Modifications and Additions

The following modifications and additions have been made, in response to the comments raised on this GSCID application:

(1) “There is not a strong justification for the choice of the particular cell line and mouse model for Chlamydia. I am sure there is evidence from published data on Chlamydia why these were chosen, but this justification is not found in this white paper.”

Justification of the chosen in vitro and in vivo culture models

(a) In vitro culture model: HEP-2 cells
Chlamydial species exhibit a broad host range despite significant genomic and gene content similarity. Similarly, Chlamydia infection has been successfully established in vitro in a wide number of host cell types, including macrophages and other immune cells, astrocytes, microglial cells, aortic artery muscle cells and endothelial cells (Moulder 1991; Hafner et al. 2008; Beagley et al. 2009). Localization of chlamydia to a wide number of cell types in vivo has also been reported by many (Moulder 1991; Hafner et al. 2008; Beagley et al. 2009).

Despite the broad range of cells that Chlamydia can infect, mucosal epithelial cells are the primary host cell in vivo. Accordingly, in vitro cell culture systems for Chlamydia use immortalized cell lines of epithelial origin. Epithelial cell lines typically cited in the literature for chlamydial in vitro infection include HEp-2, HeLa, HEC-1B, BGMK, McCoy, Chinese hamster ovary and Vero cells (Maass and Harig 1995; Davis and Wyrick 1997; Suchland et al. 2003; Guseva et al. 2007; Dessus-Babus et al. 2008). Despite these cells all being epithelial in origin, phenotypic differences with respect to infection processes have been reported.

In this application, we propose to use HEp-2 cells, a human laryngeal carcinoma epithelial cell line (Moore et al. 1955), as hosts for the in vitro series of experiments. HEp-2 cells were selected as: (1) chlamydial infection is readily established in vitro using this cell line; (2) the Bavoil laboratory uses HEp-2 cells extensively and (3) HEp-2 cells are widely used in the chlamydial research community. Other epithelial cell lines could be easily substituted without significant impact to the proposed experiments provided the same cell line is used throughout. Future experiments could explore the genotypic reasons for the observed phenotypic variations between the different popular cell lines.

(b) In vivo culture model: C. muridarum/mouse genital infection model
In this application, we propose a novel use of RNA-Seq with the C. muridarum/mouse model of chlamydial infection. Animal models are the primary tools for the study of chlamydial genital infection, particularly for examining the host response to C. trachomatis in females. Models of chlamydial infection have been established in the mouse, guinea pig, non-human primate, non-primate monkey, rat and the pig (Hafner et al. 2008). However the mouse is the preferred model for chlamydial genital infection primarily because of the availability of inbred mouse lines, transgenic and knockout mice, extensive immunological reagents, and the availability of a native murine chlamydial pathogen (Farris and Morrison 2010).

C. muridarum is a native murine pathogen - it was previously termed C. trachomatis MoPn (mouse pneumonitis) (Nigg 1942; Barron et al. 1981). The genomes of C. muridarum and C. trachomatis are virtually identical in gene content and gene order, except for a small ~50kb region around the origin of replication (Read et al. 2000; Read et al. 2003). In addition to the genomic similarity, murine infection of C. muridarum mimics the acute disease course in women (Farris and Morrison 2010). Consequently the C. muridarum mouse model of
chlamydial infection is widely used to examine the immune response to infection and as a vehicle for vaccine development (Farris and Morrison 2010). Thus, the combination of C. muridarum as a suitable analog of chlamydia infections in humans and the general utility of mice as research reagents, make the mouse model of chlamydial infection the most useful mechanism for examining in vivo infection.

(2) “I would recommend that more members of the Chlamydia scientific community be involve to discuss what experiments are top priority for using this technology in light of not unlimited funds. What experiments and data sets would be the highest value to the scientific community to jumpstart research in this area, since this is not an R01 hypothesis driven proposal and output is data sets.”

The above recommendation was raised with senior members of the chlamydial research community with both general chlamydial expertise and specific expertise in transcriptional analysis (see below). Each responded with a Letter of Support (attached), outlining their thoughts on prioritization.

<table>
<thead>
<tr>
<th>Name</th>
<th>Expertise</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professor Patrik Bavoil</td>
<td>chlamydial cell and molecular biology</td>
<td>University of Maryland</td>
</tr>
<tr>
<td>Professor Roger Rank</td>
<td>animal models of chlamydial infection</td>
<td>University of Arkansas for Medical Sciences</td>
</tr>
<tr>
<td>Professor Gerald Byrne</td>
<td>chlamydial cell and molecular biology; transcriptional regulation using high throughput methods</td>
<td>University of Tennessee Health Science Center</td>
</tr>
<tr>
<td>Professor Priscilla Wyrick</td>
<td>chlamydial cell and molecular biology</td>
<td>East Tennessee State University</td>
</tr>
<tr>
<td>Professor Peter Timms</td>
<td>chlamydial genomics and gene regulation</td>
<td>Queensland University of Technology</td>
</tr>
</tbody>
</table>

In addition to the above responses, the term “Matched Pair RNA-Seq” has been replaced with “Heterogenous RNA-Seq” throughout the application, and an Addendum describing an alternative RNA fractionation scheme has been included.

References Cited


Simultaneous Transcriptional Profiling of Obligate Intracellular Pathogens and their Host Cells

Garry S. A. Myers

1Institute for Genome Sciences and Department of Microbiology & Immunology, University of Maryland School of Medicine, 801 West Baltimore Street, Baltimore, Maryland 21201

1. Executive Summary

Whole genome sequencing efforts of the last 15 years have yielded multiple genome sequences of all major bacterial pathogens and several human genome equivalents. Combined with current and pending sequencing capabilities, this data opens the next major sequencing frontier for infectious disease research - deciphering the gene expression dynamics of pathogens and their host cells through deep, comprehensive transcriptional profiling.

Sequencing of cDNA libraries to explore the transcriptional landscape has been used for over twenty years. However, such early approaches were limited by cost, insufficient sequencing depth and experimental biases. The shift from Sanger methods to “next-generation” cDNA sequencing, popularly termed RNA-Seq, enables affordable, comprehensive sequencing of cDNA libraries with minimal bias, generating accurate gene expression profiles of the source cells. RNA-Seq also provides insights into the transcriptome beyond just sequence, yielding splicing and expression level information as well. Examples of the successful application of RNA-Seq are beginning to appear in the literature. However, as whole genome sequencing of bacterial isolates has shown, the true disruptive power of RNA-Seq will come from detailed comparative analyses of datasets that have been specifically designed to answer biological questions and application of bioinformatics tools to address the amount of data.

