



# Unbiased Deep Sequencing of RNA Viruses from Clinical Samples

## Citation

Matranga, Christian B., Adrienne Gladden-Young, James Qu, Sarah Winnicki, Dolo Nosamiefan, Joshua Z. Levin, and Pardis C. Sabeti. 2016. "Unbiased Deep Sequencing of RNA Viruses from Clinical Samples." *Journal of Visualized Experiments : JoVE* (113): 54117. doi:10.3791/54117. <http://dx.doi.org/10.3791/54117>.

## Published Version

doi:10.3791/54117

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:29407858>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

## Share Your Story

The Harvard community has made this article openly available.  
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

## Video Article

# Unbiased Deep Sequencing of RNA Viruses from Clinical Samples

Christian B. Matranga<sup>1</sup>, Adrienne Gladden-Young<sup>1</sup>, James Qu<sup>1</sup>, Sarah Winnicki<sup>1</sup>, Dolo Nosamiefan<sup>1</sup>, Joshua Z. Levin<sup>1</sup>, Pardis C. Sabeti<sup>1,2</sup>

<sup>1</sup>Broad Institute of MIT and Harvard

<sup>2</sup>Harvard University

Correspondence to: Christian B. Matranga at [matranga@broadinstitute.org](mailto:matranga@broadinstitute.org)

URL: <http://www.jove.com/video/54117>

DOI: [doi:10.3791/54117](https://doi.org/10.3791/54117)

Keywords: Medicine, Issue 113, RNA viruses, Ebola virus, Lassa virus, intra-host variants, Lassa fever, poly(rA) carrier, rRNA, RNase H, RT-PCR

Date Published: 7/2/2016

Citation: Matranga, C.B., Gladden-Young, A., Qu, J., Winnicki, S., Nosamiefan, D., Levin, J.Z., Sabeti, P.C. Unbiased Deep Sequencing of RNA Viruses from Clinical Samples. *J. Vis. Exp.* (113), e54117, doi:10.3791/54117 (2016).

## Abstract

Here we outline a next-generation RNA sequencing protocol that enables *de novo* assemblies and intra-host variant calls of viral genomes collected from clinical and biological sources. The method is unbiased and universal; it uses random primers for cDNA synthesis and requires no prior knowledge of the viral sequence content. Before library construction, selective RNase H-based digestion is used to deplete unwanted RNA — including poly(rA) carrier and ribosomal RNA — from the viral RNA sample. Selective depletion improves both the data quality and the number of unique reads in viral RNA sequencing libraries. Moreover, a transposase-based 'tagmentation' step is used in the protocol as it reduces overall library construction time. The protocol has enabled rapid deep sequencing of over 600 Lassa and Ebola virus samples-including collections from both blood and tissue isolates-and is broadly applicable to other microbial genomics studies.

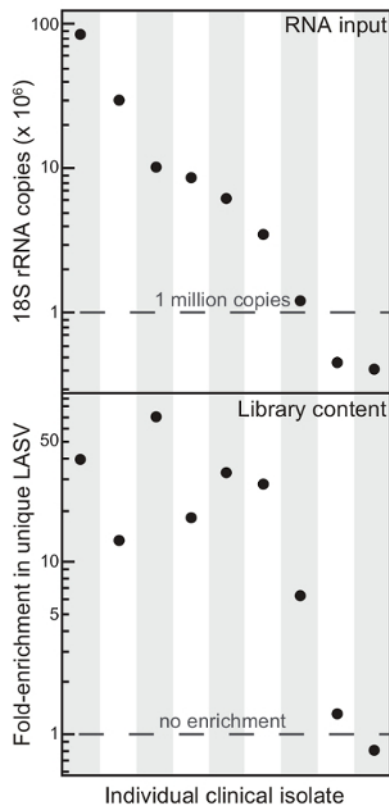
## Video Link

The video component of this article can be found at <http://www.jove.com/video/54117/>

