Biochemical and Biophysical Research Communications 436 (2013) 525–529

Contents lists available at SciVerse ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Viral categorization and discovery in human circulation by transcriptome sequencing

Weihua Wang [a,c,1], Xiaoan Zhang [a,1], Yanjuan Xu [a], Adrian M. Di Bisceglie [a,b,*], Xiaofeng Fan [a,b,*]

[a] Division of Gastroenterology and Hepatology, Department of Internal Medicine, Saint Louis University School of Medicine, St. Louis, MO 63104, USA
[b] Saint Louis University Liver Center, Saint Louis University School of Medicine, St. Louis, MO 63104, USA
[c] Wuhan Center for Tuberculosis Control, Wuhan 430030, Hubei, China

ARTICLE INFO

Article history:
Received 17 May 2013
Available online 11 June 2013

Keywords:
Viral discovery
Next-generation sequencing
Multiple displacement amplification

ABSTRACT

Serum is the most common and easily accessible patient specimen in a minimally invasive manner. As a biological resource, RNA in serum has been less explored for its clinical utilization due to prevailing concerns regarding its high degradable nature. In the current study, however, we have documented the use of human serum RNA for viral categorization and discovery through transcriptome sequencing and analysis using well-curated databases and advanced bioinformatic tools. Such an integrated approach may have an immediate application in any clinical situations concerning with viral etiology.

1. Introduction

In many clinical situations, viral infection is frequently suspected as an etiological factor. Yet unless there is an explicit viral candidate, the confirmation of putative viral infections is a difficult task. Historically viral categorization and discovery is essentially a technology-driven process. When candidate viruses are not readily grown *in vitro*, the detection of virus-encoded products or viral genomes becomes the only choice. In this setting, many methods have been developed with a focus on the high throughput nature, such as immune-based library screening [1], mass spectrometry [2], microarray [3] and next-generation sequencing (NGS) [4–6]. Among them NGS represents the most attractive approach due to its large dynamic range for gene detection and the independence of any viral sequence information [7,8]. Indeed, by taking advantage of complete decipherment of human genome sequences, a NGS-based approach, named transcriptome subtraction, had been developed and achieved initial success [4]. However, most studies, if not all, use human tissues as a starting material. In practice, tissue is not readily accessible or feasible in situations where there is no explicit target for a suspicious viral infection. Similarly, in a "hit and run" infection mode [9], there is a very narrow time window for tissue sampling. In the current study, by the integration of an enhanced amplification technique and advanced bioinformatic tools, we present a robust, sensitive and simplified NGS-based method that uses human serum as a biological source for viral categorization and discovery.

2. Materials and methods

2.1. Serum samples

In the current study, hepatitis C virus (HCV), one of the medically important RNA viruses with a single stranded RNA genome approximately at 9600 base pairs [2], was used as a model viral agent for both optimization and validation of experimental protocols.

Serum sample #1709, from a patient with chronic HCV infection, was available at large volume that allowed extensive experimental optimization. Additional serum samples, either HCV-negative or positive, were collected from patients at the Saint Louis University Hospital liver clinic. Informed consent and institutional review board approval were obtained prior to the study. All samples were stored at −80 °C until use.

2.2. Measurement of serum RNA concentration

Total RNA was extracted from 140 μL serum and eluted into 60 μL Tris buffer (pH 8.5) using QIAamp Viral RNA Mini kit (Qiagen). RNA concentration was measured with Qubit RNA BR Assay Kit in the Qubit 2.0 Fluorometer (Life Technologies). Measurement for each RNA sample was repeated three times and the mean values were used to calculate total RNA concentration in corresponding serum samples.

* Corresponding authors. Address: Division of Gastroenterology and Hepatology, Department of Internal Medicine, Saint Louis University School of Medicine, 3635 Vista Avenue, St. Louis, MO 63110, USA. Fax: +1 314 577 8125.
E-mail addresses: dibiscam@slu.edu (A.M. Di Bisceglie), fanx@slu.edu (X. Fan).
[1] These authors contributed equally to this work.

0006-291X/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.bbrc.2013.05.139

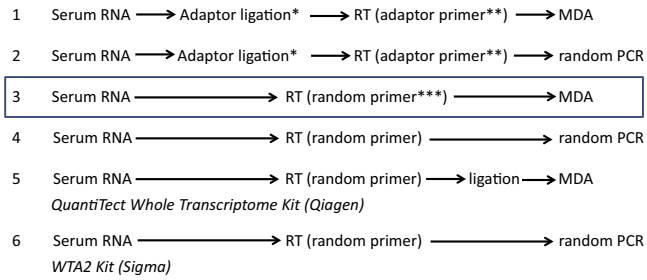

Fig. 1. A brief summary of amplification strategies. Efficient amplification of total serum RNA was estimated by six approaches, including two commercial kits (#5 and #6). The final product from each protocol was examined for robust PCR detection of HCV 5′UTR region. Only protocol #3 (framed) met this standard and was subjected to next-step optimization. *RNA ligation used a highly efficient adaptor recently developed in our lab [10]; **Adaptor primer represents the 5′ part of the adaptor [10]; ***Random primer is an exonuclease-resistant hexamer (Fidelity Systems). RT, reverse transcription; MDA, multiple displacement amplification.

2.3. Unbiased cDNA amplification from serum samples

Due to a low concentration of serum RNA, an amplification step after RT is necessary prior to NGS. There are currently no existing protocols that demonstrate an unbiased amplification from extracted serum RNA, an extremely heterogeneous sample type. In the current study, such an unbiased amplification was achieved through a two-step optimization strategy, the determination of the best approach and a further optimization of the defined approach.

