Fish & Shellfish Immunology 31 (2011) 716–724

Contents lists available at ScienceDirect

Fish & Shellfish Immunology

journal homepage: www.elsevier.com/locate/fsi

Deep sequencing of the innate immune transcriptomic response of zebrafish embryos to *Salmonella* infection

Anita Ordas [a,b], Zoltan Hegedus [c,d], Christiaan V. Henkel [e], Oliver W. Stockhammer [a], Derek Butler [f], Hans J. Jansen [e], Peter Racz [a,b], Matyas Mink [b], Herman P. Spaink [a], Annemarie H. Meijer [a,*]

[a] Institute of Biology, Leiden University, Einsteinweg 55, 2333 CC Leiden, The Netherlands
[b] Department of Genetics, Faculty of Science and Informatics, University of Szeged, 52 Kozepfasor, H-6726 Szeged, Hungary
[c] Zenon Bio Ltd., Maros u 40, H-6721 Szeged, Hungary
[d] Bioinformatics Laboratory, Biological Research Center, Hungarian Academy of Sciences, Temesvari krt. 62, H-6726 Szeged, Hungary
[e] ZF-Screens B.V., Niels Bohrweg 11, 2333 CA Leiden, The Netherlands
[f] BaseClear B.V., Einsteinweg 55, 2333 CC Leiden, The Netherlands

ARTICLE INFO

Article history:
Received 23 June 2010
Received in revised form
15 August 2010
Accepted 24 August 2010
Available online 1 September 2010

Keywords:
Zebrafish embryo
Innate immunity
Salmonella typhimurium
Tag-Seq
RNA-Seq

ABSTRACT

Salmonella enterica serovar *Typhimurium* (*S. typhimurium*) bacteria cause an inflammatory and lethal infection in zebrafish embryos. To characterize the embryonic innate host response at the transcriptome level, we have extended and validated previous microarray data by Illumina next-generation sequencing analysis. We obtained 10 million sequence reads from control and *Salmonella*-infected zebrafish embryos using a tag-based sequencing method (DGE or Tag-Seq) and 15 million reads using whole transcript sequencing (RNA-Seq), which respectively mapped to circa 65% and 85% of 28,716 known Ensembl transcripts. Both sequencing methods showed a strong correlation of sequence read counts per transcript and an overlap of 241 transcripts differentially expressed in response to infection. A lower overlap of 165 transcripts was observed with previous microarray data. Based on the combined sequencing-based and microarray-based transcriptome data we compiled an annotated reference set of infection-responsive genes in zebrafish embryos, encoding transcription factors, signal transduction proteins, cytokines and chemokines, complement factors, proteins involved in apoptosis and proteolysis, proteins with antimicrobial activities, as well as many known or novel proteins not previously linked to the immune response. Furthermore, by comparison of the deep sequencing data of *S. typhimurium* infection in zebrafish embryos with previous deep sequencing data of *Mycobacterium marinum* infection in adult zebrafish we derived a common set of infection-responsive genes. This gene set consists of known and putative innate host defense genes that are expressed both in the absence and presence of a fully developed adaptive immune system and that provide a valuable reference for future studies of host–pathogen interactions using zebrafish infection models.

1. Introduction

In the recent years zebrafish has become widely used as a model for *in vivo* studies of host–pathogen interactions. Zebrafish develop both an innate and adaptive immune system with notable similarities to that of mammals [1,2]. Zebrafish embryos can be exploited to study innate immunity separately from adaptive immune functions, since components of the innate immune system are functional already at the first day of embryogenesis contrary to the adaptive immune system that is not active during the first weeks of zebrafish development [3–6]. Furthermore, the externally developing and transparent zebrafish embryos are highly suited for real-time analysis of host–pathogen interactions, which can be combined with efficient gene knock-down analysis using antisense morpholino oligonucleotides. It has been demonstrated that the components of the main innate immune signaling pathways are strongly conserved between zebrafish and mammals [7,8] and several infection models for studying innate immune response mechanisms in zebrafish embryos have now been developed [9].

Salmonella infections, causing salmonellosis and typhoid fever, are studied in several animal models, of which the best studied is the mouse model of *S. enterica serovar Typhimurium* infection (hereafter referred to as *S. typhimurium*) [10]. The opportunity of real-time analysis led to the development of a *S. typhimurium* infection model

* Corresponding author. Tel.: +31 71 5274927; fax: +31 71 5274357.
E-mail address: a.h.meijer@biology.leidenuniv.nl (A.H. Meijer).

1050-4648/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.fsi.2010.08.022

in zebrafish embryos [11]. Intravenous infection of 1-day-old zebrafish embryos with *S. typhimurium* strain SL1027 resulted in a lethal infection with bacteria showing intracellular replication in macrophage-like cells as well as extracellular replication in micro-colonies at the epithelium of blood vessels. In contrast, lipopolysaccharide (LPS) mutants of *S. typhimurium* (Ra) were non-pathogenic in zebrafish embryos, similar as in mammalian hosts [11]. Components of the *S. typhimurium* cell wall and motility apparatus trigger innate host defense pathways, including Toll-like receptor (TLR) signaling [12]. A morpholino knock-down analysis of the common TLR-adaptor protein, MyD88, showed that zebrafish embryos lacking MyD88 function lost the ability to clear an infection with the attenuated *S. typhimurium* Ra mutant strain, demonstrating that the innate immune response of the zebrafish embryos involved MyD88-dependent signaling [13]. To characterize the zebrafish embryonic host immune response to *S. typhimurium* wild-type and Ra mutant infection a time-course microarray analysis was performed, showing the induction of genes encoding cell surface receptors, signaling intermediates, transcription factors, and inflammatory mediators, with strong similarity to host responses detected in other vertebrate models and human cells [14]. A conserved role of zebrafish Toll-like receptor 5 (TLR5) homologues in recognition of *Salmonella* flagellin was demonstrated [14]. Furthermore, similar as mammals, zebrafish embryos were shown to employ both MyD88-dependent and MyD88-independent signaling pathways during infection [14].

