Trio RNA-Seq: A Simple and Robust Method with Customizable Host Transcript Depletion for Viral Detection in Clinical Samples.

M. Peterson[1], B. Li[1], I.C. Wang[1], S. Huelga[1], D. Amorese[1], D. Dinwiddie[2], K. Schwalm[2], M. Dadlani[3] and N. Hasan[3]
[1]Tecan Genomics, Inc., 900 Chesapeake Drive, Redwood City, CA 94063, USA. [2]Department of Pediatrics, University of New Mexico Health Sciences Center, Albuquerque, NM, USA. [3]CosmosID, Rockville, MD, USA.

INTRODUCTION

Next generation sequencing (NGS) analysis of RNA isolated from clinical samples can reveal the presence of RNA viruses, expression levels of transcriptionally active DNA and RNA viruses, and identify correlations between host gene expression and viral levels, allowing for investigation of host cell changes that occur upon infection. Other current approaches and assays for viral detection in clinical specimens, such as PCR and target capture followed by sequencing, are limited by requiring previous knowledge of target sequences and by the level of multiplexing possible. While traditional RNA-Seq provides unbiased detection of all nucleic acids present in a sample, and therefore hypothesis-free data, it is limited by rather high input requirements (typically 50-100 ng of more of total RNA). In addition, sequencing data of mixed viral/host RNA-seq libraries is typically dominated by host reads. Therefore, without deep sequencing of each RNA-Seq library, viral reads may be missed altogether. We present here a simple, robust, hypothesis-free RNA-seq method that overcomes the above challenges. As little as 500 pg of total RNA is converted to cDNA and amplified with Single Primer Isothermal Amplification (SPIA®). After enzymatic fragmentation and NGS library generation, specific abundant and uninformative host transcripts are targeted for depletion, resulting in a significant reduction of the number of sequencing reads required to achieve viral detection as compared to traditional RNA-Seq methods. In this study, we present data demonstrating the utility of the Trio RNA-Seq workflow to detect viruses in various clinically relevant samples. The negative selection step to remove unwanted sequences, termed AnyDeplete™, is fully customizable, allowing users to target any class of transcript for depletion within their final libraries. Additionally, the SPIA cDNA amplification method has been used extensively for sample preparation for downstream expression arrays, including in HIV virus discovery and characterization (Malboeuf et al., NAR 2013). The combination of these technologies is a powerful tool for viral analysis in clinical samples.

TRIO RNA-SEQ TECHNOLOGY

Trio RNA-Seq incorporates three innovative technologies

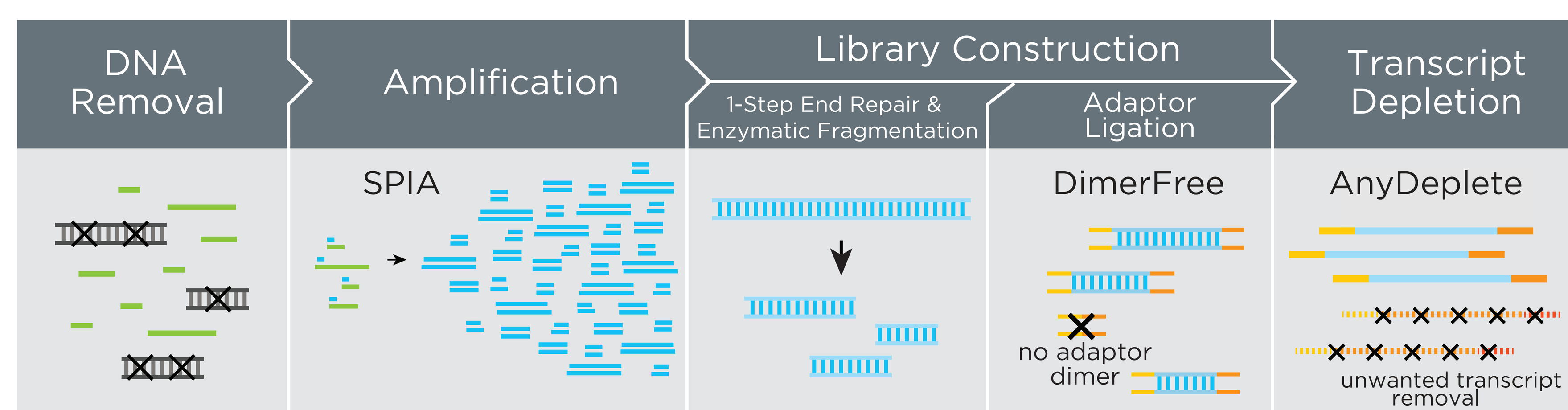

Figure 1: *Tecan's Trio RNA-Seq workflow consists of cDNA generation, SPIA amplification, enzymatic fragmentation, dimer-free barcoded-adaptor ligation, and AnyDeplete targeted depletion.*