2.4. Approaches for unbiased serum cDNA amplification

A total of six experimental approaches, including two commercial kits, were empirically decided to estimate their ability for an unbiased amplification of serum cDNA (Fig. 1). Approaches #1 and #2 had a ligation step prior to RT, which was accomplished with a powerful adaptor Linker 2 as we described previously [10]. An aliquot of 5 μL ligation product was then mixed with 15 μL RT matrix to formulate RT reaction containing 1× Mg^{2+}-free SuperScript III buffer, 5 mM DTT, 1 mM dNTPs (New England Biolabs), 16 U of Rnasein (Promega), 1 mM reverse primer HBVR1linker2 [10] and 200 U SuperScript III (Life Technologies), followed by 1-h incubation at 50 °C.

The RT product was used either for multiple displacement amplification (MDA) (approach #1) or random PCR (approach #2) (Fig. 1). The MDA was conducted in a 50-μL reaction volume consisting of 5 μL RT product, 1× phi29 DNA polymerase reaction buffer, 1 mM dNTPs, 30 U of phi29 DNA polymerase (New England Biolabs). The reaction was incubated at 30 °C for 12 h.

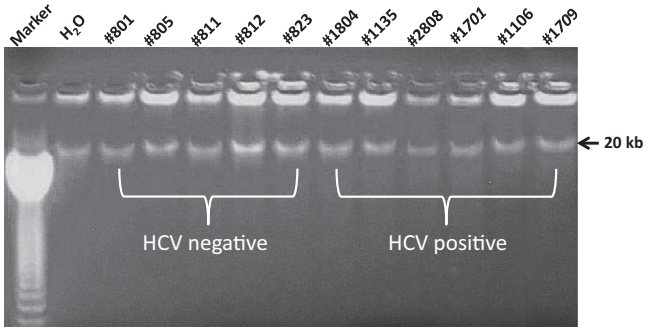

Fig. 2. A representative run of RT/MDA product from sample #1709 and additional 10 samples with or without HCV infection. A dominant band at approximate 20 kb was consistently observed for each sample as well as the negative control (H_2O).

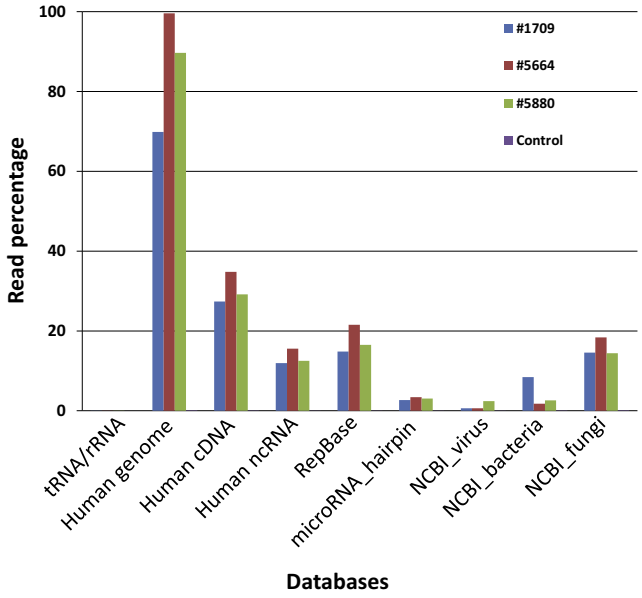

Fig. 3. Read mapping against databases by Bowtie 2 under default mapping score setting. Reads from each sample were first processed with quality control and duplicate removal. All three samples showed generally comparable mapping rates to individual databases. The negative control (H_2O) had almost negligible mapping against databases. Unlike cell-based transcriptome data, tRNA/rRNA was almost undetected. Mature microRNAs were not detected due to the size exclusion (<50 bp) in serum RNA extraction by QIAamp Viral RNA Mini kit.

In random PCR approach, entire RT product was mixed with 30-μL PCR matrix containing 3 μL of 10× DyNAzyme buffer, 0.4 μL of forward primer random 656 (5′-tac agc cta ctc cca tct ctc cac cn_7t ggc-3′) and 1 U DyNAzyme EXT DNA polymerase (ThermoFisher Scientific). Cycle parameters on DNA 480 cycler were adapted from our long RT-PCR technique except for the extension time reduced to 1 min [11]. An aliquot of 2 μL 1st round product was used for the 2nd PCR with primers 3HBVF1 and 3HBVR1 [11].

Approaches #3 and #4 were respectively similar to #1 and #2 except for the omission of RNA ligation step (Fig. 1). Two commercial kits, QuantiTect Whole Transcriptome Kit (Qiagen) and WTA2 Kit (Sigma), were also included for the estimation of unbiased cDNA amplification. The major experimental steps were outlined (Fig. 1) and their performance was basically according to instructions from the manufacturers.

The approach with robust detection of HCV 5′ non-translated domain (5′UTR) from repeated experiments was progressed into next-step optimization. Based on this selection criteria, the RT/MDA protocol (approach #3) was subjected to further optimization for an unbiased amplification of all components in serum RNA samples, as indicated by simultaneous detection in their final products for four HCV genes over entire HCV genome, i.e., 5′ NTR, Core, NS3 and NS5a (Table S1). Experimental optimization focused on the adjustment of three major parameters, including time length (from 4 to 22 h) of MDA incubation, primer concentrations and reaction buffers.

2.5. Pyrosequencing

Using the optimized RT/MDA protocol, 10 μL of extracted serum RNA was used as starting material for cDNA amplification in sample #1709 and two additional HCV-negative samples, #5664 and #5880. A negative control (water) was also included. The RT/MDA product was purified using QIAamp DNA Mini Kit (Qiagen)