This GSCID application proposes a novel use of RNA-Seq to capture the gene expression dynamics of host-pathogen interactions in vitro and in vivo by sequencing the transcriptomes of (a) HEp-2 cells infected with Chlamydia spp. under several different conditions and (b) the mouse model of chlamydial infection. Chlamydia are obligate intracellular pathogens of humans and animals, responsible for significant disease worldwide. Chlamydia cannot be grown outside of the host cell and no method of laboratory transformation or genetic manipulation exists. These limitations render Chlamydia intractable to most molecular biology techniques and they are thus ideal candidates for sequencing-based analysis. By applying RNA-Seq to infected cells, an unbiased readout of the gene expression profile of the host cell responding to chlamydial infection at that moment in time, and vice versa, can be obtained. By the very nature of an intracellular pathogen, the total infected cell transcriptome contains the transcriptome of the host cell and the transcriptome of the infecting pathogen obtained simultaneously - termed here as Heterogenous RNA-Seq.

The development of informatics tools designed to handle and visualize substantial datasets has long been a critical feature of the uptake and application of large scale sequencing analysis. Thus, to handle the immense quantity of sequence data generated by Heterogenous RNA-Seq and to rapidly identify those subsets of genes that show expression variation in each defined experiment, we propose to build a Heterogenous RNA-Seq data viewer, using the inVUE toolbox developed at IGS.

In addition to aiding basic research, the detailed view of the chlamydial/host interaction afforded by this approach would contribute significant insights to diagnostic, therapeutic and vaccine development, and surveillance efforts. Moreover, the expertise and tools created are applicable to any pathogen/host cell, providing a consistent and reproducible analysis protocol.
2. Justification
(a) *Chlamydia*

*Chlamydia* are widely distributed bacterial pathogens that cause a significant disease burden in humans. The most prevalent chlamydial pathogen is *C. trachomatis*, which causes trachoma, the primary case of infectious blindness worldwide, and genital tract infections (Thylefors 1995; Gerbase et al. 1998). Infection of the genital tract often results in pelvic inflammatory disease, ectopic pregnancy, chronic pelvic pain, epididymitis, and infant pneumoniae, and also significantly increase the risk for HIV infection. *C. pneumoniae* is responsible for approximately a tenth of pneumonia cases in industrialized countries and more recently, prior infection with *C. pneumoniae* has been linked to an increased risk of developing atherosclerosis and several other chronic diseases such as Alzheimer’s disease (Grayston 1999). Other species of the *Chlamydiaceae* cause a wide range of animal infections: for example, the common avian respiratory disease caused by *C. psittaci*, intestinal diseases caused by *C. pecorum*, or abortions in sheep caused by *C. abortus* (Bodetti et al. 2002). Zoonotic infections are frequent where there is close contact between humans and infected animals; the most well known chlamydial zoonotic infection is psittacosis, acquired from *C. psittaci*-infected birds.

*Chlamydia* are characterized by a distinctive biphasic developmental cycle. Following attachment and invasion of eukaryotic cells, chlamydiae grow within a modified intracellular vacuole (the inclusion), that does not fuse with lysosomes. A metabolically inactive infectious developmental form called the elementary body (EB) that attaches and infects the target host cell, differentiating into a metabolically active developmental form called the reticulate body (RB). After multiple divisions by binary fission, RBs differentiate back into the infectious EB forms, which are then released from the cell to initiate new rounds of infection.

Chlamydial biology and genetics continue to be poorly understood. This is a direct result of two key factors: (1) *Chlamydia* cannot be grown outside of the host cell; and (2) no general method of laboratory transformation or genetic manipulation exists. These limitations render *Chlamydia* intractable to most molecular biology techniques, making sequencing-based analysis a key methodology. Indeed, the data from high-throughput sequencing-based approaches have been the underpinning of virtually all major advances in chlamydial research over the last decade. The continuing inability to genetically modify chlamydiae, combined with the plethora of available genome sequence data and current sequencing capabilities, makes chlamydiae ideal candidates to explore the next major sequencing frontier for infectious disease research - deciphering the gene expression dynamics of pathogens and their host cells through deep, comprehensive transcriptional profiling.

(b) RNA-Seq

Sequencing of RNA has long been recognized as an efficient method for gene discovery and remains the gold standard for annotation of both coding and noncoding genes. However, while transcript sequencing has been possible for nearly 20 years, until recently it required the construction of clone libraries and was time consuming and expensive. Massively parallel sequencing of cDNA (RNA-Seq) has vastly increased the throughput of cDNA sequencing, enabling global measurement of transcript abundance; additionally, with direct sequencing of cDNA fragments, the need for cloning is obviated (Haas and Zody 2010; Nowrousian 2010). Thus, in a much smaller timeframe and at a small fraction of the cost of earlier projects, reasonably complete and unbiased coverage of a human-scale transcriptome can be achieved.

RNA-Seq also generates a wealth of information beyond that obtained from sequencing genomic DNA as it yields sequence, splicing, and expression level information leading to the identification of novel transcripts and sequence alterations, including alterations caused by alternative splicing, alternative promoters, alternative 3’ ends, and sequence variation caused by SNPs. RNA-Seq also has the potential to detect transcripts that are generated from both
alleles of a given gene. In contrast to microarrays RNA-Seq is not limited to pre-defined array elements - identification of novel transcripts and abundance estimation (expression profiling) is achieved simultaneously. In addition, expression arrays have a limited dynamic range, whereas RNA-Seq is able to measure differences in transcript expression over several orders of magnitude.

Examples of the successful application of RNA-Seq to a variety of organisms are appearing in the literature (Haas and Zody 2010; Nowrousian 2010). However, as whole genome sequencing of bacterial isolates has shown, the true power of this approach will not come from the haphazard application of a new technique but will only be revealed through detailed comparative analyses of sequence datasets that have been acquired by design to answer specific biological questions, and by the development of tools to handle, analyze and visualize this deluge of data.

