Disponible en ligne sur
ScienceDirect
www.sciencedirect.com

Elsevier Masson France
EM|consulte
www.em-consulte.com

TRANSFUSION
CLINIQUE ET BIOLOGIQUE

Transfusion Clinique et Biologique 24 (2017) 218–222

State of the art

The potential of viral metagenomics in blood transfusion safety

La métagénomique virale : un nouvel outil de surveillance des agents viraux émergents au service de la sécurité transfusionnelle

V. Sauvage *, J. Gomez , L. Boizeau , S. Laperche

Département d'études des agents transmissibles par le sang, Institut national de la transfusion sanguine (INTS), Centre national de référence risques infectieux transfusionnels, 75015 Paris, France

Available online 8 July 2017

Abstract

Thanks to the significant advent of high throughput sequencing in the last ten years, it is now possible via metagenomics to define the spectrum of the microbial sequences present in human blood samples. Therefore, metagenomics sequencing appears as a promising approach for the identification and global surveillance of new, emerging and/or unexpected viruses that could impair blood transfusion safety. However, despite considerable advantages compared to the traditional methods of pathogen identification, this non-targeted approach presents several drawbacks including a lack of sensitivity and sequence contaminant issues. With further improvements, especially to increase sensitivity, metagenomics sequencing should become in a near future an additional diagnostic tool in infectious disease field and especially in blood transfusion safety.

Keywords: Viral metagenomics; High throughput sequencing; Viral discovery; Blood-borne viruses; Emerging viruses; Blood safety

Résumé

Grâce aux avancées technologiques dans le domaine du séquençage haut débit ces dix dernières années, il est maintenant possible via la métagénomique d'établir le microbiome d'un prélèvement clinique tel que le sang. Ainsi, la métagénomique par séquençage direct apparaît comme une approche prometteuse pour l'identification et la surveillance d'agents viraux nouveaux, émergents et/ou inattendus qui pourraient impacter la sécurité transfusionnelle. Cependant, bien que cette approche sans a priori présente de considérables avantages comparés aux techniques traditionnelles d'identification des pathogènes, la métagénomique présente également plusieurs inconvénients et limitations techniques tels qu'un manque de sensibilité et la présence de nombreuses contaminants provenant notamment des réactifs. Malgré ces inconvénients qui demandent des améliorations majeures, nul doute que la métagénomique devienne dans un futur proche un outil diagnostique supplémentaire dans le domaine des maladies infectieuses, et en particulier en sécurité transfusionnelle.

Mots clés : Métagénomique ; Séquençage haut débit ; Découverte de virus ; Virus transmis par le sang ; Virus émergents ; Sécurité transfusionnelle

Thanks to the significant advent of high throughput sequencing (also referred to as deep sequencing) in the last ten years, it is now possible via metagenomics, which gives access to all nucleic acids present in a given sample, to define the spectrum of the microbial sequences present in environmental, human and animal samples. This untargeted approach (also called whole genome shotgun sequencing) has been first used to describe the microbial communities from various biological specimens (e.g. feces, urine and blood) and has rapidly become a very useful diagnostic tool for detecting unexpected or unknown infectious agents, and particularly viruses, in infectious diseases of unknown etiology.

* Corresponding author.
E-mail address: vsauvage@ints.fr (V. Sauvage).

http://dx.doi.org/10.1016/j.tracli.2017.06.018
1246-7820/

1. Principle

Metagenome analysis by high throughput sequencing comprises the following five distinct steps: extraction and purification of nucleic acids (DNA and/or RNA) from the sample, reverse transcription of RNAs into complementary DNA (cDNA) and random amplification of DNA molecules (still required for all available deep sequencing technologies when a microbiome analysis is performed), fragmentation of double-strand DNA (ds DNA) by using enzymatic or mechanical techniques prior library construction, deep sequencing and bioinformatics analysis of raw sequencing data (see review [1]). Another method can also be applied to sequence all RNA molecules present in a sample, the RNA-Seq approach, also called RNA shotgun sequencing, where RNA molecules are first fragmented, ligated to specific adapters, reverse transcribed into complementary DNA (cDNA), which are then amplified prior to sequencing. Similarly, this approach can be applied to sequence only DNA molecules (DNA-Seq) in a given sample.