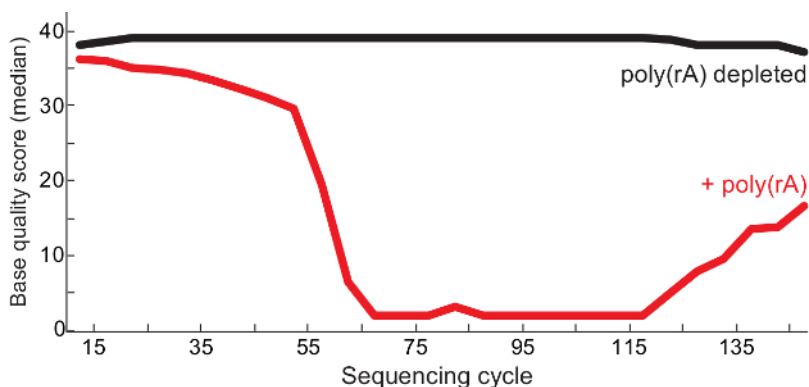
## Introduction

Next generation sequencing of viruses from clinical sources can inform transmission and the epidemiology of infections, as well as help support novel diagnostic, vaccine and therapeutic development. cDNA synthesis using random primers has allowed the detection and assembly of genomes from divergent, co-infecting or even novel viruses<sup>1,2</sup>. As with other unbiased methods, unwanted contaminants occupy many sequencing reads and negatively impact sequencing results. Host and poly(rA) carrier RNA are contaminants present in many existing viral sample collections.

The protocol describes an efficient and cost-effective way of deep sequencing RNA virus genomes based on unbiased total RNA-seq. The method utilizes an RNase H selective depletion step<sup>3</sup> to remove unwanted host ribosomal and carrier RNA. Selective depletion enriches for viral content (**Figure 1**) and improves the overall quality of sequencing data (**Figure 2**) from clinical samples. Moreover, tagmentation is applied to the protocol as it significantly reduces library construction time. These methods have been used to rapidly generate large datasets of Ebola and Lassa virus genomes<sup>2,4,5</sup> and can be used to study a wide range of RNA viruses. Lastly, the approach is not limited to human samples; the utility of selective depletion was demonstrated on tissue samples collected from Lassa-infected rodents and non-human primate disease models<sup>5,6</sup>.



**Figure 1. Total RNA Content Reflects Enrichment of Lassa Virus Content Using Selective Depletion.** Starting overall content (RNA input) and enrichment of unique Lassa virus (LASV) reads (Library content) upon rRNA depletion from nine different clinical isolates. This figure has been modified from<sup>6</sup>. [Please click here to view a larger version of this figure.](#)



**Figure 2. Higher Quality Sequencing After Carrier RNA Depletion.** Median base qualities per sequencing cycle of poly(rA)-contaminated Lassa virus libraries (red) and control (no carrier observed in library, black) from QC report<sup>13</sup>. Both read 1 and read 2 of paired end reads are merged in the library BAM file and the quality scores are shown at each base. This figure has been modified from<sup>6</sup>. [Please click here to view a larger version of this figure.](#)

The viral RNA-seq protocol details construction of libraries directly from extracted RNA collected from clinical and biological samples. To ensure personal safety, all viral serum, plasma and tissue samples should be inactivated in appropriate buffers prior to RNA extraction. In some inactivation and extraction kits, carrier poly(rA) RNA is included; this will be removed during the initial RNase H selective depletion step. Based on complete recovery, the expected concentration of carrier RNA is 100 ng/μl. In the protocol, 110 ng/μl oligo dT RNA (1.1x carrier concentration) is used for depletion. If poly(rA) carrier is not present in the sample, then oligo(dT) should not be added prior to depletion.

The following protocol is designed for 24 reactions in PCR plate format (up to 250 μl volume). An earlier version of this protocol was reported in Matranga, *et al.*<sup>6</sup>.

## Protocol

Ethics statement: Lassa fever patients were recruited for this study using protocols approved by human subjects committees at Tulane University, Harvard University, Broad Institute, Irrua Specialist Teaching Hospital (ISTH), Kenema Government Hospital (KGH), Oyo State

Ministry of Health, Ibadan, Nigeria and Sierra Leone Ministry of Health. All patients were treated with a similar standard of care and were offered the drug Ribavirin, whether or not they decided to participate in the study. For Lassa fever (LF) patients, treatment with Ribavirin followed the currently recommended guidelines and was generally offered as soon as LF was strongly suspected.