(c) Heterogenous RNA Seq
This GSCID application proposes a novel use of RNA-Seq to capture the gene expression dynamics of host-pathogen interactions by sequencing the transcriptome(s) of in vitro HEP-2 cells and in vivo mouse cells infected with Chlamydia spp. under several different conditions. By applying RNA-Seq to RNA fractionated from infected cells, an unbiased readout of the gene expression profile of the host cell responding to chlamydial infection at that moment in time, and vice versa, can be obtained. By the very nature of an intracellular pathogen, the total infected cell transcriptome contains the transcriptome of the host cell and the transcriptome of the infecting pathogen obtained simultaneously - termed here as Heterogenous RNA-Seq.

3. Experimental Rationale
(a) Application of Heterogenous RNA-Seq to Biological Questions
Four key broad categories of pathogen/host Heterogenous experiments have been identified in this proposal (Table 1); all possible biological questions that are addressable using Heterogenous RNA-Seq of infected cells will fall into one of these categories. In each case, uninfected and/or mock-infected cells would be used as controls, providing the baseline for identifying infection-derived gene expression variation in the host and pathogen. The proposed experiments in Sections 3(b) and 3(c) use these categories as a framework.

Table 1. Fundamental Categories of Heterogenous RNA-Seq Experiments

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
<th>Broad biological question</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Time series</td>
<td>How do pathogen and host gene expression change over the course of infection? How does the gene expression of each change in response to the other?</td>
</tr>
<tr>
<td>B</td>
<td>Single host cell, multiple pathogens</td>
<td>Keeping global gene expression in the host cell as a constant by using the same cell line in each infection, what are the similarities and differences of the host cell response to different infecting pathogens?</td>
</tr>
<tr>
<td>C</td>
<td>Single pathogen, multiple host cells</td>
<td>Many pathogens will infect a variety of cell types; furthermore, uninfected host immune cells will respond to infected cells. Keeping global gene expression in the pathogen as a constant, what are the similarities and differences of different host cells responding to a single pathogen?</td>
</tr>
</tbody>
</table>
(b) Proposed experiments using the Chlamydia/HEp-2 in vitro culture model (Collaboration with Dr Patrik Bavoil, UMB)

On the basis of the categories described in Table 1, four sets of Heterogenous RNA-Seq experiments using the Chlamydia/HEp-2 in vitro culture model are proposed in this GSCID application. Each has been selected to address a specific outstanding problem in chlamydial disease pathogenesis:

Experiment 1: Infection Time Course
As noted in Section 2 above, Chlamydia are characterized by a distinctive biphasic developmental cycle. A detailed understanding of the chlamydial developmental cycle in the context of the host cell is paramount, however the global regulatory signals that govern each of the key chlamydial development steps remain poorly characterized.

Heterogenous RNA-Seq over a time-course of infection (including mock infected cells as a control; 6 hours post infection (hpi), 12 hpi, 24 hpi, 36 hpi and 48 hpi) will be applied to define how C. trachomatis and the host cell respond to each other through global transcriptional changes over the chlamydial lifecycle (Category A Heterogenous experiment, Table 1).

Experiment 2: Species Specificity
C. trachomatis, C. pneumoniae and C. psittaci are the key chlamydial pathogens of humans, with C. psittaci being acquired zoonotically. Our recent work has shown that C. pneumoniae was originally a zoonotic pathogen, raising concerns about the ability of animal chlamydiae to become full-fledged human pathogens. Heterogenous RNA-Seq analysis of a strain of each species (C. trachomatis serovar E, C. pneumoniae AR39 and C. psittaci 6BC) infecting HEp-2 cells would enable a consistent and reproducible comparison of each host/pathogen pairing, allowing species-specific virulence factors to be linked with the HEp-2 host cell response and thus meaningfully compared (Category B experiment, Table 1).

Experiment 3: Serovar Specificity
Linking strain variants with disease severity is an outstanding problem in chlamydial research, exacerbated by increases in chlamydial genital tract infection rates. No consistent links between C. trachomatis virulence factors and disease outcome have yet been found, with host factors thought to be a major contributor to this failure (Byrne, 2010). Heterogenous RNA-Seq of 12 C. trachomatis variants infecting HEp-2 cells would allow simultaneous comparison of both C. trachomatis gene expression from these different isolates and the variability of the host cell response to each. As a Category B Heterogenous RNA-Seq experiment that keeps the host cell line as a constant (Table 1), this is similar to Experiment 2 but is a “finer granularity” experiment where variation of the infecting pathogen is kept to a minimum (strain variation vs species variation). In other pathogen/host pairings, genetic modification(s) of the pathogen would be the equivalent experiment. In the absence of a genetic modification system in Chlamydia, using the naturally occurring variability of chlamydial isolates is the next appropriate mechanism to map virulence responses.
(c) Proposed experiments using the Chlamydia/mouse in vivo infection model
(Collaboration with Dr Roger Rank, UAMS)

**Experiment 4: In vivo Heterogenous RNA-Seq**

We have a unique opportunity to determine the eukaryotic and chlamydial transcription profile in an actual genital tract infection where tissue can be obtained at selected time points in a synchronized infection. Dr. Rank has developed a model in which *C. muridarum* (Nigg strain) is inoculated intracervically such that the initial developmental cycle is relatively synchronous, i.e. virtually all inclusions are at the same level of maturity up to 48 hours after infection (Rank et al 2010). He has already documented the cytokine/chemokine expression profile in this model and characterized the infection at the mucosal surface by transmission electron microscopy (unpublished data). The initial experiment will be to obtain cervical tissue at 0, 24 and 36 hours post-infection from C57/Bl6 mice. A single cell suspension will be produced and the infected cells sorted from the uninfected cells by flow cytometry using an anti-LPS antibody. Methodology is available for intracellular staining which maintains the integrity of RNA and DNA (Colegan et al, 2001). By appropriate gating, polymorphonuclear leukocytes (PMNs) and mononuclear cells can be eliminated from the acquired cell suspensions. This will allow us to determine the transcription profiles of both the host cells and *Chlamydia* when the organisms are in the early reticulate body stage with only a few reticulate bodies in each inclusion versus a later stage in which the inclusions are larger and the transition to elementary bodies is occurring.

A key factor in the development of chlamydial pathology is host genetics. Therefore, we will repeat the above experiment in C3H mice. C3H mice have longer infections with *C. muridarum* and more upper tract pathology than do the more resistant C57 mice. This will allow the impact of the host strain on both host and chlamydial RNA expression to be determined.

