Universiteit
Leiden
The Netherlands

Transcriptome profiling of infectious diseases and cancer in zebrafish

Ordas, A.K.

Citation
Ordas, A. K. (2010, June 29). *Transcriptome profiling of infectious diseases and cancer in zebrafish*. Retrieved from https://hdl.handle.net/1887/15734

Version:	Not Applicable (or Unknown)
License:	Leiden University Non-exclusive license
Downloaded from:	https://hdl.handle.net/1887/15734

Note: To cite this publication please use the final published version (if applicable).

Transcriptome profiling of infectious diseases and cancer in zebrafish

PROEFSCHRIFT

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van de Rector Magnificus prof. mr. P.F. van der Heijden,
volgens besluit van het College voor Promoties
te verdedigen op dinsdag 29 juni 2010
klokke 10:00 uur

door

Anita Katalin Ordas
geboren te Szeged, Hungary
in 1980

Promotiecommissie

Promotor:	Prof. dr. H.P. Spaink
Co-promotor:	Dr. A.H. Meijer
Overige leden:	Prof. dr. C.J. ten Cate
	Prof. dr. J. den Hertog
	Prof. dr. J. Memelink
	Prof. dr. B. van de Water
	Prof. dr. N.S. Foulkes (University of Heidelberg)
	Dr. M. Mink (University of Szeged)

ISBN: 978-90-8570-579-6

Printed by Wöhrmann Print Service, Zutphen

Go where the science takes you

Contents

Chapter 1 7
General introduction

Chapter 2 29
Deep sequencing of the zebrafish transcriptome response to mycobacterium infection

Chapter 3 57
Deep sequencing of the innate immune transcriptomic response of zebrafish embryos to Salmonella infection

Chapter 4 77
MicroRNA expression during bacterial infections in zebrafish

Chapter 5 115
Liver tumor-related microRNA expression is conserved between zebrafish and human

Chapter 6 143
Summary and General discussion

Samenvatting 151

Összefoglaló 159

Curriculum vitae 165

List of publications 166

Chapter 1

General Introduction

Transcriptome analysis is of great importance in the field of genomic research. Genome-wide analysis of expressed transcripts in a given cell or tissue can reveal significant information about gene expression and regulation under different conditions such as developmental and disease processes. Recent studies have provided novel insights into the complexity of transcriptomes and the regulation of gene expression. The regulatory role for small non-coding RNAs is one of the key discoveries that have given a new dimension to the understanding of gene expression control. The latest developments in technologies for transcriptome profiling include the application of next generation sequencing systems that have made an enormous impact on biology. Here we applied these new genomics approaches to study infectious diseases and cancer using zebrafish as a model system.

Transcriptomics

The transcriptome comprises all expressed RNA transcripts in a cell or an organism. Transcriptomics measures all transcript species in a given cell, including protein-coding mRNAs and small non-coding RNAs, and gives insight into how the expression levels of these transcripts are regulated under different conditions. Analyzing the transcriptome helps to characterize the complexity of the genome. Transcriptional profiling, besides cataloging the known transcripts, can lead to the identification of novel gene models by discovering novel gene boundaries, splicing variants, additional promoters, alternative exons, 5'and 3' untranslated regions (UTRs), non-coding transcripts and antisense transcripts (Cloonan et al., 2008; Morin et al., 2008; Mortazavi et al., 2008; Pan et al., 2008; Rosenkranz et al., 2008; Sultan et al., 2008; Wang et al., 2009). Quantifying expression levels of transcripts also facilitates our understanding of the molecular background of specific developmental processes and pathological conditions (Brentani et al., 2005; Mardis and Wilson, 2009).

In the last decade several high-throughput methods have been developed for large-scale studies of gene expression levels. Basically, all these methods can be divided into two categories: microarray-based and sequencing-based approaches.

Since the mid-1990s, microarray-based methods have become generally used to analyze large-scale datasets. Using microarrays, scientists are able to detect the expression of thousands of genes simultaneously. The common basics of these methods are the hybridization of fluorescently labeled cDNA or amplified RNA to complementary oligonucleotide probes on the microarray platforms and detection of the hybridization intensity relative to the abundance of the particular mRNA species (Schena et al., 1995). They have been widely applied in molecular profiling of different developmental processes, diseases, pharmacogenomic responses, or whole genome association studies (Bertone et al., 2004; Brentani et al., 2005; Douglas and Ehlting, 2005; Scherf et al., 2000; White, 2001)). The microarray technique al-

lows to detect and measure gene expression at the mRNA level, to describe genetic variations (SNPs, copy-number variations, chromosomal changes), and to profile DNA-protein interactions. Lately, more specific types of arrays have been developed like tiling (Bertone et al., 2004), all-exon (Gardina et al., 2006) or exon-junction (Clark et al., 2002) arrays, which are used to analyze specific regions of the genome. Furthermore, such array platforms as kinome or phosphatome chips have also been developed recently to determine the phosphorylation status of expressed gene products (Arena et al., 2005; Diks et al., 2004; Hestvik et al., 2003; Parikh et al., 2009). Although microarray technology is high-throughput and relatively inexpensive it has some limitations, including the necessity of existing knowledge of the genome sequence for designing arrays or cross-hybridization problems between highly related sequences (Okoniewski and Miller, 2006; Royce et al., 2007). Additionally, issues such as poor sensitivity for low abundance-species, difficulties in comparing expression levels across different experiments, requirement for complicated normalization methods, and the measurement of relative rather than absolute expression levels, add to the limitations of microarrays (Irizarry et al., 2005; Okoniewski and Miller, 2006; Royce et al., 2007).

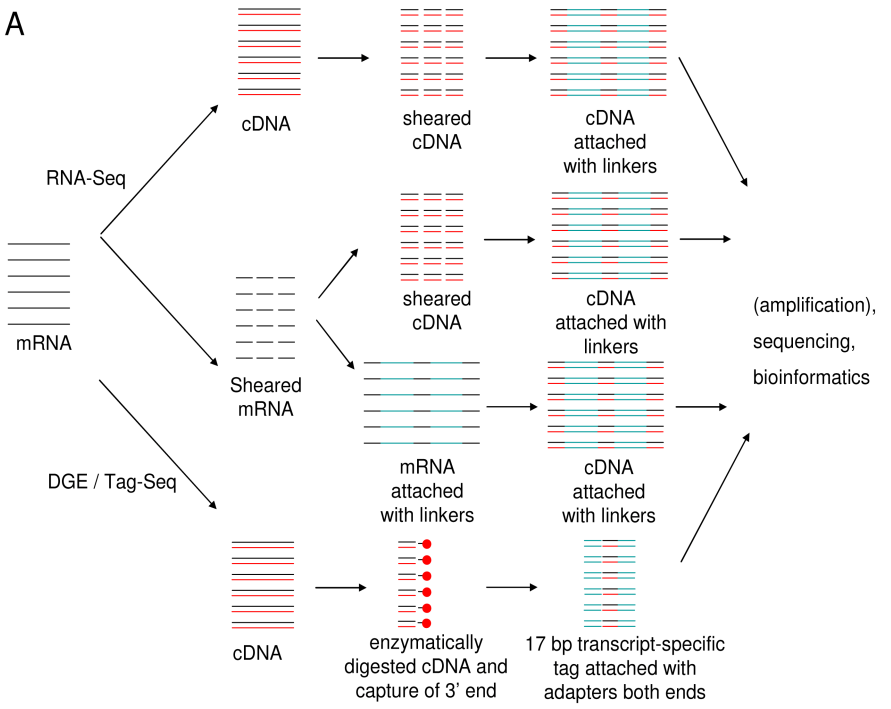
Another approach to quantify and annotate transcriptomes is the sequencing-based approach which overcomes some limitations of array-based methods (Irizarry et al., 2005; Pedotti et al., 2008; t Hoen et al., 2008; Wilhelm and Landry, 2009). Sequencing-based transcriptome profiling methods directly determine cDNA sequences. The first type of sequence-based approach was the cataloging of expressed sequence tags (ESTs) using the Sanger sequencing method. In the EST method single sequencing reads are made of the 5'or 3'ends of all cDNA clones in a cDNA library. The resulting sequences, typically 500 to 800 bp in length, give an overview of the mRNAs represented in the cDNA library. Sanger sequencing has been used for decades for deciphering complete genes and later whole genomes and it is still the most commonly used technique to date because of the long read length and low error rate. However, the Sanger-based EST method has its weaknesses namely being non-quantitative, and relatively low-throughput with high costs (Boguski et al., 1994; Gerhard et al., 2004).

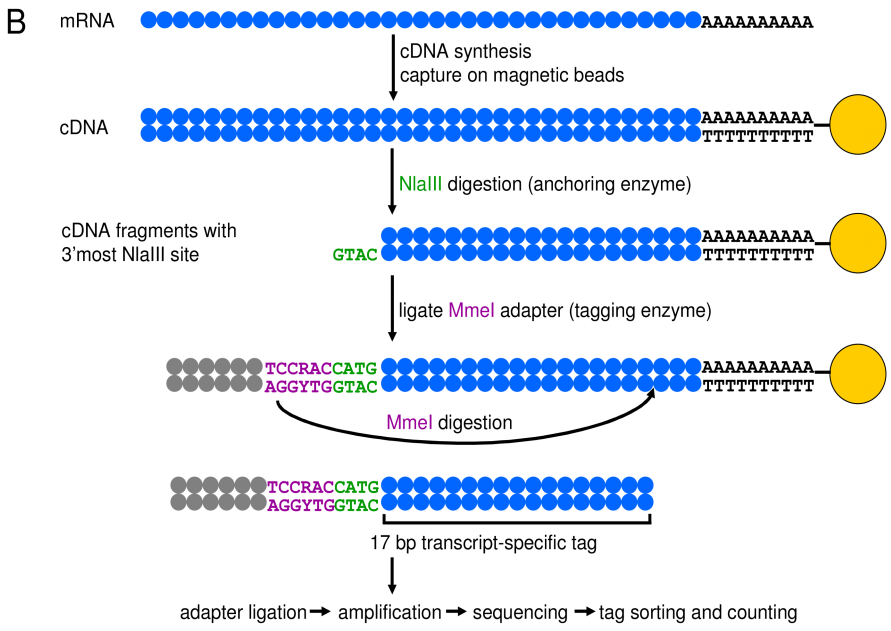
The shortcomings of the EST method triggered the generation of 'clone-and-count' (Tyagi, 2000) tag-based approaches such as serial analysis of gene expression (SAGE) (Carninci et al., 2005; Harbers and Carninci, 2005; Velculescu et al., 1995). In this profiling technique short sequence fragments representative for each cDNA are produced and concatenated to form large clones that are subsequently sequenced by the Sanger method. The sequencing step is followed by counting the number of times a short signature sequence (tag) has been observed (Morozova and Marra, 2008). Many technologies have been developed improving the basics of SAGE such as massively parallel signature sequencing (MPSS) (Brenner et al., 2000), polony multiplex analysis of gene expression (PMAGE) (Kim et al., 2007), cap analysis of gene expression (CAGE) (Carninci et al., 2005), LongSAGE (Wahl et al., 2005),

DeepSAGE (Nielsen et al., 2006) and SuperSAGE (Matsumura et al., 2005) techniques. These are medium to high throughput methods providing qualitative and quantitative data with accurate, digital output of gene expression levels compared to the analog nature of the signals generated by microarray technology. The results illustrate absolute abundances of transcripts instead of the relativity of array-based techniques and there is no requirement for prior knowledge of sequences and no bias due to a predefined array content (Harbers and Carninci, 2005). SAGE-derived sequencing methods have been applied for studying several human genetic diseases such as acute myeloid leukemia, neurodegenerative diseases, and cancer (Horan, 2009). While these methods have been used in several transcriptome analyses, the high cost, the low throughput level of sequencing and difficulties of cloning steps has limited their application.

The very recently developed next generation sequencing methods by Illumina/Solexa, ABI/SOLiD, 454/Roche, and Helicos are ultra high-throughput deep sequencing technologies that have substantially reduced the cost of sequencing. Transcriptome profiling methods based on next generation sequencing technology include the so-called RNA sequencing (RNA-Seq) and Digital Gene Expression (Fig. 1A) (DGE or Tag-Seq) that allow the simultaneous determination of millions of short sequences (Hegedus et al., 2009; Morrissy et al., 2009; Mortazavi et al., 2008; Nagalakshmi et al., 2008; Shendure, 2008; t Hoen et al., 2008; Wang et al., 2009; Wilhelm and Landry, 2009). In these methods the cloning step is eliminated and the advantages of these sequencing-based transcriptome profiling approaches over the microarray technique include the absence of background signals and large dynamic range, the improved ability to distinguish between homologous sequences, and the high accuracy, resolution and sensitivity (Shendure, 2008). In RNA-Seq methods RNA is first sheared and then converted to cDNA, or cDNA is produced first and then sheared into short fragments. Subsequently, adaptors are attached to one or both ends of the cDNA fragments, and, with or without an amplification step, these cDNA fragments

Figure 1. Next generation sequencing technologies. A) Schematic representation of RNA sequencing (RNA-Seq) and Digital Gene Expression (DGE or Tag-Seq) methods. During RNA-Seq either the mRNA is first converted to cDNA, then randomly sheared to short fragments and attached by linkers, or first mRNA is randomly sheared and then converted to cDNA and attached by linkers. An alternate option is that the fragmented mRNAs are attached by linker molecules before conversion to cDNA. During DGE the isolated mRNA is converted to DNA and through enzymatic digestions 17 bp long transcript-specific tags are generated and linked to adaptors. In both methods after amplification steps, sequencing and bioinformatic analysis are performed. Figure modified from Graveley BR (2008) Power sequencing. Nature 453: 1197. B) Schematic representation of Tag-Seq. After cDNA synthesis and capture on magnetic beads fragments are digested with an anchoring enzyme NlaIII, ligated to MmeI, a tagging enzyme cutting 21 nt downstream of the recognition sequence creating 17 bp transcript-specific tags. These tags are then ligated to adapters, amplified, sequenced, sorted and quantified.


A
RNA-Seq
mRNA
cDNA
sheared cDNA
cDNA attached with linkers
Sheared mRNA
sheared cDNA
cDNA attached with linkers
DGE / Tag-Seq
mRNA attached with linkers
cDNA attached with linkers
(amplification), sequencing, bioinformatics
cDNA
enzymatically digested cDNA and capture of 3' end
17 bp transcript-specific tag attached with adapters both ends


B
mRNA
AAAAAAAAAA
cDNA synthesis
capture on magnetic beads
cDNA
AAAAAAAAAA
TTTTTTTTTT
NlaIII digestion (anchoring enzyme)
cDNA fragments with 3'most NlaIII site
GTAC
AAAAAAAAAA
TTTTTTTTTT
ligate MmeI adapter (tagging enzyme)
TCCRACCATG
AGGYTGGTAC
AAAAAAAAAA
TTTTTTTTTT
MmeI digestion
TCCRACCATG
AGGYTGGTAC
17 bp transcript-specific tag
adapter ligation → amplification → sequencing → tag sorting and counting

are then sequenced to obtain short reads from one end (single-end sequencing) or both ends (pair-end sequencing). The reads, which are typically 30–400 bp, depending on the sequencing platform, are aligned to a reference genome or reference transcript database, or assembled *de novo* without the genomic sequence to produce a genome-scale transcription map (Wang et al., 2009). As an alternative to RNA-Seq methods, DGE or Tag-Seq involves the high-throughput sequencing of transcript-specific sequence tags similar as in conventional SAGE but without the need for tags concatenation and cloning (Fig. 1B) (Hegedus et al., 2009; Morrissy et al., 2009; t Hoen et al., 2008). While RNA-Seq methods provide more detailed information of the transcriptome complexity, DGE or Tag-Seq is more suitable for quantification and more affordable for the comparison of larger numbers of samples. In addition to the usefulness for mRNA expression profiling, next generation sequencing techology has advanced the analysis of small noncoding RNAs (Friedlander and Little, 2009; Hackenberg et al., 2009; Morin et al., 2008; Wyman et al., 2009). Furthermore, next generation sequencing technologies make an enormous contribution to the whole-genome sequencing of microbial, plant and animal genomes (Engstrand, 2009; Qi et al., 2009). Several other applications include targeted resequencing, cancer genome sequencing, HLA genotyping for immune-mediated conditions, and studies of epigenetic modifications (Bentley et al., 2009; Cloonan et al., 2008; Gabriel et al., 2009; Gargiulo and Minucci, 2009; Hurd and Nelson, 2009; Lister et al., 2008; Mardis and Wilson, 2009; Mortazavi et al., 2008; Nagalakshmi et al., 2008; Sultan et al., 2008; Wilhelm and Landry, 2009). The application of next generation technology to the analysis of transcriptomes has shown that both in prokaryotic and eukaryotic species much more of the genome is transcribed than previously known. Several recent studies revealed novel transcribed regions, the use of additional promoters, and extensive variation in splicing isoforms, gene boundaries, and 3'UTRs. In addition, there is increasing evidence for the abudant presence of antisense transcripts and other non-coding transcripts. (Cloonan and Grimmond, 2008; Morin et al., 2008; Mortazavi et al., 2008; Pan et al., 2008; Sultan et al., 2008).

Despite many advantages, sequence-based trancriptome profiling is more complex and currently still more costly than microarray analysis. Comparison of sequence-based and array-based transcriptome data has revealed overlaps as well as technology-dependent differences (Hegedus et al., 2009; t Hoen et al., 2008). Therefore, using both methods can be considered complementary to each other.

Regulation of gene expression

Transfer of biological information from DNA through RNA to proteins is the central dogma of molecular biology that was first determined by Francis Crick and his team more than 50 years ago. Transfer of the information is called gene expression. In the first step of gene expression the sequence information is transferred from DNA to messenger RNA (mRNA) by the mechanism called transcription. Followed

by several maturation steps mRNA is translated into proteins. These processes are strictly regulated, which is essential to ensure the correct macromolecular machinery of the living organism.

Gene expression is regulated on different levels, from transcription, through post-transcriptional regulation to post-translational modification of the protein. These mechanisms together tune the appropriate level of the gene product. In eukaryotes, transcription depends on the chromatin structure of the DNA, which is a complex of DNA and its binding proteins, mostly histones. Modification of this structure by methylation and acetylation of histones, and DNA-binding proteins, decreases the binding affinity of RNA polymerase and can result in gene silencing. On the transcriptional level, several proteins and complexes can regulate gene expression, acting as transcription factors (activators or repressors) or transcriptional cofactors that can modulate the activity of transcription factors or the function of RNA polymerase. After mRNA is transcribed, posttranscriptional modifications can alter the level of gene expression as well. If RNA molecules are stabilized by a 5' cap and polyadenylated tail they are protected from degradation and transported to the site of translation. The translation process provides another level of regulation either by ribosome recruitment at the initiation step of translation or by modulation of the elongation and termination steps of the protein synthesis. Secondary structure of the mRNA is also an important factor in regulation.

According to the central dogma, transcripts coding for proteins are transcribed from the protein-coding strand of the DNA that is called the sense strand. However, numerous antisense transcripts have been detected in prokaryotes and viruses lately. Antisense transcripts are molecules that are transcribed from the opposite strand to the protein-coding strand. These transcripts were found to regulate gene expression by affecting mRNA transcription, processing and translation (Wagner and Simons, 1994). Recent studies provide evidence that the majority of genes are transcribed in a bidirectional manner in eukaryotes as well (Carninci et al., 2005; Katayama et al., 2005; Yelin et al., 2003). Although the mechanism by which antisense RNAs act on sense strand expression is largely unknown, antisense transcripts have been ascribed to play roles in gene regulation, including degradation of the corresponding sense transcript (RNA interference) and in gene silencing at the chromatin level (Beiter et al., 2009; Carninci et al., 2005; Katayama et al., 2005). As an example of the antisense phenomenon, the proper dosage of the expression of PU.1, a human and murine hematopoietic transcription factor gene critical for suppression of leukemia, was shown to rely on antisense RNA modulators (Ebralidze et al., 2008).

Recently, it has become evident that small non-coding RNAs play major roles in the regulation of gene expression by binding to complementary sequences of mRNAs and destabilizing and sequestering them for degradation. Small non-coding RNAs (ncRNA) are 19-31 nucleotide long functional RNAs that are not translated into protein. There are several groups of them, such as transfer RNA (tRNA), ribosomal RNA (rRNA), small nuclear (snRNA), small nucleolar RNA (snoRNA), trans-

acting (tasiRNAs), small-scan RNAs (scnRNAs), repeat-associated siRNAs (rasiRNA), Piwi-interacting RNAs (piRNAs), microRNA (miRNA) and small interfering RNA (siRNA) (Chu and Rana, 2007). They are involved in many cellular processes both in plants and animals. Small RNAs can be conserved or specific for different species. Although there are numerous classes and sizes of ncRNAs, they all share a common function in cellular physiology: epigenetic regulation of gene expression. They behave as sequence-specific triggers for mRNA degradation, translation repression, heterochromatin formation, and transposon control. In the recent years, both microarray-based and next generation sequencing methods have been applied for profiling of non-coding transcripts and many novel small ncRNAs have been discovered using these high-throughput techniques (Morozova and Marra, 2008).

In the last decade, miRNAs have become one of the major focuses of research in molecular biology and they have one of the best understood and described roles in gene regulation among the small ncRNA molecules. MicroRNAs are involved in the post-transcriptional regulation of gene expression by regulating mRNA translation or stability (Filipowicz et al., 2008; Jackson and Standart, 2007; Nilsen, 2007). Although a large number of studies have been performed on miRNAs and many miRNAs have already been discovered, our knowledge about miRNA-directed regulation is still not complete and a large number of miRNAs have not yet been identified. It can be expected that in the coming years the application of the above-mentioned high-throughput methods and further technology developments for transcriptomic analysis will lead to the identification of additional miRNAs and will make a large contribution to the analysis of their function.

MicroRNAs

MicroRNAs (miRNAs) are small non-coding single-stranded RNAs with a length of 19-25/21-25 nucleotides (nt). They regulate gene expression by binding to their target mRNAs, and they are key negative regulators of stability and translation (Schmidt et al., 2009).

The first miRNA was detected in *C. elegans* in 1993. *Lin-4*, a 22 nucleotide long small RNA was shown to be a post-transcriptional regulator of *lin-14* and essential in post-embryonic developmental events. This phenomenon was thought to be specific, but a few years later another small RNA *let-7* was also found to be important in the development of *C. elegans* and to be conserved in many other animal species (Pasquinelli et al., 2000). This finding also led to the discovery of hundreds of miRNAs in plants and animals using bioinformatic approaches and molecular cloning. MiRNAs have been demonstrated to show high conservation between different eukaryotic species (Griffiths-Jones et al., 2005; Lim et al., 2003; Rodriguez et al., 2004).

The miRNAs show high diversity of function as they are involved in a wide variety of biological processes such as proliferation, differentiation, cell fate determination, apoptosis, signal transduction, development and disease pathogenesis (Sonkoly and

Pivarcsi, 2009). Their importance in development has been proved by blocking the production of all miRNAs in mouse and zebrafish, which resulted in developmental arrest during embryogenesis (Kloosterman et al., 2004). Regarding their involvement in diseases, numerous miRNAs have been detected to play important roles in tumorigenesis (Bellon et al., 2009). Nevertheless, the biological relevance of most miRNAs and their targets still remains to be determined.

Genes coding for miRNAs are located both in protein-coding and non-coding regions of the genome. MiRNA genes are either organized in clusters or encoded in single loci. MiRNAs in clusters are first transcribed as polycistronic primary transcripts and later processed to single miRNAs (Medina and Slack, 2008). The majority of all known mammalian miRNAs are mapped to intronic regions of either coding or non-coding transcriptional units, whereas ~10% are detected at exonic regions of non-coding transcripts.

MiRNAs are produced through a series of maturation steps (Fig. 2) (reviewed in Carissimi et al., 2009; Medina and Slack, 2008; Williams et al., 2008). They first are transcribed as a several kilobases long 5' capped and 3' polyadenylated primary transcript (pri-miRNA) from miRNA genes, mostly by RNA polymerase II. This double-stranded pri-miRNA is recognized and cleaved in the nucleus by the RNAase III enzyme Drosha /DGCR8 complex into one or more precursor miRNAs (pre-miRNA). The pre-miRNA is a 60-80 nt long stemloop structure with a 2-nucleotide 3' overhang, which is recognized and transported to the cytoplasm by exportin-5 via a Ran-GTP-dependent mechanism. In the cytoplasm, an RNase III enzyme called Dicer cleaves the loop from the stem, thus producing a mature 20-23 nt double-stranded miRNA. Dicer cleavage requires association with the transactivating response RNA-binding protein (TRBP) and Argonaut 2 (Ago2) completing the RNA Induced Silencing Complex (RISC)/ribonuclear particle (RNP) complex. Only the strand with the higher stability of the duplex in the 5' half, the guide strand, is subsequently loaded into a RISC. The other strand (passenger strand or star strand) is degraded. However, in some cases both strands of the pre-miRNA are detected as mature miRNAs. These mature miRNAs are now able to bind to their targets by recognizing sequence complementarity.

Binding by perfect complementarity commonly occurs in plants and only occasionally in animals (Tang et al., 2003). In this case the mRNA is deadenylated and subsequently degraded. In animals, the miRNAs bind more frequently to the 3' untranslated region (3' UTR) of their targets with imperfect complementary. The sequence complementarity in the so-called seed sequence of the miRNA (nucleotide 2-8 from the 5' end) seems to determine the specificity of miRNA–target RNA. Binding of the 3' end of the miRNA to the target may facilitate the regulation and can even compensate for the improper complementarity of the seed sequence (Brennecke et al., 2005; Didiano and Hobert, 2006). This results in translational inhibition, followed by a variable degree of mRNA degradation. The mechanisms by which miRNAs regulate translational silencing are either the blocking of translational initiation

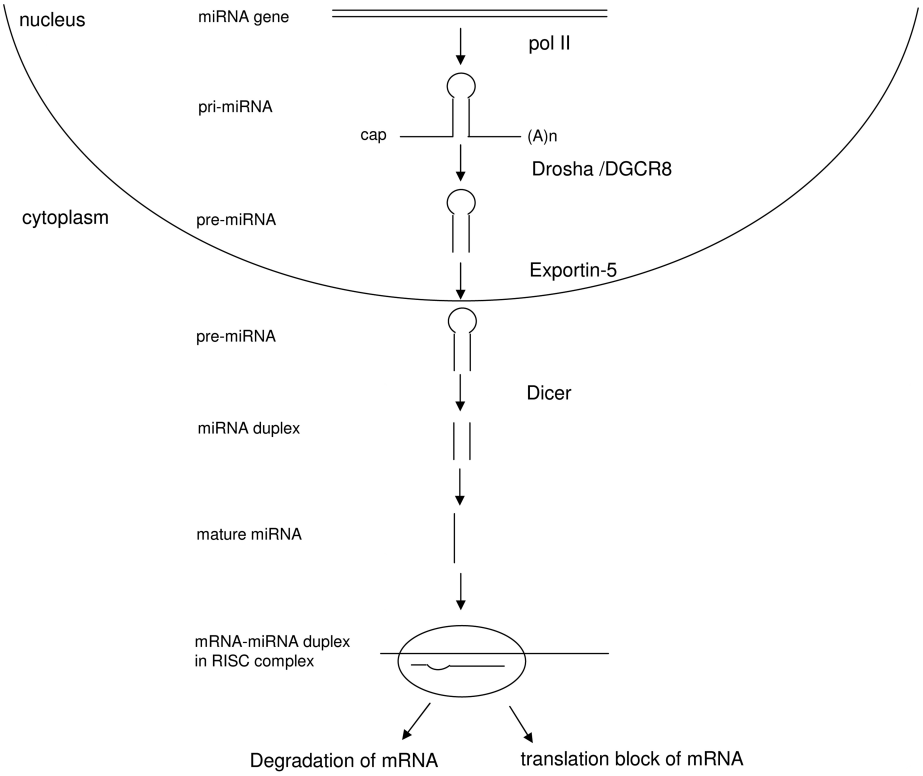


Figure 2. MicroRNA biogenesis. In the nucleus miRNA genes are first transcribed as primary miRNA (pri-miRNA) transcripts with 5′ cap and 3′ poly(A) tail by RNA polymerase II (pol II) which are recognized and cleaved by Drosha/DGRC8. The generated hairpin precursor miRNA (pre-miRNA) transcript is exported to the cytoplasm by exportin-5 and processed by Dicer RNase enzyme producing 20-23 nt double-stranded miRNA. After cleavage, only the guide strand, the mature miRNA strand is loaded into RNA Induced Silencing Complex (RISC)/ribonuclear particle (RNP) complex leading to mRNA cleavage or induction of translational repression. In some cases, also the other strand (miRNA star sequence) can be incorporated into the RISC complex with similar efficiency, and therefore can also be functional.

or elongation steps, or the sequestering of their targets in the cytoplasmic P-bodies, where mRNA storage and decay takes place (Medina and Slack, 2008). It has recently been shown that translational repression occurs prior to deadenylation and destabilization (Cannell et al., 2008).

In animals the expression of 30% or more of the genes is predicted to be regulated by miRNAs (Kapsimali et al., 2007). While some genes involved in basic cellular processes avoid miRNA regulation due to short 3'UTRs that are specifically depleted of miRNA binding sites, others are regulated by a set of miRNAs. On the other

hand, the effects of a miRNA can be pleotropic. It has been estimated that a single miRNA family can regulate as many as 200 or even more different genes (Medina and Slack, 2008). One miRNA may have several target sites in the 3'UTR of a single mRNA (Saetrom et al., 2007). Spatial and temporal expression patterns of miRNAs have been shown to be related to their conserved functions (Kapsimali et al., 2007; Lagos-Quintana et al., 2002; Sempere et al., 2004). It has also been demonstrated that changes in the level of specific miRNAs correlate with abnormal conditions of a cell, tissue or organism (Kloosterman et al., 2006).

MiRNAs in diseases

Since miRNAs play important roles in different cellular mechanisms responsible for maintaining the cell's homeostasis, it is conceivable that their aberrant expression may contribute to a wide variety of diseases. It was discovered that miRNAs are associated with neurodegenerative disorders like Alzheimer's disease (significantly decreased level of miR-107) and Parkinson's disease (very low level or absence of miR-133b in the midbrain tissue); hepatic disorders (involvement of miR-196 and miR-448); primary muscle disorders like Duchenne muscular dystrophy, Miyoshi myopathy and dermatomyositis (consistently up-regulated levels of miR-146b, miR-21, miR-155, miR-214, miR-222); or cardiovascular diseases like cardiac hypertrophy and coronary artery disease (miR-133, miR-1 over-expression) (Garofalo et al., 2008).

MiRNAs are also involved in the pathogenesis of infectious diseases and cancer as they play a role in the mechanisms of apoptosis, proliferation, stress resistance, metabolism, defense against pathogenic infections, and importantly tumorigenesis and metastasis (Medina and Slack, 2008).

There is increasing evidence that a complex network of miRNAs works within the immune system and that miRNAs themselves may also represent an ancient antiviral immune system (Lu and Liston, 2009). Complex regulatory mechanisms have been revealed where miRNAs fine-tune development and function of haematopoietic lineages, B and T cell homeostasis and response, and T cell selection in the thymus. MiR-181a has an essential role in the development of both B and T lymphocytes, miR-150 plays a critical role in B cell differentiation, miR-142 is involved in T cell development, and miR-223 is important in granulocytic and T cell differentiation (Carissimi et al., 2009; Sonkoly and Pivarcsi, 2009).

During infection, the first defense line of the organism is the innate immune response. As part of this response monocytes and macrophages recognize microbial ligands by pattern recognition receptors on their membranes, such as Toll-like receptors (TLRs) (Mogensen, 2009). During this first immune response hundreds of genes are induced to defeat the pathogen. Recently, several studies have suggested that miRNAs play an important role in the regulation of this complex response. Several miRNAs are differentially regulated by bacterial and viral infections ensur-

ing regulation of macrophage inflammatory responses. While miR-146 and miR-155 were up-regulated by bacterial ligands, miR-132 and miR-125b showed decreased expression level during LPS-treatment (Sonkoly et al., 2008). Beside these *in vitro* studies, *in vivo* experiments showed a rapid increase in the numerous miRNAs such as miR-21, miR-25, miR-27b, miR-100, miR-140, miR-142-3p, miR-181c, miR-187, miR-194, miR- 214, miR-223 and miR-224. During antiviral defense several miRNAs were over-expressed (such as miR-196, miR-296, miR-351, miR-431 and miR-448) having sequence-predicted targets within the hepatitis C virus (HCV) genomic RNA (Sonkoly and Pivarcsi, 2009).

In the past few years miRNA expression profiling has also come into focus in the field of molecular oncology. Several studies have demonstrated differentially expressed miRNAs in cancer. One class of these miRNAs, called oncomirs, act on oncogenes (oncogenic oncomirs) and/or tumor suppressor genes (tumor-suppressing oncomirs) and are direct causative agents of tumorigenesis. Other cancer-associated miRNAs appear to be indirectly involved in genomic, epigenomic or physiological changes relevant to the different types of cancer (Esquela-Kerscher and Slack, 2006; Lu et al., 2005; Medina and Slack, 2008). MiR-15 and miR-16 have decreased levels or show total absence in most patients with chronic lymphocytic leukemia, consistent with a tumor-suppressive function of these miRNAs (Calin and Croce, 2006). MiRNAs of the let-7 family also act as tumor suppressors *in vivo*, regulating RAS oncogenes commonly down-regulated in solid tumors (Garofalo et al., 2008). Increased levels of a particular miRNA, for example due to gene amplification or transcriptional disregulation, can also be a cause of cancer. Up-regulation of miR-155 in B cells is associated with high grade lymphoma pre-B leukemia (Costinean et al., 2006). Similarly, over-expression of the miR-106b-25 cluster was shown to play critical role in gastric cancer (Petrocca et al., 2008). Cancer pathogenesis may also be affected by mutations that disrupt or create new miRNA binding sites in target mRNAs (Costinean et al., 2006; Garofalo et al., 2008; Sonkoly and Pivarcsi, 2009). Differences in miRNA expression levels were found not only between normal and tumor tissues, but also between primary tumors and metastatic tissues. For example, miR-10b was shown to induce invasion and metastasis, whereas miR-335 and miR-126 are metastasis suppressors in human breast cancer (Garofalo et al., 2008).

Revealing groups of differentially expressed miRNAs under different disease conditions and demonstrating their biological functions is of great importance in understanding of pathogenesis and the development of novel strategies for disease treatment. These efforts are assisted by disease modeling using *in vitro* and *in vivo* systems.

Modeling of diseases

To study pathogenesis of different human diseases researchers use *in vitro* and animal models. *In vitro* cell-based assays are used to simplify experiments on patho-

genesis of human diseases and because of methodological and ethical issues regarding the use of mammals in experiments. Although cell culture systems are highly useful to unravel specific steps in complex physiological processes and signaling mechanisms, they cannot provide information about disease pathogenesis *in vivo* (Lieschke and Currie, 2007).

When choosing an animal model system researchers take into account the similarities and differences between the animal and human. There are practical as well as molecular and cell biological aspects to consider in selecting an animal as a model organism. Some of these are; husbandry infrastructure, cost of keeping animals, anatomical, molecular, genetic or pathological similarities, and possibility, feasibility and affordability of forward and reverse genetics (Lieschke and Currie, 2007). Several animal models have been developed to better understand different disorders including blood diseases, diabetes, muscular dystrophy, neurodegenerative disease, syndromes of angiogenesis and lipid metabolism, and also for infectious diseases and cancer.

Invertebrate animal models like *Caenorhabditis elegans* and *Drosophila melanogaster* are commonly applied, especially in the field of developmental biology based on their molecular and genetic similarities to vertebrates. However, by lacking many structures and organ systems involved in human diseases, these organisms are not very useful for studying pathogenesis. Although both *C. elegans* and *D. melanogaster* have an innate immune system – which is very primitive in the former but more complex and similar to that of vertebrates in the latter one – neither of them have an adaptive immune system, which limits their application in immunological studies (Ewbank, 2002; Hoffmann, 2003).

The closer evolutionary proximity and genomic, anatomical, cell biological and physiological homologies between human and other vertebrates makes these higher animals commonly used for disease modeling. For example, mouse and rat are excellent models in biomedical research because of the similarities they share with human and also the genetics approaches that are possible in these organisms. However, rodent and other mammalian animal models have some limitations, for example *in utero* development limits embryological studies. Furthermore, infrastructure support and high cost of experiments make large-scale screens challenging to perform.

Some of the limitations of mammalian models seem to be resolved by using zebrafish (*Danio rerio*) as a model organism. Recognizing their high developmental and cellular physiological similarities with mammals and advantages for experimental manipulation, George Streisinger pioneered the use of zebrafish as a vertebrate model system for developmental biology and embryogenesis (Kari et al., 2007). Working with zebrafish has several advantages, including that animals can be kept easier and at minimal cost compared to higher vertebrates. The zebrafish genome has been sequenced by the Sanger Center and is available to the research community. The zebrafish has many orthologous genes, and even conserved synteny with mammals (Barbazuk et al., 2000). Direct observation and manipulation of embryogenesis

is possible because of the external fertilization and development. Detecting functional and morphological changes is also facilitated by the optical transparency of the embryos and early adults. Small size, high fecundity (100-300 embryos per week) and short generation times (embryogenesis is complete by 5 days post fertilization) also make the zebrafish model suitable for experimental screens. The organs are permeable to small molecules during organogenesis, facilitating protein visualization with dye stainings and analyzing therapeutic compounds. These characteristics make zebrafish one of the major subjects of high-throughput screenings. Large scale forward and reverse genetic screens have made use of all these features of zebrafish and gained success in understanding vertebrate-specific mechanisms of development (Meeker and Trede, 2008).

