sequencing was carried out on an Illumina HiSeq-2000 sequencer (DNAVision, Charleroi, Belgium) with paired-end reads of 101 nucleotides in length. Five quality filters were successively performed on reads to obtain sequences of interest: duplicated reads introduced by PCR artefacts were suppressed, low-quality sequences and adapter sequences related to sequencing technology were trimmed, and ribosomal RNAs removed. The human genome was then filtered with SOAPaligner (available from: http://soap.genomics.org.cn) using the *Homo sapiens* hg19 reference. A number of assembly programmes dedicated to short or medium-sized reads were used to perform a *de novo* assembly that produces longer sequences called contigs: SOAPdenovo (available

Table I - Characteristics of HBV-, HCV- and HIV-infected blood donors*.

	Sample ID	Age	Gender	HBV load (IU/mL)	HBsAg	Anti-HBc	HCV load (IU/mL)	Anti-HCV	HIV load (copies/mL)	Anti-HIV
HBV-infected donors	B5507	40	M	2,080	+	+	–	–	–	–
	B5515	18	F	27	+	+	–	–	–	–
	B5583	31	M	15,737	+	+	–	–	–	–
	B5587	25	M	615	+	+	–	–	–	–
	B5590	32	F	9	+	+	–	–	–	–
	B5597	45	M	4,013	+	+	–	–	–	–
	B5664	30	F	4,026	+	+	–	–	–	–
HCV-infected donors	C3704	51	M	nd	–	–	–	+	–	–
	C3706	44	M	nd	–	–	–	+	–	–
	C3708	51	M	nd	–	–	–	+	–	–
	C3742	51	M	nd	–	+	5,455	+	–	–
	C3747	54	F	nd	–	–	569,394	+	–	–
	C3763	52	M	nd	–	–	607,006	+	–	–
	C3767	55	M	nd	–	+	–	+	–	–
HIV-infected donors	HIV/143	25	M	nd	–	–	nd	–	230	+
	HIV/38	37	M	1.5×10^8	+	+	nd	–	104	+
	HIV/110	27	M	nd	–	+	nd	+	–	+
	HIV/129	26	M	nd	–	+	nd	–	539	+
	HIV/44	33	M	nd	–	+	nd	–	73	+
	HIV/213	24	M	nd	–	–	nd	+	114	+
	HIV/131	22	M	nd	–	–	nd	–	65	+

–: negative; +: positive ; nd: not determined.
*Cobas TaqMan HBV, HCV and HIV assays from Roche Diagnostics (Meylan, France) were used to determine the viral loads of HBV DNA (limit of quantification of 6 IU/mL), HCV RNA (25 IU/mL) and HIV RNA (34 copies/mL), respectively.

from: http://soap.genomics.org.cn/; Velvet (available from: http://www.ebi.ac.uk/); and CLC Genomics Workbench (available from: http://www.clcbio.com). Taxonomic assignments of contigs were obtained through successive BlastN/X/N alignments against the NCBI NT then NR and finally WGS nucleotide collections (respectively with an e-value of 10e-3, 10 and 10e-3). Taxonomic assignments of singletons were obtained using only the BlastN algorithm against the viral subsection of the NCBI GenBank[21] nucleotide collection (e-value of 10e-3). In order to attest viral taxonomy assignation, every singleton alignment was confronted (i.e. realigned, BlastN, e-value of 10e-3) with the comprehensive NCBI NT nucleotide collection. This step can lead to an invalidation of some singleton alignments and refine the identification of viruses.

The amount of viral sequences for each pool was estimated by considering the number of reads included in contigs that match the reference sequence over 100 bp and the number of singletons with a hit match length covering at least 90 pb. On this basis, sequences sharing at least 80% nucleotide or amino acid identity with the best hit were classified as belonging to "known viruses". The other viral sequences (<80% identity) fell into the category referred to as "sequences identified as potential viral genes".