Due to the severe outbreak for Ebola Virus Disease (EVD), patients could not be consented through our standard protocols. Instead use of clinical excess samples from EVD patients was evaluated and approved by Institutional Review Boards in Sierra Leone and at Harvard University. The Office of the Sierra Leone Ethics and Scientific Review Committee, the Sierra Leone Ministry of Health and Sanitation, and the Harvard Committee on the Use of Human Subjects have granted a waiver of consent to sequence and make publically available viral sequences obtained from patient and contact samples collected during the Ebola outbreak in Sierra Leone. These bodies also granted use of clinical and epidemiological data for de-identified samples collected from all suspected EVD patients receiving care during the outbreak response. The Sierra Leone Ministry of Health and Sanitation also approved shipments of non-infectious, non-biological samples from Sierra Leone to the Broad Institute and Harvard University for genomic studies of outbreak samples.

## 1. DNase-treatment of Sample RNA (Up to 55 µl Extracted Total RNA, ~4 hr)

- Set up the DNase reaction in a 96-well PCR plate on ice in a biosafety cabinet as described in **Table 1**, Step 1.1 (total volume, 70 µl/well).  
Note: A master mix can be prepared.
- Vortex gently and thoroughly, then centrifuge at 280 x g at RT for 1 min.
- Incubate at 37 °C for 30 min.
- Cleanup using RNA Solid Phase Reversible Immobilization (SPRI) beads.
  - Warm RNA beads to RT for 30 min.
  - Gently shake RNA beads bottle to resuspend any magnetic particles that may have settled. Add 1.8x volume (126 µl) of RNA beads to DNase-treated RNA (70 µl), mix by pipette 10 times and incubate for 5 min at RT (total volume in well, 196 µl).
  - Place mixture on the magnetic station. Wait for the solution to clear (5 - 10 min).
  - Remove cleared solution while on the station by pipette and discard. While on station, wash beads by covering pellet with 70% ethanol and incubate for 1 min. Remove ethanol with pipette and discard. Repeat for a total of two washes.  
Note: Using precisely 70% freshly prepared ethanol is critical, as a higher percentage will result in inefficient washing of smaller-sized molecules, whereas <70% ethanol could cause loss of sample<sup>7</sup>.
  - Keep plate on the station and leave open to air-dry. Note: Be sure to allow the beads to dry completely until beads begin to crack.
  - Add 55 µl of nuclease-free water to the plate to elute RNA. Remove plate from the station to mix the beads and water by pipetting thoroughly. Note: Alternatively, use less water (≤ 10 µl) in order to concentrate the total RNA.
  - Place plate back on the station. Wait until solution clears to transfer by pipette to new screw-cap tube for long-term storage (-80 °C). Place 5 µl RNA in new 96-well PCR plate for depletion (step. 2.4).
  - Optional: Save and dilute 1 µl in 19 µl water (1:20) for qRT-PCR of rRNA (e.g., 18S, 28S rRNA) (**Table 2**) and viral markers<sup>5</sup>.

## 2. Selective Depletion of Ribosomal and Carrier RNA from Viral RNA Sample (~4 hr)

- Make 5x hybridization and 10x RNase H reaction buffers, and nuclease-free water with linear acrylamide carrier as described in **Table 1**.
- Set up hybridization reaction by combining RNA with rRNA depletion oligos (**Table 3**) and oligo(dT) on ice in a 96-well PCR plate as described in **Table 1**.  
Note: A master mix can be prepared. 50 femtograms (fg) of a unique synthetic RNA (ERCCs<sup>8</sup>) can be added for tracking both the viral sequencing process and potential index read cross-contamination.
  - Vortex gently and thoroughly, then centrifuge at 280 x g at RT for 1 min.  
Incubate at 95 °C for 2 min, slow ramping to 45 °C at -0.1 °C per sec. Pause the thermocycler at 45 °C.
- Set up RNase H reaction mix on ice as described in **Table 1**, then preheat at 45 °C for 2 min. Note: A master mix can be prepared.
  - Add the pre-heated RNase H mix to the hybridization reaction in plate while keeping the plate in the thermocycler at 45 °C.
  - Mix well by gentle pipetting 6 - 8 times. Incubate at 45 °C for another 30 min. Place on ice.
- Set up the DNase reaction mix on ice as described in **Table 1**. Note: A master mix can be prepared.
  - Add to the RNase H reaction in plate, vortex gently and thoroughly, then centrifuge at 280 x g at RT for 1 min. Incubate at 37 °C for 30 min.
  - Stop DNase reaction by adding 5 µl 0.5 M EDTA. Vortex gently and thoroughly, then centrifuge at 280 x g at RT for 1 min.
- Cleanup using RNA beads (see step 1.3) using a 1.8x volume (144 µl) beads. Elute in 11 µl of nuclease-free water. Note: For safe cold storage, store depleted RNA samples at -80 °C O/N.