From the 1980s zebrafish has become an excellent experimental model for studying pathogenic mechanisms of human diseases (Lieschke and Currie, 2007). Many zebrafish models of human diseases have been created through forward and reverse genetics approaches. With forward genetics approaches a set of point mutations are generated by random mutagenesis using the mutagen ethylnitrosourea (ENU) or by non-targeted retroviral- and transposon-mediated insertional mutagenesis. Phenotype-based screening of such mutant populations is a powerful approach to discover new genes or new functions for known genes. The main reverse genetics approaches include the generation of knock-outs by targeting induced local lesions in genomes (TILLING), transient gene knock-down with antisense morpholino oligonucleotides, transient gene over-expression with mRNA, and transgenesis. With the generation of specific zebrafish disease models, investigation of the molecular and cellular background of human diseases is facilitated. Zebrafish models have been generated for congenital and hereditary diseases, metabolic, endocrine, nutritional disorders, psychological and behavioral abnormalities, immunological diseases, infectious diseases, as well as carcinogenesis (Kari et al., 2007; Lieschke and Currie, 2007; Meeker and Trede, 2008).

Zebrafish as a model of infectious diseases and cancer

In the last decade, zebrafish has been developed as an *in vivo* model system for immunological research due to the remarkable similarities of its immune system to that of human (Traver et al., 2003; van der Sar et al., 2004). Similarly to mammals, zebrafish has both an innate and an adaptive immune system which developed evolutionary from the jawed fish (Kasahara et al., 2004). Zebrafish has the counterparts of most if not all determinants involved in the human immune system, for example macrophages and neutrophils, homologues of T and B lymphocytes, cytokines and signaling molecules, as well as a well-developed complement system and inflammatory proteins (Lesley and Ramakrishnan, 2008; Lieschke and Currie, 2007; Meeker and Trede, 2008). However, there are differences between the sites of maturation. One of the most important features of the zebrafish immune system is that the adap-

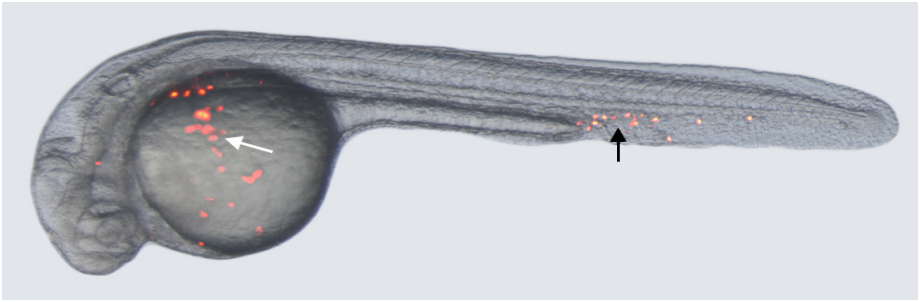


Figure 3. Real-time visualization of bacterial infection in the zebrafish embryo. The embryo was injected at 27 hours post fertilization with approximately 250 colony forming units of DsRed-labelled Salmonella typhimurium bacteria into the caudal vein and imaged at 2 hours post infection. An overlay picture of the bright field and DsRed fluorescent images is shown. Bacteria are trapped inside macrophages that are mainly observed in the bloodstream over the yolk sac (white arrow) and in the blood island posterior of the yolk sac extension (black arrow). The embryo is oriented with the anterior side to the left and dorsal side to the top.

tive immunity develops only after embryogenesis during the first several weeks of larval development, while the innate immune system is present from the first day of embryogenesis onwards. Combined with the transparency of the embryos, this characteristic allows for the real-time and intravital analysis of the role of the innate immune system in host-pathogen interactions, in separation from the role of adaptive immunity (Lesley and Ramakrishnan, 2008).

To better understand bacterial and viral pathological mechanisms several experimental infections have been developed in embryonal and adult zebrafish using Gram-negative bacteria (*Salmonella typhimurium*, *Salmonella arizonae*, *Vibrio anguillarum*, *Aeromonas salmonicida*, *Edwardsiella tarda*), Gram-positive bacteria (*Mycobacterium marinum*, *Bacillus subtilis*, *Listeria monocytogenes*, *Staphylococcus aureus*, *Streptococcus iniae*, *Streptococcus pyogenes*) and viruses (*Spring viraemia of carp virus*, *Infectious haematopoietic necrosis virus* (IHNV), *Infectious pancreatic necrosis virus*, *Viral hemorrhagic septicemia virus*) (Fig. 3) (Meeker and Trede, 2008; van der Sar et al., 2004).

Zebrafish has also been recognized as a powerful model for tumor biology and cancer drug development (Feitsma and Cuppen, 2008). It spontaneously develops almost any type of tumor with morphology similar to human tumors, most commonly in testis, gut, thyroid, liver, peripheral nerve, connective tissue, and ultimobranchial gland. Several experimental approaches to generate zebrafish cancer models have been applied, namely treatment with mutagens, forward genetic screening for mutants showing enhanced proliferation or genomic instability, the generation of knock-outs in known tumor suppressor genes, and the generation of transgenics to express (human) oncogenes (Feitsma and Cuppen, 2008). In addition, xeno-trans-

plantation models of mammalian or human cancer cells into zebrafish have been developed (Haldi et al., 2006; Nicoli et al., 2007).

Although the above-mentioned advantages make zebrafish a very practical model organism, it also has some disadvantages. One issue is that compared to mammals, zebrafish lacks certain organs like lungs and mammary glands, or shows differences in organ functions (e.g. in zebrafish the kidney marrow replaces the function of the mammalian bone marrow). In addition, there are difficulties in establishing inbred lines, cell lines and tissue cultures, there is a lack of antibody reagents, and the technology for targeted gene modification and knock-out generation is less advanced than for rodent models (Lieschke and Currie, 2007). A weakness of zebrafish as an infection model is that zebrafish is generally maintained between 26°C and 29°C, while optimal growth and expression of virulence factors of human pathogens typically occurs at 37°C. However, some models, for example the zebrafish – *Mycobacterium marinum* model for tuberculosis, use natural fish pathogens with lower permissive temperature ranges that are closely related to human pathogens (Lesley and Ramakrishnan, 2008; Meeker and Trede, 2008).

Outline of the thesis

In this study zebrafish models for infection and cancer are used to analyze the transcriptome complexity during these disease processes and to gain insight into regulatory functions of microRNAs by using array-based and sequencing-based large-scale genomics tools.

Chapter 1 is a general introduction reviewing the following topics: (i) the aim of transcriptomics and high-throughput approaches of genome-wide scale gene expression profiling, (ii) mechanisms of gene regulation, (iii) discovery of microRNAs, their biogenesis and mode of action, (iv) the role of microRNAs in diseases, (v) *in vitro* and animal modeling of diseases focusing on zebrafish as a novel model organism, and (vi) recent advances in the use of zebrafish as a model system for studying the immune system response to infectious diseases and cancer.

In **Chapter 2** the functional complexity of the adult zebrafish transcriptome in response to *Mycobacterium* infection is characterized using Solexa/Illumina's digital gene expression (DGE/Tag-Seq) system, a novel tag-based transcriptome sequencing method. DGE/Tag-seq data is compared with a previous multi-platform microarray analysis and qPCR results, and infection-induced new gene products are discovered and validated.

Chapter 3 describes deep sequencing of the transcriptome of zebrafish embryos in response to *Salmonella* infection. Data from tag-based sequencing (DGE/Tag-Seq) are compared with full RNA sequencing (RNA-Seq) data and previous microarray results. Additionally, the *Salmonella*-induced transcriptome alterations are compared with the *Mycobacterium* infection data from chapter 2.

Chapter 4 describes microRNA expression profiling analyses of infectious dis-

eases in zebrafish. Different bacterial infections are evaluated and common microRNA induction profiles are assessed. The dataset is used to compare human and zebrafish microRNA regulation, with a specific focus on the miR-146 family and their predicted target genes in the Toll-like receptor pathway. Furthermore, microRNA expression is compared with expression of predicted target genes.

In **Chapter 5** results of microRNA expression profiling using a zebrafish liver cancer model are reported. Differentially expressed microRNA in liver tumor tissue versus healthy liver tissue are categorized and, compared to microRNAs detected in human hepatocellular carcinoma and other types of cancer. Target predictions of microRNAs associated with liver cancer in zebrafish and human are made. In addition, the set of microRNAs from the liver cancer study is compared to the dataset from the infection study of chapter 4 to deduce common microRNA regulators of cancer and the immune response.

Finally, **Chapter 6** gives a summary and general discussion of the technologies used in this study and the transcriptome results in infection and cancer models.

References

Arena S., Benvenuti S. and Bardelli A. (2005) Genetic analysis of the kinome and phosphatome in cancer. *Cell Mol Life Sci* **62**, 2092-9.

Barbazuk W. B., Korf I., Kadavi C., Heyen J., Tate S., Wun E., Bedell J. A., McPherson J. D. and Johnson S. L. (2000) The syntenic relationship of the zebrafish and human genomes. *Genome Res* **10**, 1351-8.

Beiter T., Reich E., Williams R. W. and Simon P. (2009) Antisense transcription: a critical look in both directions. *Cell Mol Life Sci* **66**, 94-112.

Bellon M., Lepelletier Y., Hermine O. and Nicot C. (2009) Deregulation of microRNA involved in hematopoiesis and the immune response in HTLV-I adult T-cell leukemia. *Blood* **113**, 4914-7.

Bentley G., Higuchi R., Hoglund B., Goodridge D., Sayer D., Trachtenberg E. A. and Erlich H. A. (2009) High-resolution, high-throughput HLA genotyping by next-generation sequencing. *Tissue Antigens* **74**, 393-403.

Bertone P., Stolc V., Royce T. E., Rozowsky J. S., Urban A. E., Zhu X., Rinn J. L., Tongprasit W., Samanta M., Weissman S., Gerstein M. and Snyder M. (2004) Global identification of human transcribed sequences with genome tiling arrays. *Science* **306**, 2242-6.

Boguski M. S., Tolstoshev C. M. and Bassett D. E., Jr. (1994) Gene discovery in dbEST. *Science* **265**, 1993-4.

Brennecke J., Stark A., Russell R. B. and Cohen S. M. (2005) Principles of microRNA-target recognition. *PLoS Biol* **3**, e85.

Brenner S., Johnson M., Bridgham J., Golda G., Lloyd D. H., Johnson D., Luo S., McCurdy S., Foy M., Ewan M., Roth R., George D., Eletr S., Albrecht G., Vermaas E., Williams S. R., Moon K., Burcham T., Pallas M., DuBridge R. B., Kirchner J., Fearon K., Mao J. and Corcoran K. (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat Biotechnol* **18**, 630-4.

Brentani R. R., Carraro D. M., Verjovski-Almeida S., Reis E. M., Neves E. J., de Souza S. J., Carvalho A. F., Brentani H. and Reis L. F. (2005) Gene expression arrays in cancer research: methods and applications. *Crit Rev Oncol Hematol* **54**, 95-105.

Calin G.A. and Croce C.M. (2006) Genomics of chronic lymphocytic leukemia microRNAs as new players with clinical significance. *Semin Oncol* **33**,167-73

Cannell I. G., Kong Y. W. and Bushell M. (2008) How do microRNAs regulate gene expression? *Biochem Soc Trans* **36**, 1224-31.

Carissimi C., Fulci V. and Macino G. (2009) MicroRNAs: novel regulators of immunity. *Autoimmun Rev* **8**, 520-4.

Carninci P., Kasukawa T., Katayama S., Gough J., Frith M. C., Maeda N., Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K., Bajic V. B., Brenner S. E., Batalov S., Forrest A. R., Zavolan M., Davis M. J., Wilming L. G., Aidinis V., Allen J. E., Ambesi-Impiombato A., Apweiler R., Aturaliya R. N., Bailey T. L., Bansal M., Baxter L., Beisel K. W., Bersano T., Bono H., Chalk A. M., Chiu K. P., Choudhary V., Christoffels A., Clutterbuck D. R., Crowe M. L., Dalla E., Dalrymple B. P., de Bono B., Della Gatta G., di Bernardo D., Down T., Engstrom P., Fagiolini M., Faulkner G., Fletcher C. F., Fukushima T., Furuno M., Futaki S., Gariboldi M., Georgii-Hemming P., Gingeras T. R., Gojobori T., Green R. E., Gustincich S., Harbers M., Hayashi Y., Hensch T. K., Hirokawa N., Hill D., Huminiecki L., Iacono M., Ikeo K., Iwama A., Ishikawa T., Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H., Kitano H., Kollias G., Krishnan S. P., Kruger A., Kummerfeld S. K., Kurochkin I. V., Lareau L. F., Lazarevic D., Lipovich L., Liu J., Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L., Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K., Mottagui-Tabar S., Mulder N., Nakano N., Nakauchi H., Ng P., Nilsson R., Nishiguchi S., Nishikawa S., et al. (2005) The transcriptional landscape of the mammalian genome. *Science* **309**, 1559-63.

Chu C. Y. and Rana T. M. (2007) Small RNAs: regulators and guardians of the genome. *J Cell Physiol* **213**, 412-9.

Clark T. A., Sugnet C. W. and Ares M., Jr. (2002) Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. *Science* **296**, 907-10.

Cloonan N., Forrest A. R., Kolle G., Gardiner B. B., Faulkner G. J., Brown M. K., Taylor D. F., Steptoe A. L., Wani S., Bethel G., Robertson A. J., Perkins A. C., Bruce S. J., Lee C. C., Ranade S. S., Peckham H. E., Manning J. M., McKernan K. J. and Grimmond S. M. (2008) Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat Methods* **5**, 613-9.

Cloonan N. and Grimmond S. M. (2008) Transcriptome content and dynamics at single-nucleotide resolution. *Genome Biol* **9**, 234.

Costinean S., Zanesi N., Pekarsky Y., Tili E., Volinia S., Heerema N. and Croce C. M. (2006) Pre-B cell proliferation and lymphoblastic leukemia/high-grade lymphoma in E(mu)-miR155 transgenic mice. *Proc Natl Acad Sci U S A* **103**, 7024-9.

Didiano D. and Hobert O. (2006) Perfect seed pairing is not a generally reliable predictor for miRNA-target interactions. *Nat Struct Mol Biol* **13**, 849-51.

Diks S. H., Kok K., O'Toole T., Hommes D. W., van Dijken P., Joore J. and Peppelenbosch M. P. (2004) Kinome profiling for studying lipopolysaccharide signal transduction in human peripheral blood mononuclear cells. *J Biol Chem* **279**, 49206-13.

Douglas C. J. and Ehlting J. (2005) Arabidopsis thaliana full genome longmer microarrays: a powerful gene discov-

ery tool for agriculture and forestry. *Transgenic Res* **14**, 551-61.
Ebralidze A. K., Guibal F. C., Steidl U., Zhang P., Lee S., Bartholdy B., Jorda M. A., Petkova V., Rosenbauer F., Huang G., Dayaram T., Klupp J., O'Brien K. B., Will B., Hoogenkamp M., Borden K. L., Bonifer C. and Tenen D. G. (2008) PU.1 expression is modulated by the balance of functional sense and antisense RNAs regulated by a shared cis-regulatory element. *Genes Dev* **22**, 2085-92.
Engstrand L. (2009) How will next-generation sequencing contribute to the knowledge concerning Helicobacter pylori? *Clin Microbiol Infect* **15**, 823-8.
Esquela-Kerscher A. and Slack F. J. (2006) Oncomirs - microRNAs with a role in cancer. *Nat Rev Cancer* **6**, 259-69.
Ewbank J. J. (2002) Tackling both sides of the host-pathogen equation with Caenorhabditis elegans. *Microbes Infect* **4**, 247-56.
Feitsma H. and Cuppen E. (2008) Zebrafish as a cancer model. *Mol Cancer Res* **6**, 685-94.
Filipowicz W., Bhattacharyya S. N. and Sonenberg N. (2008) Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat Rev Genet* **9**, 102-14.
Friedlander A. M. and Little S. F. (2009) Advances in the development of next-generation anthrax vaccines. *Vaccine* **27 Suppl 4**, D28-32.
Gabriel C., Danzer M., Hackl C., Kopal G., Hufnagl P., Hofer K., Polin H., Stabentheiner S. and Proll J. (2009) Rapid high-throughput human leukocyte antigen typing by massively parallel pyrosequencing for high-resolution allele identification. *Hum Immunol* **70**, 960-4.
Gardina P. J., Clark T. A., Shimada B., Staples M. K., Yang Q., Veitch J., Schweitzer A., Awad T., Sugnet C., Dee S., Davies C., Williams A. and Turpaz Y. (2006) Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array. *BMC Genomics* **7**, 325.
Gargiulo G. and Minucci S. (2009) Epigenomic profiling of cancer cells. *Int J Biochem Cell Biol* **41**, 127-35.
Garofalo M., Condorelli G. and Croce C. M. (2008) MicroRNAs in diseases and drug response. *Curr Opin Pharmacol* **8**, 661-7.
Gerhard D. S., Wagner L., Feingold E. A., Shenmen C. M., Grouse L. H., Schuler G., Klein S. L., Old S., Rasooly R., Good P., Guyer M., Peck A. M., Derge J. G., Lipman D., Collins F. S., Jang W., Sherry S., Feolo M., Misquitta L., Lee E., Rotmistrovsky K., Greenhut S. F., Schaefer C. F., Buetow K., Bonner T. I., Haussler D., Kent J., Kiekhaus M., Furey T., Brent M., Prange C., Schreiber K., Shapiro N., Bhat N. K., Hopkins R. F., Hsie F., Driscoll T., Soares M. B., Casavant T. L., Scheetz T. E., Brown-stein M. J., Usdin T. B., Toshiyuki S., Carninci P., Piao Y., Dudekula D. B., Ko M. S., Kawakami K., Suzuki Y., Sugano S., Gruber C. E., Smith M. R., Simmons B., Moore T., Waterman R., Johnson S. L., Ruan Y., Wei C. L., Mathavan S., Gunaratne P. H., Wu J., Garcia A. M., Hulyk S. W., Fuh E., Yuan Y., Sneed A., Kowis C., Hodgson A., Muzny D. M., McPherson J., Gibbs R. A., Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A., Whiting M., Madari A., Young A. C., Wetherby K. D., Granite S. J., Kwong P. N., Brinkley C. P., Pearson R. L., Bouffard G. G., Blakesly R. W., Green E. D., Dickson M. C., Rodriguez A. C., Grimwood J., Schmutz J., Myers R. M., Butterfield Y. S., Griffith M., Griffith O. L., Krzywinski M. I., Liao N., Morin R., Palmquist D., et al. (2004) The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC). *Genome Res* **14**, 2121-7.
Griffiths-Jones S., Moxon S., Marshall M., Khanna A., Eddy S. R. and Bateman A. (2005) Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res* **33**, D121-4.
Hackenberg M., Sturm M., Langenberger D., Falcon-Perez J. M. and Aransay A. M. (2009) miRanalyzer: a microRNA detection and analysis tool for next-generation sequencing experiments. *Nucleic Acids Res* **37**, W68-76.
Haldi M., Ton C., Seng W. L. and McGrath P. (2006) Human melanoma cells transplanted into zebrafish proliferate, migrate, produce melanin, form masses and stimulate angiogenesis in zebrafish. *Angiogenesis* **9**, 139-51.
Harbers M. and Carninci P. (2005) Tag-based approaches for transcriptome research and genome annotation. *Nat Methods* **2**, 495-502.
Hegedus Z., Zakrzewska A., Agoston V. C., Ordas A., Racz P., Mink M., Spaink H. P. and Meijer A. H. (2009) Deep sequencing of the zebrafish transcriptome response to mycobacterium infection. *Mol Immunol* **46**, 2918-30.
Hestvik A. L., Hmama Z. and Av-Gay Y. (2003) Kinome analysis of host response to mycobacterial infection: a novel technique in proteomics. *Infect Immun* **71**, 5514-22.
Hoffmann J. A. (2003) The immune response of Drosophila. *Nature* **426**, 33-8.
Horan M. P. (2009) Application of serial analysis of gene expression to the study of human genetic disease. *Hum Genet*.
Hurd P. J. and Nelson C. J. (2009) Advantages of next-generation sequencing versus the microarray in epigenetic research. *Brief Funct Genomic Proteomic* **8**, 174-83.
Irizarry R. A., Warren D., Spencer F., Kim I. F., Biswal S., Frank B. C., Gabrielson E., Garcia J. G., Geoghegan J., Germino G., Griffin C., Hilmer S. C., Hoffman E., Jedlicka A. E., Kawasaki E., Martinez-Murillo F., Morsberger L., Lee H., Petersen D., Quackenbush J., Scott A., Wilson M., Yang Y., Ye S. Q. and Yu W. (2005) Multiple-laboratory comparison of microarray platforms. *Nat Methods* **2**, 345-50.
Jackson R. J. and Standart N. (2007) How do microRNAs regulate gene expression? *Sci STKE* **2007**, re1.
Kapsimali M., Kloosterman W. P., de Bruijn E., Rosa F., Plasterk R. H. and Wilson S. W. (2007) MicroRNAs show a wide diversity of expression profiles in the developing and mature central nervous system. *Genome Biol* **8**, R173.

Kari G., Rodeck U. and Dicker A. P. (2007) Zebrafish: an emerging model system for human disease and drug discovery. *Clin Pharmacol Ther* **82**, 70-80.
Kasahara M., Suzuki T. and Pasquier L. D. (2004) On the origins of the adaptive immune system: novel insights from invertebrates and cold-blooded vertebrates. *Trends Immunol* **25**, 105-11.
Katayama S., Tomaru Y., Kasukawa T., Waki K., Nakanishi M., Nakamura M., Nishida H., Yap C. C., Suzuki M., Kawai J., Suzuki H., Carninci P., Hayashizaki Y., Wells C., Frith M., Ravasi T., Pang K. C., Hallinan J., Mattick J., Hume D. A., Lipovich L., Batalov S., Engstrom P. G., Mizuno Y., Faghihi M. A., Sandelin A., Chalk A. M., Mottagui-Tabar S., Liang Z., Lenhard B. and Wahlestedt C. (2005) Antisense transcription in the mammalian transcriptome. *Science* **309**, 1564-6.
Kim J. B., Porreca G. J., Song L., Greenway S. C., Gorham J. M., Church G. M., Seidman C. E. and Seidman J. G. (2007) Polony multiplex analysis of gene expression (PMAGE) in mouse hypertrophic cardiomyopathy. *Science* **316**, 1481-4.
Kloosterman W. P., Steiner F. A., Berezikov E., de Bruijn E., van de Belt J., Verheul M., Cuppen E. and Plasterk R. H. (2006) Cloning and expression of new microRNAs from zebrafish. *Nucleic Acids Res* **34**, 2558-69.
Kloosterman W. P., Wienholds E., Ketting R. F. and Plasterk R. H. (2004) Substrate requirements for let-7 function in the developing zebrafish embryo. *Nucleic Acids Res* **32**, 6284-91.
Lagos-Quintana M., Rauhut R., Yalcin A., Meyer J., Lendeckel W. and Tuschl T. (2002) Identification of tissue-specific microRNAs from mouse. *Curr Biol* **12**, 735-9.
Lesley R. and Ramakrishnan L. (2008) Insights into early mycobacterial pathogenesis from the zebrafish. *Curr Opin Microbiol* **11**, 277-83.
Lieschke G. J. and Currie P. D. (2007) Animal models of human disease: zebrafish swim into view. *Nat Rev Genet* **8**, 353-67.
Lim L. P., Glasner M. E., Yekta S., Burge C. B. and Bartel D. P. (2003) Vertebrate microRNA genes. *Science* **299**, 1540.
Lister R., O'Malley R. C., Tonti-Filippini J., Gregory B. D., Berry C. C., Millar A. H. and Ecker J. R. (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* **133**, 523-36.
Lu J., Getz G., Miska E. A., Alvarez-Saavedra E., Lamb J., Peck D., Sweet-Cordero A., Ebert B. L., Mak R. H., Ferrando A. A., Downing J. R., Jacks T., Horvitz H. R. and Golub T. R. (2005) MicroRNA expression profiles classify human cancers. *Nature* **435**, 834-8.
Lu L. F. and Liston A. (2009) MicroRNA in the immune system, microRNA as an immune system. *Immunology* **127**, 291-8.
Mardis E. R. and Wilson R. K. (2009) Cancer genome sequencing: a review. *Hum Mol Genet* **18**, R163-8.
Matsumura H., Ito A., Saitoh H., Winter P., Kahl G., Reuter M., Kruger D. H. and Terauchi R. (2005) SuperSAGE. *Cell Microbiol* **7**, 11-8.
Medina P. P. and Slack F. J. (2008) microRNAs and cancer: an overview. *Cell Cycle* **7**, 2485-92.
Meeker N. D. and Trede N. S. (2008) Immunology and zebrafish: spawning new models of human disease. *Dev Comp Immunol* **32**, 745-57.
Mogensen T. H. (2009) Pathogen recognition and inflammatory signaling in innate immune defenses. *Clin Microbiol Rev* **22**, 240-73, Table of Contents.
Morin R., Bainbridge M., Fejes A., Hirst M., Krzywinski M., Pugh T., McDonald H., Varhol R., Jones S. and Marra M. (2008) Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. *Biotechniques* **45**, 81-94.
Morozova O. and Marra M. A. (2008) Applications of next-generation sequencing technologies in functional genomics. *Genomics* **92**, 255-64.
Morrissy A. S., Morin R. D., Delaney A., Zeng T., McDonald H., Jones S., Zhao Y., Hirst M. and Marra M. A. (2009) Next-generation tag sequencing for cancer gene expression profiling. *Genome Res* **19**, 1825-35.
Mortazavi A., Williams B. A., McCue K., Schaeffer L. and Wold B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5**, 621-8.
Nagalakshmi U., Wang Z., Waern K., Shou C., Raha D., Gerstein M. and Snyder M. (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**, 1344-9.
Nicoli S., Ribatti D., Cotelli F. and Presta M. (2007) Mammalian tumor xenografts induce neovascularization in zebrafish embryos. *Cancer Res* **67**, 2927-31.
Nielsen K. L., Hogh A. L. and Emmersen J. (2006) DeepSAGE--digital transcriptomics with high sensitivity, simple experimental protocol and multiplexing of samples. *Nucleic Acids Res* **34**, e133.
Nilsen T. W. (2007) Mechanisms of microRNA-mediated gene regulation in animal cells. *Trends Genet* **23**, 243-9.
Okoniewski M. J. and Miller C. J. (2006) Hybridization interactions between probesets in short oligo microarrays lead to spurious correlations. *BMC Bioinformatics* **7**, 276.
Pan Q., Shai O., Lee L. J., Frey B. J. and Blencowe B. J. (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* **40**, 1413-5.
Parikh K., Peppelenbosch M. P. and Ritsema T. (2009) Kinome profiling using peptide arrays in eukaryotic cells. *Methods Mol Biol* **527**, 269-80, x.
Pasquinelli A. E., Reinhart B. J., Slack F., Martindale M. Q., Kuroda M. I., Maller B., Hayward D. C., Ball E. E.,

Degnan B., Muller P., Spring J., Srinivasan A., Fishman M., Finnerty J., Corbo J., Levine M., Leahy P., Davidson E. and Ruvkun G. (2000) Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* **408**, 86-9.

Pedotti P., t Hoen P. A., Vreugdenhil E., Schenk G. J., Vossen R. H., Ariyurek Y., de Hollander M., Kuiper R., van Ommen G. J., den Dunnen J. T., Boer J. M. and de Menezes R. X. (2008) Can subtle changes in gene expression be consistently detected with different microarray platforms? *BMC Genomics* **9**, 124.

Petrocca F., Visone R., Onelli M. R., Shah M. H., Nicoloso M. S., de Martino I., Iliopoulos D., Pilozzi E., Liu C. G., Negrini M., Cavazzini L., Volinia S., Alder H., Ruco L. P., Baldassarre G., Croce C. M. and Vecchione A. (2008) E2F1-regulated microRNAs impair TGFbeta-dependent cell-cycle arrest and apoptosis in gastric cancer. *Cancer Cell* **13**, 272-86.

Qi W., Kaser M., Roltgen K., Yeboah-Manu D. and Pluschke G. (2009) Genomic diversity and evolution of Mycobacterium ulcerans revealed by next-generation sequencing. *PLoS Pathog* **5**, e1000580.

Rodriguez A., Griffiths-Jones S., Ashurst J. L. and Bradley A. (2004) Identification of mammalian microRNA host genes and transcription units. *Genome Res* **14**, 1902-10.

Rosenkranz R., Borodina T., Lehrach H. and Himmelbauer H. (2008) Characterizing the mouse ES cell transcriptome with Illumina sequencing. *Genomics* **92**, 187-94.

Royce T. E., Rozowsky J. S. and Gerstein M. B. (2007) Toward a universal microarray: prediction of gene expression through nearest-neighbor probe sequence identification. *Nucleic Acids Res* **35**, e99.

Saetrom P., Heale B. S., Snove O., Jr., Aagaard L., Alluin J. and Rossi J. J. (2007) Distance constraints between microRNA target sites dictate efficacy and cooperativity. *Nucleic Acids Res* **35**, 2333-42.

Schena M., Shalon D., Davis R. W. and Brown P. O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467-70.

Scherf U., Ross D. T., Waltham M., Smith L. H., Lee J. K., Tanabe L., Kohn K. W., Reinhold W. C., Myers T. G., Andrews D. T., Scudiero D. A., Eisen M. B., Sausville E. A., Pommier Y., Botstein D., Brown P. O. and Weinstein J. N. (2000) A gene expression database for the molecular pharmacology of cancer. *Nat Genet* **24**, 236-44.

Schmidt W. M., Spiel A. O., Jilma B., Wolzt M. and Muller M. (2009) In vivo profile of the human leukocyte microRNA response to endotoxemia. *Biochem Biophys Res Commun* **380**, 437-41.

Sempere L. F., Freemantle S., Pitha-Rowe I., Moss E., Dmitrovsky E. and Ambros V. (2004) Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation. *Genome Biol* **5**, R13.

Shendure J. (2008) The beginning of the end for microarrays? *Nat Methods* **5**, 585-7.

Sonkoly E. and Pivarcsi A. (2009) Advances in microRNAs: implications for immunity and inflammatory diseases. *J Cell Mol Med* **13**, 24-38.

Sonkoly E., Stahle M. and Pivarcsi A. (2008) MicroRNAs and immunity: novel players in the regulation of normal immune function and inflammation. *Semin Cancer Biol* **18**, 131-40.

Sultan M., Schulz M. H., Richard H., Magen A., Klingenhoff A., Scherf M., Seifert M., Borodina T., Soldatov A., Parkhomchuk D., Schmidt D., O'Keeffe S., Haas S., Vingron M., Lehrach H. and Yaspo M. L. (2008) A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* **321**, 956-60.

t Hoen P. A., Ariyurek Y., Thygesen H. H., Vreugdenhil E., Vossen R. H., de Menezes R. X., Boer J. M., van Ommen G. J. and den Dunnen J. T. (2008) Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms. *Nucleic Acids Res* **36**, e141.

Tang G., Reinhart B. J., Bartel D. P. and Zamore P. D. (2003) A biochemical framework for RNA silencing in plants. *Genes Dev* **17**, 49-63.

Traver D., Herbomel P., Patton E. E., Murphey R. D., Yoder J. A., Litman G. W., Catic A., Amemiya C. T., Zon L. I. and Trede N. S. (2003) The zebrafish as a model organism to study development of the immune system. *Adv Immunol* **81**, 253-330.

Tyagi S. (2000) Taking a census of mRNA populations with microbeads. *Nat Biotechnol* **18**, 597-8.

van der Sar A. M., Appelmelk B. J., Vandenbroucke-Grauls C. M. and Bitter W. (2004) A star with stripes: zebrafish as an infection model. *Trends Microbiol* **12**, 451-7.

Velculescu V. E., Zhang L., Vogelstein B. and Kinzler K. W. (1995) Serial analysis of gene expression. *Science* **270**, 484-7.

Wagner E. G. and Simons R. W. (1994) Antisense RNA control in bacteria, phages, and plasmids. *Annu Rev Microbiol* **48**, 713-42.

Wahl M. B., Heinzmann U. and Imai K. (2005) LongSAGE analysis significantly improves genome annotation: identifications of novel genes and alternative transcripts in the mouse. *Bioinformatics* **21**, 1393-400.

Wang Z., Gerstein M. and Snyder M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**, 57-63.

White K. P. (2001) Functional genomics and the study of development, variation and evolution. *Nat Rev Genet* **2**, 528-37.

Wilhelm B. T. and Landry J. R. (2009) RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. *Methods* **48**, 249-57.

Williams A. E., Perry M. M., Moschos S. A., Larner-Svensson H. M. and Lindsay M. A. (2008) Role of miRNA-146a in the regulation of the innate immune response and cancer. *Biochem Soc Trans* **36**, 1211-5.

Wyman S. K., Parkin R. K., Mitchell P. S., Fritz B. R., O'Briant K., Godwin A. K., Urban N., Drescher C. W., Knudsen B. S. and Tewari M. (2009) Repertoire of microRNAs in epithelial ovarian cancer as determined by next generation sequencing of small RNA cDNA libraries. *PLoS One* **4**, e5311.

Yelin R., Dahary D., Sorek R., Levanon E. Y., Goldstein O., Shoshan A., Diber A., Biton S., Tamir Y., Khosravi R., Nemzer S., Pinner E., Walach S., Bernstein J., Savitsky K. and Rotman G. (2003) Widespread occurrence of antisense transcription in the human genome. *Nat Biotechnol* **21**, 379-86.

Chapter 2

Deep sequencing of the zebrafish transcriptome response to mycobacterium infection

This chapter is based on:
Hegedus Z, Zakrzewska A, Agoston VC, Ordas A, Racz P, Mink M, Spaink HP, Meijer AH. *Deep sequencing of the zebrafish transcriptome response to mycobacterium infection*. Mol Immunol. 2009 Sep;46(15):2918-30.

Abstract

Novel high-throughput deep sequencing technology has dramatically changed the way that the functional complexity of transcriptomes can be studied. Here we report on the first use of this technology to gain insight into the wide range of transcriptional responses that are associated with an infectious disease process. Using Solexa/Illumina's digital gene expression (DGE) system, a tag-based transcriptome sequencing method, we investigated mycobacterium-induced transcriptome changes in a model vertebrate species, the zebrafish. We obtained a sequencing depth of over 5 million tags per sample with strong correlation between replicates. Tag mapping indicated that healthy and infected adult zebrafish express over 70% of all genes represented in transcript databases. Comparison of the data with a previous multi-platform microarray analysis showed that both types of technologies identified regulation of similar functional groups of genes. However, the unbiased nature of DGE analysis provided insights that microarray analysis could not have achieved. In particular, we show that DGE data sets are instrumental for verification of predicted gene models and allowed us to detect mycobacterium-regulated switching between different transcript isoforms. Moreover, genomic mapping of infection-induced DGE tags revealed novel transcript forms for which any previous EST-based evidence of expression was lacking. In conclusion, our deep sequencing analysis revealed in depth the high degree of transcriptional complexity of the host response to mycobacterial infection and resulted in the discovery and validation of new gene products with induced expression in infected individuals.

Introduction

Cellular identity and function is determined by the transcriptome: the complete repertoire of expressed RNA transcripts. Development and disease processes in multicellular organisms are governed by complex variations in transcriptional activities. Deciphering the functional complexity of transcriptomes is extremely challenging, especially since recent studies indicate that much more of the genome is transcribed than previously thought and that the majority of genes is transcribed in a bidirectional manner (Carninci et al., 2005; Katayama et al., 2005; Yelin et al., 2003). Recent advances in the development of ultra high-throughput deep sequencing technologies are making a huge impact on genomic research. These next generation sequencing systems, such as the Solexa/Illumina Genome Analyzer and the ABI/SOLiD Gene Sequencer, can sequence in parallel millions of DNA molecules derived directly from mRNA, without the need to use bacterial clones (Cloonan and Grimmond, 2008; Morozova and Marra, 2008; Wang et al., 2009). The direct sequencing yields libraries of short sequences (25–50 nucleotides), referred to as RNA-Seq data or Digital Gene Expression (DGE) data. Sequencing-based methods generate absolute rather than relative gene expression measurements and avoid many of the inherent limitations

of microarray analysis (Irizarry et al., 2005; Pedotti et al., 2008; t Hoen et al., 2008; Wilhelm and Landry, 2009), which has been the most commonly used technology for transcriptome profiling over the last decade.