Finally, it is very apparent from Dr. Rank’s current studies that PMNs have a significant impact on the course of the infection, with respect to the cytokine/chemokine response, targeting of infected cells which can be detached from the epithelium by PMNs, and by the killing of chlamydiae. To investigate this, C57 mice will be treated with antibody to PMNs to eliminate them and cervical tissue will be obtained at 24 and 36 hours after infection.

3. Core Methods and Data Production

In the proposed experiments described above, we will variously infect HEp-2 cells, a human laryngeal carcinoma epithelial cell line or obtain infected cells from the *C. muridarum* mouse infection model, using *C. trachomatis* serovar E, *C. muridarum*, *C. psittaci* 6BC or *C. pneumoniae* AR39. Each is the prototype strain for the species, has an available genome sequence and has been used in multiple studies worldwide. Infections will be performed following standard protocols in routine use in the Bavoil laboratory. All large volume cultures of *C. psittaci* will be performed under strict BSL3 conditions in accordance to Biosafety in Microbiological & Biomedical Laboratories 5th Edition (BMBL5) in the certified BSL3 facility of the UMB Dental School. Mock infected HEp-2 cells or uninfected mice will be established in parallel with each experiment as controls.

(a) The Challenge of Applying RNA-Seq to Infected Cells: Mixed RNA Moieties

Eukaryotic mRNA is relatively simple to isolate as the virtually ubiquitous poly(A) tail enables rapid purification of mRNA from cellular lysates using poly(T) capture on a solid medium, such as magnetic beads. In contrast, prokaryotic mRNA lacks the polyA tail and, furthermore, is typically present in much lower stoichiometric amounts relative to eukaryotic RNAs in the infected cell. Thus the primary experimental challenge is to separate prokaryotic mRNA from eukaryotic RNA (mRNA, tRNA and rRNA) in sufficient quantities for sequencing. In the absence of a direct selection step for prokaryotic RNA, this will be achieved through a series of
depletion steps that remove eukaryotic RNA moieties, thereby enriching for prokaryotic RNA (Figure 1, section 3(d) below).

**Figure 1.** Flow chart of prokaryotic and eukaryotic mRNA preparation, each section described in text is denoted.

**b) Total RNA preparation from Chlamydia-infected HEP-2 cells.**

For each infection time point, one dish of infected or mock-infected HEP-2 cells will be harvested by scraping the cells into a volume of RNA Later (Ambion). Cell suspensions will stored at -80°C.

Cell suspensions in RNA Later Reagent (Ambion) will be thawed at room temperature for 2 hours. Total RNA will be further purified using the RiboPure kit (Ambion) as per the manufacturer's instructions.

**c) Total RNA Preparation from Chlamydia-infected primary mouse cells**

Mouse cervical tissue will be dissected into a volume of RNA Later (Ambion) and enzymatically dissociated to single cells. A monoclonal anti-chlamydia LPS antibody will be used to label and sort infected cells by flow cytometry in RNA Later. Uninfected cells, PMNs and mononuclear cells will also be collected through appropriate flow cytometric gating. Collected infected and uninfected cells will be subjected to RNA fractionation as described in Figure 1.

**d) Fractionation of prokaryotic and eukaryotic RNAs.**

The MICROBEnrich kit (Ambion) will be used to initially separate total eukaryotic RNA from total chlamydial prokaryotic RNA. MICROBEnrich uses RNA hybridization to magnetic beads with probes that will capture eukaryotic rRNA and mRNA. Using this approach with infected cell lysates will enable enrichment of prokaryotic RNA through depletion of the major eukaryotic RNA moieties. This process will be repeated twice to maximize the depletion. Eukaryotic RNA captured will be recovered from the beads and subjected to further purification (see below). At the end of this step, two fractions will be obtained, one enriched for prokaryotic RNA (Fraction 1P) and one for eukaryotic RNA (Fraction 1E) (Figure 1); both will be further enriched.

Fraction 1P will be further processed using the MICROBExpress bacterial mRNA enrichment kit (Ambion), which uses rRNA probes to capture and remove eukaryotic rRNA. Fraction 1P will be processed twice with the MICROBExpress kit to create Fraction 2P, constituting highly enriched chlamydial mRNA. Fraction 2P will be checked for purity and quantity prior to chlamydial RNA-Seq library construction.
Fraction 1E will be further processed using the Poly(A)Purist mRNA purification kit (Ambion), a positive selection step for eukaryotic mRNA. The resulting mRNA (Fraction 2E) will be checked for purity and quantity prior to eukaryotic RNA-Seq library construction.

(e) Sequencing
Fractionation, library construction and sequencing will be conducted in the Genomics Resource Center (GRC), a high-throughput core laboratory and data analysis group within IGS. The GRC has sequenced and generated expression profiles for more than fifty human and mouse transcriptomes during the past year. Libraries will be constructed according to Illumina protocol, which includes an oligo-dT capture of poly-adenylated transcripts from purified total RNA. Each fragment library will be tagged with a unique 6bp barcode to enable multiplexed sequencing on our Illumina HiSeq2000 platform. The Illumina sequencing platform has been in operation within the GRC since 2008 and yields consistent, high-throughput, high-quality sequence data. The HiSeq2000 represents the latest model sequencer from Illumina. It uses the same, well-established and consistent chemistry as the Genome Analyzer sequencers, but includes significant hardware and design upgrades to enable much higher throughput. The HiSeq2000 can sequence two flowcells simultaneously, alternating each between chemistry and image capture steps to maximize data capture per day. It also uses a new laser excitation and line-scanning image capture method to sequence both surfaces of the flowcell. These improvements result in a 4-5x increase in yield per run compared to the Genome Analyzer IIX model.

We will sequence 100bp from both ends of each library fragment. Each lane of Illumina HiSeq sequencing at the GRC generates 80 – 100 million high-quality, passed-filter sequence read pairs. Because of this high-throughput, it is possible to multiplex two RNA-Seq libraries per flowcell lane, resulting in 40 – 50 million read pairs per library, representing 12-20Gb total base pairs. This represents approximately 150X coverage for a mammalian transcriptome and several thousand-fold coverage for a prokaryotic transcriptome (L. Tallon, Pers. Comm.). Illumina libraries prepared from the 2E and 2P fractions (Figure 1) for each time point or data point in an experiment (Table 2) will be loaded onto a single channel of a HiSeq2000 flowcell.