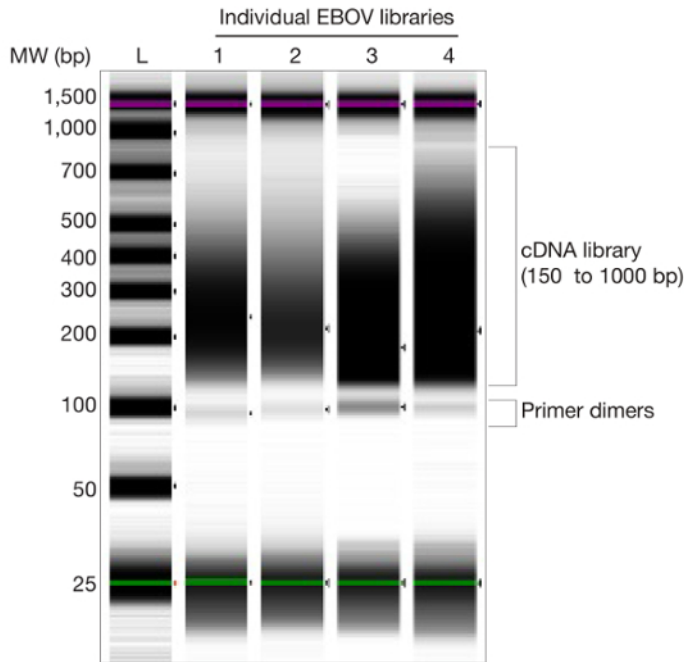
## 3. cDNA Synthesis (~6 hr)

- Mix rRNA/carrier-depleted RNA with random primers on ice in a 96-well PCR plate as described in **Table 1**, vortex gently and thoroughly, then centrifuge at 280 x g at RT for 1 min.
  - Heat the mixture to 70 °C for 10 min in a thermocycler. Immediately after heat denaturation, place the RNA on ice for 1 - 5 min. Do not allow the RNA to stand (even on ice) for longer than 5 min prior to the first-strand reaction.
- Set up first-strand synthesis reaction mix on ice as described in **Table 1**.  
Note: A master-mix may be prepared.
  - Add to RNA/random primer mix in plate, vortex gently and thoroughly, then centrifuge at 280 x g at RT for 1 min. Incubate at 22 - 25 °C for 10 min.

2. Incubate at 55 °C in an air incubator for 60 min. Place the plate on ice to terminate the reaction. Note: The use of an air incubator is recommended to create gradual warming of the first-strand reaction during which the primers anneal and the first strand begins to elongate.
3. Set up second-strand synthesis reaction mix on ice as described in **Table 1**.  
Note: A master-mix may be prepared.
  1. Add to the first-strand synthesis reaction in the plate, vortex gently and thoroughly, then centrifuge at 280 x g at RT for 1 min. Incubate for 2 hr at 16 °C (keep lid at 25 °C). Do not allow the temperature to rise above 16 °C.
  2. Place the plate on ice, then inactivate reaction by adding 5 µl of 0.5 M EDTA, mix gently and thoroughly, then centrifuge at 280 x g at RT for 1 min.
4. Cleanup with DNA beads (see step 1.3 for protocol) using 1.8x volume (153 µl) of beads. Elute in 9 µl of elution buffer (EB). Save 1 µl for quantification. Use 1 ng of cDNA for subsequent steps. If cDNA concentration is too low to detect, use 4 µl of cDNA for tagmentation (see step 4.1).
5. For safe cold storage, store double-stranded cDNA at 4 °C O/N or -20 °C for long-term storage.