METHODS

Sample Collection and RNA Isolation

Nasal or nasopharyngeal samples were collected with a Nylon Flocked Dry Swab and Puritan UniTranz-RT Transport Systems transport media. RNA was isolated with Zymo direct-zol RNA miniprep. Samples were DNase I treated (in-column) and RNA eluted into DNase/RNase-Free Water. Yield ranged between 10ng and 700ng. Residual DNA was detected in several samples. The fraction to be processed by Tecan's Trio RNA-Seq was further DNase treated immediately prior to use.

Reference Method

Libraries were generated using a capture/pull-down target enrichment method. Double stranded cDNA was generated from 5ng-350ng total RNA using random primers. Adaptors were ligated to these fragments and the resulting libraries amplified by PCR. These amplicons were denatured and annealed to biotinylated probes design to represent specific viral genomes. Following capture and elution, the libraries were re-amplified, quantitated and sequenced.

Tecan Method - Trio RNA-Seq

Ten microliters of total RNA was converted to cDNA and amplified with SPIA. The SPIA product (~100-250ng) was enzymatically fragmented followed by adaptor ligation and enriched by PCR (8 cycles). Amplicons were denatured, AnyDeplete probes (human rRNAs, mitochondrial genes) were annealed, extend, and nuclease treated. The remaining, non-targeted libraries, were amplified (8 cycles) and sequenced.