Results

Bioinformatics pipeline and taxonomic assignment

Sequencing generated 83 to 164×10^6 paired-end reads of 101 nucleotide length per pool. After quality filtering steps, including PCR artefact removal, quality trimming and human sequences elimination, an average of 17.3×10^6 reads per pool (range: 8.6-24.8$\times10^6$ reads/pool) was kept for the *de novo* assembly step. An average of 29.5% of these highly valuable reads (range: 26.8-31.7%) was extended into contigs. A taxonomic assignment could be reached for 79% of them (range: 67.1-87%). Contigs for which no sequence homology was detected were classified as "unknown". Some of these unknown contigs reached a maximum length of 1195 nucleotides (*data not shown*).

Targeted eukaryotic DNA and RNA viruses represented an average 2.0% (1.4-2.3%) of the reads that passed all filtering steps, except in HIV plasma pool in which they accounted for 11% of valuable reads (*data not shown*) due to the presence of a high concentration of HBV, as described below.

Identification of known viruses

Known viruses for which sequences were found are shown in Table II. All pools contained anelloviruses, which for a large proportion (96.1-99.4%), were distributed in the *Alphatorquevirus* genus (TTV) (Figure 1). *Betatorquevirus* (TTMV) and *Gammatorquevirus* (TTMDV/SAV) did not exceed 3.7% of all sequences. Human pegivirus (HPgV), formerly known as GB virus C (GBV-C), was also found in all pools with a particularly high number of reads covering the full length of the genome in HIV-infected-subjects.

HBV and HCV pools contained viral sequences of HBV and HCV, respectively, and sequences of human adenoviruses C, genotypes 2 (KF268310) and 5 (KF429754). The presence of these adenoviruses could not be confirmed by specific PCR due to the shortage of nucleic acid sample. HBV sequences have been detected in HCV pool and *vice versa*, although none of the blood donors were known to be HCV/HBV co-infected (Table I). The contig related to the *Circoviridae* family identified in the HBV pool showed 99% nucleotide identity with the *Circovirus*-like NI/2007 Rep sequence (strain JN837698), previously reported in a metagenomic analysis of cases of unknown tropical febrile illness[22]. Finally, two contigs closely related to *Mimivirus*-like viruses *(Megavirus chiliensis* and *Moumouvirus monve),* which are known to infect free-living amoebae[23], were detected in the HBV pool.

HIV pool analysis showed 350 HCV and more than 1,700,000 HBV sequences, while no HIV sequence was detected. The high viral titre of the subject co-infected with HBV (accounting for 2.2×10^7 IU/mL HBV DNA in the pool) explains the detection of a complete viral genome

Table II - Identification of known viruses in plasma samples of multi-transfused recipients and HBV-, HCV- and HIV-infected blood donors.

Samples	HBV pool	HCV pool	HIV pool	Multi-transfused Pool 1	Multi-transfused Pool 2
Total reads, raw data	101,985,062	83,838,650	91,368,632	164,010,226	93,808,700
Total reads after quality filtering	24,841,058 (24.36%)	18,711,255 (22.32%)	18,934,556 (20.72%)	15,440,155 (9.41%)	8,694,334 (9.27%)
Anelloviridae	0.24%[b]	0.05%	0.18%	0.115%	0.14%
Flaviviridae, human Pegivirus	0.0002%	0.0004%	0.07%	0.014%	0.004%
Flaviviridae, hepatitis C virus	0.0005%	0.15%	0.0004%	0.004%	0.0015%
Hepadnaviridae, hepatitis B virus	0.0375%	0.046%	1.90%	nd	nd
Retroviridae, human immunodeficiency virus 1 or 2	nd	nd	nd	nd	nd

[a] nd: not detected; [b] calculated according to the total raw reads.