#### 4. Library Preparation — DNA Library Construction (~4 hr)

1. Transfer 4 µl of cDNA to a 96-well plate and save the remaining cDNA for a second attempt if needed.
2. Set up the tagmentation reaction on ice as described in **Table 1**.  
Note: A master-mix may be prepared. To reduce background and overall cost, the total volume of the tagmentation reaction is reduced from 20 to 10 µl. As cDNA is the limiting factor, the amount of ATM (*i.e.*, transposome) used in the reaction is also reduced to decrease the number of integration sites.
  1. Add tagmentation mix to cDNA in the plate, vortex gently and thoroughly and centrifuge at 280 x g (at RT) for 1 min. Incubate at 55 °C for 5 min, hold at 10 °C.
  2. Once at 10 °C, immediately add 2.5 µl of Neutralize Tagment Buffer (NT) to end the reaction. Mix by pipetting up and down, then centrifuge at 280 x g (at RT) for 1 min.
  3. Incubate at RT for 5 min.
3. Set up PCR amplification reaction on ice as described in **Table 1**.
  1. Vortex gently and thoroughly, then centrifuge at 280 x g at RT for 1 min.
  2. Perform PCR on thermocycler using the conditions described in **Table 1**.  
Note: 12 cycles of PCR are suggested for 1 ng of tagmented cDNA; however, viral clinical samples often have undetectable amounts of cDNA. For low amounts of cDNA (<1 ng), use up to 18 cycles of PCR to create enough library for sequencing.
4. Library preparation — cleanup and pooling for sequencing
  1. Bring sample up to 50 µl with EB.
  2. Cleanup with DNA beads (see step 1.3 for protocol) using 0.6x volume (30 µl) beads. Elute in 15 µl EB.
  3. Determine concentration of library (**Figure 3**) by conducting region analysis (150 to 1,000 bp) using bioanalyzer software<sup>9</sup>, excluding primer dimers (~120 bp) from region analysis. Note: Alternatively, qPCR can be used to quantify libraries<sup>10</sup>.
  4. Pool libraries at the lowest molar concentration of 1 nM or greater. If library is below 1 nM, add a small volume of library to pool (~1x volume of other libraries) to capture sequence information from these libraries.
  5. Cleanup pool with 0.7x DNA beads as outlined above (see step 2). Elute in 15 µl EB. Note: Volume of beads will depend on the final volume of the pool.
  6. Analyze pool<sup>9</sup>. Determine molar concentration by conducting region analysis (150 to 1,000 bp)<sup>9</sup>. Note: Alternatively, qPCR can be used to quantify library pool<sup>10</sup>.
  7. Load sequencer at 10 pM library concentration to generate 101 bp, paired-end reads with dual barcode reads<sup>11</sup>.



**Figure 3. Libraries Constructed from Ebola Virus Clinical Samples.** Gel image of 4 representative Ebola virus (EBOV) libraries. Regions of library and primer dimers are shown. [Please click here to view a larger version of this figure.](#)

## Representative Results

The described protocol enables the generation of high quality sequencing reads from low-input viral RNA samples while enriching for unique viral content. As shown in **Figure 1**, the protocol enriched unique Lassa virus content at least five-fold in all samples (compared to non-depleted controls) with at least one million copies of 18S rRNA (~100 pg total RNA). Likewise, sequencing success also correlated with the amount of virus within a given sample. Using qRT-PCR as a surrogate for viral quantity, samples that contained ~1,000 or more viral genome copies most often created full assemblies (data not shown). Moreover, depletion of poly(rA) carrier reduces homopolymer sequences of A and T in libraries, resulting in cleaner preparations and ensuring better quality sequencing reads (**Figure 2**). Final libraries from low input viral clinical samples often have a broad fragment length from 150 to 1,000 bp (**Figure 3**).

After sequencing, to reduce sample misidentification and crosstalk between libraries within a pool<sup>12</sup>, only index reads with a base quality score of 25 (q25) and ensure zero mismatches are kept during the demultiplexing process. Viral genomes are assembled using a bioinformatics pipeline specific for divergent viruses<sup>2,4-6</sup>. These tools are available at <https://github.com/broadinstitute/viral-ngs> or through commercial cloud platforms<sup>4</sup>.

