Tuberculosis xxx (2016) 1–9

Contents lists available at ScienceDirect

Tuberculosis

journal homepage: http://intl.elsevierhealth.com/journals/tube

Sequencing-relative to hybridization-based transcriptomics approaches better define *Mycobacterium tuberculosis* stress-response regulons

Ashley V. Veatch [a, c], Tianhua Niu [d], John Caskey [e], Amanda McGillivray [a], Uma Shankar Gautam [a], Ramesh Subramanian [e], K. Gus Kousoulas [e], Smriti Mehra [b, e], Deepak Kaushal [a, c, *]

[a] Divisions of Bacteriology & Parasitology, Tulane National Primate Research Center, Covington LA, USA
[b] Divisions of Microbiology, Tulane National Primate Research Center, Covington LA, USA
[c] Department of Microbiology & Immunology, Tulane University School of Medicine, New Orleans, LA, USA
[d] Department of Biostatistics & Bioinformatics, Tulane University School of Public Health and Tropical Medicine, New Orleans LA, USA
[e] Department of Pathobiological Sciences, Louisiana State University School of Veterinary Medicine, Baton Rouge, LA, USA

Keywords:
Tuberculosis
Microarray
RNAseq
Correlation
Stress response

A B S T R A C T

Mycobacterium tuberculosis (*Mtb*) infections cause tuberculosis (TB), an infectious disease which causes ~1.5 million deaths annually. The ability of this pathogen to evade, escape and encounter immune surveillance is fueled by its adaptability. Thus, *Mtb* induces a transition in its transcriptome in response to environmental changes. Global transcriptome profiling has been key to our understanding of how *Mtb* responds to the different stress conditions it faces during its life cycle. While this was initially achieved using microarray technology, RNAseq is now widely employed. It is important to understand the correlation between the large amount of microarray based transcriptome data, which continues to shape our understanding of *Mtb* stress networks, and newer data being generated using RNAseq. We assessed how well the two platforms correlate using three well-defined stress conditions: diamide, hypoxia, and re-aeration. The data used here was generated by different individuals over time using distinct samples, providing a stringent test of platform correlation. While correlation between microarrays and sequencing was high upon diamide treatment, which causes a rapid reprogramming of the transcriptome, RNAseq allowed a better definition of the hypoxic response, characterized by subtle changes in the magnitude of gene-expression. RNAseq also allows for the best cross-platform reproducibility.

1. Introduction

Microarrays transformed our ability to study gene expression in a given system and allowed for a greater understanding of cellular systems [1]. With the arrival of whole genome sequences, our ability to study genes and thus transcriptomics was greatly expanded. Short known sequences called expressed sequence tags (ESTs) allowed for a certain number of genes to be probed [2]. This technology was expanded to serial analysis of gene expression (SAGE). SAGE allowed for all mRNA to be transcribed to cDNA, digested, tagged, and then manually and individually sequenced and matched to ESTs or the genome if it was available [3]. Transcriptomic analysis moved towards using ESTs as hybridization probes. The sequences were PCR amplified and affixed chips allowing the hybridization of tagged sequences, which could be identified based on location creating a microarray [4,5]. While the highly parallel analysis of thousands of transcripts at the same time moved the field forward significantly, microarrays still had caveats. These include background noise, low signal intensity, and rigorous statistical analysis, all of which are complicated by low technical and biological replicates. Microarray chips are also limited to the probes affixed to the chip, which reduces the flexibility of analysis to the known genes at the time [1,6,7]. In recent years the field of

Abbreviations: Mtb, *Mycobacterium tuberculosis*; EHR, Enduring Hypoxic Response; rpm, revolutions per minute.

* Corresponding author. Department of Microbiology & Immunology, Tulane National Primate Research Center, 18703 Three Rivers Rd, Covington, LA, USA, 70433.
E-mail address: dkaushal@tulane.edu (D. Kaushal).

http://dx.doi.org/10.1016/j.tube.2016.09.020
1472-9792/

transcriptomics has moved towards RNAseq. This technology addresses several shortcomings of microarrays [8], e.g., the potential to discern single transcript differences and increase the dynamic range [8]. As the field increasingly moves towards RNAseq, it is important to validate whether previous data collected using microarrays will correlate with the new data generated by RNAseq.

Mycobacterium tuberculosis (Mtb), the bacterium that leads to tuberculosis (TB), causes greater mortality than any other infectious agent, with approximately 1.4 million annual deaths [9]. This bacterium is thought to have infected nearly one third of the world's current population, and approximately 9 million individuals are newly infected with *Mtb* each year [10]. Like other highly successful pathogens, *Mtb* has developed a plethora of mechanisms to combat host immune defenses [11–14]. In order to respond quickly to environmental stress, *Mtb* has a higher ratio of sigma factors to genome size than any other obligate human pathogen [11,15]. 11 of the 13 sigma factors encoded by *Mtb* are considered part of the extracytoplasmic function (ECF) subfamily, which specifically responds to the environment [12]. The ECF sigma factors activate regulons to allow *Mtb* to combat host defenses [11]. Microarrays have been used extensively to define the regulatory pathways of *Mtb in vitro* and *in vivo* [1,16–22]. These studies have allowed us to define regulatory pathways such as the SigH regulon [16,17,23], the DosR [24] and the extended hypoxic response (EHR) regulons [20]. The first comprehensive RNAseq-based profiling of a bacterial transcriptome did not occur until 2009 [25]. RNAseq studies of *Mtb* have concentrated on profiling clinical strains [26,27] and non-coding RNA [28] rather than specific *in vitro* stress conditions. In order for us to draw conclusions across studies, we must validate and correlate data obtained from new technology with former technology. While this has been done previously using other systems [29–31], to our knowledge data obtained from *Mtb* using microarrays and RNAseq has not been analyzed for correlation.