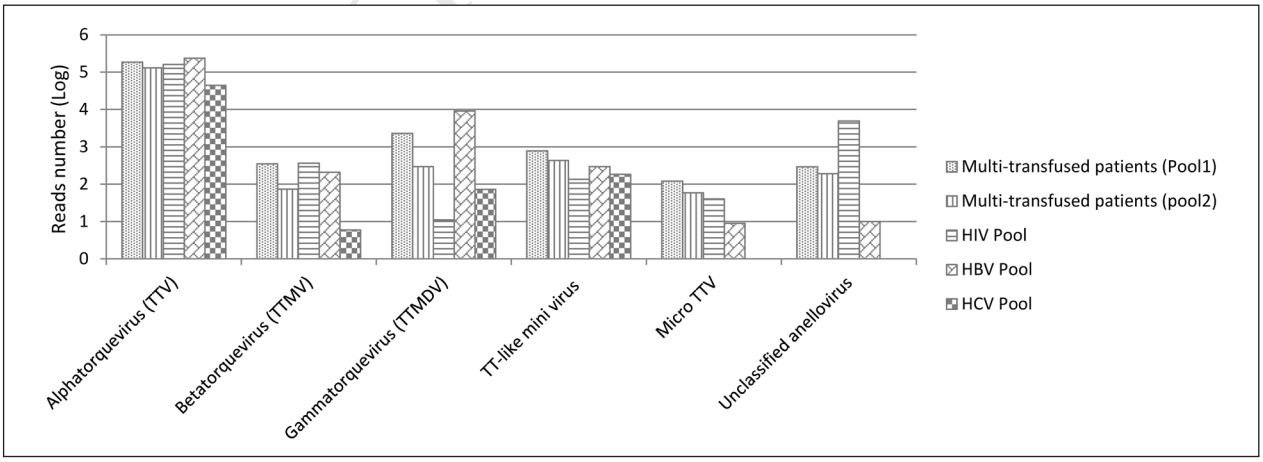

Figure 1 - Amounts of known *Anelloviruses* in plasma samples of multi-transfused recipients and HBV-, HCV- and HIV-infected blood donors.
Amounts were estimated by considering the number of reads included in contigs that match the reference sequence over 100 bp and the number of singletons with a hit match length covering at least 90 pb. Sequences sharing at least 80% nucleotide or amino acid identity with the best hit were classified as "known *Anelloviruses*".

with a high depth of coverage. The presence of HCV was not confirmed either in the HIV pool or in the original HIV samples with a PCR targeting the NS5b region (analytical sensitivity estimated to be 6×10^2 IU/mL)[24]. On the other hand, few sequences that aligned to HERV (92-100% nucleotide identity with HERV-K [Class II ERV group]), *Circoviridae* (95% nucleotide identity with *Circovirus*-like NI/2007 Rep sequence, JN837698), *Rhabdoviridae (*95% nucleotide identity with the G protein gene of various vesicular stomatitis indiana virus clones), and *Virgaviridae* families were found. Interestingly, sequences from plant viruses were found, such as contigs sharing more than 98% nucleotide identity with Tobacco mosaic virus (strains KC007531 and HE818459), together with reads that mapped to Viridiplantae (*Streptophyta phylum*) and especially to species found in human food products, including tomato (*Solanium lycopersicum*) and potato (*Solanium tuberosum*). These findings suggest that nucleic acids from ingested vegetables and their viruses may enter the bloodstream[25]. However, such reads are most likely the results of contaminations, as reported by Lusk[26], rather than valuable reads.

HCV sequences were present in the two plasma pools of multiply transfused recipients and were confirmed by specific NS5b RT-PCR. In addition, as observed in HBV and HIV pools, *Circovirus*-like NI/2007 Rep (99% nucleotide identity with JN837698) and HERV-K sequences (100% nucleotide identity) have also been found.

No sequence of the human hepegivirus 1 (HHpgV-1), a newly described virus retrieved in old serum samples of blood transfusion recipients and haemophilia patients[27], was detected in any samples investigated in this study. In addition, HHpgV-1 RNA was not detected by RT-PCR in nucleic acid extracts from HIV, HBV and HCV pools and corresponding individual plasma samples.

Sequences identified as potential viral genes

Potentially novel viral sequences belonging to the Nucleocytoplasmic Large DNA Viruses (NCLDVs) group, which constitutes the *Megavirales* order[28] including seven viral families (*Ascoviridae*, *Asfarviridae*, *Iridoviridae*, *Phycodnaviridae*, *Poxviridae*, *Mimiviridae* and the recently discovered *Marseilleviridae*[29]), were observed. The majority of contigs from this group shared only 25-50% of amino acid identities (and ≤75%) to known viruses. The other viral sequences were mostly related to anelloviruses (mainly TTV sequences with amino acid identity ranging from 60% to 80% that might be new genotypes or species) and circoviruses (sequences found in bats[30] and reclaimed water[31]). A unique contig exhibiting 34.4% amino acid identity to HERV was identified.