Table 2. Sequencing Requirements

<table>
<thead>
<tr>
<th>Description</th>
<th># channels</th>
<th>Flow Cells’</th>
<th>Expected Min. Sequence Data (Gb)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Experiment 1</strong> Time course of infection</td>
<td>24</td>
<td>3</td>
<td>36.0</td>
</tr>
<tr>
<td>• 6 pairs with technical replicates and mock infected controls at each time point</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• single species, multiple time points of infection.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• mock infected HEP-2 cells, 6, 12, 24, 36, 48 hours post infection (hpi)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>C. trachomatis</em> serovar E</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Experiment 2</strong> Host/pathogen response using various chlamydial species</td>
<td>9</td>
<td>1 1/8</td>
<td>13.5</td>
</tr>
<tr>
<td>• 3 matched pairs, with technical replicates</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• multiple species, single time point (36 hpi).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• mock infected HEP-2 cells, <em>C. trachomatis</em> serovar E, <em>C. psittaci</em> 6BC and <em>C. pneumoniae</em> AR39</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table

<table>
<thead>
<tr>
<th>Description</th>
<th># channels</th>
<th>Flow Cells</th>
<th>Expected Min. Sequence Data (Gb)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Experiment 3</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Host/pathogen responses using <em>C. trachomatis</em> strain variants</td>
<td>36</td>
<td>4 1/2</td>
<td>54.0</td>
</tr>
</tbody>
</table>
| • 12 matched pairs with technical replicates and mock infected controls for each variant
  • multiple variants, single time point (36 hpi).
  • mock infected HEp-2 cells, 12 serovars of *C. trachomatis*               |            |            |                                 |
| **Experiment 4**                                                          |            |            |                                 |
| Host/pathogen responses using an *in vivo* mouse model of chlamydia infection | 20         | 2 4/8      | 30                              |
|                                                                           |            |            |                                 |
| **Total:**                                                                 | 89         | 11 1/8     | 133.5                           |

*8 channels per HiSeq2000 flow cell; two flow cells per full HiSeq2000 run*

### 4. Data Analysis

**(a) Read Mapping and Annotation Matching**

Analysis of the Heterogenous RNA-Seq sequencing datasets will be achieved in a multistep process, beginning with mapping of sequence reads to a reference genome, either chlamydial or mammalian.

Raw data from the sequencer will be processed using Illumina’s RTA and CASAVA pipeline software, which includes image analysis, base calling, and sequence quality scoring, as well as pre-existing GRC-developed pipelines for sequence assessment and quality control. Following base-calling by Illumina pipelines, our quality control pipeline assesses base-call quality, converts Illumina quality scores (ASCII+64) to Sanger-style/Phred-like scores (ASCII+32) and truncates reads where the median Phred-like quality score falls below Q20. The pipeline also plots base composition by cycle/position to identify potential cycles with aberrant chemistry errors or sample bias. These processes are parallelized using our high-performance Sun Grid Engine compute grid, which contains more than 1000 cores and five terabytes of available RAM.

Further analysis of the resulting sequence reads will be performed using our transcriptome analysis pipeline based upon Cufflinks software package (Trapnell, et al. 2010). Cufflinks improves upon earlier RNA-Seq analysis methods, which estimated relative expression at the gene level as the fraction of reads aligning to exons after normalizing for gene length. This RPKM (reads per kilobase of exon per million mapped reads) method, like all early RNA-Seq data analysis methods, faced the challenge of ambiguous assignment of sequence reads to isoforms. The combination of paired-end sequencing, longer reads, and discontiguous read-fragment alignment algorithms significantly reduces the uncertainty in assigning reads to alternative splice variants and enables more accurate expression profiling at the transcript level. Cufflinks uses these features to extend the RPKM method to the new FPKM method (paired-read library fragments per kilobase of exon per million mapped reads).

Following alignment of sequence reads to a reference genome using Tophat (Trapnell, et al. 2009), which can produce spliced read alignments to identify reads that span putative splice junctions, Cufflinks assembles read pairs to identify incompatible pairs that indicate distinct transcripts arising from the same gene. These assemblies are used to estimate the minimum possible set of spliced transcripts that explain all aligned read pairs. Read pairs that are compatible by alignment with more than one predicted transcript are assigned to an isoform using a statistical model based upon frequency of splice detection and length of transcript. Following assignment of read pairs to isoforms, FPKM values are calculated. These transcript assemblies and abundance estimations are performed in the absence of gene annotation data.
to allow for detection of novel transcripts. The Cuffcompare utility is used to compare the predicted transcripts to known annotated transcripts and identify potentially novel isoforms.

(b) Comparative Analysis and Visualization Tool Development (Collaboration with Drs Gajer and Ravel, IGS)
The proposed experiments will generate substantial amounts of sequence data (Table 2). However, the key benefit of performing Heterogenous RNA-Seq on a specific pathogen/host pairing, such as the Chlamydia/HEp-2 model, is not to exhaustively examine every gene, but rather to identify what subsets of genes and associated features (splice site variation, operon structure etc) are changing. This will be achieved through the “overlay” of linked sets of gene expression data such as succeeding temporal time points or by comparison to a control gene expression dataset, i.e; the uninfected HEp-2 cell. This subset of changing genes can then be interpreted in the context of the biological question being posed - i.e; identification of temporal expression changes over a time course in vitro (Experiment 1) or in vivo (Experiment 4), or pairwise changes observed between the infection process across different species (Experiment 2) or serovars (Experiment 3).

Generation of sequence data on next-generation instruments and mapping of this data back to reference genomes are both robust and established technologies. To rapidly compare these large datasets in order to identify those gene subsets that are changing under the defined biological question, we propose to build a Heterogenous RNA-Seq viewer using the inVUE toolbox, an existing software platform developed at IGS by J. Ravel and P. Gajer. The toolbox is a generic scientific data viewer that runs on Unix/Linux, Windows or MacOS, is fast, flexible and does not suffer in performance when millions of data points are analyzed from multiple experiments. It has been developed under Qt, a C++ application framework, including a class library and tools for cross-platform development. While inVUE could be used as stand-alone software, it was specifically built to be used as the inner core of any analytical software. It has already been implemented in genomeMTV, a powerful analytical software for Affymetrix tiling array data (http://www.genomemtv.org/). inVUE has a built-in interface with the open-source statistical analysis program R, making it readily amenable to powerful statistical data analyses - we anticipate making use of this feature to implement algorithms that will rapidly identify regions of interest on the basis of gene expression variation. The data can be loaded into inVUE, analyzed in R, the results displayed back into inVUE and, unlike static R displays, the data and the associated gene annotation can then be dynamically viewed with regions of significant variation highlighted.