The first results of transcriptome profiling using next generation sequencing technology have recently been published (Cloonan et al., 2008; Lister et al., 2008; Mortazavi et al., 2008; Nagalakshmi et al., 2008; Sultan et al., 2008; Wilhelm et al., 2008). These RNA-Seq studies were based on different procedures starting either with mRNA fragmentation or with cDNA synthesis that is followed by fragmentation. Dependent on the protocol used, information on transcript directionality, which is useful for annotation and detection of antisense transcription, was retained or not (Shendure, 2008). RNA-Seq studies of yeasts detected transcription of 92–99% of all known genes and demonstrated that most of the genome sequence is transcribed (75% in *Saccharomyces cerevisiae* and >90% in *Saccharomyces pombe*) (Nagalakshmi et al., 2008; Wilhelm et al., 2008). Evidence of transcript heterogeneity and of novel transcribed regions, including non-coding and antisense transcripts, was obtained. Furthermore, a genome-wide regulation of splicing was revealed (Wilhelm et al., 2008). Bioinformatical analysis of RNA-Seq data needs to deal with mapping of short reads to the genome, issues of multi-mapping, and mapping of splice-crossing reads. The analysis becomes even more challenging when dealing with the more complex genomes of vertebrate species. Nevertheless, the first applications of RNA-Seq methodology to mouse and human have now already provided rich information for new or revised gene models. By indicating the presence of additional promoters, alternative exons, alternative 3'UTRs, non-coding transcripts and bidirectional transcripts, these studies have made it clear that genome annotation is far from complete (Cloonan et al., 2008; Morin et al., 2008; Mortazavi et al., 2008; Pan et al., 2008; Rosenkranz et al., 2008; Sultan et al., 2008;Wang et al., 2008).

An alternative to RNA-Seq is the use of tag-based transcriptome sequencing methods, such as serial analysis of gene expression (SAGE), which generates short signature sequences (tags) for the 3'-end regions of mRNA transcripts (Harbers and Carninci, 2005; Velculescu et al., 1995). SAGE-derived technologies include MPSS (massive parallel signature sequencing) and PMAGE (polony multiplex analysis of gene expression) that rely on amplification of 3' tag sequences on microbeads (Brenner et al., 2000; Kim et al., 2007). Other useful tag-based sequencing technologies include the 5'-SAGE or CAGE (cap analysis of gene expression) methods that determine the 5'-ends of mRNAs by oligo-capping or cap-trapping (Harbers and Carninci, 2005; Hashimoto et al., 2009). The classical LongSAGE technique is based on the cleavage of cDNA with two restriction enzymes, an anchoring enzyme (commonly NlaIII, which cuts at CATG sites) and a tagging enzyme (MmeI) cutting 17 bp downstream of the anchoring enzyme's recognition site. Although SAGE, LongSAGE and similar SuperSAGE methods have greatly contributed to transcriptome analysis, their application has previously been limited by laborious ditag formation and concatemer cloning procedures and by the costs and throughput level of sequencing

steps. However, SAGE-derived methods, which are particularly suitable for quantitative expression analysis, are back into view with next generation sequencing technology that entirely eliminates the need for tag cloning and provides a much greater sequencing depth.

Here, we used the second Illumina Genome Analyzer platform (GA II) to perform a SAGE-derived Digital Gene Expression (DGE) analysis of the zebrafish transcriptome response to mycobacterium infection. The zebrafish-mycobacterium model recapitulates hallmark features of human tuberculosis (Lesley and Ramakrishnan, 2008; Swaim et al., 2006). We have previously performed a multiplatform microarray study showing that mycobacterium-infected zebrafish express many homologs of human immune response genes and genes that have previously been implicated in the response to mycobacterial infection (Meijer et al., 2005). We also observed induction of genes with previously unknown relationship to the immune response, indicating that the use of the zebrafish mycobacterium model can assist functional annotation of genes and provide new leads in the investigation of mycobacterial pathogenesis. The previous study was limited by contents of commercially available microarray designs, whereas the DGE analysis reported here provides wide unbiased coverage of the entire transcriptome. Comparison of DGE and microarray data revealed a substantial degree of overlap in differentially expressed transcripts as well as technology-dependent differences, indicating the value of using two complementary transcriptome analysis methods. Furthermore, by mapping our DGE tag data onto transcript databases and onto the genomic sequence we could show that many alternative transcript forms and novel transcripts are disease-specifically regulated, information that could not have been obtained by microarray analysis.

Results

Basic quantitative parameters and reproducibility of DGE library sequencing

In previous microarray studies we have show that mycobacterium infection of zebrafish results in a highly reproducible host response at the transcriptome level (Meijer et al., 2005; van der Sar et al., 2009). Here we have pooled biological replicates from these previous studies to make representative samples for deep sequencing analysis. We sequenced two DGE libraries (technical duplicates) of a pool of four control fish (cont-a, cont-b) and two DGE libraries of a pool of four mycobacterium-infected fish (inf-a, inf-b). The total count number of tags per library ranged from 5.3 to 8.4 million and the number of tag entities with unique nucleotide sequences ranged from 0.8 to 1.1 million (Supplementary Table 2). As shown in Fig. 1, the distribution of tag entities and total tag counts over different tag abundance categories showed very similar tendencies for all four DGE libraries. The lowest abundant tags that were still consistently detected in all four libraries occurred at a frequency of below 1 count per million. To further investigate the reproducibility of DGE library

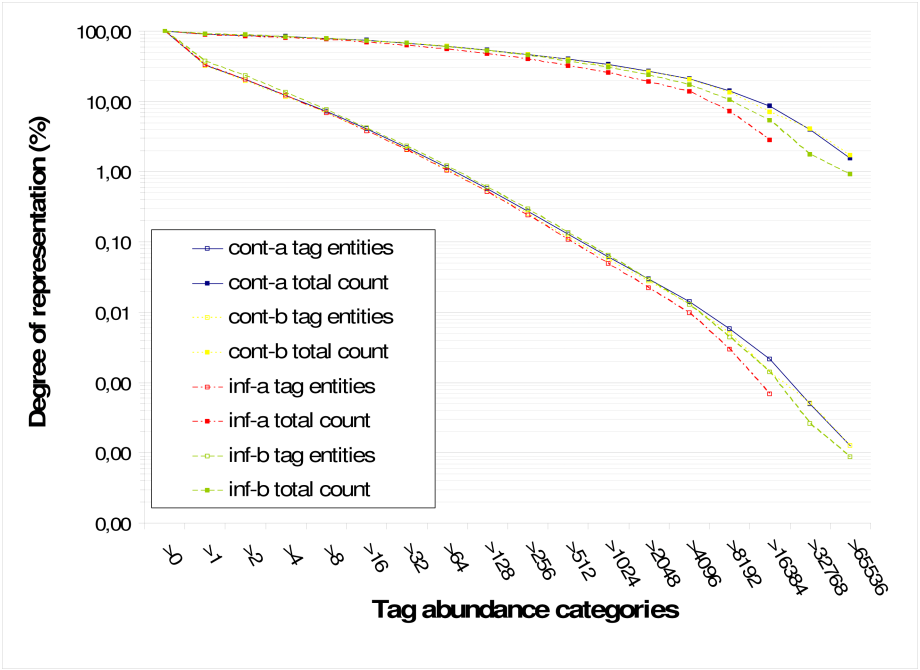


Figure 1. Distribution of tag entities and total tag counts over different tag abundance categories. Categories of tag abundance were assigned by setting the lower limit of the count number that includes a tag as a category member. The percentages of total tag counts (filled squares) and number of different tag entities (open squares) per category are plotted on a logarithmic scale.

sequencing we performed correlation analyses of all four possible sample combinations (cont-a vs. cont-b, inf-a vs. inf-b, cont-a vs. inf-a, and cont-b vs. inf-b) (Fig. 2). Pearson correlations for the parallel libraries (cont-a vs. cont-b and inf-a vs. inf-b) were 0.99 and 0.95, indicating the high technical reproducibility of the DGE method. In contrast, Pearson correlations for control vs. infected library pairs were much lower (0.77 and 0.81), consistent with a large effect of mycobacterium infection on the host transcriptome as previously observed in microarray experiments (Meijer et al., 2005; van der Sar et al., 2009).

Efficiency of tag mapping to transcript databases

Mapping the tags to known transcripts is the most efficient way to reveal the molecular events behind DGE profiles. In our study the tag sequences of the four DGE libraries were mapped to the zebrafish transcript datasets of the UniGene, RefSeq and Ensembl databases and found to match with over 70% of all sequence entries in these databases (Table 1). The mapping efficiency increased significantly as the count number of tags increased. Specifically, a 50% mapping efficiency was observed for tags with around 10 copies and the mapping efficiency increased to about 80%

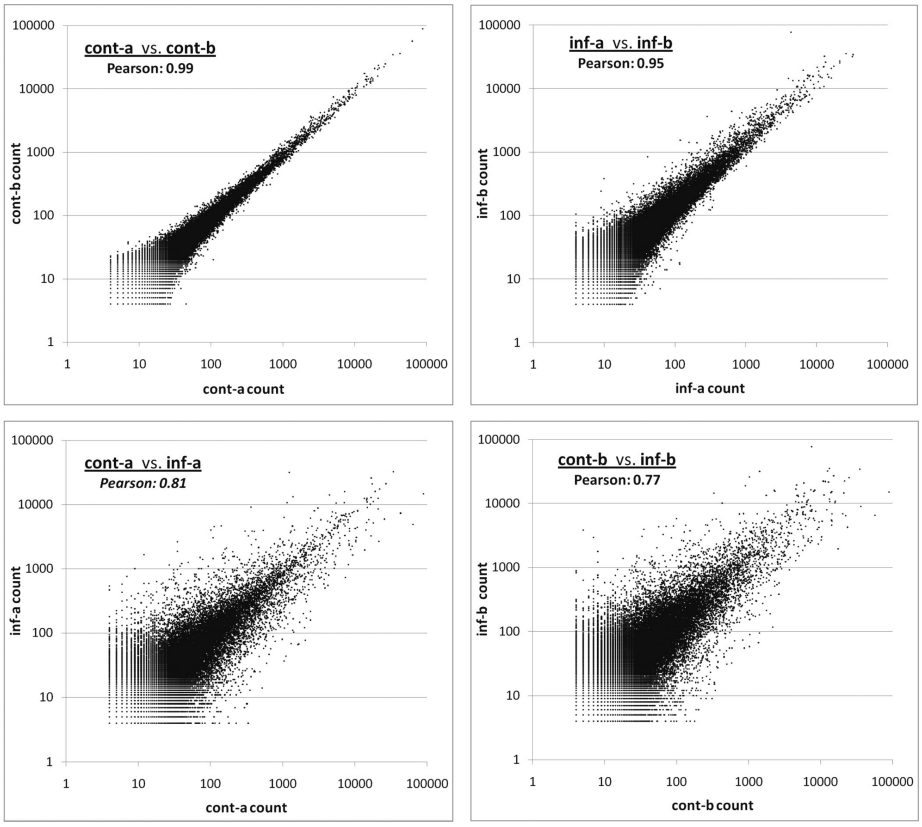


Figure 2. Correlation analysis of DGE libraries. Correlation charts are shown of the tag entity counts of the four possible combinations of DGE libraries from control fish (cont-a and cont-b) and from mycobacterium-infected fish (inf-a and inf-b). Dots in the charts indicate individual tag entities. Pearson correlation coefficients are shown in the upper left corner of each plot.

for tags with an abundance of over 200 copies (Fig. 3). This means that the most abundant tags correspond to the most highly expressed transcripts, which in turn are most likely to be found in the existing zebrafish transcript collections. Tags mapping to a unique sequence position form the most important subset of the DGE libraries as they can unambiguously identify a transcript. Up to 54% of the sequence records in the different transcript databases could be unequivocally identified by unique tag mapping (Supplementary Table 3). Examples of the most abundantly expressed genes in all libraries that could be unequivocally identified by tag mapping comprised more than ten different ribosomal subunit genes, translation initiation and elongation factor genes (*eif5a* and *ef1a*), creatine kinase genes (muscles a and b) and several cytoskeletal protein genes, including those encoding alpha and beta actin, keratin and skeletal muscle myosin proteins.

Table 1. Representation of transcript databank entries in the DGE libraries.

Transcript databank	Total number of databank entries	cont-a + cont-b libraries [a]	inf-a + inf-b libraries [a]
UniGene	56,561	73%	77%
RefSeq	37,396	69%	73%
Ensembl	31,841	74%	86%

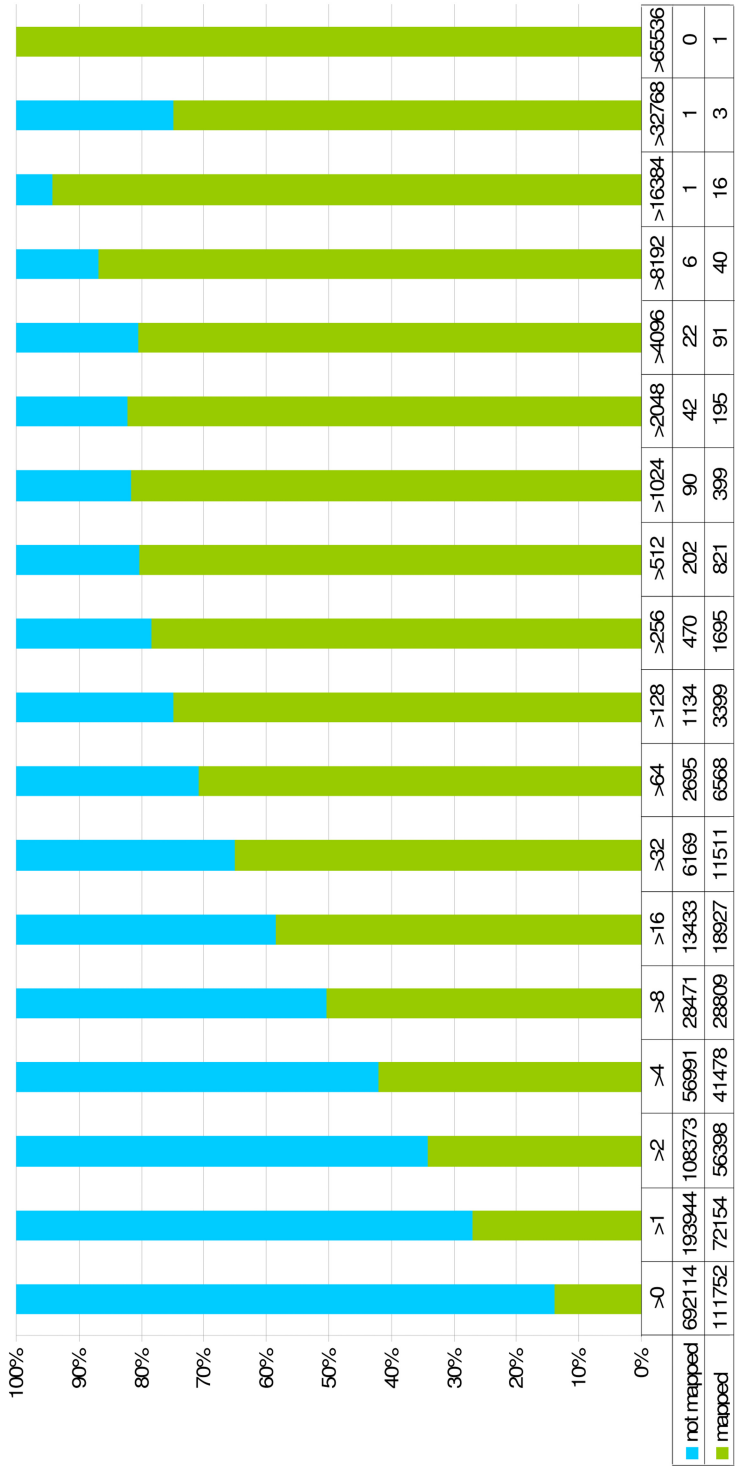
[a] Replicate DGE libraries (a and b) were merged *in silico* and indicated is the percentage of entries from the UniGene, RefSeq and Ensembl transcript databanks that were hit by mapping of DGE tags from the merged libraries.

Changes in tag profile induced by mycobacterium infection

In order to identify the tags showing a significant change in expression between the two control vs. infected library pairs we used the Bayesian approach described by Lash et al. that takes into account differences in library size (Lash et al., 2000). The cont-b vs. inf-b library pair showed the highest number of significant tag entities (Supplementary Fig. 1), which can be explained by the greater sequencing depth of inf-b library compared with the inf-a library (Supplementary Table 2). A total of 5049 significantly changed tag entities were detected in the intersection of the cont-a vs. inf-a and the cont-b vs. inf-b library pairs (Supplementary Fig. 1; Supplementary Table 4). Efficiencies of tag mapping to the UniGene, RefSeq and Ensembl transcript databases were very similar for the two parallel library pairs (Supplementary Table 5). In both cases the highest percentage of significant tags (ca. 40%) could be mapped to the UniGene database, which is the largest in size (Supplementary Table 5). The intersection of the two library pairs contained a total of 1051 different UniGene transcripts, constituting 2% of the total number of transcripts in this database (Supplementary Fig. 1). Mapping to a unique UniGene transcript was observed for 27% of the significant tags. Approximately 2-fold more significant tag entities mapped to the sense strand of the transcripts than to the antisense strand in any of the three investigated transcript databases (Supplementary Table 5). By comparison, the ratio of sense to antisense mapping of the total number of tags (significant and non-significant) was approximately 1:1 for all libraries. This suggests that in spite of the high number of antisense mapping events detected, the transcriptional regulation in the mycobacterium-induced immune response acts most strongly on the sense strand. The possibility that the host response to mycobacterium infection may also depend on the regulation of antisense expression is of great interest and will require further investigations. In the present study, we focused on the regulation of sense strand transcripts.

Comparison of DGE tag data with a reference set of mycobacterium-regulated genes

Previously we analyzed the zebrafish transcriptome response to mycobacterium infection using three different microarray platforms (Sigma-Compugen spotted oligonucleotide library, MWG spotted oligonucleotide library and Affymetrix). In



	>0	>1	>2	>4	>8	>16	>32	>64	>128	>256	>512	>1024	>2048	>4096	>8192	>16384	>32768	>65536
not mapped	692114	193944	108373	56991	28471	13433	6169	2695	1134	470	202	90	42	22	6	1	1	0
mapped	111752	72154	56398	41478	28809	18927	11511	6568	3399	1695	821	399	195	91	40	16	3	1

Figure 3. Efficiency of tag mapping to the UniGene transcript database. Bars indicate the percentages of successful (green) and non-successful (blue) tag mapping events for the different categories of tag abundance as assigned in Fig. 1. The numbers of tag entities per category are indicated below the graph.

that study we identified a set of genes whose differential expression was detected irrespective of the type of microarray used and that could therefore serve as a multi-platform reference for future transcriptome profiling studies in the mycobacterium—zebrafish model (Meijer et al., 2005). For the present study we took the exact same samples as used for DGE tag profiling and hybridized these to a fourth microarray platform (Agilent 44k). Next, a new reference set was defined consisting of 121 mycobacterium-regulated genes (55 up-regulated and 66 down-regulated) confirmed in two separate studies and by in total four microarray platforms (Supplementary Table 6). This reference set was used for comparison with our DGE tag data. As shown in Supplementary Table 6, 120 genes out of the 121 genes in the reference set were represented by transcript tags in our DGE libraries. The only gene not represented lacks the NlaIII restriction site that is required for detection by DGE. For 100 out of the 120 genes detected (83%) the DGE experiments identified significantly different tag counts between control and infected libraries, whereas the significance threshold was not met for 20 genes. 79 out of 100 significantly changing genes were unambiguously identified by the DGE tags. In 97% of these cases (77 out of 79) the direction of change positively correlated between DGE and microarray analysis (Fig. 4A; Supplementary Table 6). Identification of the remaining 21 genes was ambiguous because the tags mapped to multiple UniGene transcripts. However, in nearly all cases (20 out of 21) multiple mapping occurred to highly related transcripts, encoding a fragment or isoform of the same protein or a near-identical protein. Furthermore, the direction of change positively correlated between DGE and microarray data in 90% of the cases where multiple mapping occurred (19 out of 21) (Fig. 4A; Supplementary Table 6). Therefore, we conclude that in most cases even the multi-mapped tags, which are usually discarded in SAGE-based experimental systems, can also reveal useful information. Finally, when we extended the comparison of the direction of change between DGE and microarray analysis to include also the non-significant tags (mapping to 20 UniGenes in the reference set), we found a positive correlation in 93% of the cases (112 out of the 120 detected genes, Fig. 4B). In conclusion, comparison of our DGE tag data with the reference set of mycobacterium-regulated genes indicates a strong agreement between data obtained by different transcriptome profiling technologies.

Comparison of DGE tag data with Agilent microarray data and qPCR

The flexibility of the Agilent platform is a major advantage for microarray analyses of zebrafish since it allows continuous upgrading of custom designed probes. Agilent is also a preferred platform because previously used platforms (Meijer et al.,

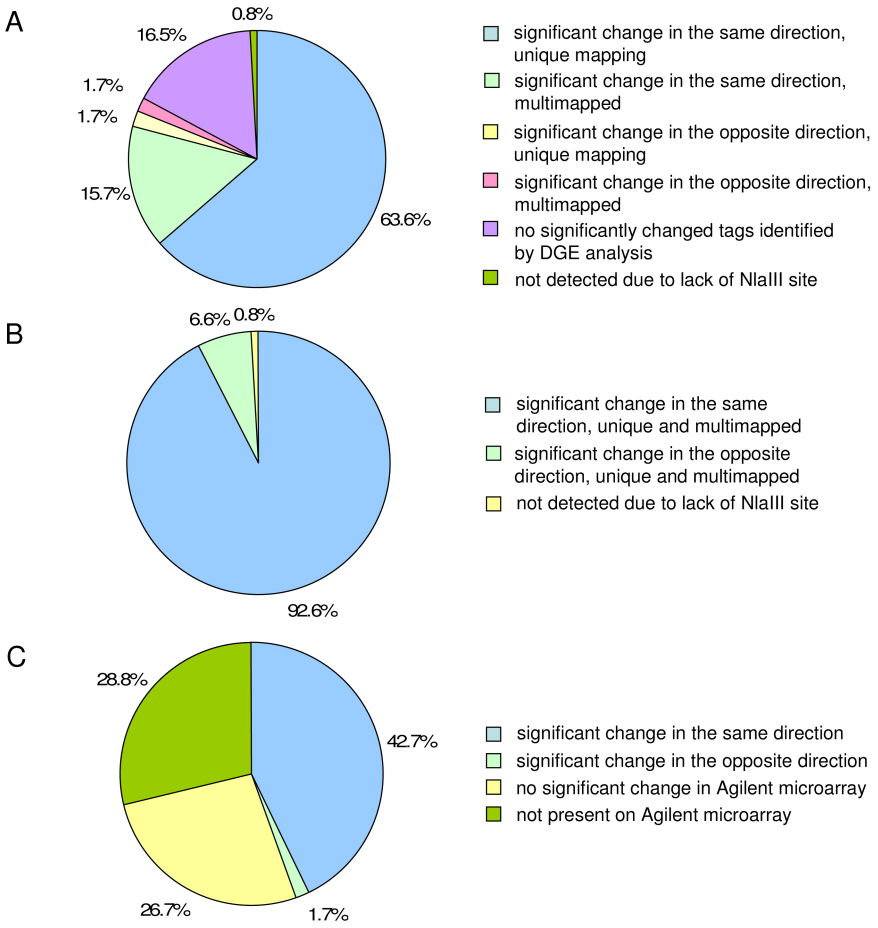


Figure 4. Correlation between DGE and microarray analysis. (A) Comparison of DGE results with a reference set of mycobacterium-regulated genes based on four microarray platforms. DGE results were evaluated by P-value with the significance threshold at 0.05. (B) The same comparison as in (A), but based on the direction of change of all the tags (significant and non-significant) mapped on the same transcript. This approach gives a raw estimation of concordance even in those cases where DGE detected non-significant mappings. (C) Comparison of DGE and Agilent microarray results.

2005) are no longer available (MWG, Sigma-Compugen) or several years outdated (Affymetrix). In order to compare our DGE tag data to Agilent microarray data we used the subset of UniGene transcripts that were identified in both control-infected library pairs by unambiguous mapping of significant tags to the sense strand. A total of 815 different UniGene identifiers met these criteria but only 580 of these UniGenes

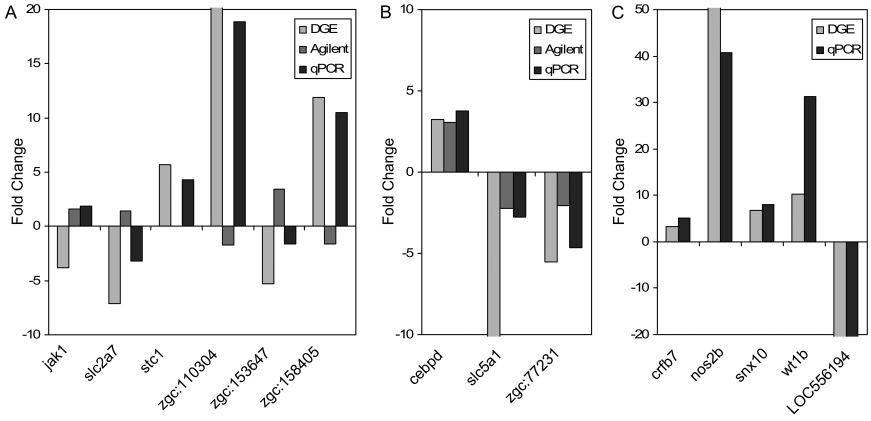


Figure 5. qPCR validation of DGE tag data. (A) Genes showing a significant change by DGE analysis and for which Agilent microarray analysis showed a significant change in the opposite direction (*jak1*, *slc2a7*, *zgc:110304*, *zgc:153647*, *zgc:158405*) or no change (*stc1*). (B) Genes showing a significant change in the same direction by DGE and Agilent microarray analysis. (C) Genes showing a significant change by DGE analysis and not represented on the Agilent microarray platform. Up-regulation of gene expression by mycobacterium infection is indicated by a positive fold change and down-regulation by a negative fold change. Bars touching the chart border indicate infinite fold changes (zero tags in either the control or infected libraries). The following selection of genes was tested with their gene description, UniGene ID and Entrez Gene ID indicated between brackets: *jak1* (janus kinase 1, Dr.74470, 30280), *slc2a7* (slc2a7 solute carrier family 2 (facilitated glucose transporter), member 7, Dr.77040, 100006415), *stc1* (stanniocalcin 1, Dr.88421, 393511), *zgc:110304* (predicted tumor-associated calcium signal transducer 1 homolog, Dr.39071, 550255), *zgc:153647* (predicted cytochrome P450, family 2, subfamily J, polypeptide 2 homolog, Dr.79897, 768288), *zgc:158405* (predicted neutrophil cytosolic factor 2 homolog, Dr.66415, 562473), *cebpd* (CCAAT/enhancer binding protein (C/EBP), delta, Dr.1280, 140817), *slc5a1* (solute carrier family 5 (sodium/glucose cotransporter), member 1, Dr.87868, 93654), *crfb7* (cytokine receptor family member b7, Dr.91624, 777651), *nos2b* (nitric oxide synthase 2b, inducible, Dr.118320, 563654) *snx10* (sorting nexin 10a, Dr.13606, 403027), *wt1b* (wilms tumor 1b, Dr.91799, 568416), *LOC556194* (hypothetical LOC556194, Dr.77733, 556194), *zgc:77231* (predicted fast skeletal myosin alkali light chain 1 isoform 1f homolog, Dr.1448, 336165).

were represented in the Agilent probe set. Agilent microarray analysis showed a significant change in the same direction as determined by DGE analysis for 348 of these genes (60%) (Fig. 4C; Supplementary Table 7). Only 14 genes showed a conflicting significant change in the opposite direction, and no significant change was detected for 218 genes (Fig. 4C; Supplementary Table 7). These results indicate a substantial overlap between DGE and Agilent microarray results in addition to technology-dependent differences.

In order to verify a subset of the DGE tag data by a third independent technology we used quantitative reverse transcriptase PCR (qPCR) analysis. The selection of genes tested included 6 genes that showed a conflict between DGE and Agilent microarray data (Fig. 5A), 3 genes that behaved similarly between the two technologies (Fig. 5B), and 5 genes whose differential expression was detected only by DGE analysis because they were not represented on the Agilent microarray (Fig. 5C). For almost all genes tested, with the exception of *jak1* only (Fig. 5A), qPCR analysis confirmed the direction of change detected by DGE analysis.

Functional annotation and human disease relationships

To achieve an unbiased functional annotation of the infection-responsive genes that were identified by DGE analysis we tested for significant enrichment of Gene Ontology (GO) groups (Supplementary Table 8). Among the up-regulated UniGene transcripts we observed specific enrichment of gene groups including immune response related transcription factors, proinflammatory cytokines and MHC class II proteins (with enriched GO-terms 'immune system process', 'response to stimulus', 'extracellular region', 'catalytic activity'). Gene groups encoding proteolytic enzymes (such as matrix metalloproteinases, cathepsins and proteasome subunits) and lysosomal proton transporting ATPases were also enriched among the up-regulated transcripts. Statistical testing of the down-regulated UniGene transcripts revealed significant enrichment of gene groups encoding enzymes involved in carbohydrate, alcohol, steroid, amino acid and lipid metabolism (with enriched GO-terms 'metabolic process' and 'catalytic activity'). Additionally down-regulated were genes encoding cytoskeletal proteins (among which skeletal muscle actin and myosin), tight junction proteins, solute carriers and fatty acid binding proteins (with enriched GO terms 'structural molecule activity', 'transporter activity', 'binding', 'envelope'). The up- and down-regulated gene groups corresponded well with those identified in our previous microarray analysis of the zebrafish host response to mycobacterium infection (Meijer et al., 2005), indicating that DGE and microarray analysis identified similar functional groups of genes.

Gene ontology analysis allowed identification of only 41% of the significantly up- or down-regulated transcripts. In order to extend and improve the annotation of the DGE data set we retrieved the putative human orthologs of the zebrafish genes by using the g:Orth function of the g:Profiler web server (http://biit.cs.ut.ee/gprofiler) and by searching the NCBI HomoloGene database. In addition, we compiled information on protein similarities from the UniGene database. As a result, we were able to annotate 73% of the up-regulated and 83% of the down-regulated transcripts, whereas the remaining transcripts lacked any similarity to genes with known function (Fig. 6; Supplementary Table 9).

Since *M. marinum* infection of zebrafish is considered as an animal model of human tuberculosis, next we analyzed how the whole set of mycobacterium-regulated genes may associate with a known human pathological condition. A search using

GeneALaCart, provided by GeneCards (www.genecards.org), linked our data set to 397 OMIM (Online Mendelian Inheritance in Man) disorders, of which 132 were additionally represented by UniProt disorders that are associated with monogenic human genetic diseases (Supplementary Table 9). This is a substantial proportion (16%) of the mycobacterium-regulated UniGene clusters identified by DGE analysis.

Genomic mapping of DGE tag sequences

Disease-related genes may be underrepresented in the zebrafish transcript databases that are mostly based on sequencing of cDNA libraries from healthy fish. Therefore, additional mycobacterium-regulated transcripts can be revealed by genomic mapping of the DGE tag sequences (Supplementary tables 2 and 5). Genomic mapping of the significant tags had a success rate of 59% (of which 84% unique mapping events), which was approximately 1.5-fold more efficient than mapping to the UniGene transcript database (Supplementary Table 5). The fact that we did not allow mismatches and that the zebrafish genome is highly polymorphic may explain that still a substantial proportion of the significant tags could not be mapped. In addition, tags extending over intron boundaries would not be picked up in genomic mapping and gaps in the zebrafish genome sequence still exist. The chromosomal distribution of the complete collection of tag entities (significant and non-significant) was roughly proportional to the chromosome length (Fig. 7A). However, the chromosomal distribution of the significant tags was less flat (Fig. 7B) and specific regions could be identified within the individual chromosomes where a number of tags clustered together, suggestive of potential genomic hotspots linked with the physiological processes induced by mycobacterium infection. In Fig. 7C, we show an example of such a region on chromosome 3 where 10 mycobacterium-regulated genes were identified. These included *grb2* and *ras* family genes (*zgc:112091*, *rac2*, *arl6ip1*) linked with immune response and hematopoietic signaling (e.g. Diebold and Bokoch, 2001; Martin et al., 2005; Pettersson et al., 2000), several genes involved in inflammation and defense (encoding granulin a, ferritin-like, nucleobindin-1 and actinoporin proteins (Aroian and van der Goot, 2007; He and Bateman, 2003; Leclerc et al., 2008; Recalcati et al., 2008), and the zebrafish homolog of the human *CORO1A* gene, which encodes the actin-binding protein coronin that has long been implicated in mycobacterial pathogenesis (Jayachandran et al., 2007; Kaul, 2008).

Verification of predicted gene models based on DGE tag information

Since DGE is a SAGE-based transcript profiling method, theoretically all the observed tags in the DGE libraries should be mapped to the so called canonical site, a sequence region located following the last NlaIII cutting site on the sense strand at the 3'-end of the transcript. In our experiments only about 40% of the sense strand mappings fell to the canonical site. Furthermore, approximately 70% of the detected transcripts were represented by more than one tag. These observations are consistent with similar results in DGE analysis of the mouse transcriptome (t Hoen et al.,

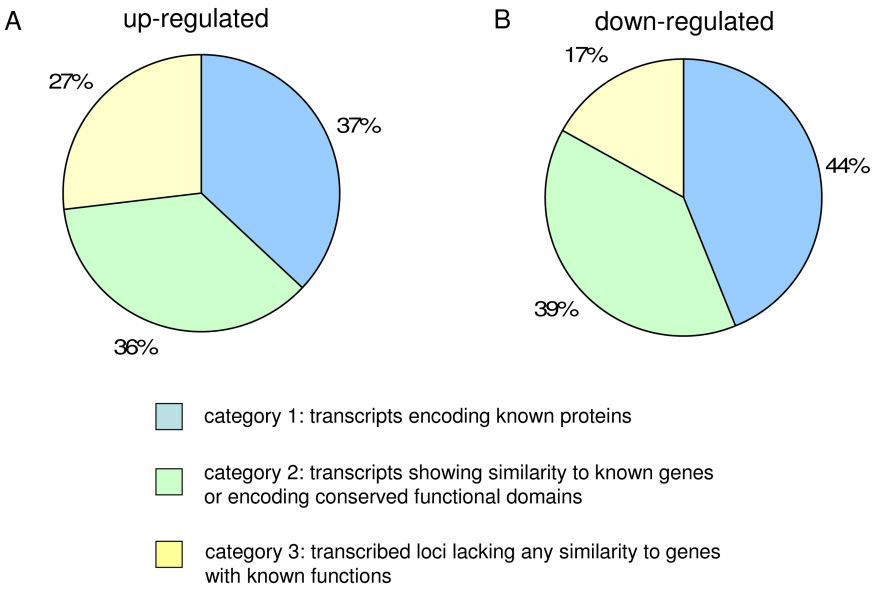


Figure 6. Distribution of transcripts over different annotation categories. (A) Transcripts up-regulated by mycobacterium infection. (B) Transcripts down-regulated by mycobacterium infection.

2008) and may be explained partly by incomplete NlaIII digestion during library preparation as well as by frequent use of alternative polyadenylation and/or alternative splicing sites (Pan et al., 2008; Wang et al., 2008). The existence of multiple tags per transcript can be conveniently utilized for the experimental verification of *ab initio* gene predictions from the Ensembl database. Many of these predicted transcripts gather more than one tag hit and many of them have count numbers over several hundred up to thousands. Detection of multiple tags with high count numbers specific for a predicted transcript indicates the reliability of the transcript sequence and the biological importance of the given gene. In a specific example (Ensembl: GENSCAN00000043834, encoding a homolog of the human olfactomedin 4 (*OLFM4*) gene induced by mycobacterium infection in our DGE study) we show how tag mapping over different Genscan exons can support predicted gene models (Fig. 8A).