Discussion

To identify new viruses that could impair blood transfusion safety we used an ultra-deep sequencing method to analyse samples collected from at risk subjects, including recipients of multiple blood transfusions and HBV-, HCV- and HIV-infected blood donors. A large number of various viral sequences were found, showing the efficacy of the method. Among them, anelloviruses and HPgV were the most frequently identified agents. The presence of hepatotropic viruses in plasma pools from subjects infected with HBV and HCV was confirmed, whereas HIV sequences were not detected in the HIV pool. No unknown agent was revealed, despite the identification of numerous short nucleotide sequences distantly related to viral sequences reported in the current databases. Although the present study confirms the utility of viral metagenomics to identify the viral diversity in blood samples, the results have also identified some limitations of this approach.

As expected, commensal anelloviruses and the highly prevalent orphan HPgV[32], both initially discovered in patients with non A-E hepatitis[33-36], were predominantly identified in all samples. In agreement with previous studies, all investigated specimens showed large number of TTV and mixed infections of *Anellovirus*, which is in accordance with the absence of pathogenicity of these viruses[37]. Furthermore, the discovery of putative novel sequences of anelloviruses confirms the high genetic diversity of this viral family[38]. The presence of HPgV is in line with the estimated prevalence of 1-5% of healthy blood donors in developed countries[32,39], and is significantly higher in populations at high risk for parenteral exposure to infectious agents (20-40% in HIV-infected individuals, 7-20% in patients receiving blood products, and 20% and 10% in HCV and HBV-infected subjects, respectively)[39,40].

Among the sequences identified as potential viral genes, sequences related to NCLDV group and *Circoviridae* were identified. To date, only a few publications have observed NCLDV sequences in human specimens such as plasma, nasopharyngeal swabs and stool[41,42], and only *Poxviridae* and *Mimiviridae* from this group have been associated with human diseases[43-45]. Recently, the use of HTS allowed the genome of two marseilleviruses to be retrieved from human samples: Senegalvirus, in a stool sample of an asymptomatic Senegalese subject[41] and Giant Blood Marseillevirus (GBM) in blood samples from asymptomatic donors[46]. Interestingly, GBM viral DNA was found in 4% of 174 studied healthy donors and in 9.1% of post-transfusion sera from thalassaemia patients, suggesting that this virus could be transmitted by transfusion[47]. In the present study, 7 short contigs with very weak similarity to marseilleviruses proteins were detected, but since

the coverage of significant portions of viral genomes was not achieved, it was not possible to conclude that such sequences represent *bona fide* marseillevirus-like viruses. In addition, in two recently reported studies (of 339[48] and 187[49] subjects, respectively), GMB viral DNA was not detected by PCR in any of the human plasma samples, suggesting that GBM might be a laboratory contaminant rather than a virus infecting humans. Similarly, we were not able to conclude that the sequences related to other NCLDV members were due to the presence of the viruses in plasma samples; it is most likely that these short virus-like sequences represent contaminants[26,50].

Control libraries (i.e. water only) tested in the process contained sequences related to *Asfavirus* and *Circovirus* (e.g. Bat circovirus ZS/YunnanChina/2009), which confirmed the presence of contaminants in the reagents. Such contaminations have been reported with a *Parvovirus*-like hybrid genome (PHV/NIH-CQV), initially described as a possible new causative agent of seronegative hepatitis[51], but subsequently found in water eluted from some nucleic acid extraction silica-binding columns[52,53]. This viral contamination has been suggested to originate from diatom algae present in silica[52] or oomycetes (*Phytophtora spp* and *Pythium spp*) introduced during water washing of silica[54]. These hypotheses are in agreement with our findings using a specific PHV/NIH-CQV PCR; we detected sequences in plasma pools and water controls when nucleic acids were extracted with Qiagen reagents (QIAamp Viral RNA kit and QIAamp DNA Blood kit), as previously described[51], but not when extracted with other reagents (High Pure System Viral Nucleic Acid kit; Roche, Mannheim, Germany). Other viral contaminants from spin columns, including circoviruses/densoviruses and iridoviruses, have been previously reported[7]. Taken together, these cases strongly suggest that contaminants from different origins can easily be introduced in the process from the nucleic acid extraction step up to the library construction. However, they clearly demonstrate that the extraordinary potency of HTS methods generates results that must be interpreted with caution before any conclusions about the discovery of a new agent can be made.