ANYDEPLETE INCREASES DETECTION OF LOW ABUNDANCE TRANSCRIPTS

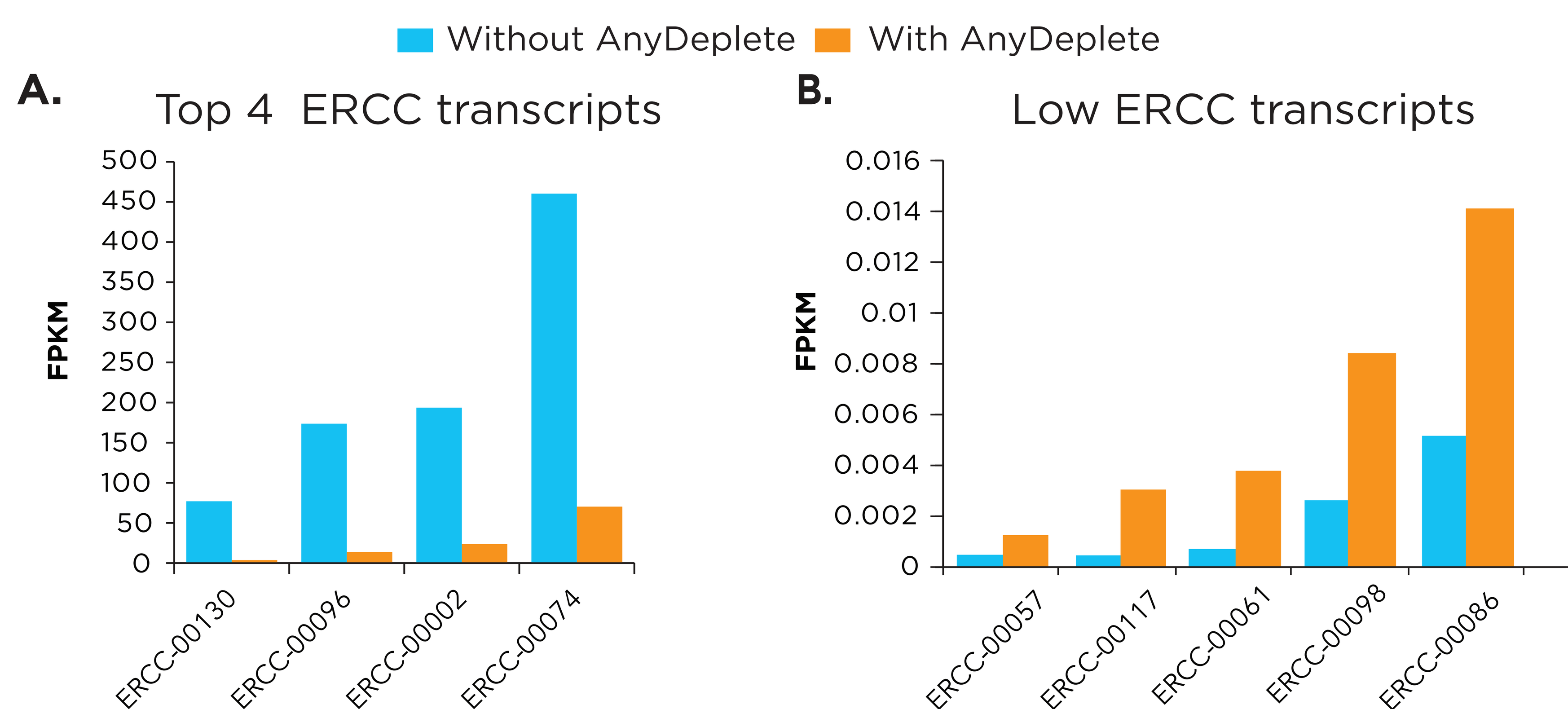

Figure 2: *(A) RNA-seq libraries were generated using 1 ng of ERCC Spike-In Mix 1 (ThermoFisher Scientific). These libraries were then treated with AnyDeplete to target the top four most abundant transcripts, and the libraries were sequenced and compared to a No AnyDeplete control to measure differential detection. Approximately 95% of the target transcripts are removed using the AnyDeplete technology. (B) FPKM values of low abundance, non-targeted ERCC transcripts were compared in the libraries treated with AnyDeplete to the no AnyDeplete control to measure changes in detection rates of low abundance transcripts, revealing a two to four fold increase in detection of low abundance transcripts after targeted transcript removal using the AnyDeplete technology.*

COMPARABLE DETECTION OF TARGET VIRUS BETWEEN TRIO AND REFERENCE METHOD

Sample	Reported Sample conc. (ng/ul)	Virus	Detected by Ref. method	Detected by Trio
Sample 1	Too Low	Human metapneumovirus / Human adenovirus C	✔	✔
Sample 2	Too Low	Human coronavirus OC43 / Human Bocavirus 1	✔	✔
Sample 3	47.5	Human metapneumovirus		✔
Sample 4	78	Human Influenza B		✔
Sample 5	43.1	Human metapneumovirus		✔
Sample 6	51	Respiratory Syncytial Virus B		✔
Sample 7	68.6	Not Detected		
Sample 8	83	Not Detected		
Sample 9	14.7	Human metapneumovirus	✔	✔
Sample 10	3.5	Human Parainfluenza 3	✔	✔
Sample 11	99	Human Parainfluenza 3	✔	✔
Sample 12	8.3	Human coronavirus 229E / Human Bocavirus 1	✔	✔

Table 1: *RNA was extracted from 12 human nasal samples, quantified by Qubit, and processed by both the reference hybrid capture method and the Trio RNA-Seq method. For each sample the RNA concentration as measured by Qubit is shown, along with a virus detected (check) or not detected (empty) by the indicated method. Trio RNA-Seq detected viruses across varying input levels, including cases where the reference method failed.*

TRIO RNA-SEQ ENABLES DISCOVERY

Sample	Human	HMPV	*S. aureus*	*E. coli*	*A. aegypti*	*A. baumannii*
T CNP_0039	92.5%	0.0%	0.0%	0.3%	0.3%	0.3%
T CNP_0041	69.0%	0.0%	13.4%	1.2%	0.6%	2.8%
T CNP_0082	94.8%	0.0%	0.0%	0.0%	0.3%	0.0%
T CNP_0090	90.9%	0.0%	0.0%	0.0%	0.5%	0.0%
T CNP_0107	64.2%	1.2%	0.1%	5.2%	0.7%	3.7%
T CNP_0116	74.9%	0.0%	0.3%	1.2%	0.4%	1.7%
T CNP_0123	50.4%	0.0%	0.8%	2.4%	0.6%	2.9%