Infection-induced transcript isoform switching

Next we have tried to identify transcripts where the infection triggered unbalanced alterations in tag expression pattern that may be caused by the selective induction or repression of particular transcript isoforms generated by alternative splicing, alternative polyadenylation or alternative transcription initiation. To this extent we

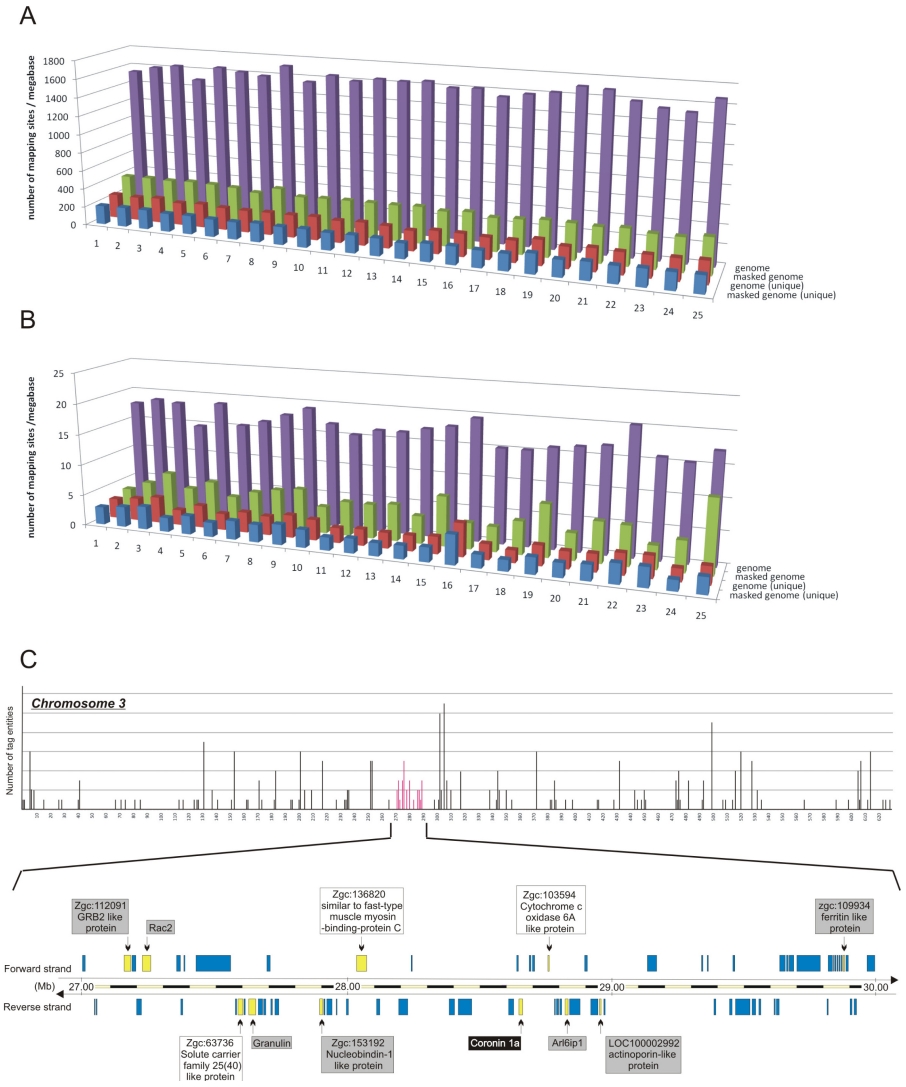


Figure 7. Distribution of genomic DGE tag mapping sites over the different chromosomes. (A) Normalized chromosomal distribution of DGE tags (significant and non-significant) per megabase. Mapping results of one DGE library (cont-a) are shown as a representative example. The total number of mapping sites and the number of mapping sites with unique sequence were determined using the unmasked and masked versions of the Zv7 zebrafish genome sequence. Both the upper and lower strand of genomic DNA was equally populated by the tags of the DGE libraries (Supplementary Table 3). (B) Normalized chromosomal distribution of the DGE tags that were significantly changed by mycobacterium infection. (C) Distribution of the unique and significant tags mapping to chromosome 3. A region between 27.2 and 29.9Mb where 29 significant tags map closely together, assigning an active chromosomal region during mycobacterium infection, is in-

dicated in red with the genomic structure shown below. Genes that are hit on the sense strand by the significant tags are shown in yellow and other genes in blue. Annotations of encoded proteins on the two strands of the genomic sequence are indicated by rectangles above and below the ruler. The *coronin 1a* gene (*coro1a; coronin, actin binding protein, 1A*) that is hit by 2 up-regulated tags has a direct link with tuberculosis infection (Jayachandran et al., 2007; Kaul, 2008). The proteins encoded by other genes that were hit by up-regulated tags are indicated in grey boxes. All these proteins have previously been linked with responses to inflammation or infection (e.g. Aroian and van der Goot, 2007; Diebold and Bokoch, 2001; He and Bateman, 2003; Leclerc et al., 2008; Martin et al., 2005; Pettersson et al., 2000; Recalcati et al., 2008). Proteins encoded by genes hit by down-regulated tags are shown in white boxes.

selected transcripts with at least two mapped tags, and we investigated how their tag expression patterns changed within the transcripts upon mycobacterium infection. Among the mycobacterium-regulated gene set we identified over 30 cases where part of the detected tag repertoire changed to a much greater extent than other tags, or where different tags for the same transcript were changed in the opposing direction. An example of a case where only part of the tag repertoire was changed is *stanniocalcin 1* (*stc1*), a gene for a calcium-regulating glycoprotein hormone. Up-regulation of this gene upon mycobacterium infection was confirmed by qPCR analysis (Fig. 5A). Five tags at the 3'-end of the transcript showed a significant change with 4–10 times increase in count number in the two control vs. infected library pairs (Fig. 8B). However, the most 3' tag of this transcript was unchanged in both library pairs. Since the last four tags are located in the 3' UTR of this transcript, this observation most probably reflects a selectively induced alternative polyadenylation event, where the infection results in the expression of a transcript variant having a shorter 3' UTR region. This hypothesis is also supported by the finding that the 3' UTR contains two polyadenylation signals, one at the very end of the sequence after the last weakly changing tag and another in a more upstream position. Alternative polyadenylation might be a mechanism for regulation of *stc1* mRNA stability, which was previously shown to be affected by external stimuli (Ellis and Wagner, 1995). An example of a case where tags for the same transcript changed in opposite directions is *cathepsin D* (*ctsd*). The tag expression pattern for this gene indicates selective down-regulation of a longer transcript form with concomitant up-regulation of a transcript with a shorter 3' UTR (Fig. 8C).

Identification of novel transcript forms induced by mycobacterium infection

Finally, we attempted to find novel mycobacterium-regulated transcript forms (splice variants) without any previous indications from different prediction methods. As above, we took advantage of the observation that the majority of known transcripts were represented by more than one tag. We anticipated that the same would

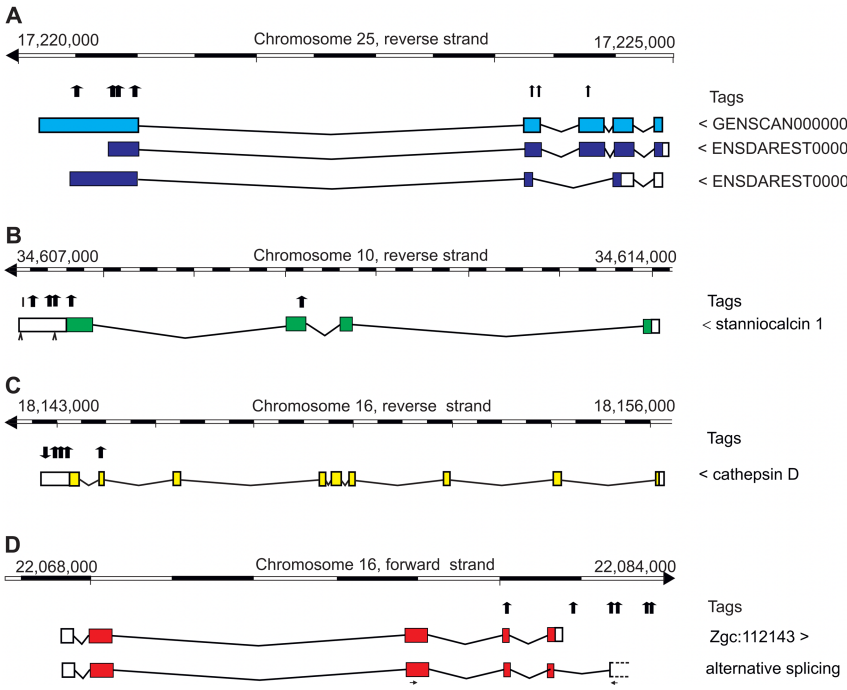


Figure 8. Genome data mining using DGE tag information. Arrows pointing up indicate the mapping positions of tags with significantly increased expression under mycobacterium-infection conditions. Arrows pointing down indicate tags significantly down-regulated by mycobacterium infection. Lines indicate non-significant tags. Bold arrows correspond to tag counts above 100 in the DGE libraries of mycobacterium-infected fish. (A) Verification of an *ab initio* gene prediction from Ensembl. Three exons of the predicted GENSCAN00000043834 gene, a homolog of the human olfactomedin 4 gene, are supported by DGE tag information. The tag distribution further supports the existence of at least two alternative splice forms, represented by ENSDAREST00000036260 and ENSDAREST00000036258. (B) Infection-dependent isoform switching. Arrowheads below the gene indicate the positions of the last two polyadenylation signals in the *stc1* transcript. The unchanged expression of the most 3′ tag vs. the significant induction of more upstream tags suggest alternative polyadenylation induced by mycobacterium infection leading to a transcript with a shorter 3′ UTR. (C) Infection-dependent up- and down-regulation of two transcript isoforms of the *ctsd* gene. (D) Infection-dependent alternative splicing leading to a novel gene product. A cluster of 5 significant tags was detected downstream of the known *zgc:112143* gene, encoding the zebrafish homolog of the tumor-necrosis factor alpha-induced protein 9. Horizontal arrows indicate the position of RT-PCR primers used to verify the expression of an alternative transcript form that comprises the tag cluster region. From sequence analysis of the amplified product we could derive alternative splicing of the 5th exon to a new exon 6, encoding a different C-terminal domain of the protein. The 3′-end of exon 6 was not determined (open rectangle).

likely be true for currently unknown transcript forms and therefore searched for tag clusters with significant tags that could not be mapped to any of the three transcript databases used in this work, but having genomic localizations close to each other. Next, we examined the neighboring genomic environment of these tag clusters, and inspected the gene annotations in that region for the presence of potential immune relevant functions. Out of 98 investigated clusters 29 proved to be located close to a gene that might have a function in immune defense (Supplementary Table 10). As an example, we experimentally confirmed the expression of a novel splice form of the *zgc:112143* gene, which is homologous to the human tumor necrosis factor alpha-induced protein 9 gene (*TNFAIP9*, also known as *STEAP4*), encoding a six transmembrane putative channel or transporter protein (Moldes et al., 2001). To this extent we performed RT-PCR on RNA samples from control and infected fish, taking a forward primer in one of the known exons of the gene (exon 3) and a reverse primer overlapping with one of the mycobacterium-induced tags located downstream of the known 3' transcript end in a region without any predicted or EST-based indications of transcription. An RT-PCR product could be amplified from infected but not from control RNA samples, in agreement with the much higher tag counts in the DGE libraries from infected fish (total count numbers: cont-a/cont-b: 3/2; inf-a/inf-b: 86/369). Sequence analysis of the resulting product revealed alternative splicing in the known exon 5 resulting in a novel transcript form comprising a 6th exon that is the source of the significant DGE tags that we identified in our study (Fig. 8D). Alternative splicing changes the C-terminal domain of the TNFAIP9 homolog, downstream of the six predicted transmembrane domains of this protein.

Discussion

Major advances in transcriptomics have become possible as a result of novel technology developments in deep sequencing (Cloonan and Grimmond, 2008; Morozova and Marra, 2008; Wang et al., 2009; Wilhelm and Landry, 2009). Here we report on the first deep sequencing study of the vertebrate host response to infectious disease. We chose the mycobacterium-zebrafish infection model as a case study because of its relevance to human tuberculosis, a disease that is characterized by an intricate interaction between host and pathogen (Lesley and Ramakrishnan, 2008; Pieters, 2008). Furthermore, multi-platform microarray data sets of infection in this model were available for validation purposes (Meijer et al., 2005). Transcriptome alterations underlying a complex process like infectious disease are characterized not only by massive gene induction and repression responses, but also by subtle changes in transcript levels and presumably in the expression of alternative transcript forms (Vos et al., 2007). Our results demonstrate that deep sequencing is a major improvement over microarray analysis for detection of transcriptional changes over such a wide range. Moreover, the unbiased nature of deep sequencing data gives a fundamental advantage for gene discovery and genome annotation.

In this study we used Solexa/Illumina's Digital Gene Expression (DGE) system, which is essentially a SAGE-based tag profiling approach. We could reach a sequencing depth of 5–8 million tags per library and confirmed the reproducibility of replicate DGE library constructions by correlation analysis. Using a Bayesian method for statistical evaluation of our DGE tag data from control and mycobacterium-infected animals, we found over 5000 sequence tags to be differentially expressed. Count numbers of the significant tags spanned a dynamic range of three orders of magnitude. An example of one of the most abundant tag sequences represented skeletal muscle actin alpha 1, which was detected by over 40,000 copies in the controls and down-regulated to around 10,000 copies upon mycobacterium infection. As an example of the sensitivity of the method, a tag sequence specific for the transcript encoding peptidoglycan recognition protein 1 was undetectable in the DGE libraries from control fish and was significantly up-regulated to around 20 copies in the DGE libraries from infected fish. The sensitivity of DGE profiling is further demonstrated by the fact that our analysis of a single developmental stage of the zebrafish (adult stage) could detect over 70% of all transcripts in the UniGene, RefSeq and Ensembl databases. The significantly changed tag sequences mapped to approximately 2% of the transcripts in these databases.

Since only one other Illumina-based DGE study has presently been reported (t Hoen et al., 2008), it was important to validate our DGE results by independent transcriptome profiling technologies. Our DGE data showed strong correlation (over 90%) with a reference set of mycobacterium-regulated genes that had previously been confirmed in a multi-platform microarray study. Comparison with microarray data obtained from a single microarray platform (Agilent) also showed a substantial level of correlation (60%). Furthermore, gene ontology analysis showed significant enrichment of similar functional groups of genes in DGE and Agilent microarray data sets. In most of the investigated cases where DGE and microarray analysis gave conflicting results, qPCR analysis supported the DGE results. Conflicts between DGE and microarray results may occur for technical reasons, but may also result from the fact that microarray probes will often detect a mixture of different transcript isoforms, whereas DGE analysis can discriminate between specific transcript isoforms (t Hoen et al., 2008). Both methods have their intrinsic limitations, as discussed extensively by t Hoen et al. (2008). In short, important drawbacks of microarray analysis concern limited sensitivity, cross-hybridization problems and inadequate probe design. On the other hand, DGE will fail to identify some transcripts that lack a unique tag sequence or cutting site for the DGE anchoring enzyme (NlaIII in this study). Therefore, microarray analysis and DGE can be considered complementary to each other. DGE compares favorably to microarray analysis in standardization between laboratories (Irizarry et al., 2005; Marioni et al., 2008; Pedotti et al., 2008; t Hoen et al., 2008). The most important advantage of DGE analysis is that the method is not limited by predefined array content. In our study, we detected differential expression of many transcripts that were not represented on the available microarray platforms.

By qPCR analysis we confirmed the differential expression of a subset of these transcripts, with tag abundances in the range of 20–400 copies.

Classical SAGE studies (Velculescu et al., 1995), based on conventional sequencing of cloned tags, have focused on analysis of the canonical tags, which are the tags resulting from the most 3' located NlaIII restriction site on the sense strand of the transcript. The drawback of mapping only the canonical tags is that any information on alternative polyadenylation or alternative splicing is lost. In addition to biological mechanisms, partial digestion may also contribute to non-canonical mappings. As long as partial digestion is carefully controlled for by simultaneous preparation of parallel libraries with the same batch of reagents it does not present a problem. Rather, as a result of the enormous sequencing depth achieved in DGE analysis, the more 5' located non-canonical tags can provide a useful source of additional information. Furthermore, valuable insights can be obtained from genomic mapping of those tags that fail to hit the transcript databases. Here we have shown that the information obtained from genomic mapping of the entire set of DGE tags can be used (i) to investigate genomic clustering of co-regulated transcriptome responses, (ii) to verify the expression of *ab initio* predicted genes, (iii) to detect switching between alternative transcript forms induced by mycobacterial infection, and (iv) to detect and map novel mycobacterium-regulated transcript forms for which any previous EST-based evidence of expression was lacking. Altogether, we found a 50% higher efficiency for genomic mapping of the significant tag entities than for mapping to the UniGene transcript database. This shows the limitations of the use of cDNA sequence data for disease studies, a problem that is solved by deep sequencing strategies.

Contrary to tag-based DGE analysis, full RNA sequencing procedures (RNA-Seq), in which mRNA or cDNA is fragmented mechanically, result in overlapping short fragments that cover the entire transcriptome. Clearly this approach is even more powerful for unraveling transcriptome complexity. However, tag sampling methods, which produce a less complex data mass, are currently more suitable and affordable for comparative expression studies of larger numbers of samples. For the mouse it has been estimated that 40 million RNA-Seq reads are required to achieve coverage of most transcripts and to allow accurate analysis of expression differences between samples (Mortazavi et al., 2008). In contrast, approximately 2 million transcript-specific DGE tags from mouse were shown to be sufficient to reliably detect low abundant genes (t Hoen et al., 2008). Whereas the quantification of RNA-Seq results may be complicated by unequal coverage of transcripts by sequence reads, the quantitative comparison of DGE tag libraries can build directly on the vast knowledge of statistical evaluation of conventional SAGE experiments (Lash et al., 2000; t Hoen et al., 2008; Zhu et al., 2008).

Unlike some of the RNA-Seq protocols, SAGE-derived DGE analysis provides information on transcript directionality. In our study approximately 40% of the mapping events were detected on the antisense strand of UniGene, RefSeq or Ensembl database transcripts. This is comparable to data reported for mouse DGE analysis,

where evidence for bidirectional transcription was found for 51% of all detectable UniGene clusters (t Hoen et al., 2008). Similar as in their study, we found no correlation between the abundance of sense and antisense tags for the same transcript. To what extent antisense transcripts might result from transcriptional noise is currently unknown. However, there is accumulating evidence for the widespread occurrence of antisense transcription and its biological relevance (Beiter et al., 2009; Carninci et al., 2005; Katayama et al., 2005). For example, the proper dosage of expression of the human and murine hematopoietic transcription factor PU.1, critical for suppression of leukemia, was shown to rely on antisense RNA modulators (Ebralidze et al., 2008).We found that mycobacterium infection significantly induced both sense and antisense DGE tags of the zebrafish homolog of the *PU.1* gene (*spi1*), suggesting that antisense regulation of the expression of this gene may be an ancient evolutionary mechanism.

The mycobacterium-infected zebrafish analyzed in this study were at the late stage of chronic tuberculosis. As previously observed in microarray analysis, our DGE study showed massive changes in the expression levels of known immune response genes and of genes that have been implicated in human tuberculosis (Meijer et al., 2005). Many of the mycobacterium-regulated genes that we could functionally annotate were homologous to human genes that have been linked with genetic diseases. These included inflammatory and hematologic disorders, for example, anemia, which is commonly associated with tuberculosis (Lee et al., 2006). We also identified many interesting genes that had not been linked to tuberculosis in previous studies. For example, we demonstrated infection-dependent up-regulation of an *ab initio* predicted homolog of olfactomedin 4 (*OLFM4*), also known as granulocyte colony stimulating factor stimulated clone-1 (*hGC-1*). This gene encodes a glycoprotein of unknown function, suggested to be involved in cell adhesion, cancer progression and myeloid development (Chin et al., 2008; Liu et al., 2008). We also observed induction of a specific transcript variant of the zebrafish homolog of *stanniocalcin 1* (*stc1*), a gene suggested to be involved in macrophage chemotaxis and transendothelial migration of inflammatory cells (Chakraborty et al., 2007; Kanellis et al., 2004). Furthermore, by genomic tag mapping we demonstrated induction of a novel alternative splice form of a gene homologous to the human tumor necrosis factor alpha-induced protein 9 gene (TNFAIP9, also known as STEAP4). This gene encodes a six transmembrane putative channel or transporter protein implicated in inflammatory responses and cancer progression (Korkmaz et al., 2005; Wellen et al., 2007). The above mentioned genes are merely given as examples to illustrate how our deep sequencing-based systems approach can quickly lead from mapping of total transcriptome responses to detailed functional annotation. More importantly, the entire description of all mycobacterial regulated genes provides an unbiased basis for subsequent candidate gene approaches that was not possible with prior microarray-based transcriptome profiling. In future studies we will use deep sequencing technology to compare the present data set of the late stage of tuberculosis with

earlier stages of disease progression, including granuloma formation in the zebrafish embryo model that is more amenable to experimental manipulation (Lesley and Ramakrishnan, 2008).

Conclusions

We have demonstrated here that deep sequencing analysis, using Solexa/Illumina's digital gene expression (DGE) system, provides a robust, sensitive and unbiased alternative to microarray analysis with major advantages for detection of the broad range of transcriptional responses that occur during the process of an infectious disease, as exemplified in this study using the zebrafish mycobacterium model. First, by mapping of DGE sequence tags to transcript databases we showed that mycobacterium infection induced quantitative changes in the expression levels of many genes, including those previously implicated in human tuberculosis. These data showed over 90% concordance with a reference set of mycobacterium-regulated genes that had previously been confirmed in a multi-platform microarray study. Second, by genomic mapping of the DGE sequence tags we could reveal transcriptional responses that microarray analysis would have failed to detect, such as the switching between alternative transcript isoforms, the expression of novel splice products not present in the current transcript and EST databases, and a high level of antisense transcription. Our DGE study substantiates recent RNA sequencing results in other model species indicating a much larger extent of genome transcription than previously thought. Furthermore, it demonstrates the advantages of a deep sequencing approach for gene discovery and genome annotation and provides new leads for functional studies of candidate genes involved in host–pathogen interaction.

Materials and methods

Zebrafish husbandry and infection experiments

Zebrafish were handled in compliance with the local animal welfare regulations and maintained according to standard protocols (http://ZFIN.org). The infection experiment was approved by the animal welfare committee (DEC) of Leiden University. Adult male zebrafish were infected by intraperitoneal inoculation with approximately 1×10^3 *Mycobacterium marinum* bacteria as previously described (Meijer et al., 2005). Four of the RNA samples used for this study were identical to those from our previously published chronic infection study (control fishes c1, c2 and infected fishes i1, i2) (Meijer et al., 2005). Four additional RNA samples (control fishes c3, c4 and infected fishes i3, i4) were from an independent *M. marinum* E11 infection experiment performed under similar conditions. All four infected fish were sacrificed when they showed overt signs of fish tuberculosis, including lethargy, skin ulcers and extensive granuloma formation in organs, such as liver and kidney. Histological examination of fish from the same experiments confirmed that the pathology of infected fish cor-

responded to fish tuberculosis (Swaim et al., 2006; van der Sar et al., 2004) and that no characteristics of the disease were present in the control fish.

RNA isolation

Fish were snap frozen in liquid nitrogen, stored at −80 °C, and homogenized in liquid nitrogen using amortar and pestle. Portions of 50–100µg of powdered tissue were used for extraction of total RNA with 1ml of TRIZOL® Reagent (Invitrogen) according to the manufacturer's instructions. The RNA samples were incubated for 20min at 37°C with 10 units of DNaseI (Roche Applied Science) to remove residual genomic DNA prior to clean up using RNeasy columns (Qiagen). The integrity of the RNA was confirmed by Lab-on-chip analysis using the 2100 Bioanalyzer (Agilent Technologies). The samples used had an average RIN value of 9.5 and a minimum RIN value of 8.9.

Digital gene expression-tag profiling (DGE)

For DGE analysis RNA samples from the four control fish (c1, c2, c3, c4) were pooled, and RNA samples from the four infected fish (i1, i2, i3, i4) were pooled. Before pooling the individual RNA samples had been checked by microarray analysis for correlation between biological replicates. From each pool duplicate libraries for tag sequencing were prepared in order to assess technical reproducibility. Tag library preparation was performed using the DGE: Tag Profiling for NlaIII Sample Prep kit from Illumina according to the manufacturer's instructions. In brief, 1 µg of total RNA was used for mRNA capture using magnetic oligo(dT)beads. First- and second-strand cDNA was synthesized and bead-bound cDNA was subsequently digested with NlaIII. Fragments other than the 3' cDNA fragments attached to oligo(dT) beads were washed away and a GEX NlaIII adapter was ligated to the free 5'-end of the digested bead-bound cDNA fragments. The GEX NlaIII adapter contains a restriction site for MmeI which cuts 17–18 bp downstream from the NlaIII site, thereby releasing 21–22 bp tags starting with the NlaIII recognition sequence, CATG. A second adapter (GEX adapter 2) was ligated at the site of MmeI cleavage, and the adapter-ligated cDNA tags were enriched using PCR-primers that anneal to the adaptor ends. The resulting 85 bp fragments were purified from a 6% acrylamide gel. Purity and yield were checked by Lab-on-chip analysis using the 2100 Bioanalyzer (Agilent Technologies). A total of 6 pmol of cDNA per tag library was used for cluster generation on individual lanes of Illumina's 1.4 mm channel flow cell, and sequencing by synthesis was performed using the Illumina Genome Analyzer II system (ServiceXS, 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. The raw data (tag sequences and counts) were deposited in the GEO database under submission number GSE14782.

Statistical evaluation of DGE libraries

The tag entities and count numbers of DGE libraries from control and infected fish were collected and summarized by custom perl and PHP scripts. Statistical comparison was performed using the Bayesian method described by Lash et al. (2000) with the software tool available from the SAGEmap resource. We accepted the change of a tag expression as significant if the chance of a false positive hit was less than 5% ($P < 0.05$). The correlation of the detected count numbers between parallel libraries was statistically assessed by calculation of Pearson correlation coefficient using the built-in function of Microsoft Excel.

Mapping of DGE tags

For mapping of DGE tags to different transcript databases or to the zebrafish genome we created virtual libraries containing all the possible 17 bases length sequences of these resources located next to an NlaIII restriction site. For transcript mapping we used the UniGene (Danio rerio build 105), Refseq (2007_07) and Ensembl (ZFISH7.49) transcript databases. For genomic mapping we used the native and the masked form of the zebrafish genome version Zv7, which were downloaded from the FTP server of the Ensembl database. For mapping against UniGene transcripts we used the file containing the sequence with the longest region of high-quality sequence data from each UniGene cluster. For monitoring the mapping events on both strands, both the sense and the complementary antisense sequences were included in the data collection. Information on the position of polyadenylation signals and the length of polyadenylation tails was also collected from the transcript databases. Virtual tag libraries and the DGE libraries were uploaded into an in house developed data warehouse using a relational database engine for mapping of DGE tags onto virtual libraries. Only perfect matches over the entire 21 bp length of the 17 bp tag plus the 4 bp NlaIII recognition site were allowed. During the mapping process the system tracked the so-called multiple mapping events where tags detected in the experiments could be assigned to more than one transcript or to more than one position in the genome.

Microarray analysis

Agilent microarray analysis was performed using a custom designed platform (GEO submission number GPL7735) with previously described conditions for labeling, hybridization, scanning and feature extraction (Stockhammer et al., 2009). Samples from control fish (c1–c4) were labeled with Cy3 dye and samples from infected fish (i1–i4) with Cy5 dye. After feature extraction, data were imported into Rosetta Resolver 7.1 (Rosetta Biosoftware, Seattle, Washington) and subjected to default ratio error modeling. Ratio results from four control fish vs. four infected fish (c1 vs. i1, c2 vs. i2, c3 vs. i3, c4 vs. i4) were combined using the default ratio experiment builder. Data were analyzed at the level of UniGene clusters (UniGene build #105). Microarray data were submitted to the GEO database under series GSE14782.

RT-PCR

Real-time quantitative RT-PCR (qPCR) was performed as previously described (Stockhammer et al., 2009). All reactions were done in duplicate. For normalization *ppial* (peptidylprolyl isomerase A like), which was unaffected by mycobacterium infection, was taken as reference. Sequences of forward and reverse primers are shown in Supplementary Table 1. RT-PCR verification of the mycobacterium-induced novel transcript variant of the *zgc:112143* gene was performed using the SuperScript III One-Step RT-PCR System with Platinum Taq DNA Polymerase (Invitrogen). A primer overlapping with one of the significant tags located downstream of the known 3' transcript end was used for the reverse transcription reaction (CCAGACACTCAAACAGACATG). The same primer was subsequently used in the PCR amplification step, in combination with a forward primer in exon 3 of the known transcript (ATGAGCGAGTTACCAACGGA). The resulting PCR product was sequenced using the sequencing service of ServiceXS (Leiden, The Netherlands) and submitted to GenBank under accession number FJ754358.

Acknowledgements

We thank Peter-Bram 't Hoen (Leiden University Medical Center, the Netherlands) for helpful discussions, Astrid van der Sar and Wilbert Bitter (Vrije Universiteit Medical Center, Amsterdam) for the *M. marinum* E11 strain, and Susan Wijting and Ulrike Nehrdich for help with infection experiments. This work was supported by the European Commission 6th framework project ZF-TOOLS (LSHGCT- 2006-037220).

Supplementary data

Supplementary data associated with this chapter can be found at doi:10.1016/j.molimm.2009.07.002.

References

Aroian, R., van der Goot, F.G. (2007) Pore-forming toxins and cellular non-immune defenses (CNIDs). Curr. Opin. Microbiol. 10, 57–61.

Beiter, T., Reich, E., Williams, R.W., Simon, P. (2009) Antisense transcription: a critical look in both directions. *Cell. Mol. Life Sci.* **66**, 94–112.

Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D.H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M., Roth, R., George, D., Eletr, S., Albrecht, G.,Vermaas, E., Williams, S.R., Moon, K., Burcham, T., Pallas, M., DuBridge, R.B., Kirchner, J., Fearon, K., Mao, J., Corcoran, K. (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat. Biotechnol.* **18**, 630–634.

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B.,Wells, C., et al. (2005) The transcriptional landscape of the mammalian genome. *Science* **309**, 1559–1563.

Chakraborty, A., Brooks, H., Zhang, P., Smith, W., McReynolds, M.R., Hoying, J.B., Bick, R., Truong, L., Poindexter, B., Lan, H., Elbjeirami, W., Sheikh-Hamad, D. (2007) Stanniocalcin-1 regulates endothelial gene expression and modulates transendothelial migration of leukocytes. *Am. J. Physiol. Renal Physiol.* **292**, F895–F904.

Chin, K.L., Aerbajinai, W., Zhu, J., Drew, L., Chen, L., Liu, W., Rodgers, G.P. (2008) The regulation of OLFM4 expression in myeloid precursor cells relies on NF-kappaB transcription factor. *Br. J. Haematol.* **143**, 421–432.

Cloonan, N., Forrest, A.R., Kolle, G., Gardiner, B.B., Faulkner, G.J., Brown, M.K., Taylor, D.F., Steptoe, A.L.,Wani, S., Bethel, G., Robertson, A.J., Perkins, A.C., Bruce, S.J., Lee, C.C., Ranade, S.S., Peckham, H.E., Manning, J.M., McKernan, K.J., Grimmond, S.M. (2008) Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat. Methods* **5**, 613–619.

Cloonan, N., Grimmond, S.M. (2008) Transcriptome content and dynamics at single nucleotide resolution. *Genome Biol.* **9**, 234.

Diebold, B.A., Bokoch, G.M. (2001) Molecular basis for Rac2 regulation of phagocyte NADPH oxidase. *Nat. Immunol.* **2**, 211–215.

Ebralidze, A.K., Guibal, F.C., Steidl, U., Zhang, P., Lee, S., Bartholdy, B., Jorda, M.A.,Petkova, V., Rosenbauer, F., Huang, G., Dayaram, T., Klupp, J., O'Brien, K.B., Will, B., Hoogenkamp, M., Borden, K.L., Bonifer, C., Tenen, D.G. (2008) PU.1 expression is modulated by the balance of functional sense and antisense RNAs regulated by a shared cis-regulatory element. *Genes Dev.* **22**, 2085–2092.

Ellis, T.J.,Wagner, G.F. (1995) Post-transcriptional regulation of the stanniocalcin gene by calcium. *J. Biol. Chem.* **270**, 1960–1965.

Harbers, M., Carninci, P. (2005) Tag-based approaches for transcriptome research and genome annotation. *Nat. Methods* **2**, 495–502.

Hashimoto, S., Qu, W., Ahsan, B., Ogoshi, K., Sasaki, A., Nakatani, Y., Lee, Y., Ogawa, M., Ametani, A., Suzuki, Y., Sugano, S., Lee, C.C., Nutter, R.C., Morishita, S., Matsushima, K. (2009) High-resolution analysis of the 5'-end transcriptome using a next generation DNA sequencer. *PLoS ONE* **4**, e4108.

He, Z., Bateman, A. (2003) Progranulin (granulin-epithelin precursor, PC-cell-derived growth factor, acrogranin) mediates tissue repair and tumorigenesis. *J. Mol.Med.* **81**, 600–612.

Irizarry, R.A., Warren, D., Spencer, F., Kim, I.F., Biswal, S., Frank, B.C., Gabrielson, E., Garcia, J.G., Geoghegan, J., Germino, G., Griffin, C., Hilmer, S.C., Hoffman, E., Jedlicka, A.E., Kawasaki, E., Martinez-Murillo, F.,Morsberger, L., Lee, H., Petersen, D., Quackenbush, J., Scott, A.,Wilson, M., Yang, Y., Ye, S.Q., Yu,W. (2005) Multiple laboratory comparison of microarray platforms. *Nat. Methods* **2**, 345–350.

Jayachandran, R., Sundaramurthy, V., Combaluzier, B., Mueller, P., Korf, H., Huygen, K., Miyazaki, T., Albrecht, I., Massner, J., Pieters, J. (2007) Survival of mycobacteria in macrophages is mediated by coronin 1-dependent activation of calcineurin. *Cell* **130**, 37–50.

Kanellis, J., Bick, R., Garcia, G., Truong, L., Tsao, C.C., Etemadmoghadam,D., Poindexter, B., Feng, L., Johnson, R.J., Sheikh-Hamad, D. (2004) Stanniocalcin-1, an inhibitor of macrophage chemotaxis and chemokinesis. *Am. J. Physiol. Renal Physiol.* **286**, F356–F362.

Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C.C., Suzuki, M., Kawai, J., Suzuki, H., Carninci, P., Hayashizaki, Y., Wells, C., Frith, M., Ravasi, T., Pang, K.C., Hallinan, J., Mattick, J., Hume, D.A., Lipovich, L., Batalov, S., Engstrom, P.G., Mizuno, Y., Faghihi, M.A., Sandelin, A., Chalk, A.M., Mottagui-Tabar, S., Liang, Z., Lenhard, B.,Wahlestedt, C. (2005) Antisense transcription in the mammalian transcriptome. *Science* **309**, 1564–1566.

Kaul, D. (2008) Coronin-1A epigenomics governs mycobacterial persistence in tuberculosis. *FEMS Microbiol. Lett.* **278**, 10–14.

Kim, J.B., Porreca, G.J., Song, L., Greenway, S.C., Gorham, J.M., Church, G.M., Seidman, C.E., Seidman, J.G. (2007) Polony multiplex analysis of gene expression (PMAGE) in mouse hypertrophic cardiomyopathy. *Science* **316**, 1481–1484.

Korkmaz, C.G., Korkmaz, K.S., Kurys, P., Elbi, C.,Wang, L., Klokk, T.I., Hammarstrom, C., Troen, G., Svindland, A., Hager, G.L., Saatcioglu, F. (2005) Molecular cloning and characterization of STAMP2, an androgen-regulated six transmembrane protein that is overexpressed in prostate cancer. *Oncogene* **24**, 4934–4945.