<b>Step 1.1: DNase reaction</b>	
Reagent	Volume per reaction (µl)
10x DNase buffer	7
Nuclease-free water	6
Extracted viral RNA	55
DNase (2 U/µl)	2
Total volume	70
<b>Step 2.1: 5x Hybridization buffer</b>	
Reagent	Volume for 1 ml (µl)
5 M NaCl	200
1 M Tris-HCl (pH 7.4)	500
Nuclease-free water	300
Total volume	1,000
<b>Step 2.1: 10x RNase H reaction buffer</b>	
Reagent	Volume for 1 ml (µl)
5 M NaCl	200

1 M Tris-HCl (pH 7.5)	500
1 M MgCl <sub>2</sub>	200
Nuclease-free water	500
Total volume	1,000
<b>Step 2.1: Water with linear acrylamide</b>	
Reagent	Volume for 1 ml buffer (μl)
Nuclease-free water	992
Linear acrylamide (5 mg/ml)	8
Total volume	1,000
<b>Step 2.2: Hybridization reaction for selective depletion</b>	
Reagent	Volume per reaction (μl)
5x Hybridization Buffer	2
rRNA-depletion oligo mix (100 μM)	1.22
Oligo(d)T (550 ng/μl)	1
DNase-treated total RNA	up to 5
Spike-in RNA (This is optional)	0.5
Water (with linear acrylamide)	bring up to 10 total
Total volume	10
<b>Step 2.3: RNase H reaction for selective depletion</b>	
Reagent	Volume per reaction (μl)
10x RNase H Reaction Buffer	2
Water (with linear acrylamide)	5
Thermostable RNase H (5 U/μl)	3
Total volume	10
<b>Step 2.4: DNase reaction post selective depletion</b>	
Reagent	Volume per reaction (μl)
10x DNase Buffer	7.5
Water (with linear acrylamide)	44.5
RNase inhibitor (20 U/μl)	1
RNase-free DNase I (2.72 U/μl)	2
Total volume (with RNase H reaction)	75
<b>Step 3.1: cDNA synthesis, random primer hybridization</b>	
Reagent	Volume per reaction (μl)
rRNA/carrier-depleted RNA	10
3 μg random primer	1
Total volume	11
<b>Step 3.2: First strand cDNA synthesis reaction</b>	
Reagent	Volume (μl)
5x First-Strand Reaction Buffer	4
0.1 M DTT	2
10 mM dNTP mix	1
RNase inhibitor (20 U/μl)	1
Reverse transcriptase (add last)	1
Total volume (with RNA above)	20
<b>Step 3.3: Second strand cDNA synthesis reaction</b>	

Reagent	Volume (µl)
RNase-free water	43
10x Second-Strand Reaction Buffer	8
10 mM dNTP mix	3
<i>E. coli</i> DNA Ligase (10 U/µl)	1
<i>E. coli</i> DNA Polymerase I (10 U/µl)	4
<i>E. coli</i> RNase H (2 U/µl)	1
Total volume (with 1 <sup>st</sup> strand reaction)	80
<b>Step 4.2: Tagmentation reaction</b>	
Reagent	Volume (µl)
Amplicon Tagment Mix (ATM)	1
Tagment DNA Buffer (TD)	5
Total volume (with cDNA)	10
<b>Step 4.3: Library PCR reaction</b>	
Reagent	Volume (µl)
PCR Master Mix (NPM)	7.5
Index 1 primer (i7)	2.5
Index 2 primer (i5)	2.5
Total volume (with tagmented cDNA)	25
<b>Step 4.3.2: Library PCR conditions</b>	
72 °C, 3 min	
95 °C, 30 sec	
up to 18 cycles-10 sec at 95 °C, 30 sec at 55 °C, 30 sec at 72 °C	
72 °C, 5 min	
10 °C, forever	

**Table 1: Reaction set-up and buffers.** Step-by-step tables with contents of all buffers and reaction mixes.

Oligo Name	Sequence (5' to 3')
Ebola KGH FW	GTCGTTCCAACAATCGAGCG
Ebola KGH RV	CGTCCCGTAGCTTTRGCCAT
Ebola KULESH FW	TCTGACATGGATTACCACAAGATC
Ebola KULESH RV	GGATGACTCTTTGCCGAACAATC
Lassa SL FW	GTA AGC CCA GCD GYA AAB CC
Lassa SL RV	AAG CCA CAG AAA RCT GGS AGC A
18S rRNA FW	TCCTTTAACGAGGATCCATTGG
18S rRNA RV	CGAGCTTTTAACTGCAGCAACT

**Table 2: qRT-PCR Primers Sequences.** Primers used for measuring host (18S rRNA) and viral (Ebola and Lassa) content. 'KGH' is Kenema Government Hospital in Sierra Leone, where the Ebola primers were tested<sup>2</sup>. 'Kulesh' is the investigator who designed the primer set<sup>14</sup>.