Table 2: *Samples from 7 children presenting with respiratory difficulties were negative for all viral targets assayed in the reference method. Unbiased RNA-Seq using the Trio RNA-seq kit allowed for the detection and exploration of additional viral as well as bacterial sequences present in these samples.*

INCREASED DISCOVERY OF CLINICALLY RELEVANT BACTERIAL SPECIES IN TRIO RNA-SEQ LIBRARIES

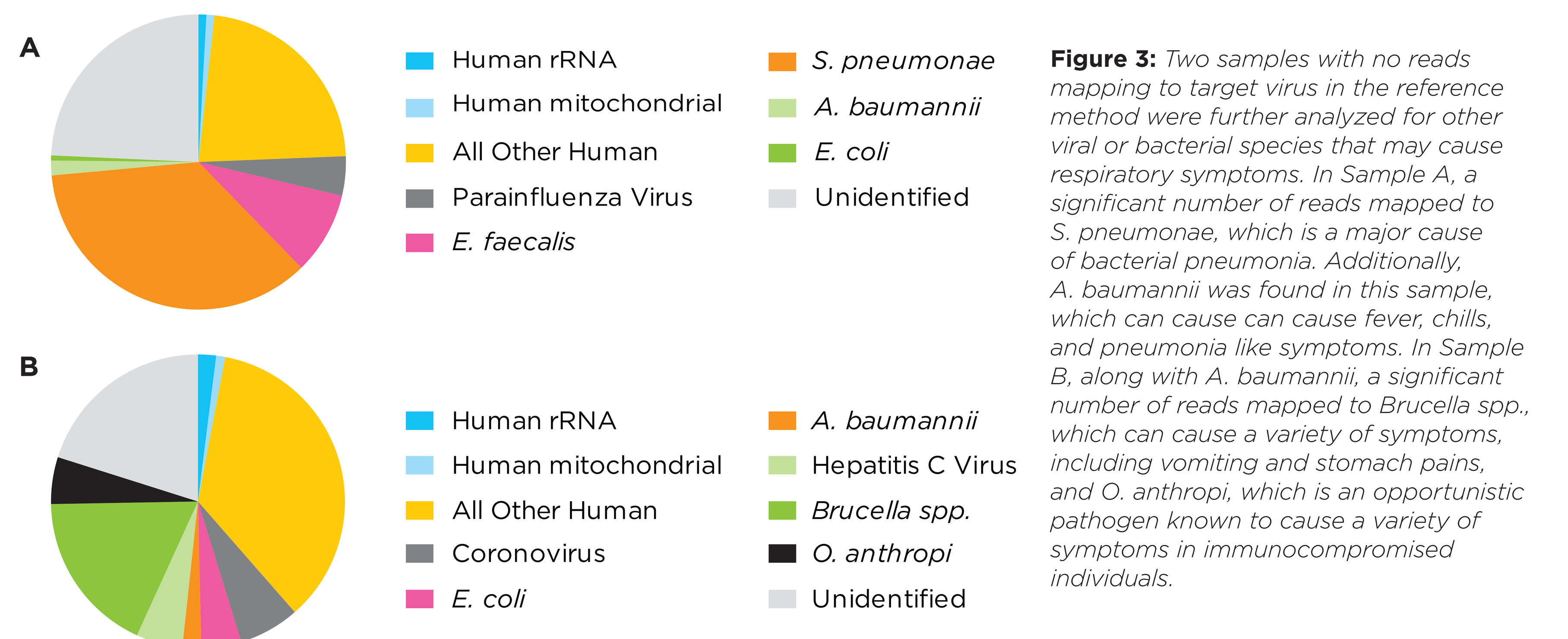

Figure 3: *Two samples with no reads mapping to target virus in the reference method were further analyzed for other viral or bacterial species that may cause respiratory symptoms. In Sample A, a significant number of reads mapped to S. pneumonae, which is a major cause of bacterial pneumonia. Additionally, A. baumannii was found in this sample, which can cause can cause fever, chills, and pneumonia like symptoms. In Sample B, along with A. baumannii, a significant number of reads mapped to Brucella spp., which can cause a variety of symptoms, including vomiting and stomach pains, and O. anthropi, which is an opportunistic pathogen known to cause a variety of symptoms in immunocompromised individuals.*