Despite the inclusion of at risk individuals, and the efficacy of this pipeline to discover new viruses[15,55-57], no unknown virus was found. However, identification of new viruses relies on sequence homologies with known viruses, meaning that viruses very distant to those present in the database might escape identification. This can be seen from a recent study that identified a new bacteriophage in the "unknown" sequences, emphasising the importance of developing new approaches and computational tools to explore this part of metagenomes[58]. More importantly, we have demonstrated the limitations of the pipeline to detect very low viral loads, as seen by the absence of HIV sequences in the HIV-positive pool. The pooling strategy was unable to detect low levels of genomic sequences present in the individual plasma samples. However, even if the sample size used by Cheval *et al.*[19] had been applied (i.e. nucleic acids extraction from 150 µL of individual plasma sample and DNA amplification on the nucleic acids corresponding to 12 µL of the original sample), such low viral concentrations (around 2 genome copies final) would still not have been identified, despite the higher sequencing depth used. This shows that sample preparation remains one of the key steps to enhance the detection of low-copy-number viruses, especially when the scope of viral discovery concerns blood safety. Thus, with this purpose, the efficiency and the robustness of different viral enrichment (i.e. ultracentrifugation, polyethylene glycol [PEG] precipitation, membrane ultrafiltration and/or nucleic acid extraction from a large plasma volume[59,60]) and random amplification methods, warrant investigation.

The identification of HBV sequences in HCV pool and *vice versa*, in spite of the absence of documented co-infections, suggested that cross-contaminations have occurred during the experiment. Biological cross-contaminations may have occurred during one or several steps of the sample preparation, including nucleic acid extraction, amplification, and library construction. Based on the almost identical quantitative and qualitative viral genome coverage patterns observed in different samples (e.g. HBV and HPgV genome coverage in HBV and HCV pools, and HCV genome coverage in HIV and HBV pools; *data not shown*), it is highly probable that, in our study, biological cross-contaminations occurred following fragmentation during library preparation. This is in not uncommon when sensitive and non-specific techniques are used, such as HTS, especially when samples containing viruses at a high load are included. Biological cross-contaminations may potentially affect quantitative and comparative analyses of virus populations between samples; however, these will have no impact on viral discovery.

Conclusions

Metagenomics is a powerful approach to identify the viral diversity present in human plasma, irrespective of the fact that, in the present study, there was no evidence of a novel virus infecting deferred French blood donors and multi-transfused recipients. It is unlikely that the methodological approach used here would have failed to identify a virus if present at high titre, but non-detection of low titre viruses cannot be excluded. Consequently, specific technical issues and

potential biases in metagenome preparation should be carefully addressed to promote its potential use for the discovery and surveillance of viruses that are of concern in issues of blood safety. In addition, it has to be noted that HTS methods give no information about the capability of viruses to cause infections since they only detect nucleic acids. This would be of concern if HTS was used, for example, as quality control methods after pathogen inactivation process. However, this is not only a specific limitation of HTS, but of all molecular tests based on nucleic acid detection. The methodology used is currently being modified for use in clinical diagnostics that requires the establishment of quality controls and protocols ensuring a high consistency and traceability of results. Overall, this study demonstrates the positive contribution of HTS in the detection of viral sequences in blood donations.

Acknowledgements

We thank Dr Girot from the Tenon hospital (Paris, France) for providing additional epidemiological data on multi-transfused patients. We also thank Daniel Candotti (INTS, Paris) for reading the manuscript and for his helpful comments.

Authorship contributions

ME, J-JL and SL designed the study, EM, MD and LB performed the experiments, VS, JC and CH performed analyses, JC, EM, MD, LB and FL contributed reagents/materials/analysis tools, VS, SL and ME drafted the manuscript. All Authors contributed to the conception of the project and approved the final manuscript.