**Table 3: Ribosomal RNA (rRNA) Depletion Oligos.** 195 50-nucleotide long sequences complementary to human rRNA for selective depletion step<sup>6</sup>. [Please click here to download this file.](#)

## Discussion

The outlined approach enables robust, universal, rapid sequencing and was used for sequencing Ebola virus during the 2014 outbreak<sup>2,4</sup>. By coupling selective depletion and cDNA synthesis with tagmentation library construction, the overall process time was reduced by ~2 days from previous adapter ligation methods. More recently, this protocol was employed by international collaborators and others with great success<sup>15,16</sup> and will be deployed to labs in West Africa to support local genomics-based research studies and diagnostics<sup>17</sup>.



The protocol described here uses random primers to prepare cDNA for viral RNA-seq libraries. Unlike previous viral RNA-seq approaches, it requires no a priori knowledge of sequence data or elaborate and time-consuming primer design for a specific virus or clade. The method can be applied to any viral RNA sample. For example, it was used to generate viral content from both Ebola and Lassa samples<sup>6</sup>. The protocol may also be used for host transcriptomic, metagenomic and pathogen discovery sequencing projects<sup>1</sup>.

A critical step of the protocol is targeted RNase H digestion, a high-throughput, low cost method for removing unwanted carrier and host RNA from viral samples. The selective depletion step of the protocol uses many components and requires skill and accuracy. Extra time and care should be taken during the initial setup.

As most clinical serum and plasma samples often have very little nucleic acid material, contamination and sample loss are common. To avoid these issues, special care should be taken when using this protocol. First, RNA is highly susceptible to degradation; therefore all areas should be clean and free of nucleases. Second, to identify samples suitable for use in this protocol, qRT-PCR assays for both host RNA and virus should be used for quantification<sup>5,6</sup>. When comparing input amounts with sequencing results from the protocol, sequencing success (*i.e.*, generation of sufficient data for full viral assembly) correlated with samples that contained at least 100 pg total RNA and 1,000 copies of virus. Third, exposure to environmental sources of nucleic acids should be avoided. The protocol outlined here is done in a biosafety cabinet for safety precautions and for limiting environmental contaminants. Moreover, our group and others have noticed that commercial enzymes may be another source of contaminating bacterial nucleic acids in low input samples<sup>6,18</sup>. Use of a clean workspace (*e.g.*, PCR hood, biosafety cabinet) and negative controls (*e.g.*, water or buffer) will help alleviate and track contamination, respectively. For samples with <100 pg of total RNA, only poly(rA) carrier RNA, not rRNA, should be depleted to ensure high quality sequencing results while limiting loss of material. For very low input samples, cDNA-amplification methods may be more suitable<sup>19</sup>, although poly(rA) carrier should be removed prior to the cDNA synthesis.

The depletion of host rRNA enriches for viral content in sequencing libraries and is applicable to different sample collections including serum or plasma, and multiple types of tissues from rodents and non-human primates<sup>5,6</sup>. In non-human organisms, reads aligning to 28S rRNA remained after depletion, suggesting 28S rRNA is less conserved between humans and other species<sup>6,20</sup>. When using this method with non-human isolates, it may be necessary to supplement with DNA oligos complementary to the divergent rRNA sequences of the specific host<sup>3,21</sup>.

Since the protocol is unbiased, viral reads may represent only a small fraction of total library content. Although rRNA is the most abundant species of host RNA and only a small percentage of rRNA reads (<1%) are found after selective depletion, all other host RNA (*e.g.*, mRNA) will remain after depletion and may account for many sequencing reads from the sample. Therefore "oversampling" (*i.e.*, oversequencing) individual libraries is required in order to have enough coverage for viral assembly and variant calls. For our studies, we attempt to sequence ~20 million reads per sample to have enough depth for analysis of viral genomic and associated variants as well as metagenomic content<sup>2,5</sup>. For metagenomic and pathogen discovery studies, it is important to note that contaminating host DNA is removed by DNase digestion. Therefore viruses and other pathogens that contain DNA genomes may be lost during the process, however RNA intermediates may still be sequenced.

