

Supplementary Material

Differential microRNA analyses of *Burkholderia pseudomallei*- and *Francisella tularensis*-exposed hPBMCs reveal potential biomarkers

Regina Z. Cer^{1,2*}, J. Enrique Herrera-Galeano^{1,2 γ} , Kenneth G. Frey^{1,2}, Kevin L. Schully^{1,2}, Truong V. Luu^{1,2}, John Pesce^{1,2, \S} , Vishwesh P. Mokashi^{1 \pm} , Andrea M. Keane-Myers^{1,2, \dagger} , Kimberly A. Bishop-Lilly^{1,2}

* **Correspondence** Regina Cer: Regina.Z.Cer.ctr@mail.mil

¹Genomics and Bioinformatics Department, Naval Medical Research Center, Biological Defense Research Directorate, Fort Detrick, MD, USA

²Henry M. Jackson Foundation for the Advancement of Military Medicine, Bethesda, MD, USA

***Correspondence:** Regina Cer

Regina.Z.Cer.ctr@mail.mil

Current address:

^{γ} KCE services and consulting, LLC, Columbia MD, USA

^{\S} Division of Microbiology and Infectious Diseases, National Institute of Allergy and Infectious Diseases, Bethesda, MD, USA

^{\dagger} Immunology, National Institute of Health, Bethesda, MD, USA

^{\pm} Navy Drug Screening Laboratory, Jacksonville, FL, USA

Supplementary Equations

Creation of Time Series

Averaged raw sequence count (avgRSC) of each miRNA from hPBMCs alone at t_0 , referred to as Pt_0 , was subtracted from that at three later time points, 30 min (Pt_{30}), 60 min (Pt_{60}), and 120 min (Pt_{120}) (Eq. 1).

$$Pt_{xt_0} = Pt_x - Pt_0 \quad (1)$$

where:

Pt_0 =avgRSC of miRNAs in hPBMCs alone at t_0

Pt_x =avgRSC of miRNAs in hPBMCs alone at t_x where $x=30, 60, \text{ or } 120$

After Pt_{30t_0} , Pt_{60t_0} and Pt_{120t_0} were calculated, these values were subtracted again from bacterial exposed hPBMCs for the corresponding time points (Eq. 2).

$$Ot_{xt_s} = Ot_x - Pt_{xt_0} \quad (2)$$

where:

O=any of the three organisms, *E. coli*, *B. pseudomallei* or *F. tularensis*

Ot_x =avgRSC of miRNAs in hPBMCs exposed with bacteria O at t_x where $x=30, 60, \text{ or } 120$

Pt_{x,t_0} = obtained in Eq. 1

For example, in the case of *B. pseudomallei*-exposed hPBMCs at time point 60 min (t_{60}), the calculations would be $Bpt_{60t_s} = Bpt_{60} - Pt_{60t_0}$ where $Bp = B. pseudomallei$ and $Pt_{60t_0} = Pt_{60} - Pt_0$.

qPCR confirmatory analysis

For each organism, the fold regulations were obtained between hPBMCs alone versus bacteria-exposed hPBMCs for each time point. For example, fold regulations between hPBMCs alone vs. *F. tularensis*-exposed hPBMCs at t30 (Eq. 3), t60 (Eq. 4), and t120 (Eq. 5) were collected.

$$x = Pt_{30} \text{ vs. } Ot_{30} \quad (3)$$

$$y = Pt_{60} \text{ vs. } Ot_{60} \quad (4)$$

$$z = Pt_{120} \text{ vs. } Ot_{120} \quad (5)$$

Raw C_t values were converted to fold regulations, and data were collected for each experiment at each time point for each organism in the format miRNA = x, y, z where x, y, z are fold regulations calculated from Equations 3, 4, and 5. To perform the appropriate statistical test, a sign base log transformation (SBLT) that involves the integration of the 2^x function producing a function of $1/\log dx$ was implemented. That is, after log transformation was applied to data, they were multiplied by a vector that preserved the numerical sign of the original fold regulation, as shown in Equations 6 and 7.