Lash, A.E., Tolstoshev, C.M., Wagner, L., Schuler, G.D., Strausberg, R.L., Riggins, G.J., Altschul, S.F. (2000)

SAGEmap: a public gene expression resource. *Genome Res.* **10**, 1051–1060.
Leclerc, P., Biarc, J., St-Onge, M., Gilbert, C., Dussault, A.A., Laflamme, C., Pouliot, M. (2008) Nucleobindin co-localizes and associates with cyclooxygenase (COX)-2 in human neutrophils. *PloS ONE* **3**, e2229.
Lee, S.W., Kang, Y.A., Yoon, Y.S., Um, S.W., Lee, S.M., Yoo, C.G., Kim, Y.W., Han, S.K., Shim, Y.S., Yim, J.J. (2006) The prevalence and evolution of anemia associated with tuberculosis. *J. Korean Med. Sci.* **21**, 1028–1032.
Lesley, R., Ramakrishnan, L. (2008) Insights into early mycobacterial pathogenesis from the zebrafish. *Curr. Opin. Microbiol.* **11**, 277–283.
Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., Ecker, J.R. (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* **133**, 523–536.
Liu,W., Liu, Y., Zhu, J.,Wright, E., Ding, I., Rodgers, G.P. (2008) Reduced hGC-1 protein expression is associated with malignant progression of colon carcinoma. *Clin. Cancer Res.* **14**, 1041–1049.
Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M., Gilad, Y. (2008) RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* **18**, 1509–1517.
Martin, M., Del Valle, J.M., Saborit, I., Engel, P. (2005) Identification of Grb2 as a novel binding partner of the signaling lymphocytic activation molecule-associated protein binding receptor CD229. *J. Immunol.* **174**, 5977–5986.
Meijer, A.H., Verbeek, F.J., Salas-Vidal, E., Corredor-Adamez, M., Bussman, J., van der Sar, A.M., Otto, G.W., Geisler, R., Spaink, H.P. (2005) Transcriptome profiling of adult zebrafish at the late stage of chronic tuberculosis due to *Mycobacterium marinum* infection. *Mol. Immunol.* **42**, 1185–1203.
Moldes, M., Lasnier, F., Gauthereau, X., Klein, C., Pairault, J., Feve, B., Chambaut- Guerin, A.M. (2001) Tumor necrosis factor-alpha-induced adipose-related protein (TIARP), a cell-surface protein that is highly induced by tumor necrosis factor-alpha and adipose conversion. *J. Biol. Chem.* **276**, 33938–33946.
Morin, R., Bainbridge, M., Fejes, A., Hirst, M., Krzywinski, M., Pugh, T., McDonald, H., Varhol, R., Jones, S., Marra, M. (2008) Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. *BioTechniques* **45**, 81–94.
Morozova, O., Marra, M.A. (2008) Applications of next-generation sequencing technologies in functional genomics. *Genomics* **92**, 255–264.
Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., Wold, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Method*s **5**, 621–628.
Nagalakshmi,U.,Wang, Z.,Waern, K., Shou, C.,Raha, D., Gerstein, M., Snyder,M. (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**, 1344–1349.
Pan, Q., Shai, O., Lee, L.J., Frey, B.J., Blencowe, B.J. (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **40**, 1413–1415.
Pedotti, P., t Hoen, P.A., Vreugdenhil, E., Schenk, G.J., Vossen, R.H., Ariyurek, Y., de Hollander, M., Kuiper, R., van Ommen, G.J., den Dunnen, J.T., Boer, J.M., de Menezes, R.X. (2008) Can subtle changes in gene expression be consistently detected with different microarray platforms? *BMC. Genomics* **9**, 124.
Pettersson, M., Bessonova, M., Gu, H.F., Groop, L.C., Jonsson, J.I. (2000) Characterization, chromosomal localization, and expression during hematopoietic differentiation of the gene encoding Arl6ip, ADP-ribosylation-like factor-6 interacting protein (ARL6). *Genomics* **68**, 351–354.
Pieters, J. (2008) *Mycobacterium tuberculosis* and the macrophage: maintaining a balance. *Cell Host Microbe* **3**, 399–407.
Recalcati, S., Invernizzi, P., Arosio, P., Cairo, G. (2008) Newfunctions for an iron storage protein: the role of ferritin in immunity and autoimmunity. *J. Autoimmun.* **30**, 84–89.
Rosenkranz, R., Borodina, T., Lehrach, H., Himmelbauer, H. (2008) Characterizing the mouse ES cell transcriptome with Illumina sequencing. *Genomics* **92**, 187–194.
Shendure, J. (2008) The beginning of the end for microarrays? *Nat. Methods* **5**, 585–587.
Stockhammer, O.W., Zakrzewska, A., Hegedus, Z., Spaink, H.P., Meijer, A.H. (2009) Transcriptome profiling and functional analyses of the zebrafish embryonic innate immune response to Salmonella infection. *J. Immunol.* **182**, 5641–5653.
Sultan, M., Schulz, M.H., Richard, H., Magen, A., Klingenhoff, A., Scherf, M., Seifert, M., Borodina, T., Soldatov, A., Parkhomchuk, D., Schmidt, D., O'Keeffe, S., Haas, S., Vingron, M., Lehrach, H., Yaspo, M.L. (2008) A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* **321**, 956–960.
Swaim, L.E., Connolly, L.E., Volkman, H.E., Humbert, O., Born, D.E., Ramakrishnan, L. (2006) *Mycobacterium marinum* infection of adult zebrafish causes caseating granulomatous tuberculosis and is moderated by adaptive immunity. *Infect. Immun.* **74**, 6108–6117.
t Hoen, P.A., Ariyurek, Y., Thygesen, H.H., Vreugdenhil, E., Vossen, R.H., de Menezes, R.X., Boer, J.M., van Ommen, G.J., den Dunnen, J.T. (2008) Deep sequencing-based expression analysis shows major advances in robustness, resolution and interlab portability over five microarray platforms. *Nucleic Acids Res.* **36**, e141.
van der Sar, A.M., Abdallah, A.M., Sparrius, M., Reinders, E., Vandenbroucke-Grauls, C.M., Bitter, W. (2004) *Mycobacterium marinum* strains can be divided into two distinct types based on genetic diversity and virulence. *Infect. Immun.* **72**, 6306–6312.
van der Sar, A.M., Spaink, H.P., Zakrzewska, A., Bitter,W., Meijer, A.H. (2009) Specificity of the zebrafish host tran-

scriptome response to acute and chronic mycobacterial infection and the role of innate and adaptive immune components. *Mol. Immunol.* **46**, 2317–2332.

Velculescu, V.E., Zhang, L., Vogelstein, B., Kinzler, K.W. (1995) Serial analysis of gene expression. *Science* **270**, 484–487.

Vos, J.B., Datson, N.A., Rabe, K.F., Hiemstra, P.S. (2007) Exploring host–pathogen interactions at the epithelial surface: application of transcriptomics in lung biology. *Am. J. Physiol. Lung Cell Mol. Physiol.* **292**, L367–L377.

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., Burge, C.B. (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476.

Wang, Z., Gerstein, M., Snyder,M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63.

Wellen, K.E., Fucho, R., Gregor, M.F., Furuhashi, M., Morgan, C., Lindstad, T., Vaillancourt, E., Gorgun, C.Z., Saatcioglu, F., Hotamisligil, G.S. (2007) Coordinated regulation of nutrient and inflammatory responses by STAMP2 is essential for metabolic homeostasis. *Cell* **129**, 537–548.

Wilhelm, B.T., Landry, J.R. (2009) RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. *Methods* **48**, 249–257.

Wilhelm, B.T., Marguerat, S., Watt, S., Schubert, F., Wood, V., Goodhead, I., Penkett, C.J., Rogers, J., Bahler, J. (2008) Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* **453**, 1239–1243.

Yelin, R., Dahary, D., Sorek, R., Levanon, E.Y., Goldstein, O., Shoshan,A., Diber,A., Biton, S., Tamir, Y., Khosravi, R., Nemzer, S., Pinner, E.,Walach, S., Bernstein, J., Savitsky, K., Rotman, G. (2003) Widespread occurrence of antisense transcription in the human genome. *Nat. Biotechnol.* **21**, 379–386.

Zhu, J., He, F., Wang, J., Yu, J. (2008) Modeling transcriptome based on transcript sampling data. *PLoS ONE* **3**, e1659.

Chapter 3

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

This chapter is based on:
Ordas A, Hegedus Z, Henkel C, Stockhammer OW, Butler D, Jansen HJ, Racz P, Mink M, Spaink HP, Meijer AH. Deep sequencing of the innate immune transcriptomic response of zebrafish embryos to Salmonella infection, in preparation

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. Comparison of tag-based sequencing (DGE or Tag-Seq) with full transcript sequencing (RNA-Seq) showed a strong correlation of sequence read counts per transcript and an overlap of 241 transcripts differentially expressed in response to infection. A slightly 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 anti-microbial 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 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.

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 (Traver et al., 2003; Trede et al., 2004). 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 (Davidson and Zon, 2004; Herbomel et al., 1999; Lam et al., 2004; Willett et al., 1999). 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 (Meijer et al., 2004; Stein et al., 2007) and several infection models for studying innate immune response mechanisms in zebrafish embryos have now been developed (Meeker and Trede, 2008). *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*) (Santos et al., 2001). The opportunity of real time analysis led to the development of a *S. typhimurium* infection model in zebrafish embryos (van der Sar et al., 2003). 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 (van der Sar et al., 2003). Components of the *S. typhimurium* cell wall and motility apparatus trigger innate host defense pathways, including Toll-like receptor (TLR) signaling (Salazar-Gonzalez and McSorley, 2005). A morpholino knockdown 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 (van der Sar et al., 2006). 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 (Stockhammer et al., 2009). A conserved role of zebrafish Toll-like receptor 5 (TLR5) homologs in recognition of Salmonella flagellin was demonstrated (Stockhammer et al., 2009). Furthermore, similar as mammals, zebrafish embryos were shown to employ both MyD88-dependent and MyD88-independent signaling pathways during infection (Stockhammer et al., 2009).

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 (Stockhammer et al., 2009). 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) (Hegedus et al., 2009; t Hoen et al., 2008) also named Tag-Seq (Morrissy et al., 2009). 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 full 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.

Results

Identification of Salmonella-responsive genes by tag sequencing

Previously we have performed a microarray analysis of the innate immune response of one day old zebrafish embryos to *Salmonella typhimurium* infection (Stockhammer et al., 2009). Over the first 8 hours of the infection we observed induction of an increasing number of specific gene groups including transcription factors, signaling molecules, and inflammatory mediators. Here we took the 8 hours post infection (hpi) time point for a deep sequencing analysis. Using the digital gene expression (DGE) procedure previously described (Hegedus et al., 2009), hereafter referred to as Tag-Seq (Morrissy et al., 2009), we obtained around 10 million sequence-specific tags from control and infected embryos each. A total of 2471 tag entities showed significantly different expression between the libraries from control and infected fish. Of all significant tags 66% (1630 tags) could be mapped to the UniGene database. The majority of these tags (95%) showed up-regulated expression during infection, while only 5% was down-regulated. Mapping of the significantly changed tags identified a total of 959 different UniGene transcripts when the tags mapping to multiple transcripts were excluded (and 2049 including the tags with multiple mapping). A total of 815 of these transcripts were identified only by tags mapping to the sense strand, 121 transcripts by tags mapping only to the antisense strands, and 23 transcripts collected significant tags mapping to both the sense and antisense strands. Gene ontology (GO) analysis of the up-regulated transcripts with tags mapping to the sense strand showed enrichment of the GO-term response to stimulus ($P<0.01$), consistent with earlier microarray results (Stockhammer et al., 2009).

Validation of Salmonella-responsive genes by comparative analysis of tag sequencing and microarray data

Approximately 60% of the UniGene transcripts that were identified by up- or down-regulated tags in the Tag-Seq analysis (sense strand mapping) were present on the custom Agilent microarray platform used in our previous study (Stockhammer et al., 2009). From those UniGene transcripts, 111 were changed in the same direction (108 up-regulated and 3 down-regulated), 1008 were not significantly changed in the microarray study, and 10 showed a change in the opposite direction between deep sequencing and microarray data (Fig. 1A). Conversely, in the microarray analysis 1254 genes in total were significantly changed (929 up-regulated and 325 down-regulated with fold change >1.5), of which 1133 were not significantly changed in the Tag-Seq analysis. The limited overlap between microarray and deep sequencing data can be explained by differences in methodology, experimental design (e.g. number of parallels), and data processing workflows that are based on different statistical methods to select significant hits. From a biological point of view, the statistical method used to accept expression changes in Tag-Seq analysis as significant (Lash et al., 2000) appeared to be more stringent, as it relied on fold changes of at least a fac-

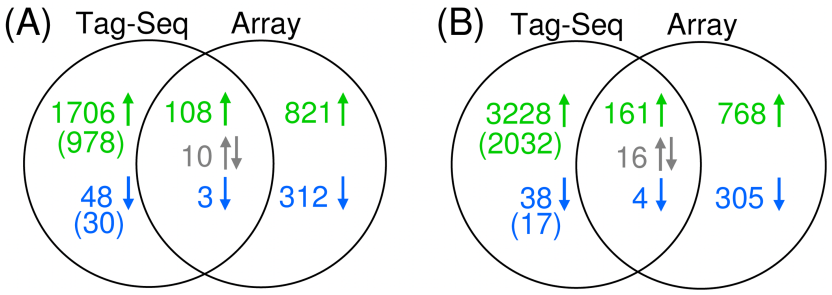


Figure 1. Comparison of Tag-Seq and microarray data of *Salmonella* infection of zebrafish embryos. The microarray data set of 1-day-old zebrafish embryos at 8 hours after infection with *Salmonella typhimurium* is taken from Stockhammer et al. (2009) and compared here with Tag-Seq data of the same samples. (A) Venn diagram showing the overlap of the microarray data set with the Tag-Seq dataset based on statistical evaluation according to Lash et al., (2000). (B) Venn diagram showing the overlap of the microarray data set with the Tag-Seq dataset based on sCTDI calculation. For the transcripts changed only in Tag-Seq analysis (left part of the Venn-diagrams) the numbers in brackets indicate presence on the microarray platform. In the overlapping sectors of the Venn-diagrams the numbers of up-regulated transcripts are indicated in green, numbers of down-regulated transcripts in blue, and numbers of transcripts changed in opposite direction in Tag-Seq and microarray analysis are indicated in grey.

tor 2 for the more abundant tags and over 5 fold change for the lower abundant tags. As an alternative to this statistical method we developed the Cumulative Transcript Detection Index (CTDI) as a means to evaluate differential expression in Tag-Seq datasets. As previously noted (Hegedus et al., 2009), in Tag-Seq analysis the majority of transcripts are represented by more than one tag in the sequence data. The CTDI value reflects the accumulated information from all tags that map to the same transcript, giving increased weight to tags with higher significance in statistical testing (Lash et al., 2000) and giving lower weight to those transcripts where tags are present that change in the opposite direction. We separately calculated the CTDI values for tags mapping to the sense strands (sCTDI) and antisense strands (asCTDI) of the UniGene database transcripts (Fig. 2) and used the sCTDI data for comparison with our microarray data (Fig. 1B). We found that 165 of the transcripts that were differentially expressed according to sCTDI calculation were changed in the same direction in our microarray analysis (161 up-regulated and 4 down-regulated; Supplementary Table 1). Therefore, the responsiveness of these transcripts to *Salmonella* infection in embryos has been confirmed by two independent transcriptome analysis methods.

Validation of Salmonella-responsive genes by comparative analysis of Tag-Seq and RNA-Seq data

Validation of the Tag-Seq data by comparison with microarray data was limited

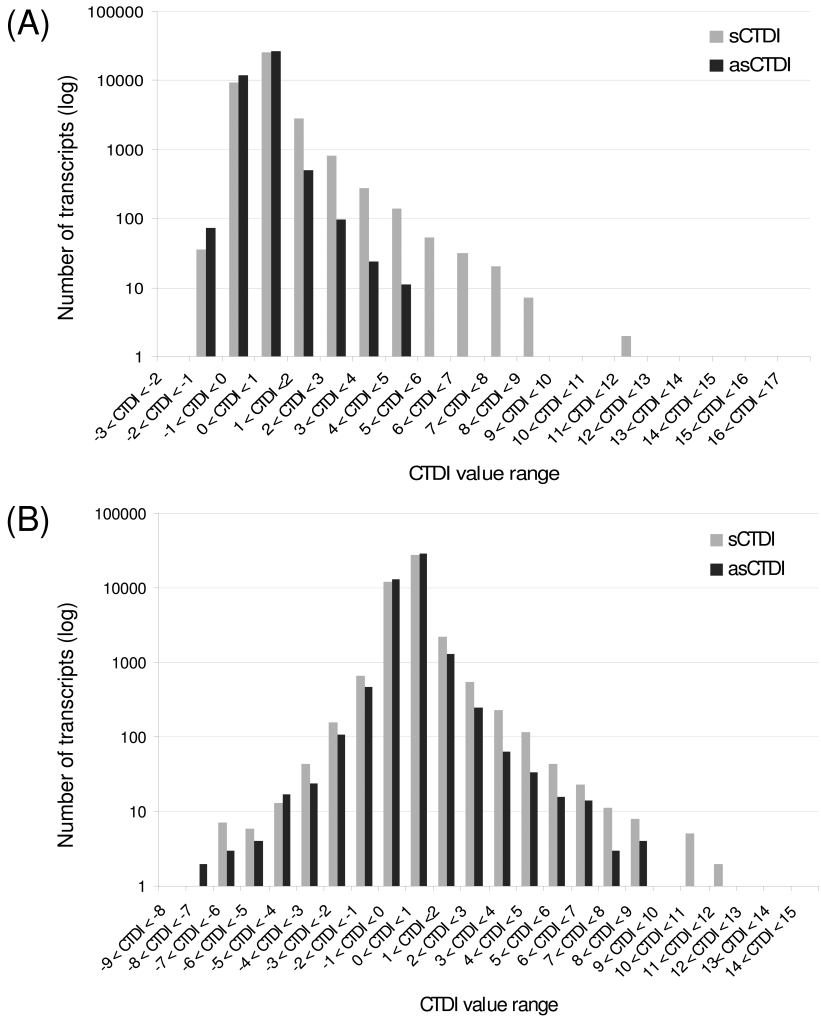


Figure 2. Evaluation of Tag-Seq data by Cumulative Transcript Detection Index (CTDI) calculation. The distribution of the number of transcripts over the range of CTDI values is plotted on a logarithmic scale. Data for tags mapping to the sense (sCTDI) and antisense (asCTDI) strands of the UniGene transcript database are plotted separately. Transcripts represented by at least one tag with a significant expression change based on statistical evaluation according to Lash et al. (2000) have an absolute CTDI value larger than 0.95. For details of CTDI calculation see the Materials and Methods section. (A) CTDI plot of *Salmonella* infection of zebrafish embryos. sCTDI values ranged between -2 and +15 and asCTDI values ranged between -2 and +7 and were in similar range when tags were mapped to the RefSeq (-4<sCDTI<11, -2<asCTDI<4) or Ensembl (-5<sCDTI<10, -3<asCTDI<4) databases. (B) CTDI plot of Tag-Seq data from *Mycobacterium* infection of adult zebrafish (Hegedus et al., 2009).

by the fact that only about one third of all transcripts in the UniGene database were represented on the microarray platform. To extend the validation of our Tag-Seq dataset we therefore performed a comparison with RNA-Seq, in which whole cDNA transcripts are fragmented and deep sequenced. For RNA-Seq analysis, control and *Salmonella*-infected embryo samples from an independent experiment were used. Approximately 15 million reads were obtained for both samples subjected to RNA-Seq. For comparison between RNA-Seq and Tag-Seq the reads were mapped to the Ensembl transcript database based on the Zv8 genome sequence. In both cases approximately half of the sequence reads could be mapped to Ensembl transcripts if no mismatches were allowed. When we allowed 1 mismatch for mapping of the longer RNA-Seq reads (51 nucleotides in RNA-Seq as compared to 17 nucleotides in Tag-Seq) the mapping efficiency could be increased to approximately two-thirds of the total reads. Mapping of RNA-Seq reads identified 85-86% of the known Ensembl transcripts, while 63-69% of transcripts were identified by mapping of Tag-Seq reads (Fig. 3A). Pearson correlations for RNA-Seq versus Tag-Seq libraries of control and infected embryos were 0.84 and 0.81 respectively, which is relatively high especially considering that RNA-Seq and Tag-Seq libraries were constructed from different biological samples (Fig. 3B). As can be observed in the correlation plots, the agreement between RNA-Seq and Tag-Seq data was better for the moderate to highly expressed transcripts than for the low abundant transcripts (Fig. 3B). Statistical evaluation showed that 1244 transcripts were differentially expressed in Tag-Seq analysis and 976 transcripts in RNA-Seq analysis. In both cases more transcripts were up-regulated than down-regulated (Fig. 3C, 3D). However, detection of down-regulated transcripts appeared more efficient with RNA-Seq, where 17% of all differentially expressed transcript were down-regulated, than with Tag-Seq, where the down-regulated transcripts comprised only 3% of all differentially expressed transcripts. Gene ontology (GO) analysis of the up-regulated transcripts of both Tag-Seq and RNA-Seq analysis showed enrichment of the GO-term response to stimulus ($P<0.05$). In total 160 up-regulated and 7 down-regulated transcripts were overlapping between Tag-Seq and RNA-Seq analysis, while 23 transcripts showed changes in opposite directions between the two methods (up in Tag-Seq and down in RNA-Seq) (Fig. 3D). When the Tag-Seq data were evaluated using the above described CTDI algorithm for cumulative detection of tags mapping to the same transcript, the overlap between Tag-Seq and RNA-Seq increased to 233 up-regulated and 8 down-regulated transcripts (Fig. 3E, Supplementary Table 1). In conclusion, Tag-Seq and RNA-Seq were similarly suitable to derive quantitative information on infection-responsive gene expression and the overlap of differentially expressed transcripts between these two deep sequencing methods was slightly better than the overlap between Tag-Seq and microarray analysis (compare Fig.1A,B, Fig.3D,E) .

Annotation of the validated Salmonella-responsive gene set

After collapsing UniGene and Ensembl transcript IDs to single genes, the total set

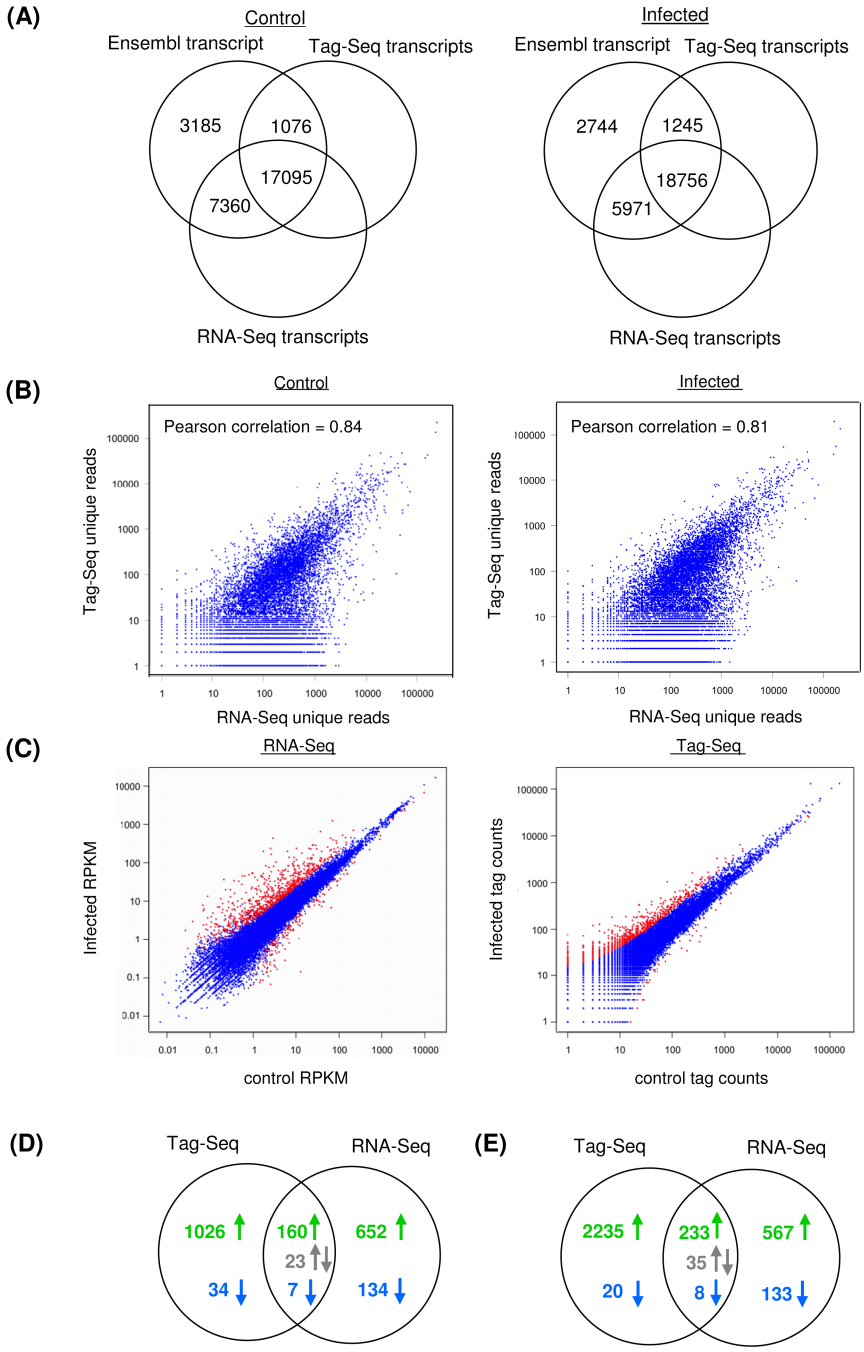

(A)
Control
Ensembl transcript
Tag-Seq transcripts
3185
1076
17095
7360
RNA-Seq transcripts
Infected
Ensembl transcript
Tag-Seq transcripts
2744
1245
18756
5971
RNA-Seq transcripts
(B)
Control
Pearson correlation = 0.84
Tag-Seq unique reads
RNA-Seq unique reads
Infected
Pearson correlation = 0.81
Tag-Seq unique reads
RNA-Seq unique reads
(C)
RNA-Seq
Infected RPKM
control RPKM
Tag-Seq
Infected tag counts
control tag counts
(D)
Tag-Seq
RNA-Seq
1026
160
652
23
34
7
134
(E)
Tag-Seq
RNA-Seq
2235
233
567
35
20
8
133

Figure 3. RNA-Seq analysis of *Salmonella* infection of zebrafish embryos. (A) Efficiency of transcript detection by Tag-Seq and RNA-Seq. The Venn-diagrams show the overlap between all Ensembl transcripts and the transcript identified by mapping of reads from Tag-Seq or RNA-Seq analysis of control and infected embryos. One mismatch was allowed for mapping of the 51 nucleotides long RNA-Seq reads and no mismatches were allowed for mapping of the 17 nucleotides long tag entities in Tag-Seq. (B) Correlation between Tag-Seq and RNA-Seq data. Scatter plots show the counts per Ensembl transcript for both methods. Only reads mapping to a single transcript were included in the analysis. (C) Tag-Seq and RNA-Seq detection of differential expression between control and infected embryos. Scatter plots show the counts per Ensembl transcript (RNA-Seq) or counts per tag (Tag-Seq) for libraries of control versus infected embryos. Transcripts (RNA-Seq) or tags (Tag-Seq) with significantly different expression between control and infected libraries are indicated in red. In RNA-Seq the reads were first mapped to Ensembl transcripts and subsequently differential expression was statistically evaluated according to Lash et al. (2000). The RNA-Seq scatter plot shows RPKM values, which are the total read counts per kilobase per million mapped reads (Mortazavi et al., 2008). In Tag-Seq differential expression was statistically evaluated at the level of the individual tags. (D) Venn diagram showing the overlap of the Tag-Seq and RNA-Seq data sets based on statistical evaluation according to Lash et al. (2000). (E) Venn diagram showing the overlap of the Tag-Seq and RNA-Seq data sets, based on sCTDI calculation for the Tag-Seq data. Tag-Seq reads mapping to the antisense strands of Ensembl transcripts were excluded from all analyses. In RNA-Seq, information on transcript directionality is not obtained and therefore all read mapping data are included in the analyses.

of *Salmonella*-responsive genes confirmed by two transcriptome analysis methods consisted of 317 genes, of which 163 were overlapping between Tag-Seq and microarray analysis, 230 between Tag-Seq and RNA-Seq, and 76 by all three methods of gene expression profiling. This set of 317 genes (309 up-regulated and 8 down-regulated) was taken for a detailed annotation (Supplementary Table 1, Fig. 4). We categorized the differentially expressed genes into 5 categories: 1A - annotated genes previously implicated in the vertebrate immune response (82 up + 4 down), 1B - novel/hypothetical genes with similarity to genes previously implicated in the vertebrate immune response (31 up), 2A - annotated genes not previously linked to the immune response (100 up + 2 down), 2B - novel/hypothetical genes with similarity to genes not previously linked to the immune response (50 up + 2 down), and 3 - genes with unknown function and for which we could not derive any functional prediction (46 up). The genes in the 1A/B and 2A/B categories were ordered by (predicted) functions based on literature and gene ontology data (Supplementary Table 1, Fig. 4). The annotated or novel/hypothetical genes with homology to human immune-related genes (categories 1A/B) together comprised 22 infection-induced genes with functions in transcription activation or repression, including transcription factors of the ATF, AP-1(JUN/FOS), CEBP, ETS, IRF, MYB, MYC, NFκB, and STAT families. Categories 1A/B also contained 21 up-regulated genes involved in immune-related signal transduction pathways, including MAP kinase (MAPK), ERBB2, interleukin

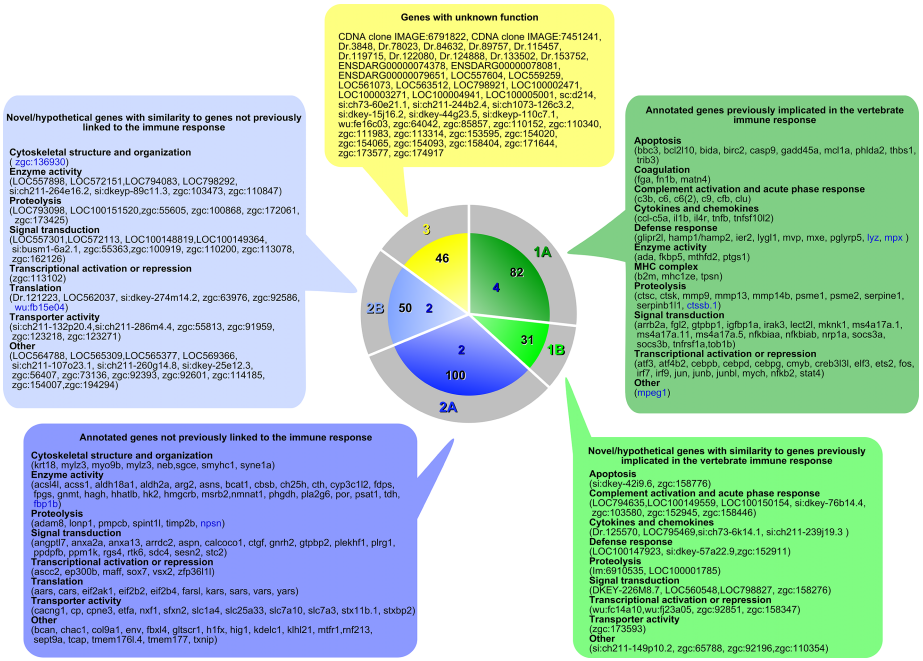


Figure 4. Annotation of the gene set responsive to *Salmonella* infection in zebrafish embryos. Differential expression of genes in the diagram was confirmed by Tag-Seq and microarray analysis, or by Tag-Seq and RNA-Seq analysis, or by all three transcriptome profiling methods as indicated in Supplementary Table 1. Genes were grouped into five categories: category 1A - annotated genes previously implicated in the vertebrate immune response based on GO annotations of the zebrafish genes and their human homologues, on PubMed abstracts and on overlap with the common host response defined by Jenner and Young (2005), 1B - novel/hypothetical genes with similarity to genes previously implicated in the vertebrate immune response, 2A - annotated genes not previously linked to the immune response, 2B - novel/hypothetical genes with similarity to genes not previously linked to the immune response, 3 - genes with unknown function. The number of genes in each category is indicated in the pie diagram and the corresponding genes are ordered by functional groups. Genes up-regulated by *Salmonella* infection (309 genes) are indicated in black, and genes down-regulated (8 genes) are indicated in blue. Gene descriptions and accession numbers are given in Supplementary Table 1.

1 receptor (IL1R) and Toll-like receptor (TLR) signaling. The genes involved in TLR signaling included several negative regulators of the pathway, including *irak3*, *socs3a* and *socs3b*, and the NFκB inhibitor genes *nfkbiaa* and *nfkbiab*, as previously noted in Stockhammer et al. (2009). The signal transduction group also contained 4 members of the membrane-spanning 4-domains subfamily. The up-regulated genes in categories 1A/B further comprised 9 genes with (predicted) cytokine or chemokine activity, 13 genes related to complement activation and the acute phase response, 12 genes

involved in apoptosis and 11 genes with proteolytic functions, including proteasome activator subunit genes, cathepsins, serpins and matrix metalloproteinases (*mmp9*, *mmp13* and *mmp14b*). Furthermore, categories 1A/B contained smaller groups of up-regulated genes encoding MHC complex proteins, coagulation factors, enzymes (e.g. the prostaglandin biosynthetic gene *ptgs1*) and several genes that we classified under defense response, such as the antimicrobial hepcidin (*hamp1/2*) gene, the peptidoglycan recognition (*plyrp5*) gene, the immediate early response gene, *ier2*. The categories 1A/B contained only four down-regulated genes of which two, *mpx* (myeloperoxidase) and *lyz* (lysozyme C), are also clearly linked to the defense response. The down-regulation of these genes might suggest a Salmonella-specific mechanism to counteract host defense. The down-regulated genes further included a cathepsin gene (*ctssb.1*) and the macrophage expressed gene 1 (*mpeg1*), which encodes a perforin-like protein whose precise function is unknown. In contrast, two other cathepsins (*ctsc*, *ctsk*) and an *mpeg1*-like gene (*zgc:110354*) were present in the up-regulated gene set. The 150 up-regulated and 4 down-regulated genes of the 2A/B categories that have not previously been directly linked to immune response were associated to GO terms such as cytoskeletal structure and organization, enzyme activity (including several metabolic genes), translation, and transporter activity (including several solute carriers). The 2A/B categories also contained up-regulated genes associated with proteolysis, signal transduction and transcription factor or cofactor functions, which were not previously linked specifically to host defense.

Comparison of different infection studies

Previously we have used Tag-Seq analysis to investigate the host response of adult zebrafish at the end stage of *Mycobacterium marinum* infection ((Hegedus et al., 2009); chapter 2). The end stage of *Mycobacterium* infection in adult fish is associated with a strong inflammatory response, similar to what we observed during *Salmonella* infection of zebrafish embryos. Therefore, we decided to compare the specific gene groups regulated in both infection studies. To this extent we re-examined our previous Tag-Seq data set from *Mycobacterium* infection by CTDI calculation (Fig. 2B). The CTDI values for tags mapping to the sense strands (sCTDI) of the UniGene database transcripts ranged between –7 and +12, whereas these values ranged between -2 and +15 in the *Salmonella* Tag-Seq data set, indicating that more transcripts were down-regulated in the *Mycobacterium* infection study. This is consistent with our previous analyses that showed a large number of genes encoding metabolic enzymes and muscle proteins to be down-regulated during *Mycobacterium* infection when the fish become strongly emaciated (Hegedus et al., 2009; Meijer et al., 2005). Next, we determined the overlap between the up- and down-regulated transcripts of the *Salmonella* and *Mycobacterium* infection studies (Fig. 5A, Supplementary Table 2). We found 228 and 3 commonly up- or down-regulated transcripts that corresponded to 206 and 2 up- or down-regulated genes. These genes were functionally annotated and grouped into the 5 categories described above, showing that the com-

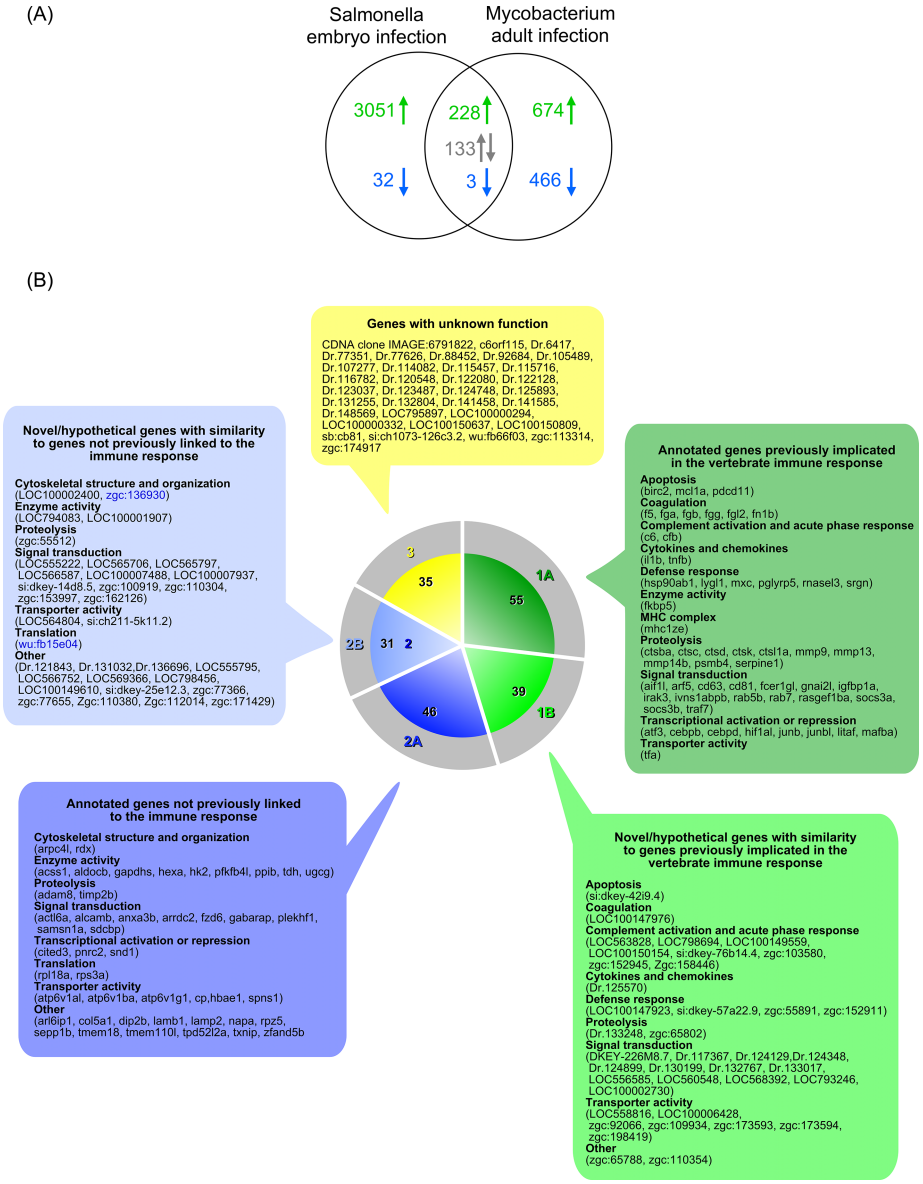


Figure 5. Comparison of infection responses in adult and embryonic zebrafish. The Tag-Seq data of *Salmonella* infection of zebrafish embryos reported here were compared with previously published DGE data of the end stage of *Mycobacterium* infection in adult zebrafish (Hegedus et al., 2009). (A) Venn diagram showing the overlap between DGE data of the different infection studies based on sCTDI calculation. In the overlapping sector of the Venn-diagram the numbers of up-regulated transcripts are indicated in green, numbers of down-regulated transcripts in blue, and the numbers of transcripts changed in opposite direction in the two different infection studies is

indicated in grey. (B) Annotation of the overlap group of up-regulated genes in the different infection studies. Genes were grouped into 5 categories as in Figure 4. The number of genes in each category is indicated in the pie diagram and the corresponding genes are ordered by functional groups. Genes up-regulated by *Salmonella* and *Mycobacterium* infection (206 genes) are indicated in black, and genes down-regulated by both infections (2 genes) are indicated in blue. Gene descriptions and accession numbers are given in Supplementary Table 2.

monly up-regulated genes included 94 annotated or novel/hypothetical genes with homology to human immune-related genes (categories 1A/B), 77 annotated or novel/hypothetical genes that have not been directly implicated in host defense in previous studies (categories 2A/B), and 35 genes for which we could make no functional prediction (category 3). Only two genes were identified as commonly down-regulated between the different infection studies, encoding a predicted intermediate filament protein and a ribosomal protein. Commonly up-regulated genes with previous links to host defense (categories 1A/B) were mainly associated with apoptosis, coagulation, complement activation and acute phase response, cytokine and chemokine activity, defense response, proteolysis, signal transduction, transcriptional activation and repression, and transporter activity. Commonly up-regulated genes in categories 2A/B were mainly associated with cytoskeletal structure and organization, enzyme activity, proteolysis, signal transduction, transcriptional activation and repression, translation and transporter activity.