INCREASED COVERAGE OF VIRAL GENOMES

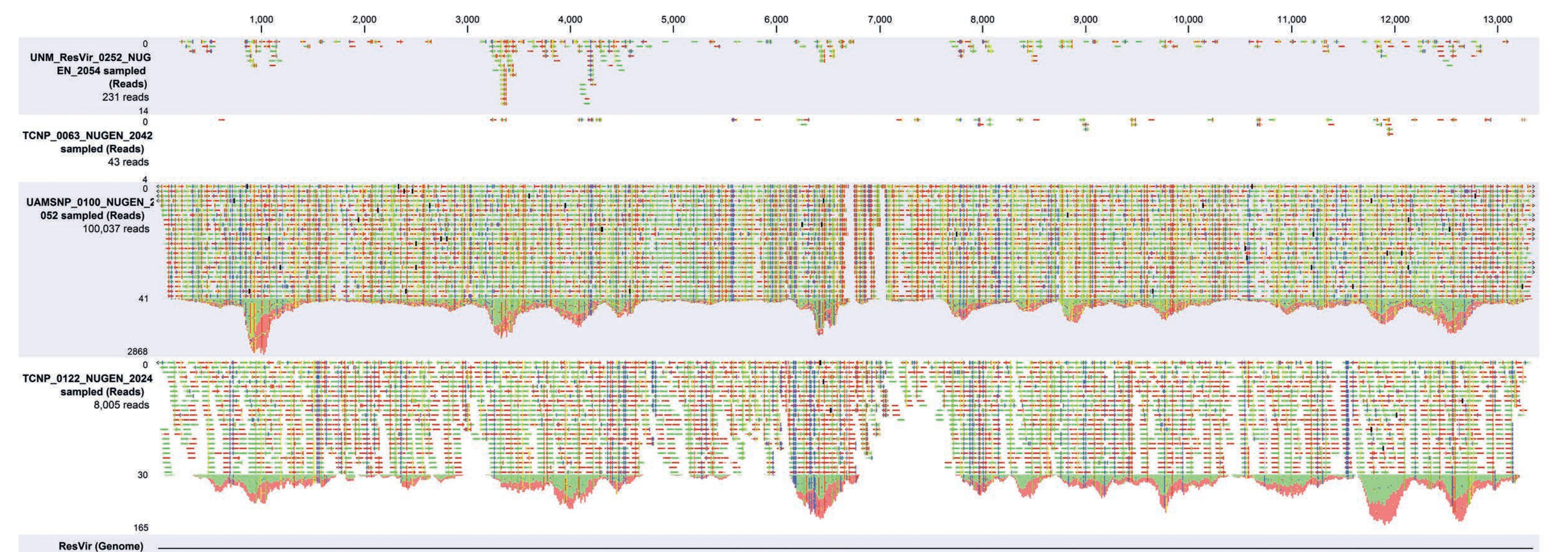

Figure 4: *Six samples were found to contain Human Metapneumovirus including 2 not detected by the reference method. To confirm that these were not a result of sample contamination, reads from these libraries were aligned to a reference virus and viewed in IGV. Four representative traces are displayed. The top 2 traces are from samples with amounts of RNA too low to measure by Qubit, they cover ~45 % and ~14% of the viral genome. While coverage was indeed sparse, reads mapped to several locations and did not appear to align with reads from other libraries or correspond to most abundant reads from libraries having more viral reads. The sample with the best coverage (bottom 2 traces) had the largest amounts of starting material and cover >99% of the genome. The sample in the bottom trace was not detected by the reference method.*

CONCLUSIONS

1) The hypothesis independent Trio RNA-seq method exhibits high correlation to a target capture method for viral discovery in clinical samples, while offering additional discovery opportunities for clinically relevant non-targeted virus and bacterial species.
2) The Trio RNA-seq method produces libraries with a high level of viral genome coverage from minimal RNA input.
3) Removal of abundant host and other uninformative transcripts improves detection of viral sequences.

www.tecan.com

For research use only. Not for use in diagnostic procedures.
For disclaimer and trademarks please visit www.tecan.com · © 2019 Tecan Trading AG, Switzerland, all rights reserved.

TECAN.