The Heterogenous RNA-Seq viewer will enable multiple expression profiles and the matching annotation of both eukaryotic and prokaryotic genomes to be viewed simultaneously as stacked “tracks”. Each track can be locked to other tracks, allowing rapid visualization of expression changes. In addition, this style of visualization will have the ancillary but potentially significant benefits of validating existing annotation, promoter identification, splice site variation (in eukaryotic expression profiles) and operon structure.

(c) Potential difficulties and limitations
The greatest experimental limitation could arise during the early time points of infection (<20 hpi) as there will be relatively little chlamydial RNA compared to eukaryotic RNA. Thus, it is possible that a linear amplification step for the chlamydial RNA will be required at these early time points to ensure sufficient starting material for sequencing. Alternatively, the amount of infected cells used as input for total RNA extraction could be increased to avoid an additional amplification step. However, even without sufficient prokaryotic mRNA in the early time points, the effects of chlamydial infection will still be discernible through the changes in eukaryotic gene expression compared to the mock infected control.
(d) **Future research directions**

We have identified four broad categories of Heterogenous RNA-Seq experiments (Table 1) applicable both *in vitro* and *in vivo*. We propose four experiments across three categories using an *in vitro* *Chlamydia*/HEp-2 infection model and an *in vivo* *Chlamydia*/mouse infection model. In each proposed experiment, future work would expand the “granularity” of the experiment by increasing the number of time points or the numbers of isolates examined. For the *in vivo* studies, other studies include, but are not limited to: comparison of *C. muridarum* Weiss versus Nigg strains; following reinfection to determine impact of adaptive immunity, and analyzing infection in a variety of knockout mice, e.g. Y-IFN, TLR 2, to investigate the effect on chlamydial pathogenesis. In addition, our approach described here will generate many gene targets for further investigation and validation through laboratory molecular biology experiments.

While we have focused on *Chlamydia*, the approach and tools will also be broadly applicable to **any pathogen that interacts with a host cell**. For example, keeping the HEp-2 cell as a constant, the Heterogenous gene expression profile of every bacterial pathogen that can infect the HEp-2 cell could be determined, enabling delineation of the similarities and differences using the HEp-2 transcriptional responses as a common baseline.

Alternatively, many significant pathogens will infect more than just one cell type. Heterogenous RNA-Seq could be applied to determine the response of many host cell types to the selected pathogen, lend significant insight to diagnostic, therapeutic and vaccine development to many different pathogens in a consistent and reproducible manner.

Finally, infection of a host cell is essentially a perturbation of the natural state, either *in vivo* or *in vitro*. The Heterogenous RNA-Seq protocol described here could be applied to any cellular system where a perturbation could be applied (physical, chemical or biological) with the expectation of detecting a transcriptional response.

(e) **Timetable**

The scale of the proposed work necessitates a two year timeframe (Figure 2). We anticipate performing all infections and RNA extraction in the first 3 months with sequencing and analysis over the next 18 months. Each Illumina run takes approximately one week (2 flowcells per run) with an estimated 12 flowcells required (Table 2). Tool development based on inVUE and subsequent analysis of these datasets with this tool, together with read mapping and annotation, will be a significant bioinformatic task. Tool development will begin immediately in order to be available for use with the first dataset. Data release will be made at multiple intervals as each sequence dataset is completed.

**Figure 2. Timeline of Proposed Work**

<table>
<thead>
<tr>
<th>Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) inVUE-based RNA-Seq Tool Development</td>
</tr>
<tr>
<td>2) <em>in vitro</em> experiments (Experiments 1, 2 and 3)</td>
</tr>
<tr>
<td>2.1) Infect cells</td>
</tr>
<tr>
<td>2.2) Prepare total RNA</td>
</tr>
<tr>
<td>2.3) RNA fractionation</td>
</tr>
<tr>
<td>2.4) Sequencing</td>
</tr>
<tr>
<td>2.5) Read mapping, annotation, data release</td>
</tr>
<tr>
<td>3) <em>in vivo</em> experiments (Experiments 4)</td>
</tr>
<tr>
<td>4.1) Infect mice</td>
</tr>
<tr>
<td>4.2) Prepare total RNA</td>
</tr>
<tr>
<td>4.3) RNA fractionation</td>
</tr>
<tr>
<td>4.4) Sequencing</td>
</tr>
<tr>
<td>4.5) Read mapping, annotation, data release</td>
</tr>
</tbody>
</table>

GSCID White Paper Application, October 2010
5. **Source and Availability of DNA, Strains and Data for Public Dissemination**

All strains selected in this study are already available for deposition into a NIAID-approved culture collection. Gene expression data will be uploaded to the NCBI GEO data repository as produced. The invUE-based tools will be made available on the IGS website as open source software.

6. **References**


Addendum: Simultaneous Transcriptional Profiling of Obligate Intracellular Pathogens and their Host Cells

Garry S. A. Myers

1Institute for Genome Sciences and Department of Microbiology & Immunology, University of Maryland School of Medicine, 801 West Baltimore Street, Baltimore, Maryland 21201

An Alternative Approach to Fractionation of Prokaryotic and Eukaryotic mRNA

In the original white paper, the key challenge of capturing different mRNA moieties was addressed by successive rounds of mRNA enrichment and/or rRNA depletion to produce distinct eukaryotic and prokaryotic mRNA fractions respectively (Figure 1A). These fractions were then subjected to RNA-Seq analysis.

While this approach remains valid, the scale of next-generation sequencing suggests an alternative approach which would reduce sample manipulation and retain prokaryotic and eukaryotic mRNA in the same tube, and thus reduce potential sources of bias. This approach also reduces the number of RNA-Seq libraries per sample.

Serial Depletion of rRNA

Serial depletion of eukaryotic and prokaryotic rRNA from the sample (Figure 1B) avoids the need to physically separate prokaryotic and eukaryotic mRNA preparations into different tubes. Instead eukaryotic rRNA (28S, 18S, 5.8S and 5S) and prokaryotic rRNA (23S, 16S, and 5S) are removed successively by the appropriate use of Ribo-Zero™ rRNA Removal Kits (Epicenter Biotechnologies). A small amount of rRNA is likely to remain in the sample, however the sheer scale of next-generation sequencing means contaminating rRNA (readily identified bioinformatically) will not significantly reduce sequence coverage.