$$\vec{d} = \{30,60,120\} \quad (6)$$

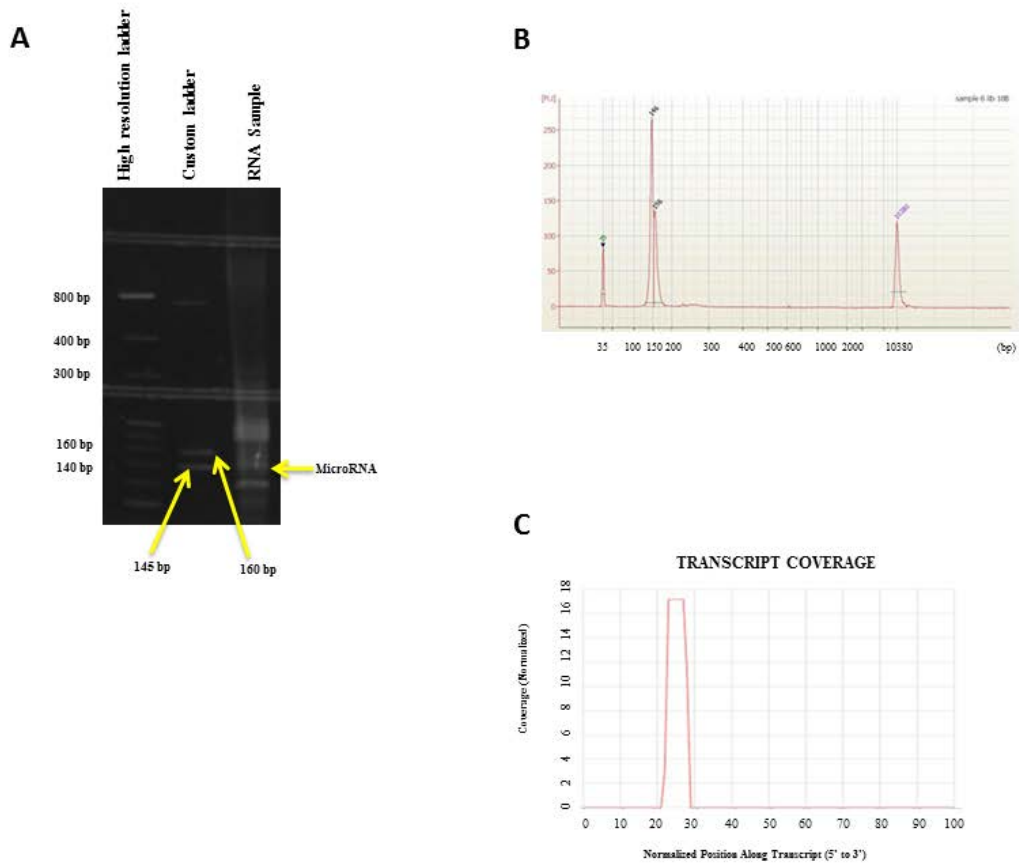
$$t = \left(\int_0^\infty g(|\vec{d}|) \right) * \vec{S} \quad (7)$$

where $\vec{S} = f(\vec{d})$, $g(x) = 2^x$ and $f(x)=1$ if $x>0$, -1 if $x<0$.