Discussion

In this study we used the *S. typhimurium*-zebrafish infection model for a deep sequencing analysis of the embryonic innate host response to an inflammatory infection. Using both tag-based (Tag-Seq) and full transcript (RNA-Seq) sequencing approaches we extended and validated previous microarray data of this infection model (Stockhammer et al., 2009). The combined sequencing-based and microarray-based transcriptome data resulted in an annotated reference set of *Salmonella*-responsive genes in zebrafish embryos, including those homologous to human immune-related genes as well as many known or novel genes not previously linked to the immune response. Furthermore, comparison of the deep sequencing data of *Salmonella* infection in zebrafish embryos with previous deep sequencing data of *Mycobacterium* infection in adult zebrafish (Hegedus et al., 2009), defined a common set of innate host defense genes that are expressed both in the absence and presence of a fully developed adaptive immune system.

For deep sequencing analysis we chose *S. typhimurium* infected 1-day-old zebrafish embryos at 8 hpi, which time point was based on the strong induction of inflammatory genes that we detected in our previous microarray analysis (Stockhammer et al., 2009). Furthermore, we previously showed that induction of inflammatory genes such as *il1b* and *mmp9* relied on MyD88-dependent signaling

at this time point (Stockhammer et al., 2009). Consistent with the microarray results, more up-regulated than down-regulated transcripts were detected by Tag-Seq and RNA-Seq deep sequencing. However, for reasons currently unknown, the percentage of down-regulated transcripts was lower in Tag-Seq (3%) than in RNA-Seq (17%) and microarray analysis (26 %). Gene ontology analysis of the up-regulated gene sets of all three transcriptome analysis methods showed enrichment of the GO-term response to stimulus. Although less than 20% of the genes previously found to be up-regulated in microarray analysis could be confirmed by deep sequencing, a commonly responsive set of 165 transcripts could be defined encoding transcription factors, signal transduction proteins, cytokines and chemokines, complement factors, proteins involved in apoptosis and proteolysis, and proteins with anti-microbial activities. The overlap between Tag-Seq and RNA-Seq was slightly higher (241 transcripts) and confirmed the differential expression of these gene groups. The total overlap between the different transcriptome profiling methods might seem limited. As previously discussed (Hegedus et al., 2009; t Hoen et al., 2008), this can be attributed not only to technical differences, but also to differences in data processing and statistical evaluation, or to differences between the methods in discriminating between expression of different transcript isoforms. Importantly, the Tag-Seq and RNA-Seq data sets contained many differentially expressed transcripts that were not included in the microarray platform, and these data sets are therefore highly useful to extend and improve the microarray design.

In this study we only analyzed sequence reads mapping to UniGene, RefSeq or Ensembl transcript databases. However, about one third of the significant Tag-Seq reads did not map to transcript databases. Similarly, transcript mapping failed for one third of the RNA-Seq reads, even when one mismatch was allowed. While polymorphisms and reads extending over intron boundaries may at least partly account for mapping failures, these observations suggests that a large set of infection-responsive genes is still unknown, which can be of great interest for further studies.

We have previously shown that Tag-Seq data can be used to detect selective induction or repression of different transcript isoforms generated by alternative splicing, alternative polyadenylation or alternative transcription initiation (Hegedus et al., 2009). Such events may be detected when different tags for the same transcript show significant changes in opposite directions. In the present study we did not exploit this aspect of the deep sequencing technique, since our main objective was to define a robust marker set of inflammatory genes as a reference for infection studies in the zebrafish embryo model. For this reason, we focused specifically on those transcripts whose corresponding tags consistently changed in the same direction and developed the Cumulative Transcript Detection Index (CTDI). This index reflects the accumulated information from all tags that map to the same transcript, giving increased weight to tags with higher significance in statistical testing. The CTDI calculation proved useful for the comparison of Tag-Seq with microarray and RNA-Seq and increased the overlap between the data sets by approximately 30% in both cases as

compared to the use of the conventional Bayesian statistical evaluation method for the analysis of tag sequence data (Lash et al., 2000).

Previous studies (Mortazavi et al., 2008; t Hoen et al., 2008) have estimated that RNA-Seq analysis requires in the order of 10-fold more sequence reads for accurate quantification of expression differences between samples. Here, for the first time, a comparative Tag-Seq and RNA-Seq analysis of the host response to infection was performed. We found that 10 million Tag-Seq reads and 15 million RNA-Seq reads were both sufficient to detect around 1000 differentially expressed Ensembl transcripts during infection based on Bayesian statistical evaluation (Lash et al., 2000). Furthermore, Pearson correlation coefficients (>0.8) showed a linear relationship between the sequence read counts per Ensembl transcript in Tag-Seq and RNA-Seq, indicating the comparable performance of both methods in quantifying the transcriptome response to infection. Clearly RNA-Seq data have proved superior to Tag-Seq data for unraveling transcriptional landscapes (Wang et al., 2009). An advantage of Tag-Seq as compared to the RNA-Seq method used here, is that it allows discrimination between the expression of sense and antisense transcripts, which is of interest in view of the increasing evidence for the widespread occurrence and biological relevance of antisense transcription (Beiter et al., 2009; Carninci et al., 2005; Katayama et al., 2005). A substantial proportion of the differentially expressed sequence tags (10%) mapped to the antisense strands of known or predicted transcripts; however, this was much lower than previously observed in Tag-Seq analysis of *M. marinum* infection in adult zebrafish (40%; Hegedus et al., 2009).

Similar to *S. typhimurium* infection of zebrafish embryos, the end stage of *M. marinum* infection of adult zebrafish is also associated with a strong inflammatory response (Meijer et al., 2005; van der Sar et al., 2009). The similarity between these inflammatory responses at the level of gene expression was demonstrated here by an overlap of 206 up-regulated genes between the Tag-Seq data of *S. typhimurium* infection and the previously reported Tag-Seq data of *M. marinum* infection (Hegedus et al., 2009). This common set of infection-responsive genes included transcription factors and signaling components involved in the innate host defense, as well as genes not previously linked to the immune response of interest for further study in zebrafish models. The transcriptome data of both infection models provide a valuable reference for future studies of host-pathogen interactions in zebrafish.

Materials and methods

DGE (Tag-Seq) library construction and sequencing

The RNA samples for DGE analysis were identical to those used in Stockhammer et al., 2009. In brief, zebrafish embryos were infected with *Salmonella typhimurium* (strain SL1027) by microinjection of DsRED-labeled bacteria into the caudal vein close to the urogenital opening after the onset of blood circulation (27 hpf). An equal volume of PBS was injected in the control group. RNA samples were collected at 8

hours 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 (Hegedus et al., 2009). 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*.

RNA-Seq library construction and sequencing

Samples used for full RNA sequencing were the infected and uninfected control groups from a morpholino knock-down study to be reported elsewhere (Stockhammer et al., unpublished results). The procedure of *S. typhimurium* infection and the time-point of analysis (8 hpi) were identical as for the DGE analysis and previous microarray study (Stockhammer et al., 2009). 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).

DGE (Tag-Seq) data analysis

Mapping of tag sequences to transcript databases or to the zebrafish genome was performed as previously described (Hegedus et al., 2009). 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 (Stockhammer et al., 2009). 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 (Lash et al., 2000). 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 datasets we developed the Cumulative Transcript Detection Index (CTDI), which accumulates data from all tags that map to the same transcript:

$$\sum_{i=0}^{n} D_i P_i^2 \left| \frac{\sum_{i=0}^{n} D_i P_i^2}{\sum_{i=0}^{n} \left| D_i P_i^2 \right|} \right|$$

where: *n* is the number of the detected tag entities in a transcript
P is the significance of tags (Lash et al., 2000)
D is the coefficient for the direction of change (1 for increase or -1 for de crease).

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 opposite direction. In other words, the CTDI measure reflects the extent of experimental confidence regarding the direction of the transcript expression change after infection. CTDI values were calculated separately for tags mapping to the sense strand (sCTDI) and the antisense strands (asCTDI) of transcripts in the database. All transcripts represented by minimally one tag with a significant expression change based on statistical evaluation according to Lash et al. (2000) have an absolute CTDI value larger than 0.95. For comparison with microarray data we considered all transcripts with an absolute sCTDI value larger than 0.95 and *P* larger than 0.8 for at least one of the tags in the CTDI calculation.

RNA-Seq data analysis

Sequence reads were mapped to Ensembl transcripts (Zv8.55) using the CLCbio Genomics Workbench version 3.6.5 (www.clcbio.com). RPKM values (read counts corrected for library size and transcript length) were calculated according to Mortazavi et al. (2008). Differential expression between control and infected samples was calculated based on total read counts per transcript using the method of Lash et al. (2000) as described above under Tag-Seq analysis.

Gene annotation

Gene Ontology (GO) analysis of significantly expressed *Salmonella*-responsive genes was performed using the GeneTools *e*GOn v2.0 web-based gene ontology analysis software (www.genetools.microarray.ntnu.no) at the level of the UniGene clusters (*D. rerio* UniGene build no. 105) (Beisvag et al., 2006). Overlapping genes

of Tag-Seq, microarray and RNA-Seq datasets were functionally annotated using the National Center for Biotechnology Information (NCBI) Gene Ontology Annotation (GOA), Entrez Gene, and HomoloGene databases and the AmiGO (http://amigo.geneontology.org/cgi-bin/amigo/go.cgi) gene ontology database. Human homologues of zebrafish genes were identified by Ensembl's BioMart data mining tool (http://www.ensembl.org/biomart/martview/) and using the NCBI HomoloGene database. The previous implication of genes in the vertebrate immune response was based on the GO data of the zebrafish genes and their human homologues supplemented by PubMed abstract searches and on overlap with the common host response gene list defined by Jenner and Young (Jenner and Young, 2005).

Supplementary data

Supplementary data associated with this chapter can be found at http://apo.szbk.u-szeged.hu/transfer/A_ORDAS/.

References

Beisvag V., Junge F. K., Bergum H., Jolsum L., Lydersen S., Gunther C. C., Ramampiaro H., Langaas M., Sandvik A. K. and Laegreid A. (2006) GeneTools--application for functional annotation and statistical hypothesis testing. *BMC Bioinformatics* **7**, 470.

Beiter T., Reich E., Williams R. W. and Simon P. (2009) Antisense transcription: a critical look in both directions. *Cell Mol Life Sci* **66**, 94-112.

Carninci P., Kasukawa T., Katayama S., Gough J., Frith M. C., Maeda N., Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K., Bajic V. B., Brenner S. E., Batalov S., Forrest A. R., Zavolan M., Davis M. J., Wilming L. G., Aidinis V., Allen J. E., Ambesi-Impiombato A., Apweiler R., Aturaliya R. N., Bailey T. L., Bansal M., Baxter L., Beisel K. W., Bersano T., Bono H., Chalk A. M., Chiu K. P., Choudhary V., Christoffels A., Clutterbuck D. R., Crowe M. L., Dalla E., Dalrymple B. P., de Bono B., Della Gatta G., di Bernardo D., Down T., Engstrom P., Fagiolini M., Faulkner G., Fletcher C. F., Fukushima T., Furuno M., Futaki S., Gariboldi M., Georgii-Hemming P., Gingeras T. R., Gojobori T., Green R. E., Gustincich S., Harbers M., Hayashi Y., Hensch T. K., Hirokawa N., Hill D., Huminiecki L., Iacono M., Ikeo K., Iwama A., Ishikawa T., Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H., Kitano H., Kollias G., Krishnan S. P., Kruger A., Kummerfeld S. K., Kurochkin I. V., Lareau L. F., Lazarevic D., Lipovich L., Liu J., Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L., Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K., Mottagui-Tabar S., Mulder N., Nakano N., Nakauchi H., Ng P., Nilsson R., Nishiguchi S., Nishikawa S., et al. (2005) The transcriptional landscape of the mammalian genome. *Science* **309**, 1559-63.

Davidson A. J. and Zon L. I. (2004) The 'definitive' (and 'primitive') guide to zebrafish hematopoiesis. *Oncogene* **23**, 7233-46.

Hegedus Z., Zakrzewska A., Agoston V. C., Ordas A., Racz P., Mink M., Spaink H. P. and Meijer A. H. (2009) Deep sequencing of the zebrafish transcriptome response to mycobacterium infection. *Mol Immunol* **46**, 2918-30.

Herbomel P., Thisse B. and Thisse C. (1999) Ontogeny and behaviour of early macrophages in the zebrafish embryo. *Development* **126**, 3735-45.

Jenner R. G. and Young R. A. (2005) Insights into host responses against pathogens from transcriptional profiling. *Nat Rev Microbiol* **3**, 281-94.

Katayama S., Tomaru Y., Kasukawa T., Waki K., Nakanishi M., Nakamura M., Nishida H., Yap C. C., Suzuki M., Kawai J., Suzuki H., Carninci P., Hayashizaki Y., Wells C., Frith M., Ravasi T., Pang K. C., Hallinan J., Mattick J., Hume D. A., Lipovich L., Batalov S., Engstrom P. G., Mizuno Y., Faghihi M. A., Sandelin A., Chalk A. M., Mottagui-Tabar S., Liang Z., Lenhard B. and Wahlestedt C. (2005) Antisense transcription in the mammalian transcriptome. *Science* **309**, 1564-6.

Lam S. H., Chua H. L., Gong Z., Lam T. J. and Sin Y. M. (2004) Development and maturation of the immune system in zebrafish, Danio rerio: a gene expression profiling, in situ hybridization and immunological study. *Dev Comp Immunol* **28**, 9-28.

Lash A. E., Tolstoshev C. M., Wagner L., Schuler G. D., Strausberg R. L., Riggins G. J. and Altschul S. F. (2000) SAGEmap: a public gene expression resource. *Genome Res* **10**, 1051-60.

Meeker N. D. and Trede N. S. (2008) Immunology and zebrafish: spawning new models of human disease. *Dev Comp Immunol* **32**, 745-57.

Meijer A. H., Gabby Krens S. F., Medina Rodriguez I. A., He S., Bitter W., Ewa Snaar-Jagalska B. and Spaink H. P. (2004) Expression analysis of the Toll-like receptor and TIR domain adaptor families of zebrafish. *Mol Immunol* **40**, 773-83.

Meijer A. H., Verbeek F. J., Salas-Vidal E., Corredor-Adamez M., Bussman J., van der Sar A. M., Otto G. W., Geisler R. and Spaink H. P. (2005) Transcriptome profiling of adult zebrafish at the late stage of chronic tuberculosis due to Mycobacterium marinum infection. *Mol Immunol* **42**, 1185-203.

Morrissy A. S., Morin R. D., Delaney A., Zeng T., McDonald H., Jones S., Zhao Y., Hirst M. and Marra M. A. (2009) Next-generation tag sequencing for cancer gene expression profiling. *Genome Res* **19**, 1825-35.

Mortazavi A., Williams B. A., McCue K., Schaeffer L. and Wold B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5**, 621-8.

Salazar-Gonzalez R. M. and McSorley S. J. (2005) Salmonella flagellin, a microbial target of the innate and adaptive immune system. *Immunol Lett* **101**, 117-22.

Santos R. L., Zhang S., Tsolis R. M., Kingsley R. A., Adams L. G. and Baumler A. J. (2001) Animal models of Salmonella infections: enteritis versus typhoid fever. *Microbes Infect* **3**, 1335-44.

Stein C., Caccamo M., Laird G. and Leptin M. (2007) Conservation and divergence of gene families encoding components of innate immune response systems in zebrafish. *Genome Biol* **8**, R251.

Stockhammer O. W., Zakrzewska A., Hegedus Z., Spaink H. P. and Meijer A. H. (2009) Transcriptome profiling and functional analyses of the zebrafish embryonic innate immune response to Salmonella infection. *J Immunol* **182**, 5641-53.

t Hoen P. A., Ariyurek Y., Thygesen H. H., Vreugdenhil E., Vossen R. H., de Menezes R. X., Boer J. M., van Ommen G. J. and den Dunnen J. T. (2008) Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms. *Nucleic Acids Res* **36**, e141.

Traver D., Herbomel P., Patton E. E., Murphey R. D., Yoder J. A., Litman G. W., Catic A., Amemiya C. T., Zon L. I. and Trede N. S. (2003) The zebrafish as a model organism to study development of the immune system. *Adv Immunol* **81**, 253-330.

Trede N. S., Langenau D. M., Traver D., Look A. T. and Zon L. I. (2004) The use of zebrafish to understand immunity. *Immunity* **20**, 367-79.

van der Sar A. M., Musters R. J., van Eeden F. J., Appelmelk B. J., Vandenbroucke-Grauls C. M. and Bitter W. (2003) Zebrafish embryos as a model host for the real time analysis of Salmonella typhimurium infections. *Cell Microbiol* **5**, 601-11.

van der Sar A. M., Spaink H. P., Zakrzewska A., Bitter W. and Meijer A. H. (2009) Specificity of the zebrafish host transcriptome response to acute and chronic mycobacterial infection and the role of innate and adaptive immune components. *Mol Immunol* **46**, 2317-32.

van der Sar A. M., Stockhammer O. W., van der Laan C., Spaink H. P., Bitter W. and Meijer A. H. (2006) MyD88 innate immune function in a zebrafish embryo infection model. *Infect Immun* **74**, 2436-41.

Wang Z., Gerstein M. and Snyder M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**, 57-63.

Willett C. E., Cortes A., Zuasti A. and Zapata A. G. (1999) Early hematopoiesis and developing lymphoid organs in the zebrafish. *Dev Dyn* **214**, 323-36.

Chapter 4

MicroRNA expression during bacterial infections in zebrafish

This chapter is based on:
Ordas A, Zakrzewska A, Stockhammer OW, Carvalho R, van der Sar AM, Zhang Y, Verbeek FJ, Mink M, Spaink HP, Meijer AH.
MicroRNA expression during bacterial infections in zebrafish, in preparation

Abstract

MicroRNAs (miRNAs) are a class of small non-coding RNAs that play important roles in controlling mRNA stability and translation. Several miRNAs have been implicated in development and function of the immune system and their aberrant expression has been linked to pathogenesis of human diseases. Recently, zebrafish infection models have been developed to study the vertebrate immune system. Signaling pathways of the immune response were found to be conserved between human and zebrafish, but the interaction between miRNAs and the zebrafish immune system has not yet been addressed. Here, we used a custom-designed microarray platform to evaluate miRNA expression in zebrafish infection models. By infection studies with two pathogens, *Mycobacterium marinum* and *Salmonella typhimurium*, modeling human tuberculosis and salmonellosis, respectively, we identified a set of common infection-responsive miRNAs, including members of the miR-21, miR-146 and miR-181 families, previously linked to immunity and cancer in human and mammalian models. The context of innate immunity was sufficient for the infection response of these miRNAs, as their differential expression occurred not only in adult fish but also in embryos prior to development of adaptive immunity. We also determined that predicted target genes of the strongly infection-inducible miR-146 family are conserved between human and zebrafish. These include key signaling intermediates of the TLR pathway that is pivotal for pathogen recognition and activation of the innate immune response, suggesting a miR-146-mediated feed back mechanism to properly control the inflammatory response. Together, our results uncovered overlap in human and zebrafish immune-related miRNAs as well as miRNA target genes, indicating the usefulness of zebrafish infection models to increase our knowledge of miRNA functions in the vertebrate immune system.

Introduction

The discovery of microRNAs (miRNA) has added a new level of complexity to how biological processes are controlled. MicroRNAs are small non-coding single stranded RNAs with a length of 19-25/21-25 nucleotides (nt) that control gene expression by regulating mRNA stability and translation in a unique tissue-specific, developmental stage-specific and disease-specific manner (Schmidt et al., 2009). They are involved in the regulation of variety of biological processes, including cell cycle, differentiation, development, metabolism and disease pathogenesis (Sonkoly and Pivarcsi, 2009). MiRNA genes are evolutionarily conserved and located in introns and exons of protein-coding genes and non-coding regions of the genome. They are transcribed mostly by RNA polymerase II into several kilobases long, double-stranded primary transcripts (pri-miRNA). Pri-miRNA is subsequently cleaved in the nucleus by the Drosha /DGCR8 complex into one or more ~ 70 nt-long precursor miRNA strands (pre-miRNA). Pre-miRNA is exported to the cytoplasm by

exportin-5, where it is bound to RNase Dicer and to the RNA-induced silencing complex (RISC) and processed into a mature miRNA molecule. Only the active or mature strand is retained in the RISC complex, the passenger strand is removed and degraded. However, in some cases both strands of the miRNA duplex (the primary miRNA and its star sequence) are incorporated in the RISC complex with similar efficiency. In animals, the mature miRNAs usually bind to complementary sequences in the 3' untranslated region of target mRNAs, particularly with their seed sequences (nucleotide 2-8 from the 5' end). This results in repression of gene expression by translational inhibition and mRNA degradation. Under certain conditions miRNAs can also induce the translation of target mRNAs (Vasudevan et al., 2007) or transcriptionally silence gene expression (Kim et al., 2008). The physiological impact of this is still unknown. MiRNA biogenesis is controlled by extensive transcriptional and post-transcriptional regulation.

Proteins in miRNA biogenesis as well as several miRNAs have been implicated in the immune system, both in the process of haematopoiesis and in the function of haematopoietic cell lineages. The miR-17~92 cluster has been shown to be a ubiquitous regulator of B-cell, T-cell and monocyte development. MiR-142 was shown to be a regulator of T cell development, miR-150 has been implicated in B cell differentiation, miR-155 was shown to be involved in T- and B-cell maturation, miR-181a in B-cell differentiation and CD4+ T-cell selection, activation and sensitivity, and miR-223 is specifically expressed cells of the granulocytic lineage (Baltimore et al., 2008; Carissimi et al., 2009; Lindsay, 2008; Lu and Liston, 2009; Sonkoly et al., 2008; Taganov et al., 2006; Xiao et al., 2007).

MiRNAs have not only been linked to development of immune cells but also to the process of infection or inflammation. There is increasing evidence that they have important roles in regulating innate immune responses, the first line of defence to bacteria, viruses, and other pathogens. As part of the first immune response, monocytes and macrophages recognize microbial ligands with their pattern recognition receptors, of which the Toll-like receptors (TLRs) form one of the major families. Due to activation of the downstream signaling pathways hundreds of genes are induced to defeat the pathogen. Several miRNAs are induced or repressed during bacterial and viral infections. After exposure of (human or mouse) monocytes to bacterial lipopolysaccharide (LPS) miR-132, miR-146, and miR-155 showed significant upregulation (O'Connell et al., 2007; Taganov et al., 2006; Tili et al., 2007). In addition, miR-146 is also induced by proinflammatory stimuli, including interleukin-1 (IL-1) and tumor necrosis factor (TNF), in a NF-κB-dependent manner, and miR-155 was found to be stimulated by interferon (IFN) or TNF (O'Connell et al., 2007; Taganov et al., 2006). MiR-146 and miR-155 might be components of negative feedback loops attenuating TLR signaling pathways, where miR-146 limits IL-1 receptor-associated kinase 1 (IRAK1) and TNF receptor-associated factor 6 (TRAF6) expression, and miR-155 suppresses FADD, RIP and IKK (Pedersen and David, 2008). In contrast, miR-125b, which targets *TNF* mRNA, is down-regulated following LPS-induced TLR

signaling, thereby stimulating the production of TNF (Pedersen and David, 2008; Tili et al., 2007). There is also evidence that miRNAs play role in viral infections either by contributing to the antiviral defence or facilitating the replication of the pathogen.

As miRNAs regulate the normal development and function of the immune system, their aberrant expression have been suggested to be involved in pathogenesis of diseases. A large number of recent studies demonstrated deregulated miRNAs in immune-related disorders like inflammatory or autoimmune diseases. In inflammatory diseases such as psoriasis and atopic eczema, miR-146 and miR-125b were found to be deregulated (Carissimi et al., 2009). In rheumatoid arthritis miR-146 and miR-155 are commonly up-regulated. (Nakasa et al., 2008; Stanczyk et al., 2008). Many other miRNAs (miR-17~92, miR-21, miR-129, miR-203, miR-451 etc) have been detected in other types of immune-related disorders supporting the fact that these miRNAs play critical role in inflammation (Carissimi et al., 2009; Sonkoly and Pivarcsi, 2009). Since the immune system is tightly connected with cancer, not surprisingly the expression of many of the immunity-related miRNAs is also changed in a variety of tumors (Bhaumik et al., 2008; Costinean et al., 2006; Esquela-Kerscher and Slack, 2006; Garzon et al., 2009; Hurst et al., 2009; Lu et al., 2005; Medina and Slack, 2008).

Although many recent studies have contributed to the discovery of novel miRNAs and have aimed to shed light on their role in gene expression, only a limited number of the mRNA targets of miRNAs have been experimentally validated and the contribution of miRNA regulation to the pathogenesis of human diseases is far from understood. As miRNAs are highly evolutionary conserved across vertebrates, studies using different model organisms can provide valuable information about their functions in development of human diseases. In the last decade, zebrafish, which has an innate and adaptive immune system similar to that of mammals, has been commonly applied as model system for immunological research (Traver et al., 2003; van der Sar et al., 2004). Zebrafish models have been developed for different infections, including *Mycobacterium marinum* infection as a model for tuberculosis and *Salmonella typhimurium* infection as a model for food poisoning (Davis et al., 2002; Meijer et al., 2004; Meijer et al., 2005; Stockhammer et al., 2009; van der Sar et al., 2003; van der Sar et al., 2009). Zebrafish is also a convenient model to study miRNA function and its miRNA family has been well described (Davis et al., 2002; Kloosterman et al., 2007; Kloosterman et al., 2006; Schier and Giraldez, 2006; Wienholds et al., 2005). However, the functions of miRNAs in the immune system of zebrafish have not yet been addressed.

Here, we have used a custom-designed microarray platform to evaluate miRNA expression in adult zebrafish and zebrafish larvae infected with different bacterial pathogens. Common microRNA induction profiles from different bacterial infections are assessed and compared to human miRNA expression profiles in the literature. We have focused on miR-146 family and their predicted target genes in the

TLR pathway, and have compared microRNA expression with previously published expression data of predicted target genes (Stockhammer et al., 2009; van der Sar et al., 2009). Our results show that there are infection-responsive miRNAs highly conserved between human and zebrafish and also that there is considerable overlap in miRNA target genes, indicating that the zebrafish can be a useful model to dissect functions of miRNAs in the vertebrate immune system.

Results

Microarray analysis of microRNA profiles in response to infection

In order to study the effects of bacterial infection on microRNA (miRNA) expression profiles we used RNA samples from different infection experiments in adult and embryonal zebrafish that we had previously analyzed for gene expression differences (Stockhammer et al., 2009; van der Sar et al., 2009). Bacteria tested included *Salmonella typhimurium,* which causes a lethal infection in 1-day-old zebrafish embryos (van der Sar et al., 2003) and *Mycobacterium marinum* strains Mma20 and E11, which respectively cause acute disease or chronic granulomatous tuberculosis in zebrafish adults (van der Sar et al., 2004) and also induce granuloma formation in zebrafish larvae (Davis et al., 2002). We analyzed samples from embryos injected into the blood at 27 hpf with *S. typhimurium* bacteria (Stockhammer et al., 2009). In addition, embryos at 48 hpf were infected with *M. marinum* E11 using a novel infection system, in which bacterial suspensions with polyvinylpyrrolidone (PVP) as a carrier are injected into the yolk sac (Spaink et al., unpublished results). For comparison a non-pathogenic bacterium *Lactobacillus casei* Shirota was used in the yolk infection system and the effect of PVP carrier was determined as a control. In adult infection experiments, *M. marinum* strains Mma20 and E11 were injected into the intraperitoneal cavity. Details of the experimental set-up of the embryonic and adult infection studies are shown in Fig. 1.

To quantify microRNA gene expression profiles we used a custom-designed 8x15k Agilent zebrafish array. In the infection experiments with embryos, the total number of probes that showed more than 1.5-fold induction or repression of the expression level ($P < 1.00E{-}4$) was highest for the *M. marinum* E11 yolk-infected embryos (Fig. 2A). The number of differentially expressed probes was 8 times lower when similar doses of non-pathogenic *L. casei* Shirota were injected into the yolk and 18 times lower in control injections with PVP carrier. Similarly, the number of differentially expressed probes was 10 times lower in *S. typhimurium* caudal vein infections than in *M. marinum* E11 yolk-infected embryos. The numbers of differentially expressed probes in adult infection experiments with *M. marinum* were in a similar range as in the *M. marinum* E11 yolk-infected embryos (Fig.2A). Comparing the two *M. marinum* strains, a higher number of probes were differently expressed in infections with the Mma20 strain that causes acute disease than in the case of the E11 strain that causes chronic disease. For both strains, at 1 dpi 1.5 to 2-fold more probes were up-

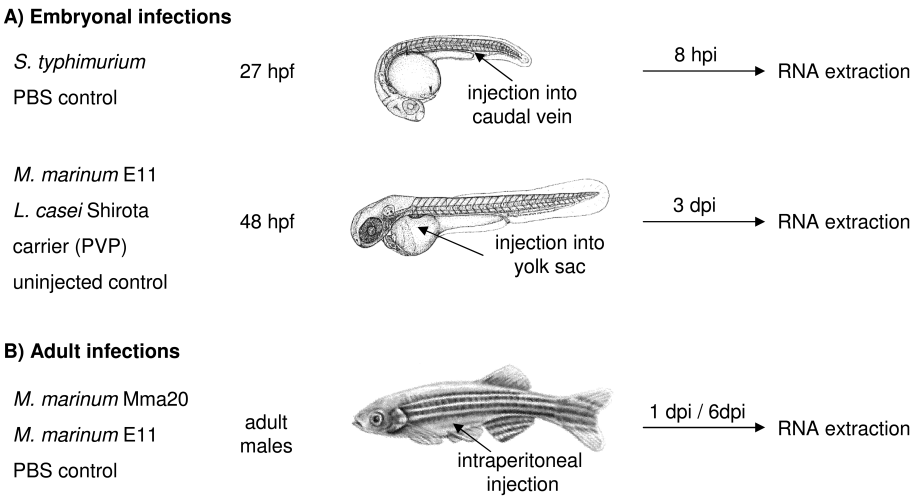


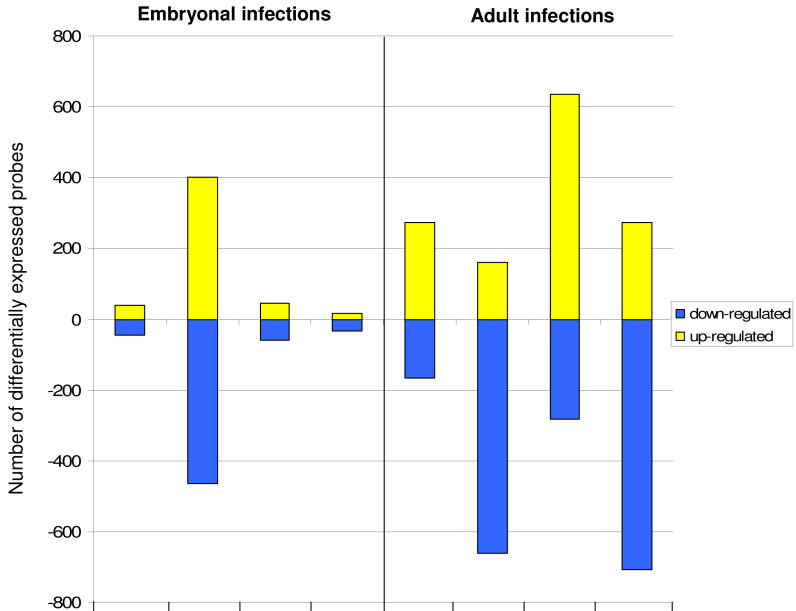
Figure 1. Overview of the infection experiments. (A) For infection of embryos with *S. typhimurium*, bacteria were microinjected into the caudal vein close to the urogenital opening after the onset of blood circulation (27 hpf). An equal volume of PBS was likewise injected for the control group (Stockhammer et al., 2009). RNA samples for transcriptome analysis were collected at 8 hours post infection (hpi). For yolk infection experiments, embryos were staged at 48 hpf and *M. marinum* E11 bacteria were injected in PVP suspension. For comparison a non-pathogenic bacterium *Lactobacillus casei* Shirota was injected in PVP suspension at similar concentrations as used for *M. marinum*. As controls PVP-injected and uninjected embryos were employed. For gene expression analysis we extracted RNA samples at 3 days post infection (dpi). (B) In adult infection experiments, *M. marinum* strains Mma20 and E11 were injected into the intraperitoneal cavity of male zebrafish and PBS-injected males were used as a control. For microarray analysis we collected RNA samples from two time points, 1 and 6 days post infection (dpi).

regulated than down-regulated, while at 6dpi experiments 3 to 4 times more probes were down-regulated than up-regulated.

Annotation of differentially expressed microRNAs

The custom-designed microarray platform contained probes for both known and predicted miRNAs, the latter classified into three categories: C1 - novel confident miRNAs for which there is compelling experimental evidence, C2 - candidate miRNAs with clear hairpin structure but less experimental evidence, and C3 - additional hairpins identified in the zebrafish genome sequence that are less likely to be miRNAs. The distribution of differentially expressed probes over the known and predicted miRNA categories in the different embryonal and adult infection experiments is shown in Fig. 2B. The total number of known miRNAs regulated during infection was highest for the *M. marinum* yolk infections in embryos (181 out of a total of 496 probes), much lower for the other embryonal infections (13-20 probes),

A



B

	St	E11	LcS	PVP	E11 1dpi	E11 6dpi	Mma20 1dpi	Mma20 6dpi
known up (496)	11	23	5	1	6	24	16	38
known down (496)	9	158	8	11	38	15	66	20
C1 up (24)	0	3	1	1	1	4	3	3
C1 down (24)	0	9	0	1	3	1	8	3
C2 up (18)	0	4	0	0	1	0	2	1
C2 down (18)	0	2	0	0	2	2	1	2
C3 up (5912)	27	371	39	14	266	132	615	231
C3 down (5912)	37	296	51	21	122	644	208	682

Figure 2. Differentially expressed miRNA probes in adult and embryonal infections. (A) Total number of up-regulated and down-regulated probes. (B) Number of up-regulated and down-regulated probes in four annotation categories: known miRNAs, and predicted miRNAs, C1, C2, and C3, with decreasing level of confidence as further detailed in the text. The total number of probes on the array per category is: 546 probes in known, 62 probes in C1, 54 probes in C2 and 6942 probes in C3 category. Embryonal infections were performed by injection of *S. typhimurium* (St) into the caudal vein at 27 hpf or by yolk injection of *M. marinum* (E11) or *L. casei* Shirota (LcS) at 48 hpf. PVP is the carrier control for yolk injections. Adult infections were performed with *M. marinum* strains E11 (E11 1dpi, E11 6dpi) and Mma20 (Mma20 1dpi, Mma20 6dpi). For further details of the experimental set-up of the infection studies see Figure1.

and intermediate for the adult *M. marinum* infections (39-82 probes). Furthermore, in the *M. marinum* yolk-infected embryos a particularly high number of known miRNAs fell into the down-regulated category (158) compared with the other infections. In comparison with the known miRNAs, the number of predicted miRNAs

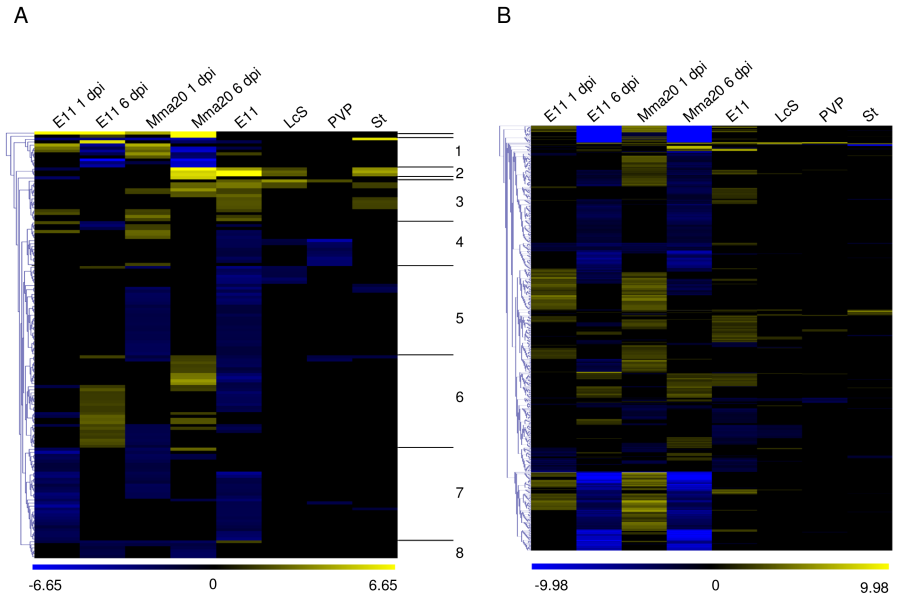


Figure 3. Hierarchical cluster analysis of differentially expressed miRNA probes in adult and embryonic infection experiments. A) The set of significantly up- and down-regulated known miRNAs, C1, and C2 category probes were analyzed together by one-dimensional hierarchical cluster analysis using MultiExperiment Viewer with settings for average linkage method with Euclidean distance. Induced genes are indicated by increasingly brighter shades of yellow, and down-regulated genes are indicated by increasingly brighter shades of blue. Eight clusters were defined in this set of probes named 1-8. Clusters containing less than 3 miRNAs were excluded. B) Cluster analysis of significantly up- and down-regulated C3 category probes.