Disclosure of conflicts of interest

JC, EM, MD, CH are employees and ME is the chairman of PathoQuest. The remaining Authors declare no conflicts of interest.

References

1) Barros AE, Gomes AM, do Bonfim CV. Viral inactivation in hemotherapy: systematic review on inactivators with action on nucleic acids. Rev Bras Hematol Hemoter 2012; **34**: 231-5.
2) Hauser L, Roque-Afonso AM, Beylouné A, et al. Hepatitis E transmission by transfusion of Intercept blood system-treated plasma. Blood 2014; **123**: 796-7.
3) Stramer SL, Hollinger FB, Katz LM, et al. Emerging infectious disease agents and their potential threat to transfusion safety. Transfusion 2009; **49** (Suppl 2): 1S-29S.
4) Zhang W, Ke L, Changqing L, et al. Parvovirus B19V DNA contamination in Chinese plasma and plasma derivatives. J Transl Med 2012; **10**: 194.
5) Ng TF, Marine R, Wang C, et al. High variety of known and new RNA and DNA viruses of diverse origins in untreated sewage. J Virol 2012; **86**: 12161-75.
6) Minot S, Sinha R, Chen J, et al. The human gut virome: inter-individual variation and dynamic response to diet. Genome Res 2011; **21**: 1616-25.
7) Lysholm F, Wetterbom A, Lindau C, et al. Characterization of the viral microbiome in patients with severe lower respiratory tract infections, using metagenomic sequencing. PLoS One 2012; **7**: e30875.
8) Foulongne V, Sauvage V, Hebert C, et al. Human skin microbiota: high diversity of DNA viruses identified on the human skin by high throughput sequencing. PLoS One 2012; **7**: e38499.
9) Phan TG, Kapusinszky B, Wang C, et al. The fecal viral flora of wild rodents. PLoS Pathog 2011; **7**: e1002218.
10) Shan T, Li L, Simmonds P, et al. The fecal virome of pigs on a high-density farm. J Virol 2011; **85**: 11697-708.
11) Feng H, Shuda M, Chang Y, Moore PS. Clonal integration of a polyomavirus in human Merkel cell carcinoma. Science 2008; **319**: 1096-100.
12) Palacios G, Druce J, Du L, et al. A new arenavirus in a cluster of fatal transplant-associated diseases. N Engl J Med 2008; **358**: 991-8.
13) Xu B, Liu L, Huang X, et al. Metagenomic analysis of fever, thrombocytopenia and leukopenia syndrome (FTLS) in Henan Province, China: discovery of a new bunyavirus. PLoS Pathog 2011; **7**: e1002369.
14) Victoria JG, Wang C, Jones MS, et al. Viral nucleic acids in live-attenuated vaccines: detection of minority variants and an adventitious virus. J Virol 2010; **84**: 6033-40.
15) Gagnieur L, Cheval J, Gratigny M, et al. Unbiased analysis by high throughput sequencing of the viral diversity in fetal bovine serum and trypsin used in cell culture. Biologicals 2014; **42**: 145-52.
16) Gagnieur L, Cheval J, Cochet M, et al. Analysis by high throughput sequencing of Specific Pathogen Free eggs. Biologicals 2014; **42**: 218-9.
17) Barzon L, Lavezzo E, Costanzi G, et al. Next-generation sequencing technologies in diagnostic virology. J Clin Virol 2013; **58**: 346-50.
18) Lipkin WI, Firth C. Viral surveillance and discovery. Curr Opin Virol 2013; **3**: 199-204.
19) Cheval, J, Sauvage V, Frangeul L, et al. Evaluation of high-throughput sequencing for identifying known and unknown viruses in biological samples. J Clin Microbiol 2011; **49**: 3268-75.
20) Woyke T, Sczyrba A, Lee J, et al. Decontamination of MDA reagents for single cell whole genome amplification. PLoS One 2011; **6**: e26161.
21) Benson DA, Cavanaugh M, Clark K, et al. GenBank. Nucleic Acids Res 2013; **41**: D36-42.
22) Yozwiak NL, Skewes-Cox P, Stenglein MD, et al. Virus identification in unknown tropical febrile illness cases using deep sequencing. PLoS Negl Trop Dis 2012; **6**: e1485.
23) Yoosuf N, Yutin N, Colson P, et al. Related giant viruses in distant locations and different habitats: Acanthamoeba polyphaga moumouvirus represents a third lineage of the Mimiviridae that is close to the megavirus lineage. Genome Biol Evol 2012; **4**: 1324-30.
24) Sandres-Sauné K, Deny P, Pasquier C, et al. Determining hepatitis C genotype by analyzing the sequence of the NS5b region. J Virol Methods 2003; **109**: 187-93.
25) Spisák S, Solymosi N, Ittzés P, et al. Complete genes may pass from food to human blood. PLoS One 2013; **8**: e69805.
26) Lusk RW. Diverse and widespread contamination evident in the unmapped depths of high throughput sequencing data. PLoS One 2014; **9**: e110808.
27) Kapoor A, Kumar A, Simmonds P, et al. Virome Analysis of Transfusion Recipients Reveals a Novel Human Virus That Shares Genomic Features with Hepaciviruses and Pegiviruses. MBio 2015; **6**: e01466-15.
28) Colson P, De Lamballerie X, Yutin N, et al. "Megavirales", a proposed new order for eukaryotic nucleocytoplasmic large DNA viruses. Arch Virol 2013; **158**: 2517-21.