2. Advantages

Metagenomics sequencing offers considerable advantages compared to other methods of pathogen identification such as polymerase chain reaction (PCR) amplification or DNA microarrays. Firstly, compared to PCR amplification, this molecular approach is able to detect simultaneously multiple organisms without a priori knowledge of the pathogen's nucleic acid sequences. Consequently, there is no need to design a set of sequence-specific primers able to target multiple pathogens (and their different variants), and with the capacity to function without any interference or competition to produce nonspecific amplifications. Furthermore, nucleic acid detection by PCR amplifications can be laborious, time consuming and expensive in the case where no pathogen is clearly suspected. Compared to DNA microarrays, which contain representative sequences from all known viruses (e.g. Virochip [2]) and/or bacteria [3], deep sequencing detects the full spectrum of pathogens with a higher sensitivity and offers the possibility to get full genomes. Metagenomics sequencing also allows analyzing simultaneously and in parallel several samples in a same sequencing run, decreasing significantly the cost of this approach. One of the most attractive advantages is the capacity to identify sequences of new microorganisms including the description of novel species [4] and even novel genera [5]. The power of this approach to discover new pathogens rests in large part on a robust in depth bioinformatics analysis that requires a skilled analyst, and thus represents a major limitation for numerous laboratories with limited resources. To date, the bioinformatics analysis still remains a critical step in metagenomics.

3. Limits

Like all molecular techniques, metagenomics sequencing presents several drawbacks. In virology, the application of metagenomics to clinical samples is made difficult by the fact that viral sequences represent a very low proportion compared to the host DNA, ribosomal RNA and bacterial sequences. To improve the detection of viruses, different steps can be included to enrich viral sequences during sample preparation such as filtering out cellular material, removal of ribosomal RNA and/or host cell-free DNA/RNA sequences or viral genomes capture [6]. Once viral nucleic acids are extracted, a random amplification is required both to optimize the detection of small amount of viruses and to obtain the amount of DNA required by the sequencing library preparation kits (1 μg to 1 ng depending on library preparation kits and sequencing platform). This step also amplifies remaining host nucleic acids leading to a significant proportion of host reads after sequencing. To get around this major disadvantage, at least in part, a high depth of sequencing is required for good sample coverage and to detect viruses present at low level. The question is then, how many reads are needed to obtain the analytical sensitivity desired? As the sensitivity of the approach depends of the sample matrix and is the result of all successive steps of the workflow, from the nucleic acid extraction method to the bioinformatics analysis, a rigorous pilot study on spiked clinical samples must be imperatively carried out before implementation of a metagenomics pipeline in a laboratory [7]. Very recently, the National Institute for Biological Standards and Control (NIBSC, Hertfordshire, UK) produced a reagent containing 25 target viruses (Virus Multiplex Reagent 11/242-001) intended for use as a reference materials in adventitious virus detection assays employing deep sequencing technology. This first common material reference constitutes a very useful control to ensure satisfactory assay performance and sensitivity to enable comparisons between laboratories [8]. Reagent and laboratory contaminants also represent a real problem in metagenomics that could lead to misinterpretations of data. The most representative example is the discovery of a novel parvovirus-like hybrid genome, which was described as a possible new causative agent of seronegative hepatitis [9] and was in fact a contaminant found in Qiagen nucleic acid extraction columns [10,11]. Other studies have reported viral contaminants from extraction columns including circoviruses/densoviruses and iridoviruses [12] or from reagents used for random amplifications including plant viruses [13,14]. To address this issue, it is recommended to sequence in parallel of samples, a negative control, from which the abundance and the nature of reads is subtracted by the bioinformatics pipeline. This measure is only available if samples and negative controls (e.g. water samples) are prepared with reagents from a same batch number. On the other hand, if this measure, expensive given the cost of sequencing, is fully justified for microbiome analyses, this one can be discussed for viral metagenomics. Indeed, viral contaminant sequences from reagents and environment being much less abundant than bacterial contaminants, it is less expensive and equally effective, to investigate the presence of the detected viral sequence(s) using a specific PCR assay in negative extraction and random amplification controls. Contaminations occurring after ligation of adapters onto the insert fragment do not constitute an issue since they cannot be sequenced and then, detected. The technique is also particularly propitious to cross-contaminations between samples and especially with samples containing high viral-titers. Using different barcoded adapters for each sample can help to