In the current study, we validated previously obtained data from *Mycobacterium tuberculosis* (*Mtb*) specific microarray chips to RNAseq. We used microarray data previously generated by one individual and correlated it with RNAseq data currently generated by another individual. The microarray data was published previously [18,32].

2. Materials and methods

2.1. Culture conditions

Mtb CDC1551 cultures were grown using Middlebrook 7H9 agar supplemented with 10% albumin dextrose catalase (ADC), 0.25% glycerol and 0.2% tween-80 for liquid cultures and Middlebrook 7H10 supplemented with 10% ADC, 0.5% glycerol for growth on solid medium. Liquid cultures were incubated at 37 °C at 220 rpm except where noted and colonies were grown for three weeks at 37 °C [18,23,32,33].

2.2. Stress conditions

For redox stress, triplicate cultures were grown to mid-log phase (OD_{260} 0.39–0.41). 25 mL of culture was reserved for RNA extraction for the untreated control and the remaining culture was treated with diamide at a final concentration of 10 mM as previously described [23]. After 1 h of treatment 25 mL of culture was reserved for RNA extraction.

Hypoxic stress was induced using the modified Wayne's model [34]. Briefly, triplicate cultures were grown to mid-log phase (OD_{260} 0.39–0.41) and 25 mL of culture was reserved for RNA extraction for the untreated control. The remaining cultures were transferred to 50 mL conical tubes with a headspace of 0. Cultures were sealed and incubated without disruption at 37 °C for 5 days. After 5 days, 25 mL of culture was reserved for RNA extraction. After 7 days of hypoxia, cultures were transferred to a disposable Erlenmeyer vented flask with a 1:5 vol to headspace ratio and allowed to re-aerate with 220 rpm of shaking for 6 h. 25 mL of culture was reserved for RNA extraction.

2.3. Transcriptomics

For microarrays, RNA was phenol extracted from lysed cells using TRizol (Ambion) followed by bead beating as described previously [23]. 2 mg of RNA was used for *Mtb*-specific DNA microarrays (MYcroarrays, Biodiscovery LLC.) as described previously [23] and analyzed using Spotfire DecisionSite for Microarray Analysis [35].

For RNAseq, cells were lysed using TRizol (Ambion) followed by bead beating. RNA was extracted using RNAeasy column purification kit (QAIGEN). RNA was then depleted of DNA using TURBO DNA-free DNase (Ambion) and confirmed by 40 cycles of PCR. DNA-free RNA was enriched for mRNA by removing rRNA using the RiboZero bacteria kit (Illumina). cDNA libraries were generated using the Ion Total RNAseq kit v2 (Ion Torrent). Ion PI Hi-Q OT2 kit (Ion Torrent) was used for template preparation and Ion PI Hi-Q Sequencing 200 kit (Ion Torrent) was used for sequencing on the Ion Proton system. Sequences were aligned using the CDC1551 reference genome from Pathosystems Resources Integration Center (PATRIC) and analyzed using the R program DESeq2.

2.4. Comparison of microarray and RNAseq

Correlation analyses were conducted using Statistical Analysis Software (SAS) version 9.3 (SAS Institute Inc., Cary, NC) and statistical significance was assessed using Pearson's correlation coefficient (r). Pearson's r was chosen due to the large sample size of the groups. A 2 fold cut off was used to determine genes up- and down-regulated by a particular stress condition for both microarray and RNAseq as compared to baseline. Heat maps were generated using Spotfire. Correlations for regulon data sets were calculated using GraphPad (La Jolla, CA) Prism v6.0. Venn diagrams were generated using the online tool BioVenn (http://www.cmbi.ru.nl/cdd/biovenn/) [36].

3. Results

3.1. Complete profile of all possible genes of Mtb identified by microarray and RNAseq

Three *in vitro* stress conditions were used for this experiment: diamide, hypoxia, and re-aeration. Diamide induces thiol specific oxidation of small molecular weight thiols and proteins [37] and has been used numerous times in the *Mtb* field [17,18,23,38]. The modified Wayne's model has long been considered to recapitulate in-vivo hypoxia [34], while the re-aeration protocol was developed more recently by Sherrid et al. [39]. We subjected the cultures to 1 h of 10 mM diamide treatment, 5 days of hypoxia, and 6 h of re-aeration after 7 days of hypoxia. The duration of 1 h of diamide treatment was chosen from the available microarray time points as the most commonly used in our lab [18,23]. Day 5 of hypoxia was chosen from the previous microarray data because it had the largest number of genes that exhibited perturbed expression using a 2-fold up and down cut-off. The duration of 6 h of re-aeration after 7 days of hypoxia was chosen to match the conditions of the previous experiment and a time at which a large number of genes was perturbed using the 2-fold up and down cut-off [32].

Please cite this article in press as: Veatch AV, et al., Sequencing-relative to hybridization-based transcriptomics approaches better define *Mycobacterium tuberculosis* stress-response regulons, Tuberculosis (2016), http://dx.doi.org/10.1016/j.tube.2016.09.020