29) Colson P, Pagnier I, Yoosuf N, et al. Marseilleviridae, a new family of giant viruses infecting amoebae. Arch Virol 2012; **158**: 915-20.
30) Ge X, Li J, Peng C, et al. Genetic diversity of novel circular ssDNA viruses in bats in China. J Gen Virol 2011; **92**: 2646-53.
31) Rosario K, Duffy S, Breitbart M. Diverse circovirus-like genome architectures revealed by environmental metagenomics. J Gen Virol 2009; **90**: 2418-24.
32) Stapleton JT, Foung S, Muerhoff AS, et al. The GB viruses: a review and proposed classification of GBV-A, GBV-C (HGV), and GBV-D in genus Pegivirus within the family Flaviviridae. J Gen Virol 2011; **92**: 233-46.
33) Kew MC. Hepatitis viruses (other than hepatitis B and C viruses) as causes of hepatocellular carcinoma: an update. J Viral Hepat 2013; **20**: 149-57.
34) Nishizawa T, Okamoto H, Konishi K, et al. A novel DNA virus (TTV) associated with elevated transaminase levels in posttransfusion hepatitis of unknown etiology. Biochem Biophys Res Commun 1997; **241**: 92-7.
35) Simons JN, Leary TP, Dawson GJ, et al. Isolation of novel virus-like sequences associated with human hepatitis. Nat Med 1995; **1**: 564-9.
36) Linnen J, Wages J Jr, Zhang-Keck ZY, et al. Molecular cloning and disease association of hepatitis G virus: a transfusion-transmissible agent. Science 1996; **271**: 505-8.
37) Ninomiya M, Takahashi M, Nishizawa T, et al. Development of PCR assays with nested primers specific for differential detection of three human anelloviruses and early acquisition of dual or triple infection during infancy. J Clin Microbiol 2008; **46**: 507-14.
38) Peng YH, Nishizawa T, Takahashi M, et al. Analysis of the entire genomes of thirteen TT virus variants classifiable into the fourth and fifth genetic groups, isolated from viremic infants. Arch Virol 2002; **147**: 21-41.
39) Lefrère JJ, Loiseau P, Pawlotsky JM. [The G/GBV-C virus: from hepatology to hematology]. Hématologie 1996; **2**: 537-44. [In French.]
40) Lefrère JJ, Roudot-Thoraval F, Morand-Joubert L, et al. Prevalence of GB virus type C/hepatitis G virus RNA and of anti-E2 in individuals at high or low risk for blood-borne or sexually transmitted viruses: evidence of sexual and parenteral transmission. Transfusion 1999; **39**: 83-94.
41) Colson P, Fancello L, Gimenez G, et al. Evidence of the megavirome in humans. J Clin Virol 2013; **57**: 191-200.
42) Law J, Jovel J, Patterson J, et al. Identification of hepatotropic viruses from plasma using deep sequencing: a next generation diagnostic tool. PLoS One 2013; **8**: e60595.
43) Haller SL, Peng C, McFadden G, Rothenburg S. Poxviruses and the evolution of host range and virulence. Infect Genet Evol 2014; **21C**: 15-40.
44) La Scola B, Marrie TJ, Auffray JP, Raoult D. Mimivirus in pneumonia patients. Emerg Infect Dis 2005; **11**: 449-52.
45) Saadi H, Pagnier I, Colson P, et al. First isolation of Mimivirus in a patient with pneumonia. Clin Infect Dis 2013; **57**: e127-34.