A. Original Approach

- Extract total RNA
  - RiboPure
- Separate eukaryotic & chlamydial RNA
  - MICROBEEnrich
- Eukaryotic RNA enriched
  - Fraction 1E
  - Poly(A) Purist
  - Eukaryotic mRNA
    - Fraction 2E
- RNA-Seq library
  - Illumina sequencing
  - Bioinformatics pipeline

B. Simplified Depletion Approach

- Extract total RNA
  - RiboPure
- Deplete eukaryotic rRNA
  - Ribo-Zero (Human/Mouse/Rat)
  - MICROBEExpress
- Eukaryotic RNA depleted
  - Fraction 1P
- Chlamydia enriched RNA
  - Fraction 2P
- RNA-Seq library
  - Illumina sequencing
  - Bioinformatics pipeline
Dr. Garry Myers  
Department of Microbiology and Immunology & Institute for Genome Sciences  
University of Maryland School of Medicine

RE: support for your Genome Sequencing Center for Infectious Diseases proposal  

Dear Dr. Myers,

Your proposal to do RNA-sequencing on *Chlamydia*-infected cells is brilliant! Further, I agree wholeheartedly on the 4 categories of your experimental rationale. They are well thought out and focus on the key issues for better understanding chlamydial pathogenesis and the insidiousness of chlamydial infections. Your thinking in considering a whole transcriptome comparison of chlamydiae-infected polarized cells versus non-polarized cells in order to begin to understand the difference between in vitro mode of infection versus the more in vivo mode is critical. We have spent some 25 years studying this very issue and the results to date are impressive. Briefly, genital *C. trachomatis* infection in polarized cells: (i) first identified EB entry in human epithelial cells by clathrin-coated pit-mediated endocytosis; (ii) proper exit of this luminal pathogen via the apical domain versus the basal domain, reflecting infection-directed spread; (iii) by a novel apical lift procedure, identified the role of protein disulfide isomerase (PDI), as well as estrogen, in chlamydial attachment/entry processes; (iv) showed that the azithromycin MBC<sub>90</sub> was considerably lower in polarized cells (<0.125 mg/L) versus non-polarized cells (0.5 mg/L); (v) confirmed the delivery of azithromycin by PMNs transmigrating through a polarized monolayer; (vi) proved that EB progeny are more infectious than EB progeny released from non-polarized cells; (vii) if from sufficiently differentiated genital epithelia, retain hormone responsiveness and show that estrogen dramatically enhances *C. trachomatis* infection; and that (viii) tryptophan-induced persistence of chlamydiae is prolonged because tryptophan-uptake is doubled in polarized epithelia and they retain larger pools of tryptophan than non-polarized epithelia. To name a few differences. Let me know if I can help!!

Respectfully submitted,

Priscilla B. Wyrick, Ph.D.  
Professor and Chair  
AAM; Fellow, AAAS  
Past President, CBRS
November 24, 2010

NIH, NIAID
Bethesda, MD.

Dear Sir or Madam:

I am writing to indicate my full and enthusiastic support for Dr. Garry Myer’s proposal to simultaneously obtain global transcriptional profiles of *Chlamydia*, their host cells and the host in both cell culture and in vivo environments. I think that this represents an opportunity to advance our fund of knowledge dramatically as we consider the unique relationship that chlamydiae have evolved with their host cells as well as during an infection. These studies will be key in improving our understanding of why this pathogen is as successful as any on the planet. Dr. Myers has the technological expertise, informatic tools and appropriate infectious disease research background to maximize the impact of applying RNA-seq technology to the study of chlamydial development at the level of the host cell and the infected host.

I think Dr. Myers has proposed a number of sensible approaches as he embarks on this exciting new research endeavor. Application of his approaches also will enable a number of additional comparative studies in cell culture models and in vivo. For example in cell culture fundamentally important studies could easily be done to compare host and pathogen profiles under various growth conditions. We normally set growth conditions to maximize successful completion of the developmental cycle by including host cell inhibitory reagents such as cycloheximide. How does the chlamydial profile change when the host cell is allowed to compete for essential nutrients in the absence of host-specific protein inhibitors? There are a number of environmental iterations that can be tested to learn more about chlamydial growth status differences as the growth environment is altered. Another basic area of investigation that could readily be accomplished includes an opportunity to improve our understanding how chlamydial gene regulation is modulated as new clinical isolates are adapted to cell culture systems. It is very clear that new isolates grow better as they adapt to cell culture growth conditions, and Dr. Myers will be able to provide definitive information about how this happens. Finally, all too often chlamydiae strains are handled as heterogeneous populations rather than as clonal isolates. Dr. Myer will be able to monitor differences in clonal isolates obtained either by limiting dilution methods or by plaque purification. How do isolates compare? What is their potential for developing genetic heterogeneity? Dr. Myers proposes to examine profiles of representatives of different chlamydial species and different strains within species in different host cells. Including altered environmental
conditions; adaptations of clinical isolates to in vitro environments and examination of clonal isolates from a number of perspectives will enrich the in vitro portion of his work plan.

Dr. Myer also plans to do in vivo work by comparing C. muridarum growth in two mouse strains with differing susceptibility to chlamydiae. These studies are important and could be expanded to include sets of recombinant inbred mice (e.g. BXD mice or collaborative cross strains). This would greatly enrich the host-profiling portion of the study and will likely provide additional data with respect to host patterns of genetic susceptibility and resistance and add depth to the study of requirements for protective immunity and separating protective from pathologic immune responses.

The opportunity to simultaneously evaluate chlamydiae and host gene expression during the course of either cell culture growth or infection of experimental animals is hugely exciting, timely and do-able. I look forward to learning more about chlamydiae and their relationship with the host cell and the host as the study progresses.