## Disclosures

The authors have no competing financial interests.

## Acknowledgements

This work has been funded in part with Federal funds from the National Institutes of Health, Office of Director, Innovator (No.: DP2OD06514) (PCS) and from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Contracts (No:HHSN272200900018C, HHSN272200900049C and U19AI110818).

## References

1. Stremlau, M. H. *et al.* Discovery of novel rhabdoviruses in the blood of healthy individuals from West Africa. *PLoS Negl Trop Dis* **9**, e0003631 (2015).
2. Gire, S. K. *et al.* Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science* **345**, 1369-1372 (2014).
3. Morlan, J. D., Qu, K., & Sinicropi, D. V. Selective depletion of rRNA enables whole transcriptome profiling of archival fixed tissue. *PLoS One*. **7**, e42882 (2012).
4. Park, D. J. *et al.* Ebola Virus Epidemiology, Transmission, and Evolution during Seven Months in Sierra Leone. *Cell* **161**, 1516-1526 (2015).
5. Andersen, K. G. *et al.* Clinical Sequencing Uncover Origins and Evolution of Lassa Virus. *Cell* **162**, 738-750 (2015).
6. Matranga, C. B. *et al.* Enhanced methods for unbiased deep sequencing of Lassa and Ebola RNA viruses from clinical and biological samples. *Genome Biol* **15**, 519 (2014).
7. Tang, F. *et al.* RNA-Seq analysis to capture the transcriptome landscape of a single cell. *Nat Protoc* **5**, 516-535 (2010).
8. Jiang, L. *et al.* Synthetic spike-in standards for RNA-seq experiments. *Genome Res* **21**, 1543-1551 (2011).
9. Agilent Technologies, [http://www.agilent.com/cs/library/usermanuals/Public/G2946-90004\\_Vespucchi\\_UG\\_eBook\\_\(NoSecPack\).pdf](http://www.agilent.com/cs/library/usermanuals/Public/G2946-90004_Vespucchi_UG_eBook_(NoSecPack).pdf). (2015).
10. Kapa Biosystems, <https://www.kapabiosystems.com/product-applications/products/next-generation-sequencing-2/library-quantification/>. (2015).
11. Illumina Technologies, [https://support.illumina.com/content/dam/illumina-support/documents/documentation/system\\_documentation/miseq/preparing-libraries-for-sequencing-on-miseq-15039740-d.pdf](https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/miseq/preparing-libraries-for-sequencing-on-miseq-15039740-d.pdf). (2015).
12. Kircher, M., Sawyer, S., & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3 (2012).
13. Andrews, S. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. (2010).
14. Trombley, A. R. *et al.* Comprehensive panel of real-time TaqMan polymerase chain reaction assays for detection and absolute quantification of filoviruses, arenaviruses, and New World hantaviruses. *Am J Trop Med Hyg* **82**, 954-960 (2010).

15. Hu, Y. *et al.* Serial high-resolution analysis of blood virome and host cytokines expression profile of a patient with fatal H7N9 infection by massively parallel RNA sequencing. *Clin Microbiol Infect* **21**, 713.e1-4 (2015).
16. Simon-Loriere, E. *et al.* Distinct lineages of Ebola virus in Guinea during the 2014 West African epidemic. *Nature* **524**, 102-104 (2015).
17. Folarin, O. A., Happi, A. N., & Happi, C. T. Empowering African genomics for infectious disease control. *Genome Biol.* **15**, 515 (2014).
18. Blainey, P. C., & Quake, S. R. Digital MDA for enumeration of total nucleic acid contamination. *Nucleic Acids Res.* **39**, e19 (2011).
19. Malboeuf, C. M. *et al.* Complete viral RNA genome sequencing of ultra-low copy samples by sequence-independent amplification. *Nucleic Acids Res* **41**, e13 (2013).
20. Gonzalez, I. L., Sylvester, J. E., Smith, T. F., Stambolian, D., & Schmickel, R. D. Ribosomal RNA gene sequences and hominoid phylogeny. *Mol Biol Evol.* **7**, 203-219 (1990).
21. Adiconis, X. *et al.* Comparative analysis of RNA sequencing methods for degraded or low-input samples. *Nat Methods* **10**, 623-629 (2013).