(total of the C1-C3 categories) that showed differential expression in the infection experiments was between 3 and 20-fold higher. In particular, in all *M. marinum* embryo or adult infections the number of differentially expressed probes in the C3 category was very high (ranging between 388 and 913 out of a total of 5912 probes). In the much smaller C1 and C2 categories (consisting of 24 and 18 probes, respectively), the highest number of changes was observed in the *M. marinum* embryo infection (C1 12 probes, C2 6 probes). In the adult infection experiments with the *M. marinum* E11 and Mma20 strains, C3 probes were mostly up-regulated at 1dpi, and down-regulated at 6dpi. In contrast, the known miRNAs were more frequently down-regulated at 1dpi and up-regulated at 6dpi.

Common microRNA induction profiles between different bacterial infections

To further analyze the common response to different bacterial infections, we performed a two-dimensional hierarchical cluster analysis. To limit the complexity of

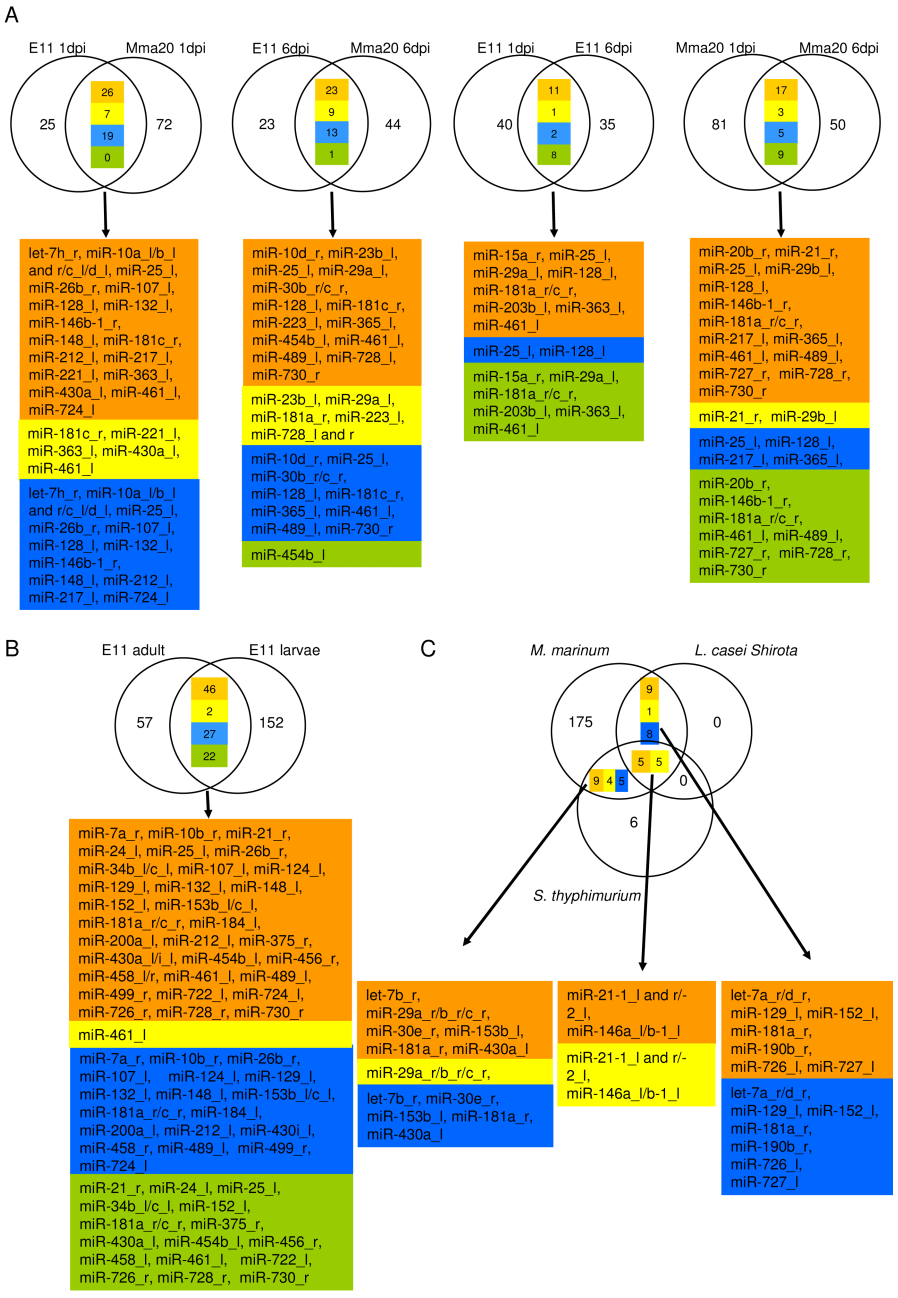


Figure 4. Venn diagrams showing the overlap between differentially expressed miRNA probes in adult and embryonic infection experiments. Orange - total number of overlapping probes; yellow - overlapping probes showing up-regulation in both groups; blue - overlapping

probes showing down-regulation in both groups; green - overlapping probes showing regulation in opposite directions. Only the probes in the category of known miRNAs are denominated in the colored boxes below the Venn diagrams. Probes correspond to the left (l) or right (r) arms of the pre-miRNA hairpin structure. A) Comparison of differentially expressed known miRNAs, C1, and C2 category probes in adult infections with the Mma20 and E11 strains. The Venn diagrams show the overlap in responsive miRNAs between the strains and at the two time points of analysis (1 and 6 dpi). B) Comparison of known miRNA, C1, and C2 category probe expression patterns in adult (1dpi and 6 dpi combined) and embryonal *M. marinum* E11 infection experiments. Numbers in yellow/blue/green boxes do not add up to 46 due to the fact that miRNAs can be changed in different directions at the 1dpi and 6 dpi time points. C) Comparison of known miRNAs, C1, and C2 category probes in embryonal infections with *M. marinum* E11, *S. typhimurium* and *L. casei* Shirota.

the cluster analysis we included only those probes that were differentially expressed in two or more infection experiments. We divided the probes into two main groups for which we performed separate cluster analysis (Fig. 3A,B). In the first set we combined the probes for known miRNAs and probes for miRNAs in the small C1 and C2 categories that include candidate miRNAs supported by experimental evidence. In the second set only C3 category probes that are less likely to represent miRNAs were evaluated. The cluster analysis of the C3 probes showed that in adult infection with the *M. marinum* E11 and Mma20 strains at 1 dpi a largely overlapping set of probes was up-regulated (Fig. 3B). In addition, the large numbers of down-regulated C3 probes observed with both strains at 6 dpi was also largely overlapping. For further analysis we concentrated on the set with known miRNAs and C1 and C2 predicted miRNAs. The clustering pattern of this probe set was notably different from that of the C3 probe set, with much less overlap between the *M. marinum* adult infection samples. We could distinguish eight major clusters of probes showing different trends in expression over the embryonal and adult infection experiments (Fig. 3A). MiRNAs in cluster 1 exhibited significant changes in several adult infection samples but not in embryonal infections. In contrast, miRNAs in clusters 2 and 3 were commonly induced in several samples of both adult and embryonal infections. Cluster 4 includes a non-specifically regulated group of miRNAs that was down-regulated in embryonic yolk injections with *M. marinum* but also by the PVP carrier used for these infections. In cluster 5, miRNAs are commonly down-regulated in *M. marinum* Mma20 infection of adults (1 dpi) and in *M. marinum* E11 embryonal yolk infection. Cluster 6 miRNAs were up-regulated at 6 dpi in adult *M. marinum* infection (Mma20 or E11 strains or both), but not in the embryonal *M. marinum* E11 infection. In cluster 7 miRNAs were commonly decreased in adult (1 dpi) and embryonal *M. marinum* infections, while cluster 8 miRNAs were commonly down-regulated in adult *M. marinum* infections at 6 dpi but not in embryonal infection.

In addition to the cluster analysis, we used Venn diagrams to show the common responsive miRNAs in different infection experiments. As shown in Fig.4A, in the *M. marinum* adult infections several tens of regulated miRNAs were overlapping

Table 1. Comparison of miRNA regulation between zebrafish and human or mammalian models.

		Infection experiment										
		Embryonal				Adult						
Zebrafish miRNA	Human miRNA	St	E11	LcS	PVP	E11 1dpi	E11 6dpi	Mma20 1dpi	Mma20 6dpi	Function/associated diseases in human or mammalian animal models	Demonstrated target genes	Reference
dre-miR-20b* (miR-20b_r)	hsa-miR-20b*	-	-2.2	-	-2	-	-	-2.2	1.7	overexpression in human T-cell leukemias, in c-Myc induced mouse mammary tumors, in undifferentiated gastric cancer	*Mylip, Hipk3, Rbp1-like*	Landais S et al 2007 Sun Y et al 2009 Katada T et al 2009 Guo J et al 2009
dre-miR-21 (miR-21-1_l)	hsa-miR-21	1.7	3.2	1.9	-	-	-	-	2.7	acquisition of invasive and metastatic properties by colon and breast cancer cell lines; overexpressed in advanced clinical stage and lymph node metastasis in human breast cancer, glioblastoma, breast, lung, colon, and pancreatic cancers, in invasive cancer cells, and in tumour metastases; associated with HCC	*PTEN, PDCD4, TPM1, MASPIN, IL12-p35, WNT1, JAG1, MARCKS, TIMP3, RECK*	Zhu S et al 2008 Asangani I et al 2008 Yan LX et al 2008 Baffa R et al 2009 Connolly E et al 2008 Zhang Z et al 2008
dre-miR-21* (miR-21-1_r)	hsa-miR-21*	3.8	4.2	1.8	-	-1.9	-1.4	-	8.3			
dre-miR-21 (miR-21-2_l)	hsa-miR-21	1.6	3.1	1.9	-	-	-	-	2.6			
dre-miR-21* (miR-21-2_r)	hsa-miR-21*	-	1.7	1.4	-	-	-	1.8	4.3			
dre-miR-128* (miR-128-2_l)	hsa-mir-128*	5.9	-	-	-	-3	-3.2	-4.1	-4.3	up-regulation inhibits *Reelin* and *DCX* expression and reduces neuroblastoma cell motility and invasiveness; Targeting of the Bmi-1 oncogene/stem cell renewal factor inhibits glioma proliferation and self-renewal; inhibits glioma cell proliferation by targeting transcription factor E2F3a; Inhibition of *SNAP25* expression by HIV-1 Tat; significantly overexpressed in acute lymphoblastic leukemia	*Reelin, DCX, Bmi-1, E2F3a, UBE2W*	Evangelisti C et al 2009 Godlewski J et al 2008 Zhang Y et al 2009 Mi S et al 2007 Fulci V et al 2009

		Infection experiment										
		Embryonal				Adult						
Zebrafish miRNA	Human miRNA	St	E11	LcS	PVP	E11 1dpi	E11 6dpi	Mma20 1dpi	Mma20 6dpi	Function/associated diseases in human or mammalian animal models	Demonstrated target genes	Reference
dre-miR-146a (miR-146a_l)	hsa-miR-146a	4.3	8.8	2.6	-	-	-	-	5.5	regulates the acute innate immune response following activation via the TIR superfamily, negatively regulate the release of interleukin (IL)-1β-induced IL-8 and RANTES during inflammatory stimulation of lung alveolar epithelial cells; up-regulated during viral infection, it is a negative regulator of the RIG-I-dependent antiviral pathway by targeting TRAF6, IRAK1, and IRAK2 in mouse macrophages;implicated in the metastatic and proliferative response associated with different cancer types (cervical, ovarian, breast, pancreatic and prostate cancer)	*TRAF6, IRAK1, IRAK2, IRF5, STAT-1, CXCR4, LMP1, ROCK1*	Lindsay MA et al 2008 Perry MM et al 2008 Hou J et al 2009 Tang Y et al 2009 Pauley KM et al 2008 Williams AE et al 2008
dre-miR-146a* (miR-146a_r)	hsa-miR-146a*	-	-	-	-	-	-	-	6.1			
dre-miR-146b (miR-146b-1_l)	hsa-miR-146b-5p	3.3	7.7	2.3	-	-	1.3		5.7	negative regulator of innate immunity by targeting TRAF6, IRAK1, IRF5; regulates cancer cell invasiveness by targeting MMPs; decreased in ATL cells may lead to increased inflammation and decreased T-reg functions, characteristics observed in HTLV-I–infected patients	*TRAF6, IRAK1, IRF5, AKT2, KRAS,*	Perry et al 2008 Taganov KD et al 2006 Bellon M et al 2009 Xia H et al 2009 Fulci V et al 2009
dre-miR-146b* (miR-146b-1_r)	hsa-miR-146b-3p	-	-	-	-	-2.2	-	-2.2	2.9			
dre-miR-152* (miR-152_l)	hsa-mir-152*	-	-3.2	-1.9	-	-	1.5	-2.3	-	decreased expression in endometrial serous adenocarcinomas, correlates with poor survival;affected by epigenetic inactivation: aberrant hypermethylation and subsequent transcriptional repression of miR-152 in human breast cancer; associated with apoptosis;	*MLH1, LTBP-4, ENPP2*	Hiroki E et al 2009 Lehmann U et al 2008 Kannan M et al 2009 Tan Z et al 2007 Visone R et al 2009

		Infection experiment										
		Embryonal				Adult						
Zebrafish miRNA	Human miRNA	St	E11	LcS	PVP	E11 1dpi	E11 6dpi	Mma20 1dpi	Mma20 6dpi	Function/associated diseases in human or mammalian animal models	Demonstrated target genes	Reference
dre-miR-181a* (miR-181a-1_r)	hsa-miR-181a*	-1.7	-1.9	-	-1.7	-	1.7	-1.6	1.6	specifically expressed and dynamically regulated in hematopoietic cells, expression in hematopoietic stem cells associated with increased B-cell development; decreased in HTLV-I, which dampens TCR signaling and T-cell activation to persist in the host; Down-regulation contributes to an aggressive leukemia phenotype through mechanisms associated with the activation of innate immunity pathways mediated by toll-like receptors and interleukin-1beta; miR-181a promotes early CD4 and CD8 double-positive T cell development;positvely regulated by protein kinase Akt1; critical player in EpCAM-positive hepatic cancer stem cells	*AKT3, CDX2, GATA6, NLK*	Chen CZ et al 2004 Bellon M et al 2009 Debernardi S et al 2007 Larson RA. 2009 Fulci V et al 2009 Liu G et al 2008 Androulidaki A et al 2009 Ji J et al 2009
dre-miR-181a-2* (miR-181a-2_r)	hsa-miR-181a-2*	-	-1.8	-1.3	-	-1.8	1.6	-	1.5			
dre-miR-181c* (miR-181c_r)	hsa-miR-181c*	-	-1.8	-	-	3	-2	3.1	-4.9			
dre-miR-461* (miR-461_r)	not in human	-	1.6	-	-	2.7	-3.9	2.5	-5.4	not in human and other mammalians		
dre-miR-728 (miR-728_r)	not in human	-	-2.1	-	-	-	2.1	-1.9	1.5	not in human and other mammalians		
dre-miR-730* (miR-730_r)	not in human	-	1.7	-	-	-	-2.2	3.8	-2.5	not in human and other mammalians		
(block1581404_C1_r)		-	-1.7	-1.3	-	-1.8	1.8	-1.9	1.4			
(sblock96274_C1_r)		-	-			10	17	4.6	18.4			

The miRNAs that showed more than 1.5-fold differential expression (P<10-4) in four or more experimental samples from the infection studies are listed in the table. MiRNA names were derived from miRBase and star sequences (not in all cases present in miRBase) were named accordingly. Probe names on the custom microarray are indicated in brackets behind the zebrafish (dre) miRNA names, and the corresponding human (hsa) miRNA names are given. Data on function, association to diseases in human and in mammalian animal models were collected from the literature. The literature data are for primary miRNAs, not for the star sequences, and literature data for miR-181a and miR-181c are combined. Experimentally demonstrated target genes for the primary miRNAs are also reported in the table (human gene symbols are in uppercase letters, mouse gene symbols begin with an uppercase letter followed by lowercase letters). In summary, these miRNAs are involved in several immune response related mechanisms (immune cell development, regulation of innate immunity, cytokine regulation, inflammation), and implicated in multiple cancer related mechanisms in a number of types of cancer. Yellow indicates up-regulation, blue indicates down-regulation. Light yellow and blue indicate significant fold changes (P<10-4) but below the 1.5-fold threshold. The numbers indicate the fold change values for S. typhimurium (St), M. marinum (E11), L. casei Shirota (LcS), and PVP carrier control injections of embryos and M. marinum (E11 and Mma20) adult infections, with experimental set-up as indicated in Fig. 1.

between the E11 and Mma20 strains and between the 1dpi and 6dpi time points. However, regulation of other miRNAs appeared to be strain or time point dependent. For example, miR-15a and miR-363 were up-regulated by E11 at 1 dpi and down-regulated by E11 at 6dpi, but were not regulated by Mma20 at both time-points. Similarly, miR-203b and miR-29a were not regulated by Mma20, but down-regulated by E11 at 1dpi and up-regulated at 6 dpi. On the other hand, miR-146b* (146b-1-r), miR-20b, miR-21 (miR-21-2-r), miR-217, miR-29b, miR-365, miR-489, miR-727, miR-728, and miR-730 were regulated by Mma20 at both time points but not by E11 at both time points. We also found miRNAs that changed in the opposite direction at 1 and 6 dpi time points in infections with both the E11 and Mma20 strains, such as miR-181c (down at 1 dpi, up at 6 dpi) and miR-461 (up at 1 dpi, down at 6 dpi). In addition, we identified that 46 known or C1/C2 category miRNAs were commonly regulated in adult and embryonal *M. marinum* E11 infections, of which 2 were up-regulated in the same direction, and 27 down-regulated in the same direction (Fig.4B). Comparison of all embryonal infection experiments showed that miR-21, miR-146a, and miR-146b were commonly up-regulated by the pathogenic *M. marinum* and *S. typhimurium* bacteria as well as by the non-pathogenic *L. casei* Shirota bacteria (Fig.4C). In conclusion, we observed common aspects in the response of miRNAs to different bacterial infections as well as differences related to the bacterial species or strains, the time points of analysis, and the adult or embryonal stage of the zebrafish host.

Functions and disease associations of commonly infection responsive miRNAs

To further narrow down the list of miRNAs that are commonly responsive to bacterial infections in zebrafish we took those miRNAs that occurred in more than four infection experiments. These miRNAs include miR-20b, miR-21, miR-128, miR-146a, miR-146b, miR-152, miR-181a, miR-181c, miR-461, miR-728, miR-730 and two predicted miRNAs in the C1 category (Table 1). We searched the literature for data on the function of the known miRNAs and their association to diseases in human and in mammalian animal models (Table 1). Reviewing the literature revealed that these miRNAs are involved in several processes related to the immune response, such as immune cell development, regulation of innate immunity, cytokine regulation, and inflammation. The miR-146 family, whose members miR-146a and miR-146b we find commonly up-regulated in infections of zebrafish, has previously been shown to affect the TLR signaling pathway that is involved in the innate immune response to bacterial infection (Taganov et al., 2006; Williams et al., 2008). Down-regulation of miR-181 family members, which we observed in four infection experiments, has been shown to contribute to an aggressive leukemia phenotype through mechanisms associated with the activation of innate immunity pathways mediated by TLRs and interleukin-1beta. The involvement of miR-181 in T-cell development has also been demonstrated (Chen et al., 2004; Larson, 2009). The interleukin-12 p35 subunit gene is a demonstrated target of the miR-21 family (Lu et al., 2009), whose members were

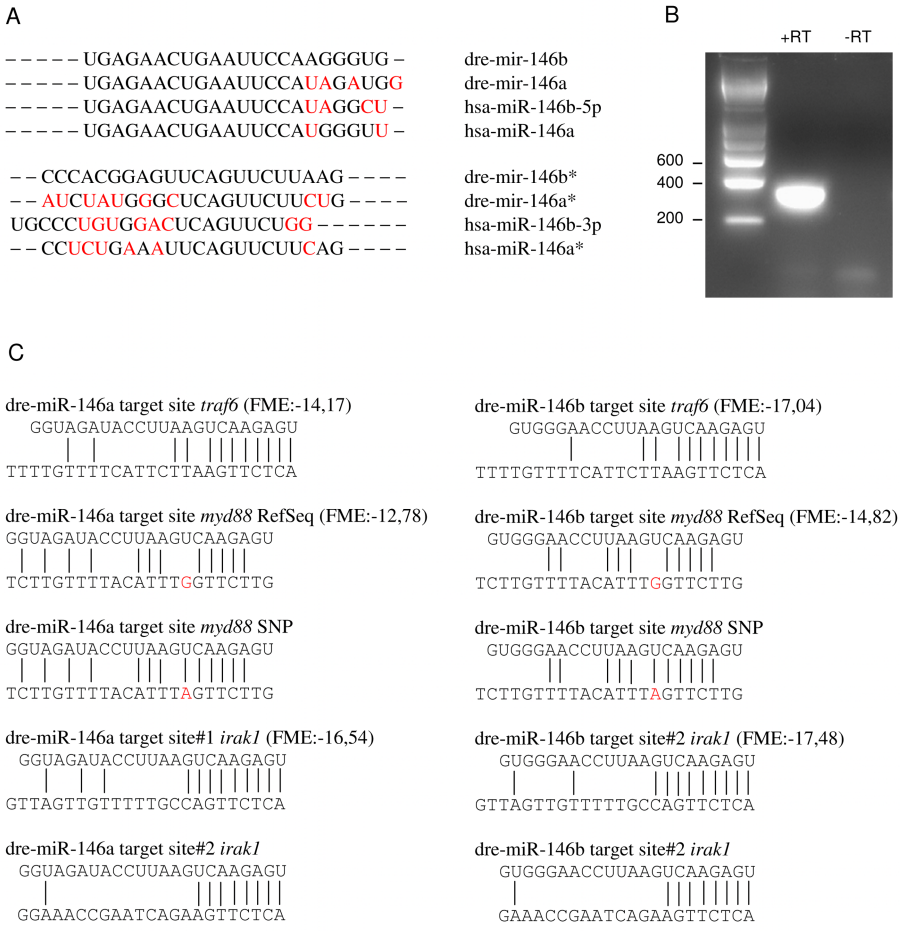


Figure 5. Members of the miR-146 family in human and zebrafish and their target sites in zebrafish TLR and IL-1R pathway genes. A) Sequence alignment of the members of the human (hsa) and zebrafish (dre) miR-146 family. Nucleotide differences with the dre-miR-146b/b* sequences are indicated in red. B) Confirmation of the presence of miR-146a/b target sites in the 3'UTR of the zebrafish *irak1* gene. A target site for miR-146a and miR-146b is located in the Zv8 genomic sequence between nucleotides 255-276 downstream of the stop codon of the *irak1* transcript. To check if this target site is included the 3'UTR of the *irak1* transcript RT-PCR was performed on RNA from *S. typhimurium*-infected embryos. RT-PCR reactions were performed with (+RT) and without (-RT) the reverse transcription step to verify that genomic DNA contamination was removed. With inclusion of the RT step a 316 bp fragment was amplified that includes the miR-146a/b target site. C) Alignments of predicted target sites of zebrafish miR-146a and b in TLR and IL-1R pathway genes, *traf6* (NM_199821), *myd88* (NM_212814) *and irak1* (Zv8 genome). We compared target sites of *myd88* transcript NM_212814 from the RefSeq database and the *myd88* transcript from the Ensembl database (ENSDARTooooo011143) where we detected a SNP (G > A polymorphism, indicated in red)

that increases the complementarity with the miR-146a/b sequences. In all alignments the miR sequences are shown on top in the 3′ to 5′ direction. Free minimum energy values (FME) for the interactions are indicated next to the alignments.

up-regulated in four zebrafish infections and down-regulated in one. MiR-152 has also been implicated in the immune system (Hiroki et al., 2009; Kannan et al., 2009; Lehmann et al., 2008). Furthermore, in previous studies most of the commonly infection responsive miRNAs (miR-20b, miR-21, miR-128, miR-146b, miR-152, miR-181a, miR-181c) have been implicated in various processes related to the development of cancer, such as apoptosis, proliferation, invasion, metastasis, and cell motility. Aberrant expression of these miRNAs has been detected in different types of cancer, including breast, liver, lung, and colon cancer, and leukemia (Table 1). Considering the common expression characteristics in adult and embryonal zebrafish infection experiments and the clear link with the immune response in human and other animal models, we focused our further analyses on family members of miR-146.

Targeting of the TLR and IL-1R pathways by the miR-146 family in human and zebrafish

MiR-146 was one of the most consistently induced miRNAs in our infection studies in zebrafish and has previously been linked to innate immune response in mammalian systems (Hou et al., 2009; Perry et al., 2009; Taganov et al., 2006). In human, the miR-146 family members are derived from two genes, *MIR146A* and *MIR146B*, located on chromosomes 5 and 10, respectively. The miRNAs processed from the left arm of the stem loop precursor miRNAs, miR-146a and miR-146b-5p, differ in their mature sequence only by two nucleotides at the 3' end (Taganov et al., 2007). In addition, mature miRNAs, miR-146a* and miR-146b-3p, are also processed from the right arm of the precursor miRNAs. Recently, it has been reported that there is a rapid increase in miR-146a and miR-146b expression in immune cells following activation of members of the TLR and IL-1 receptor families, activated by microbial ligands and pro-inflammatory cytokines, respectively. The induction of most TLR pathways and IL-1 signaling is initiated by the recruitment of adaptor protein MyD88 to the receptor. MyD88 transmits the signal to IL-1 receptor associated kinase 1 (IRAK1) that then recruits TNF receptor-associated factor 6 (TRAF6) into the complex. This activates IKB kinase and JNK and, in turn, the downstream NF-κB and AP-1 transcription factors and results in up-regulation of several hundreds of immune-responsive genes. IRAK1 and TRAF6 have been shown to represent potential targets of the miR-146 family, and overexpression of miR-146a and b resulted in down-regulation of IRAK1 and/or TRAF6 (Taganov et al., 2006). It has also been shown that miR-146a expression is regulated via NF-KB and JNK-1/2, whilst miR-146b expression was mediated via MEK-1/2 and JNK-1/2 (Perry et al., 2009). Together, these data suggest a role for the miR-146 family in negative feed back regulation of the innate immune response.

According to the data in miRBase (*www.miRbase.org/)* the zebrafish miR-146 fam-

A

miR-146a

	dre	mmu	hsa	gga	rno
dre	701	34	38	23	43
mmu	-	836	102	61	234
hsa	-	-	857	38	82
gga	-	-	-	589	49
rno	-	-	-	-	797

miR-146b

	dre	mmu	hsa	gga	rno
dre	681	40	27	28	-
mmu	-	828	88	53	-
hsa	-	-	812	41	-
gga	-	-	-	602	-
rno	-	-	-	-	-

B

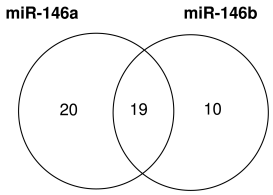


Figure 6. **Conservation of predicted targets of the miR-146 family.** Target genes of miR-146 family members in zebrafish (dre), human (hsa), mouse (mmu), rat (rno) and chicken (gga) were predicted using MicroCosm Targets. The predicted targets from zebrafish, mouse, rat, and chicken, were converted to human Ensembl Gene ID homologues and were compared across different species. If the predicted targets had the same human homologues, then these targets were considered to be potentially conserved across these species. The numbers of overlapping targets between species are indicated in the table. B) Comparison of miR-146a and miR-146b predicted targets. The Venn diagram shows the overlap between the targets of miR-146a and miR-146b that were commonly predicted for human and zebrafish.

ily consists of two members, dre-miR-146a and dre-miR-146b, derived from genes located on chromosomes 13 and 21, respectively. Experimental evidence is indicated in miRBase only for mature miRNAs derived from the left stem loop precursor arms. However, in our microarray analysis we also detected differential expression of probes for miRNAs derived from the right stem loop precursor arms, which we named dre-miR-146a* and dre-miR-146b*. The alignment of mature sequences of human and zebrafish miR-146 family members is shown in Fig. 5A.

To determine if the specificity of miR-146 members for the *irak1* and *traf6* genes in the TLR/IL-1R pathways might be conserved in zebrafish we used different target prediction programs. MiRanda prediction identified *traf6* as a putative target of both miR-146a and b from zebrafish, however *irak1* was not identified. We checked *irak1* mRNAs/ESTs in the database, but only few sequences were present, of which only two ESTs contained a very short 3'UTR. Upon inspection of the *irak1* genomic sequence we identified two perfect miR-146 target sites present at nucleotide locations 293-316 and 903-924 downstream of the *irak1* stop codon. By reverse transcriptase PCR on RNA from *S. typhimurium*-infected zebrafish embryos we confirmed that the *irak1* transcript contains a longer 3'UTR containing at least the first target site (Fig. 5B). MiRanda analysis predicted *myd88* as a third gene in the TLR pathway that could be a potential target of zebrafish miR-146 a and b. However, it was not predicted for human miR-146. Interestingly, zebrafish *myd88* is polymorphic for the pre-

Table 2. Ranking of predicted miR-146 target sites in the TLR pathway genes *myd88*, *irak1* and *traf6*.

	miRanda ranking			RNAhybrid ranking			TargetScan ranking			Integration algorithm		
	myd88	***irak1***	***traf6***	***myd88***	***irak1***	***traf6***	***myd88***	***irak1***	***traf6***	***myd88***	***irak1***	***traf6***
dre-miR-146a	10	16	1	13	44	7	-	6	3	14	2	2
dre-miR-146a*	-	-	-	190	65	2	-	5	2	102	93	-
dre-miR-146b	12	14	3	28	35	3	-	4	1	13	1	1
dre-miR-146b*	-	-	-	24	-	-	-	-	-	20	-	-
hsa-miR-146a	-	17	2	567	189	25	-	8	9	570	8	1
hsa-miR-146a*	-	130	26	-	451	153	-	-	60	-	159	23
hsa-miR-146b-5p	-	33	5	592	352	82	-	7	11	575	23	2
hsa-miR-146b-3p	61	-	-	143	37	-	57	74	-	31	158	-

Three different target site prediction methods, MiRanda, RNAhybrid and TargetScan, were used to identify putative miR-146 family target genes in transcript databases downloaded from the MicroCosm Targets database (formerly known as miRBase Targets). As the zebrafish *irak1* transcript in this database contains an incomplete 3'UTR we replaced this sequence with the longer sequence derived from the Zv8 genome. The predicted target genes were ranked by calculating a weight for the prediction based on different parameters used by the target site prediction methods. In miRanda the weight is calculated from local alignment score, thermodynamic stability, and the number of predicted target sites in the transcript. In RNAhybrid the weight is based on *P*-value for the significance of increasing free energy and the number of predicted target sites. In TargetScan the weight is based on site type (matching to 5' seed region and contribution of 3' pairing) and the number of predicted target sites. The integration algorithm calculates a final weight based on the three prediction methods. Numbers in the table show the position of the gene in the weight ranking of all predicted target genes.

dicted target site, which contains a SNP (single nucleotide polymorphism) in the 5' seed region that is most important for the miRNA interaction (Fig. 5C). Alignments of miR-146 a and b with their predicted target sites in all three TLR pathway genes, *traf6*, *irak1*, and *myd88* are shown in Fig. 5C.

To further investigate the conservation of miR-146 target sites in the *myd88*, *irak1* and *traf6* genes we employed two target site prediction methods, RNAhybrid and TargetScan, in addition to miRanda, and also used an integration algorithm to rank these predictions (Table 2). This showed that miR-146 target sites in *traf6* are in the top 2 raking for miR-146a and b from both human and zebrafish. MiR-146a and b target sites in zebrafish *irak1* (containing the longer 3'UTR from genomic prediction) ranked in the top 2 and for human *IRAK1* ranked on 8th and 23rd position using the integration algorithm. MiR-146 target sites in zebrafish *myd88* were in the top 20, but for human *MYD88* were predicted only by the RNAhybrid method and with low ranking. Target sites in the TLR pathway genes are also predicted for the miR-146 star/3p sequences but with lower ranking.

Other predicted targets of the human and zebrafish miR-146 family

To predict further conserved targets we downloaded miRanda-predicted target gene lists of zebrafish (dre-miR-146a, -146b), human (hsa-miR-146a, -146a*,-146b-5p, -146b-3p), mouse (mmu-miR-146a, -146b, -146b*), rat (rno-miR-146a) and chicken (gga-miR-146a, -146b, -146b*) miR-146 family members from MicroCosm Targets (formerly known as miRBase Targets). To compare targets across different species, we converted the predicted targets from zebrafish, mouse, rat and chicken to the homologous human Ensembl Gene IDs and HGNC gene symbol (Fig. 6A). Among 701 predicted target genes for miR-146a in zebrafish, we found 38 overlapping between zebrafish and human, of which 7 were also conserved in at least one other species (Fig. 6 and Table 3). For miR-146b, 27 out of 681 predicted target genes in zebrafish were overlapping with human, with 6 conserved in at least one other species (Fig. 6 and Table 3). Next, we compared the results for miR-146a and b showing that 19 genes were predicted to be commonly targeted by miR-146a and b in both human and zebrafish, 19 genes only targeted by miR-146a, and 10 genes only targeted by miR-146b (Fig. 6B). To this list of 48 genes we added those predicted targets that overlapped between zebrafish and at least two other species but not human (Table 3). This list of in total 64 genes was used for functional annotation and analysis of biological pathways (Table 3). These analyses revealed that most of the predicted targets in the list have immune related functions and were previously shown to be associated with immunological diseases or cancer. Genes were mostly classified into signaling in the immune system (*ARHGAP17, COMMD7, HSPA1A, IRAK1, IRF5, LCK, PIAS2, TRAF6, TRPC1*), haemostasis (*FGB, MMP11*), apoptosis (*ARD1, ARSG, CTSB, DAXX, FAS*), cell cycle signaling (*CHEK1, PPP2R4, ASPM*), gene expression (*ARID3B, BCOR, FBXO8, IARS, POLR2A, PUM2, SMYD1, SNRPC, TGIF, YEATS4, ZNF143, ZNF326*), and metabolic processes (*AMARC, B3ALT2, GNE, NUDT4, PHKA1, POLQ, PTGES2,*

Table 3. Pathway annotation of predicted target genes of the miR-146 family.