Sincerely yours,

Gerald I. Byrne, Ph.D.
Professor and Chair
November 22, 2010

Garry Myers, Ph.D.
Assistant Professor
Institute for Genomic Sciences
University of Maryland School of Medicine
Baltimore, MD

Dear Garry:

I am pleased to provide you with a letter of support for your proposed experimentation to develop the technology for the simultaneous transcription profiling of chlamydiae and the cells that they are infecting. As you know, I have been worked in the area of the host response to chlamydial infections for approximately 35 years, with emphasis on the host immune response and how it relates to protection but also the pathologic response. Currently, we have been studying the inflammatory response to the infection and particularly the mechanisms by which the organism elicits the chemokine and cytokine response. We have developed the methodology to produce a synchronous infection through the first developmental cycle and have been able to determine the host gene expression during this time. While this has yielded some interesting kinetics data, one problem with the current technology is that we have no idea from which cells is the message derived or even how much it is diluted by using a whole tissue homogenate; so we cannot relate the host message directly to the signals initiated by chlamydiae or through their interaction with the host cell which they are parasitizing.

However, if the technology can be developed to assess the transcription profile of the very same cell which is being infected, then this would be a remarkable step forward in understanding the relationship of the organism at its different steps in the developmental cycle to events in the host cell, whether initiated directly by signals from the organism or by the response of the host cell to infection with “foreign body.” If this could be accomplished in vivo by the isolation of infected host cells versus uninfected cells from the same animal, we could learn not only about the interaction of the organism with the host cell but whether there is any bystander effect, i.e. signaling from the infected cell which has an impact on the adjacent cells. Moreover, it would open up a tremendous new area because one could perform the experiment in mice lacking a particular cell population, chemokine, cytokine, or signal
transduction molecule. It would take our understanding of the events in the local cellular milieu to a new level.

In reviewing your proposal, I feel that before moving to the animal, which will clearly have problems to overcome such as the amount of host cell and bacterial RNA which can be collected, we need to have evidence that the procedure is viable. The obvious first step is to use cells from tissue culture which are infected with a high MOI is ensure a high percentage of infection. It is also easy to take cultures at various times after infection to determine what changes occur in the course of the developmental cycle. Collecting sufficient host cell and bacterial RNA should not be a problem. After the parameters have been established in vitro, then one can move to the more difficult in vivo experiments. Actually, there is going to be a substantial number of problems that can be investigated in vitro, e.g. the profiles of different chlamydial species, infection with various defined mutants, and infection in the presence of specific inhibitors such as an inhibitor of type III secretion.

I commend you for taking this on and know of no one else who has the capability to develop this technology. I look forward to seeing the first set of data and to helping you ultimately transfer this technology to the mouse model.

Best wishes!

Roger G. Rank, Ph.D.
Professor
Re: Letter of Support for NIAID Genome Sequencing Centre for Infectious Diseases application by Dr Garry Myers et al.

I wish to take this opportunity to provide this Letter of Support for the “Simultaneous transcriptional profiling of obligate intracellular pathogens and their host cells” project, led by Dr Garry Myers at The Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, USA.

This is a very exciting and important project. While researchers have been thinking about transcriptional profiling of the intracellular parasite and its host cell at the same time, relatively little good data has yet emerged. This proposed project will make use of cutting edge technology to achieve this aim and will thereby make this more broadly available to other researchers in the same or related fields. Overall, the project is very well designed, is very likely to be successful and will provide some very interesting and useful data.

As an established researcher in the Chlamydia field, I would like to offer some suggestions for additional areas of investigation or focus of resources. A very powerful and therefore fruitful area of investigation is the comparison of closely related chlamydioid strains. By comparing strains that have very similar genomes, but show differences in their biological properties, it is possible to rapidly determine some of the key aspects responsible for host specificity and pathogenicity. While the current proposal has some aspects of this, it would perhaps be useful to expand this area of investigation. One example would be the analysis of several closely related Chlamydia pneumonii isolates (4 human and one animal C.pneumonii full genomes are already available) which could be cultured in several relevant cell lines (eg human derived and animal derived). These cultures could then be perturbed with key components, such as interferon-gamma or other cytokines. By performing heterogenous RNA-seq on these samples, some very powerful data could be obtained.

Another area of possible expansion is the use of natural physiological conditions to perturb the growth of Chlamydia in vitro (and potentially in vivo). It has long been known that for women, infection with C.trachomatis at different stages of the menstrual cycle varies significantly. This is no doubt a result of the delicate parasite-host cell interface. By growing C.trachomatis strains in vitro with varying levels of hormones, and then performing heterogenous RNA-seq, key information will be obtained.

While these suggestions could potentially improve the project plan, I want to emphasise that it already is a very powerful plan to achieve some cutting edge information about an important pathogen-host interaction.

Regards,

[Signature]

Professor Peter Timms
School of Life Sciences
Institute of Health and Biomedical Innovation
Queensland University of Technology, Kelvin Grove Campus, Brisbane, Australia, 4059
Tel : +61 7 3138 6199  Fax : +61 7 3138 6030  Email : p.timms@qut.edu.au
Garry Myers  
Institute for Genome Science  
University of Maryland Baltimore  

11 December 2010  

Dear Garry,  

I am very happy to provide you with a letter of support for your proposed series of experiments of simultaneous transcriptional profiling of the *Chlamydia*-infected cell. I have worked on chlamydial pathogenesis since 1983, and my expertise ranges from the role of *Chlamydia* type III secretion in pathogenesis and development, the molecular biology and impact on disease of *Chlamydia* phages, the role of the polymorphic membrane protein family of *C. trachomatis* in infection and disease to comparative genomics within the *Chlamydiaceae*.  

As you know, the defining feature of *Chlamydia* in the laboratory is how difficult it is to work with. The lack of a genetic transformation system and the need to grow *Chlamydia* within a host cell places severe constraint on the tools available to researchers. One of the ways that you and I and others have worked around this is by applying the techniques of whole genome sequencing and comparative genomics. Because of the highly conserved nature of the Chlamydia genome and its small size, this approach has proven to be a primary mover for the field.  

However, it is obvious we cannot examine chlamydial pathogenesis without considering the host cell. If the technology can be developed to assess the transcription profile of the very same cell which is being infected, then this would be a major step forward in understanding the relationship of the organism to events in the host cell, whether initiated directly by signals from the organism or by the response of the host cell to infection.  

Thus, I strongly support the development and application of “Heterogeneous RNA-Seq to *Chlamydia*-infected cells” as described in your GSCID white paper. I am committed to support your experimental and any other needs you may have for this project. I look forward to participating in this ground breaking endeavor.  

Sincerely,  

Patrik M. Bavoil, PhD