46) Popgeorgiev N, Boyer M, Fancello L, et al. Marseillevirus-like virus recovered from blood donated by asymptomatic humans. J Infect Dis 2013; **208**: 1042-50.
47) Popgeorgiev N, Colson P, Thuret I, et al. Marseillevirus prevalence in multitransfused patients suggests blood transmission. J Clin Virol 2013; **58**: 722-5.
48) Sauvage V, Livartowski A, Boizeau L, et al. No evidence of Marseillevirus-like virus presence in blood donors and recipients of multiple blood transfusions. J Infect Dis 2014; **210**: 2017-18.
49) Phan TG, Desnues C, Switzer WM, et al. Absence of giant blood Marseille-like virus DNA detection by polymerase chain reaction in plasma from healthy US blood donors and serum from multiply transfused patients from Cameroon. Transfusion 2015; **55**: 1256-62.
50) Laurence M, Hatzis C, Brash DE. Common contaminants in next-generation sequencing that hinder discovery of low-abundance microbes. PLoS One 2014; **9**: e97876.
51) Xu B, Zhi N, Hu G, et al. Hybrid DNA virus in Chinese patients with seronegative hepatitis discovered by deep sequencing. Proc Natl Acad Sci USA 2013; **110**: 10264-9.
52) Naccache SN, Greninger AL, Lee D, et al. The Perils of Pathogen Discovery: Origin of a Novel Parvovirus-Like Hybrid Genome Traced to Nucleic Acid Extraction Spin Columns. J Virol 2013; **87**: 11966-77.
53) Smuts H, Kew M, Khan A, Korsman S. Novel hybrid parvovirus-like virus, NIH-CQV/PHV, contaminants in silica column-based nucleic acid extraction kits. J Virol 2014; **88**: 1398.
54) Zhi N, Hu G, Wong S, et al. Reply to Naccache, et al: Viral sequences of NIH-CQV virus, a contamination of DNA extraction method. Proc Natl Acad Sci USA 2014; **111**: E977.
55) Sauvage V, Cheval J, Foulongne V, et al. Identification of the first human gyrovirus, a virus related to chicken anemia virus. J Virol 2011; **85**: 7948-50.
56) Sauvage V, Foulongne V, Cheval J, et al. Human polyomavirus related to African green monkey lymphotropic polyomavirus. Emerg Infect Dis 2011; **17**: 1364-70.
57) Sauvage V, Ar Gouilh M, Cheval J, et al. A member of a new Picornaviridae genus is shed in pig feces. J Virol 2012; **86**: 10036-46.
58) Dutilh BE, Cassman N, McNair K, et al. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. Nat Commun 2014; **5**: 4498.
59) Hall RJ, Wang J, Todd AK, et al. Evaluation of rapid and simple techniques for the enrichment of viruses prior to metagenomic virus discovery. J Virol Methods 2014; **195**: 194-204.
60) Thurber RV, Haynes M, Breitbart M, et al. Laboratory procedures to generate viral metagenomes. Nat Protoc 2009; **4**: 470-83.

Arrived: 22 June 2015 - Revision accepted: 18 January 2016
Correspondence: Syria Laperche
Institut National de la Transfusion Sanguine
Département d'Etudes des Agents Transmissibles par le Sang
Centre National de Référence des Hépatites Virales B et C et du VIH en Transfusion
6, rue Alexandre Cabanel
F-75015 Paris, France
e-mail: slaperche@ints.fr