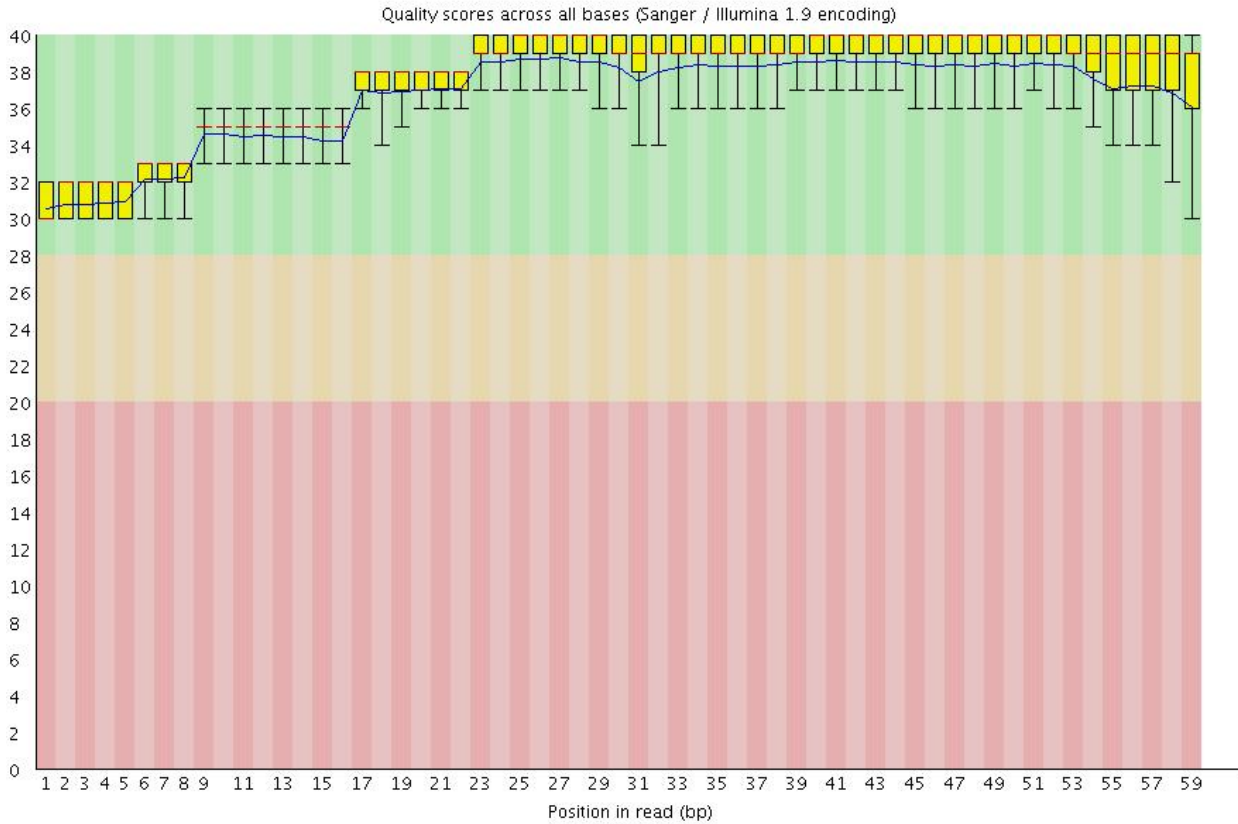
Supplementary Figures



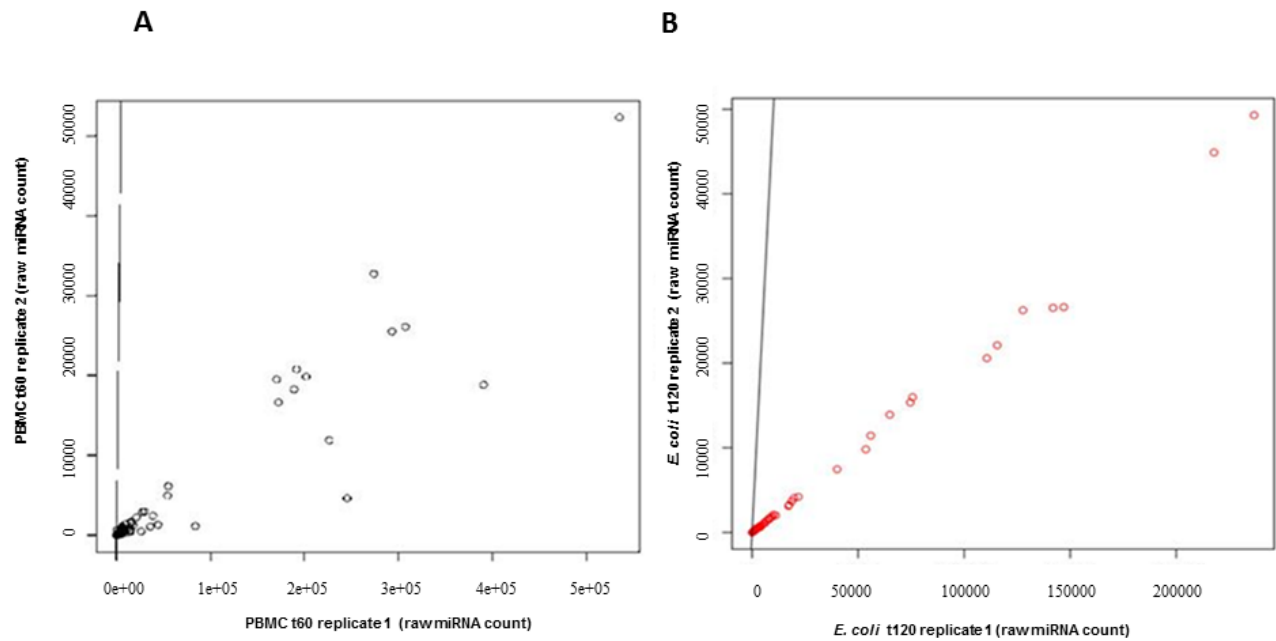
Supplementary Figure 1 | Detailed steps used for processing FASTQ files generated from MiSeq runs. Artificially introduced 3' adapter sequences were trimmed, and post-trimmed reads that were a minimum of 15 base pairs were filtered against contaminants. Reads that did not match to contaminants were screened for mature miRNA species (black box) which were further analyzed for statistical significance.



Supplementary Figure 2 | RNA purity and quality. (A) The microRNA is seen between 145 nt and 160 bp on a 6% Norvex gel. The ladders used are TruSeq High Resolution Ladder and Custom Ladder (B) Agilent trace shows a good range of small RNA from 145 nt-160 nt after ligation (C) Transcript coverage shows the majority of transcript ranges between 20 and 30 bp



Supplementary Figure 3 | FASTQC analysis. An overview of the range of quality values across all bases at each position in the FastQ file. For each position, a BoxWhisker type plot is drawn. The y-axis on the graph shows the quality scores and in this run, the quality scores range from 30 to 40.



Supplementary Figure 4 | Correlation plots of 2 replicates. (A) in hPBMC alone at time point 60 minutes **(B)** *E. coli*-exposed hPBMCs at time point 120 minutes