As demonstrated by our previous microarray analysis, the *S. typhimurium*-zebrafish model presents a useful case study for the embryonic innate host response to an inflammatory bacterial infection [14]. Here we have extended this microarray study by a deep sequencing analysis using the previously described tag-based sequencing method known as digital gene expression (DGE) [15,16] also named Tag-Seq [17]. We determined the overlap between deep sequencing and microarray data and report a detailed annotation of the *S. typhimurium*-responsive gene set validated by both analysis methods. Furthermore, we compared the tag-based sequencing approach with whole transcript sequencing (RNA-Seq), and based on the overlap between the data sets demonstrate the usefulness of both deep sequencing approaches for transcriptome quantitation during infection. We compared the data with our previous deep sequencing analysis of *Mycobacterium marinum* infection in adult zebrafish and annotated the gene set commonly induced in both infection models. These annotated gene sets provide a valuable reference for future studies using zebrafish infection models.

2. Materials and methods

2.1. DGE (Tag-Seq) library construction and sequencing

The RNA samples for DGE analysis were identical to those used for previous microarray analysis [14]. In brief, zebrafish embryos were infected with *Salmonella typhimurium* (strain SL1027) by microinjection of approximately 250 colony forming units of DsRED-labeled bacteria into the caudal vein close to the urogenital opening after the onset of blood circulation (27 h post fertilization). An equal volume of PBS was injected in the control group. RNA samples were collected at 8 h post infection (hpi) and samples from triplicate infection experiments were pooled. DGE libraries from the RNA pools (1 μg) of *Salmonella*-infected and control embryos were prepared using the DGE:Tag Profiling for NlaIII Sample Prep kit from Illumina as previously described [15]. The libraries were sequenced in duplicate using 2 and 3 pmol of cDNA. Sequencing was performed using the Illumina Genome Analyzer II System (BaseClear B.V., Leiden, The Netherlands) according to the manufacturer's protocols. Image analysis, base calling, extraction of 17 bp tags and tag counting were performed using the Illumina pipeline. Tag counts from duplicate libraries were merged *in silico*. The raw data were deposited in the GEO database under submission number GSE22472.

2.2. RNA-Seq library construction and sequencing

Samples used for whole RNA transcript sequencing were the infected and uninfected control groups from a morpholino knock-down study to be reported elsewhere [28]. The procedure of *S. typhimurium* infection and the time-point of analysis (8 hpi) were identical as for the DGE analysis described above and previous microarray study [14]. Total RNA was isolated using the Qiagen miRNeasy kit according to the manufacturer's instructions (QIAGEN GmbH, Hilden). RNA-Seq libraries were made from 4 μg of each sample, using the Illumina mRNA-Seq Sample Preparation Kit according to the manufacturer's instructions (Illumina, Inc. San Diego). An amount of 4 pmol of each library was sequenced in one lane with a read length of 51 nt using the Illumina Genome Analyzer II System (BaseClear B.V., Leiden, The Netherlands). The raw data were deposited in the GEO database under submission number GSE21024.

2.3. DGE (Tag-Seq) data analysis

Mapping of tag sequences to transcript databases or to the zebrafish genome was performed as previously described [15]. For transcript mapping we used the Ensembl *Danio rerio* Zv8.55 database, the RefSeq database (2009-09-14), and the *Danio rerio* UniGene build 105 and 117 databases. For comparison of Tag-Seq and RNA-Seq data the Ensembl transcript database derived from the latest version of the zebrafish genome was used. For comparison of Tag-Seq and microarray data we used the UniGene build 105 database, since this database was used in previous microarray analysis [14]. For genomic mapping the native and masked form of the zebrafish genome version Zv8 were downloaded from the FTP server of the Ensembl database. Statistical comparison of DGE/Tag-Seq data from *Salmonella*-infected and control embryos was performed using the Bayesian method described by Lash et al. (2000) with the software tool available from the SAGEmap resource [18]. Briefly, the method performs a key-by-key comparison of two key-count distributions by generating a probability that the frequency of any key in the distribution differs by more than a given fold factor from the other distribution. For two Tag-Seq libraries, the algorithm performs a differential, tag-by-tag count comparison, with correction for the total size of the library. In our analysis we used a 2-fold factor difference of transcript expression level as the subject of the Bayesian statistical evaluation. The algorithm returns a probability value (P) for each tag describing the chance that the detected count numbers represent a fold difference of the tag concentration between the investigated samples greater than or equal to 2. The change of a tag expression was accepted as significant if P was above 0.95. As an alternative means to evaluate differential expression in Tag-Seq data sets we developed the Cumulative Transcript Detection Index (CTDI), which accumulates data from all tags that map to the same transcript:

$$\mathrm{CTDI} = \sum_{i=1}^{n} D_i P_i^2 \left| \frac{\sum_{i=1}^{n} D_i P_i^2}{\sum_{i=1}^{n} |D_i P_i^2|} \right|$$

where n is the number of the detected tag entities in a transcript, P is the significance of tags [18], D is the coefficient for the direction of change (1 for increase or −1 for decrease).

In the CTDI calculation P^2 is used for giving increased weight for tags with higher significance, while the formula gives lower weight to those transcripts where tags are present that show changes in the