Ensembl Gene ID	Gene symbol	# species conserved	dre	hsa	mmu	rno	gga	Biological pathways	Description	Reference
ENSG00000242110	*AMACR*	3	a/b	-	-	a	a/b	Steroid metabolism	Alpha-methylacyl-CoA racemase	Mubiru JN et al 2005 Marx A et al 2008 Li W et al 2008
ENSG00000102030	*ARD1A*	2	a/b	a/b	-	-	-	Apoptosis:Akt-HIF1 signaling, Metabolism: amino acid	N-terminal acetyltransferase complex ARD1 subunit homolog A	Ametzazurra A et al 2008 Lim JH et al 2008
ENSG00000140750	*ARHGAP17*	3	a/b	-	a/b	a	-	Signaling by Rho GTPases, by EGFR	-	
ENSG00000179361	*ARID3B*	2	a	a/b	-	-	-	Regulation of transcription	AT rich interactive domain 3B (BRIGHT-like)	Cowden Dahl KD et al 2009 Kim D et al 2007 Kobayashi K et al 2006 Takebe A et al 2006
ENSG00000141337	*ARSG*	2	a/b	a/b	-	-	-	Apoptosis:Sphingolipid Metabolism	Arylsulfatase G Precursor	Frese MA et al 2008
ENSG00000066279	*ASPM*	4	a/b	b	a	a	-	Negative regulation of asymmetric cell division	Asp (abnormal spindle) homolog, microcephaly associated	Bond J et al 2002 Hagemann C et al 2008 Lin SY et al 2008 Wu SC et al 2008
ENSG00000162630	*B3GALT2*	2	a	a/b	-	-	-	Sphingolipid metabolism	Beta-1,3-galactosyltransferase 2	Felner KM et al 1998 Artrip JH et al 1999 Valli M et a 1998
ENSG00000183337	*BCOR*	2	a/b	a/b	-	-	-	Negative regulation of transcription,DNA-dependent, chromatin modification	BCL6 co-repressor	Huynh KD et al 2000 Wamstad JA et al 2008 Srinivasan RS et al 2003
ENSG00000105583	*C19ORF56*	2	a/b	a	-	-	-	-	UPF0139 membrane protein C19orf56	Aoki H et al 2000
ENSG00000165698	*C9ORF9*	2	a/b	b	-	-	-	-	Uncharacterized protein C9orf9	Boelens MC 2009 Amira N et al 2004
ENSG00000140743	*CDR2*	2	a/b	a/b	-	-	-	-	Cerebellar degeneration-related protein 2 (Paraneoplastic cerebellar degeneration-associated antigen)	Albert ML et al 1998
ENSG00000149554	*CHEK1*	2	a/b	a/b	-	-	-	Cell Cycle Checkpoints	Serine/threonine-protein kinase Chk1	Jamil S et al 2008 Matsuura K et al 2008 Barré B, Perkins ND 2007
ENSG00000184908	*CLCNKB*	2	a/b	a/b	-	-	-	Chloride transport	Chloride channel protein ClC-Kb (Chloride channel Kb)	Embark HM et al 2004
ENSG00000149600	*COMMD7*	2	b	a/b	-	-	-	Negative regulation of NF-kappaB transcription factor activity	COMM domain-containing protein 7	Maine GN, Burstein E 2007 de Bie P et al 2006
ENSG00000212710	*CTAGE1*	3	a/b	-	a/b	a	-	-	Cutaneous T-cell lymphoma-associated antigen 1	Usener D et al 2003 Eichmuller S et al 2001
ENSG00000225932	*CTAGE4*	3	a/b	-	a/b	a	-	-		
ENSG00000150527	*CTAGE5*	3	a/b	-	a/b	a	-	-		
ENSG00000244693	*CTAGE6*	3	a/b	-	a/b	a	-	-		
ENSG00000164733	*CTSB*	3	a/b	-	a/b	a	-	Regulation of apoptosis	Cathepsin B	Johnson EM et al 2009 Rose PP et al 2007 Yan S, Sloane BF 2003 Li-Korotky HS et al 2004
ENSG00000102390	*CXORF26*	2	a	a	-	-	-	-	-	
ENSG00000204209	*DAXX*	2	a	a/b	-	-	-	Apoptosis (JNK, MAPK, SUMO signalng)	Death domain-associated protein 6	Yang X et al 1997 Tavalai N et al 2008 Kitamura T et al 2009

Ensembl Gene ID	Gene symbol	# species conserved	dre	hsa	mmu	rno	gga	Biological pathways	Description	Reference
ENSG00000165487	*EFHA1*	3	a/b	-	a/b	a	-	Calcium ion binding	EF-hand domain family, member A1	
ENSG00000026103	*FAS*	3	a/b	a/b	a	-	-	Apoptosis	Tumor necrosis factor receptor superfamily member 6 Precursor (FASLG receptor)(Apoptosis-mediating surface antigen FAS)(Apo-1 antigen)(CD95 antigen)	
ENSG00000164117	*FBXO8*	2	a	a/b	-	-	-	Regulation of ARF protein signal transduction	F-box only protein 8	Winston JT et al 1999
ENSG00000171564	*FGB*	2	a/b	a	-	-	-	Hemostasis, Formation of Platelet plug	Fibrinogen beta chain Precursor	Xie FJ et al 2009
ENSG00000159921	*GNE*	3	a/b	-	-	a	a/b	Carbohydrate metabolic process, glycolysis, gluconeogenesis	Glucosamine (UDP-N-acetyl)-2-epimerase/N-acetylmannosamine kinase	Giordanengo V et al 2004
ENSG00000204389	*HSPA1A*	2	a/b	a/b	-	-	-	Signaling in Immune system, Antigen processing and presentation, Proteolysis Putative ubiquitin pathway	Heat shock 70kDa protein 1A	Merkley MA et al 2009 Wheeler DS et al 2009 Wang X et al 2008 Figueiredo C et al 2008 Gabai VL et al 2008
ENSG00000196305	*IARS*	2	a	a/b	-	-	-	Gene Expression, Cytosolic tRNA aminoacylation	Isoleucyl-tRNA synthetase, cytoplasmic , isoleucine tRNA synthetase	Nichols RC et al 1995 Lee SW et al 2004
ENSG00000184216	*IRAK1*	4	a/b	a/b	a/b	a		Signaling in Immune system, TCR signaling	Interleukin-1 receptor-associated kinase 1	Bannerman DD et al 2002 Ahmad R et al 2007 Mizobe T et al 2007 Liu G et al 2008
ENSG00000128604	*IRF5*	2	a/b	a	-	-	-	Toll-like receptor signaling pathway	Interferon regulatory factor 5	Yang L et al 2009 Pandey AK et al 2009 Hu G and Barnes BJ 2009 Balkhi MY et al 2009 Yanai H et al 2007
ENSG00000152315	*KCNK13*	3	a/b	b	-	a	-	Potassium transporters: inward current	Potassium channel, subfamily K, member 13	Theilig F et al 2008 Fearon IM et al 2006 Nogueira EF et al 2009
ENSG00000111731	*KIAA0528*	2	b	b	-	-	-	-	Hypothetical protein LOC9847	
ENSG00000108244	*KRT23*	2	a/b	a	-	-	-	-	Keratin 23 (histone deacetylase inducible)	Birkenkamp-Demtroder K et al 2007 Wang K et al 2009 Zhang JS et al 2001
ENSG00000182866	*LCK*	2	b	a/b	-	-	-	Signaling in Immune system, TCR signaling, hemostasis, HIV Infection	Proto-oncogene tyrosine-protein kinase LCK (Lymphocyte cell-specific protein-tyrosine kinase)(p56-LCK)(LSK)(T cell-specific protein-tyrosine kinase)	Wagner MJ, Smiley JR 2009 Strasner AB et al 2008 Soldaini E 2003 Seitzer U et al 2001 Lund TC et al 1997 Kadena T et al 1997
ENSG00000198934	*MAGEE1*	4	a/b	a/b	a/b	a	-	-	Melanoma-associated antigen E1 (MAGE-E1 antigen)(Hepatocellular carcinoma-associated protein 1)	De Plaen E et al 1994
ENSG00000099953	*MMP11*	2	a/b	a	-	-	-	Hemostasis: Leukocyte Extravasation, Bladder Cancer Signaling	Matrix metalloproteinase-11	Peruzzi D et al 2009 Nakopoulou L et al 2007
ENSG00000236761	-	3	a/b	-	a/b	a	-	-	-	
ENSG00000242426	-	3	a/b	-	a/b	a	-	-	.	
ENSG00000173598	*NUDT4*	3	a	-	-	a	a	Cyclic nucleotide metabolic process, regulation of RNA export from nucleus	Nudix (nucleoside diphosphate linked moiety X)-type motif 4	

Ensembl Gene ID	Gene symbol	# species conserved	dre	hsa	mmu	rno	gga	Biological pathways	Description	Reference
ENSG00000173598	*NUDT4P1*	3	a	-	-	a	a	-	Nudix (nucleoside diphosphate linked moiety X)-type motif 4 pseudogene 1	
ENSG00000157881	*PANK4*	2	a	a	-	-	-	Pantothenate and CoA biosynthesis	Pantothenate kinase 4 Pantothenic acid kinase 4	Li Y et al 2005 Xiang RL et al 2007 Wilfred BR et al 2007
ENSG00000104883	*PEX11G*	2	a/b	a/b	-	-	-	Peroxisome fission	Peroxisomal membrane protein 11C (Peroxin-11C)(PEX11gamma)	Tanaka A, et al 2003 Thoms S, Erdmann R 2005 Dharancy et al 2009
ENSG00000067177	*PHKA1*	2	b	a/b	-	-	-	Glucose metabolism	Phosphorylase kinase, alpha 1 (muscle)	Prakash S et al 2005 Wuyts W et al 2005 McDonell N et al 2000
ENSG00000078043	*PIAS2*	2	a/b	a	-	-	-	Immune response IL-12 signaling pathway, P53 signaling pathway, SUMO-1 pathway, Proteolysis, Cancer, Jak-STAT signaling pathway	E3 SUMO-protein ligase PIAS2 (Protein inhibitor of activated STAT2)(DAB2-interacting protein)(DIP)(Androgen receptor-interacting protein 3)(ARIP3)	Saito M et al 2009 Herold S et al 2008 van den Akker E et al 2005 Liu ST et al 2006 Kim J et al 2005
ENSG00000051341	*POLQ*	4	b	-	b	a	b	Metabolisom of nucleotides	Polymerase (DNA directed), theta	Kawamura K et al 2004
ENSG00000181222	*POLR2A*	2	a	a/b	-	-	-	Gene Expression, MicroRNA (miRNA) Biogenesis, viral Infection	DNA-directed RNA polymerase II subunit RPB1	Mook OR et al 2009 Fluiter K et al 2002 Nguewa PA et al 2008 Saviozzi S et al 2006
ENSG00000119383	*PPP2R4*	2	a/b	a	-	-	-	Cell cycle Regulation of G1/S transition (part 1), Small cell lung cancer	Protein phosphatase 2A activator, regulatory subunit 4	Janssens V et al 2000 Blalock WL et al 2009 Bae D, Ceryak S 2009 Teft WA et al 2009 Cornell TT et al 2009 Bernsmeier C et al 2008 Gushiken FC et al 2008
ENSG00000148334	*PTGES2*	3	a	-	a/b	a	-	Lipid metabolism	Prostaglandin E synthase 2	Meng Q et al 2005 Seo T et al 2009 Mattila S et al 2009
ENSG00000055917	*PUM2*	3	a/b	-	a/b	a	-	Regulation of translation	Pumilio homolog 2 (Drosophila)	Cao Q et al 2009 Moore FL et al 2003
ENSG00000075826	*SEC31B*	2	a/b	a	-	-	-	Clathrin-dependent protein traffic	Protein transport protein Sec31B (SEC31-related protein B)	Stankewich MC et al 2006 Tang BL et al 2000
ENSG00000139974	*SLC38A6*	2	a	a/b	-	-	-	Amino acid transport	Probable sodium-coupled neutral amino acid transporter 6	White JK et al 2005 Awomoyi AA 2007 Lykouras D et al 2008 Stober CB et al 2007
ENSG00000115593	*SMYD1*	3	a/b	-	a/b	a	-	Negative regulation of transcription	SET and MYND domain containing 1 (BOP)	Gottlieb PD et al 2002 Wanaka A 2000 Hwang I, Gottlieb PD 1995 Kawamura S et al 2008 Tan X et al 2006
ENSG00000124562	*SNRPC*	3	a/b	a/b	a/b	-	-	Spliceosomal snRNP assembly	U1 small nuclear ribonucleoprotein C (U1 snRNP protein C)(U1C protein)	Knoop LL, Baker SJ 2000
ENSG00000227624	*SNRPEL1*	2	b	a/b	-	-	-		Small nuclear ribonucleoprotein polypeptide E-like 1	Argast GM et al 2009 Borysov SI, Guadagno TM 2008 Li Z, Pützer BM 2008

Ensembl Gene ID	Gene symbol	# species conserved	dre	hsa	mmu	rno	gga	Biological pathways	Description	Reference
ENSG00000177426	*TGIF1*	3	a/b	a/b	-	a	-	Negative regulation of transcription from RNA polymerase II promoter	Homeobox protein TGIF1 (5'-TG-3'-interacting factor 1)	Chen F et al 2003
ENSG00000168286	*THAP11*	2	a/b	a/b	-	-	-	Regulation of transcription	THAP domain-containing protein 11	Zhu CY et al 2009
ENSG00000133872	*TMEM66*	3	b	a/b	a/b	-	-	-	Transmembrane protein 66 Precursor (Protein FOAP-7)(hepatitis B virus-HBV X-transactivated gene 3 protein)	Romanuik TL et al 2009 Mannherz O et al 2006 Fujii Y et al 1999
ENSG00000175104	*TRAF6*	4	a/b	a/b	a/b	a	-	TLR, TNFR and TCR signaling	TNF receptor-associated factor 6 (Interleukin-1 signal transducer)(RING finger protein 85)	
ENSG00000144935	*TRPC1*	2	a/b	a/b	-	-	-	Signaling in Immune system: CREB pathway, MEF2 in T lymphocytes	Short transient receptor potential channel 1	Takahashi Y et al 2007 Pani B et al 2006 Marasa BS et al 2006 Beck B et al 2008 Bair AM et al 2009
ENSG00000137267	*TUBB2A*	2	a/b	a/b	-	-	-	Metabolism of proteins, Protein/Tubulin folding	Tubulin beta-2A chain	Yeh TS et al 2004
ENSG00000135763	*URB2*	2	a/b	a/b	-	-	-	-	Unhealthy ribosome biogenesis protein 2 homolog	Mnaimneh S et al 2004
ENSG00000127337	*YEATS4*	4	a/b	a/b	a/b	a	-	Positive regulation of transcription, DNA-dependent	YEATS domain-containing protein 4 (Glioma-amplified sequence 41)(Gas41)(NuMA-binding protein 1) (NuBI-1)	Park JH, Roeder RG 2006 Italiano A 2008
ENSG00000166478	*ZNF143*	2	a/b	a	-	-	-	Transcription: RNA Polymerase III Transcription	Zinc finger protein 143 (Selenocysteine tRNA gene transcription-activating factor)(hStaf)(SPH-binding factor)]	Mach CM et al 2002 Schuster C et al 1995 Grossman CE et al 2004 Myslinski E et al 2007 Gérard MA et al 2007 Myslinski E et al 2006
ENSG00000162664	*ZNF326*	2	a/b	a	-	-	-	Regulation of transcription	Zinc finger protein 326, ZAN75	Lee JY et al 2007, 1998

The table includes the most reliable predicted targets based on conservation among multiple species and all targets that were overlapping between zebrafish and human. Analysis of biological pathways was performed using the Reactome pathway database, GeneAlacart (www.genecards.org) and AmiGO pathways (amigo.geneontology.org/cgi-bin/amigo/go.cgi). Literature references are given that link these genes to immune functions. Species analyzed included zebrafish (dre), human (hsa), mouse (mmu), rat (rno) and chicken (gga). a - miR-146a predicted target gene, b - miR-146b predicted target gene, a/b - predicted target gene of both miR-146a and b. An extended version of this table with details on immune functions reported in the literature is included as a supplementary file.

TUBB2A), all pathways previously linked to immune functions.

Comparison between miRNA expression and the expression of the predicted target genes

Next we compared miR-146 expression with expression of predicted target genes from Table 3 under infection conditions. For comparison we took previously published data from van der Sar et al. 2009, where the zebrafish host transcriptome response to *M. marinum* Mma20 and E11 was determined using the same samples as for miRNA analysis in the present study. In addition, we used data from Stockhammer et al. 2009, where a time-course transcriptome profiling of the zebrafish embryonic innate immune response to salmonella infection was performed. We found that approximately half (35) of the predicted target genes in Table 3 showed significant ($P<10^{-4}$) expression changes (up- or down-regulation) in either of these infection studies (Table 4). It is possible that the number of infection-responsive genes might still be higher since not all predicted target genes (13) were present on the microarray platform used in the expression studies. Responsiveness to both *M. marinum* and *S. typhimurium* infection was observed for 10 predicted target genes, 17 were responsive only to *M. marinum* infection, 8 only to *S. typhimurium* infection, and 13 predicted target genes represented on the microarray were not responsive to both infections. *Irak1*, one of the genes of the TLR and IL-1R pathways that are experimentally validated targets in human, was not present on the microarray. *Traf6* the other experimentally validated target in TLR and IL-1R signaling pathways was up-regulated at the end stage of *S. typhimurium* infection in embryos (24hpi). The interferon-responsive factor 5 (*IRF5*) gene is another key immune response gene that has been experimentally validated in mammalians and for which a miR-146 a and b target sites are also predicted in zebrafish. Expression of *irf5* was up-regulated in all samples of the *M. marinum* infection study. In conclusion, many of the predicted miR-146 target genes that are conserved between human and zebrafish can be linked to bacterial infection processes by expression changes.

Discussion

To increase knowledge of miRNAs involved in the vertebrate host response to bacterial infections we have analyzed miRNA expression profiles of zebrafish infected with two model pathogens, *S. typhimurium* and *M. marinum* (Stockhammer et al., 2009; van der Sar et al., 2009). While adult zebrafish have complex innate and adaptive immunity similar to the human immune system, zebrafish embryos rely solely on the function of their innate immune response. Therefore, by using both embryos and adult zebrafish as hosts for bacterial infection, we could link miRNA expression profiles to the innate and adaptive components of the immune response. We identified a set of commonly responsive miRNAs in embryonic and adult infection experiments that includes highly conserved miRNAs with strong associations to

Table 4. Responsiveness of predicted target genes of miR-146 to different infections.

Human gene symbol	Unigene Code	Zebrafish gene symbol	E11 1 dpi	Mma20 1 dpi	E11 6 dpi	Mma20 6 dpi	St-5h	St 8h	St 24h
AMACR	Dr.77673	*c1qtnf3*				-1.64948			
ASPM	Dr.96472	*aspm*	1.47073	1.34309	1.78499	2.45658			
B3GALT2	Dr.79277	*b3galt2*							-1.41604
C19ORF56	Dr.117007	*zgc:73111*			-1.34411				
CHEK1	Dr.83316	*chek1*	1.38266	-1.51612	-1.42253				
CLCNKB	Dr.78662	*zgc:64141*				-2.01377	-1.25287	-1.64969	-1.58357
CTAGE1	Dr.105440	*ctage5*				1.28439			
CTAGE5	Dr.105440	*ctage5*				1.28439			
CTSB	Dr.3374	*ctsba*				2.82477			3.31843
CXORF26	Dr.78321	*zgc:92611*			-1.33198				
DAXX	Dr.73516	*daxx*				1.5038			
EFHA1	Dr.77336	*zgc:113196*		1.3113					
FAS	Dr.82180	*fas*			-2.1401	1.5653		3.6617	12.54386
FBXO8	Dr.83230	*fbxo8*		-1.41323					2.19152
FGB	Dr.8505	*fgb*		1.37155		3.47284		1.3093	
GNE	Dr.132714	*gne*		-1.16012		-1.15201			
HSPA1A	Dr.115994	*hsp70l*							15.85009
IARS	Dr.5185	*iars*							3.64018
IRF5	Dr.7758	*irf5*	1.4546	1.34987	1.45303	2.67239			
LCK	Dr.91511	*lck*				-1.52648			
-	Dr.105440	*ctage5*				1.28439			
NUDT4P1	Dr.134107	*zgc:101062*							-2.15537
PANK4	Dr.132622	*zgc:66285*				-1.3341			
PIAS2	Dr.25643	*pias2*						-1.1846	
PPP2R4	Dr.9399	*ppp2r4*						1.34357	
PTGES2	Dr.16120	*ptgesl*		-1.27843		-1.98551			-1.70173
SLC38A6	Dr.45587	*slc38a6*	1.43176			2.09576			
SMYD1	Dr.21307	*smyd1a*				-2.47098			
SNRPC	Dr.18318	*snrpc*		-1.39496	-1.58672				-1.65606
SNRPEL1	Dr.117027	*snrpe*	-1.33143	-1.44041	-1.69165	-1.14326			-1.32162
TGIF1	Dr.139	*tgif1*							1.42374
THAP11	Dr.75932	*zgc:65871*		-1.15521	-1.28304				-1.45774
TMEM66	Dr.11537	*zgc:153286*		1.30393		1.33354			7.13477
TRAF6	Dr.74618	*traf6*							2.27878
TUBB2A	Dr.46855	*zgc:112335*	-1.1894			-2.91869			

Predicted miR-146 target genes from Table 3 are shown that showed infection-dependent expression changes in previously reported data sets (Stockhammer et al., 2009; van der Sar et al., 2009). Zebrafish UniGene IDs are from build #105. Significant fold change values (P<10-4) are indicated for infections of adult zebrafish with *M. marinum* E11 and Mma20 strains at 1 and 6 dpi (van der Sar et al., 2009) and for infection of zebrafish embryos with S. *typhimurium* at 5, 8 and 24 hpi (Stockhammer et al., 2009). Yellow - up-regulation, blue - down-regulation.

the immune system and development of cancer in human and mammalian models. We also determined that predicted target genes of the strongly infection-inducible miR-146 family are conserved between human and zebrafish, including key signaling intermediates of the TLR pathway that is pivotal for the innate immune response to bacterial infections (Hou et al., 2009; Perry et al., 2009; Taganov et al., 2006).

The set of commonly infection-responsive miRNAs included the highly conserved miRNAs, miR-20b, miR-21, miR-128, miR-146a/b, miR-152, and miR-181a/c. The mammalian counterparts of these miRNAs are differentially expressed in many types of cancer and have been implicated in tumor growth, invasion and metastasis (Debernardi et al., 2007; Evangelisti et al., 2009; Hiroki et al., 2009; Lei et al., 2009; Perry et al., 2009; Shi et al., 2008; Zhu et al., 2008). The mammalian counterparts of miR-21, miR-146 and miR-181 have also been linked to the immune system. Our study is the first that connects these miRNAs with mycobacterium and salmonella

infections. It is well known that the signaling pathways involved in the immune response and inflammation are strongly overlapping with those involved in the development of cancer, and it has previously been suggested that changes in expression of miRNAs might link these processes (Williams et al., 2008). In agreement, deregulated expression of miRNAs, for example miR-146a, has been observed in several inflammatory diseases, such as psoriasis, rheumatoid arthritis and osteoarthritis, as well as in different types of cancer such as papillary thyroid carcinoma, cervical cancer, ovarian cancer, breast cancer, pancreatic cancer and prostate cancer (Nakasa et al., 2008; Perry et al., 2009; Xia et al., 2009). The infection data reported here further support the notion that many miRNAs implicated in cancer processes are also connected to regulation of the immune response.

In connection to the immune response, the miR-146 family is one of the most interesting. We found that expression of miR-146a and b was commonly induced by infection in adult zebrafish as well as in embryos, indicating that the context of innate immunity is sufficient for the induction of the miR-146 miRNAs. This is consistent with results in human an animal models that have implicated the miR-146 family in regulation of innate immunity signaling pathways (Hou et al., 2009; Perry et al., 2009; Taganov et al., 2006). Induction of miR-146a and b was observed in human immune cells stimulated with microbial components (LPS) and proinflammatory mediators IL-1β, TNFα (Taganov et al., 2006). Furthermore, viral infections and *Helicobacter pylori* infections also increased expression of miR-146 (Belair et al., 2009; Hou et al., 2009; Motsch et al., 2007). NF-κB, the central transcription factor of the immune response, was shown to control miR-146 transcription (Taganov et al., 2006). Furthermore, two key intermediates of the common TLR and IL-1R signaling pathway, *TRAF6* and *IRAK1*, were shown to be potential molecular targets of the miR-146 family members (Taganov et al., 2006). Based on these results, the miR-146 miRNAs were proposed to fine-tune the innate immune response by negative feedback regulation of the TLR and IL-1R signaling pathways (Taganov et al., 2006). Consistent with this hypothesis, miR-146a expression directly or indirectly down-regulated the IL-1*β*-induced release of chemokines, IL-8 and RANTES (Williams et al., 2008) and the LPS-induced production of IFN-γ and nitric oxide (Dai et al., 2008). In addition to a role in fine-tuning TLR and IL-1R signaling, miR-146a was also proposed as a negative regulator of the RIG-I-dependent antiviral pathway by targeting *TRAF6, IRAK1*, and *IRAK2* (Hou et al., 2009). In contrast to these data, miR-146b was found to be down-regulated in an *in vivo* model of acute inflammation triggered by LPS infusion of healthy human volunteers, indicating that this miRNA might respond differently in leukocytes *in vivo* than in *in vitro* studies or that miR-146b levels might be oscillating during inflammation owing to a possible sensing mechanism (Schmidt et al., 2009). In our *in vivo* infection studies, miR-146a and b were consistently induced, but during early stages of mycobacterium infection in adult zebrafish we observed down-regulation of the miR-146 star sequences that currently have unknown function. Thus, it will be of interest to study the dynamic

pattern of miR-146 family members in greater detail using *in vivo* models. Besides its role in innate immunity, miR-146a was recently suggested to also modulate adaptive immunity (Schmidt et al., 2009). The zebrafish model, in which innate and adaptive immunity can be studied in separation, can therefore contribute to dissecting the diverse functions of the miR-146 family.

Similar as for the miR-146 family, we also observed increased expression of miR-21 primary miRNAs and star sequences during several conditions of infection in zebrafish embryos and adults. MiR-21 was also found to be up-regulated in cultured human gastric epithelial cells upon *H. pylori* infection, which is a major risk factor for gastric cancer (Zhang et al., 2008). Significantly increased levels of miR-21 were observed in several human gastric cancer tissues and cell lines as well, and the over-expression of miR-21 was shown to promote cell proliferation and migration and inhibit apoptosis in a gastric cancer cell line (Zhang et al., 2008). The up-regulation of miR-21 expression during *H. pylori* infection was found to be mediated by AP-1 and STAT3, which in turn are activated by NF-κB activation and IL-6 secretion in the gastric mucosa (Belair et al., 2009; Loffler et al., 2007). In addition, the miR-21 gene was found to be transactivated by the NF-κB p65 subunit following infection with *Cryptosporidium parvum*, a protozoan parasite that elicits strong innate immune responses of the gastrointestinal epithelium (Zhou et al., 2009). MiR-21 also showed elevated expression during Epstein-Barr virus infection (Cameron et al., 2008) and was shown to regulate interleukin-12 p35 expression during allergic airway inflammation. Thus, like for miR-146, the infection-responsiveness of miR-21 in zebrafish indicates an evolutionary conserved role for this miRNA family in the vertebrate immune response.

Another interesting miRNA family that we found to be regulated in zebrafish infection experiments is the miR-181 family that has an important role in normal haematopoiesis, B-cell differentiation and T-cell development in mammals and has been associated with aggressive leukemia (Chen and Lodish, 2005; Li et al., 2007; Pekarsky et al., 2006). Little is known of the role of miR-181 in bacterial infections, but LPS stimulation of macrophages positively regulated miRNA-181c expression in a protein kinase Akt1-dependent manner (Androulidaki et al., 2009). Like miR-146-a/b, the miR-181 family may function in negative regulation of the innate immune response, as expression levels of its members were inversely correlated with expression levels of predicted targets in TLR and IL-1R signaling (Marcucci et al., 2008). In our experiments miR-181a and c family members were down-regulated during bacterial infections of zebrafish, but showed opposite patterns of regulation in adult zebrafish. While miR-181a miRNAs were down-regulated at early stages and up-regulated at later stages of mycobacterium infection with two different strains, miR-181c expression showed the exact inverse pattern. In further studies the zebrafish-mycobacterium model may prove useful to dissect the functions of these members of the miR-181 family in the immune response.

Besides several hundreds of known miRNAs or predicted miRNAs with experi-

ment support (C1/C2 categories), our custom microarray used in the zebrafish infection studies contained probes for several thousands of additional hairpin sequences identified in the zebrafish genome sequence (C3 category). Many of these probes were infection responsive. Interestingly, there was a striking common regulation of C3 probes during mycobacterium infections of adult fish with the *M. marinum* E11 and Mma20 strains. We examined the response to these stains at 1 and 6 days after infection and observed a strongly overlapping pattern at both stages, despite the fact that there is a major difference in disease phenotypes elicited by the two strains. While E11-infected fish develop a chronic infection and show no symptoms of disease, Mma20-infected fish develop acute infection and are close to the end point of the disease at 6 dpi. The common regulation of C3 probes is in sharp contrast with what we have seen for gene expression where in a 2D cluster analysis of the data set the strains rather than the time points clustered together (van der Sar et al., 2009). In addition, the clustering pattern for the C3 category (Fig.3B) was very different from the clustering pattern of the known and C1/C2 category miRNAs (Fig. 3A) that showed much less overlap between the strains and time points. These observations warrant further study, as it is currently unknown if the C3 probes represent miRNAs or might correspond to other types of non-coding or coding RNAs.

The knowledge of target genes is crucial for the understanding of miRNA function. We focused on the infection-inducible miR-146 family to investigate possible conservation of target genes between zebrafish and mammals. As discussed above, mammalian miR-146a and b are thought to fine-tune TLR and IL-1R signaling by targeting *IRAK1* and *TRAF6*. Here we show the presence of conserved miR-146 target sites in the 3'UTRs of zebrafish *irak1* and *traf6*. Additionally, we found a miR-146-a/b target site in the 3'UTR of MyD88, which functions upstream of *irak1* and *traf6* as the common adaptor of several TLRs and IL-1R. However, targeting of MyD88 by miR-146 in zebrafish awaits experimental confirmation and appears not to be conserved in human or rodents. We have previously shown that several TLR pathway genes and downstream effector genes are induced during infections in zebrafish (Stockhammer et al., 2009; van der Sar et al., 2009). In addition, we observed that many negative regulatory genes of the TLR pathway, for example *irak3*, are induced during salmonella infection (Stockhammer et al., 2009). The induction of miR-146, as a putative negative regulator of several genes in TLR pathway, is in line with these findings. Together these gene expression and miRNA expression data suggest the importance of feed back mechanisms to properly control the inflammatory response that might in some cases be more damaging than the infection itself.

We attempted to predict other targets than the TLR pathway genes for miR-146. The available target prediction programs generate lists of hundreds to thousands of candidate target genes probably containing many false positive hits. We reasoned that predicted targets common between multiple species or distant species like human and zebrafish could be the most likely candidates for being true *in vivo* targets. We found 64 targets that appeared conserved between zebrafish and human, of which 27

were conserved also in mouse, rat and/or chicken. The majority of these genes were previously linked to immune response processes with gene ontology or pathway annotations related to apoptosis, regulation of NF-κB transcription factor activity, haemostasis, infections, T-cell development and function, and TLR signaling (Table 3). In addition, many of these genes are linked to tumor progression, consistent with the previously proposed link between miR-146 function in cancer and the immune response (Williams et al., 2008). One of the predicted targets conserved between human and zebrafish is interferon response factor 5 (*IRF5*), which is a critical transcription factor in the immune response downstream of TLR signaling and also has been linked to tumor suppression (Balkhi et al.; Hu and Barnes, 2009; Pandey et al., 2009; Yang et al., 2009). Targeting of human *IRF5* by miR-146 is supported by experimental data (Tang et al., 2009), consistent with the possibility that other predicted targets conserved between zebrafish and human may also be bona fide target genes. We also examined how the expression of the predicted miR-146 target genes responded to infection in zebrafish. In general, it cannot be predicted if and how the induction of miR-146 might affect transcript levels of its target genes. MiR-146 induction might affect only translation of its targets, or it might down-regulate their transcription or dampen their induction. However, it was notable that many of the predicted targets could be linked by expression changes to the infection process. This, together with the fact that most of the predicted targets have functional annotations linked to immunity, suggests that our list of conserved predicted targets may be a good starting point for experimental validation studies.

In conclusion, we have shown here that infection-responsive miRNAs are highly conserved between human and zebrafish. It is likely, as we have shown for the miR-146 family, that there is also considerable overlap in miRNA target genes, indicating that the zebrafish can be a useful model to dissect functions of miRNAs in the vertebrate immune system. Several miRNAs are not only responsive to infection in zebrafish adults, but also in embryos prior to the onset of adaptive immunity. In the embryo model, miRNAs can be easily over-expressed by micro-injection of synthetic miRNA duplexes, or knocked down using morpholinos (Schier and Giraldez, 2006). In addition, target protecting morpholinos can be used to experimentally validate predicted target genes *in vivo* (Choi et al., 2007). In future studies these strategies will be applied to analyze the function of infection-responsive miRNAs in the zebrafish embryo model.

Materials and methods

Bacterial strains and growth conditions

M. marinum strains E11 and Mma20 have been described before (van der Sar et al., 2004). Bacteria were grown at 30 °C in Middlebrook 7H9 medium supplemented with Middlebrook oleic acid-albumin-dextrose-catalase (BD Biosciences) and 0.05% Tween-80. *S. typhimurium* wild type strain SL1027 bacteria containing

DsRed expression vector pGMDs3 were used for the infection of zebrafish embryos (Stockhammer et al., 2009; van der Sar et al., 2003). Bacteria were freshly grown at 37 °C overnight on Luria-Bertani agar plates supplemented with 100 µg/ml carbenicillin. *M. marinum* and *S. typhimurium* bacterial numbers were determined by measuring the optical density at 600 nm and by plating and CFU determination. *L. casei* Shirota bacteria were taken from Yakult fermented milk drink (Yakult Europe B.V., Almere) specified to contain 10^8 bacteria/ml. Bacteria were collected by centrifugation and washed extensively with PBS.

Zebrafish husbandry and infection experiments

Zebrafish (*Danio rerio*) were handled in compliance with the local animal welfare regulations and maintained according to standard protocols (http://ZFIN.org). The infection experiment of adult male zebrafish with *M. marinum* E11 and Mma20 strains has been previously described (van der Sar et al., 2009). The analysis of miRNA expression profiles was done using powdered tissue from the same fish as used for mRNA expression analysis in the previous study (van der Sar et al., 2009). As before, three fish per strain and time point (1 and 6 days post infection) were used and PBS-injected fish were used as a control. Infection experiments at the embryonic stage were performed using mixed egg clutches from different pairs of wild type zebrafish. Embryos were grown at 28.5–30°C in egg water (60µg/ml Instant Ocean sea salts) and staged by morphological criteria (Kimmel et al., 1995). For the duration of bacterial injections embryos were kept under anaesthesia in egg water containing 0.02% buffered 3-aminobenzoic acid ethyl ester (tricaine; Sigma-Aldrich). For infection with *M. marinum* E11 embryos were staged at 48 h post fertilization (hpf). Bacteria were resuspended at desired concentrations in a 2% suspension of polyvinylpyrrolidone (PVP, average mol wt 40000, Calbiochem, San Diego) in PBS and approximately 1 nl of this mixture was injected into the yolk sac of 48 hpf embryos. PVP was chosen as the vehicle for yolk injections as it is commonly used in clinical medicine and experimental animals. The amount of injected M. marinum per embryo totalled 2,000 or 20,000 CFU. As we observed only a minor concentration effect on miRNA expression, these samples were treated as biological duplicates in the microarray analysis. *L. casei* Shirota was used as a non-pathogenic bacterium for comparison and injected in PVP suspension at similar concentrations as used for *M. marinum*. PVP-injected and uninjected embryos were employed as controls. Pools of 10–30 embryos from each group were taken at 72 h post infection (hpi). For the *S. typhimurium* infection study embryos were staged at 27 hpf and approximately 250 CFU of *S. typhimurium* bacteria in PBS were injected into the caudal vein close to the urogenital opening. As a control an equal volume of PBS was likewise injected. Triplicate pools of 20–40 embryos were collected at 8hpi. In all infection experiments, bacterial injections were controlled using a Leica MZ Fluo 3 stereomicroscope with epifluorescence attachment, a Femtojet microinjector (Eppendorf) and a micromanipulator with pulled microcapillary pipettes.

RT-PCR analysis of irak1

RNA from *S. typhimurium*-infected 1-day-old embryos at 24 hpi (Stockhammer et al. 2009) was used to amplify the 3' UTR of *irak1*. Traces of genomic DNA were removed by incubation with DNA-free solution (Ambion). RT-PCR analysis was performed using the SuperScript III One-Step RT-PCR System with Platinum Taq DNA Polymerase (Invitrogen) using forward (5'AGAAAGAAACGCACAACAGATA TGTTG3') and reverse (5'GTGCTGTTTAACAAGGCCGTAATC3') primers, located respectively at nucleotides 1-27 and 293-316 downstream of the *irak1* stop codon. To confirm the identity of the amplified sequence, the PCR product was cloned in pCRII-TOPO vector (Invitrogen) and sequenced using the sequencing service of ServiceXS (Leiden, The Netherlands).

Microarray design

Custom-designed 8x15k microarray slides were ordered from Agilent Technologies. The 15k custom design was obtained from Edwin Cuppen and Eugene Berezikov, Hubrecht Institute, Utrecht, The Netherlands and will be submitted into the Gene Expression Omnibus database. The 15k design contained a duplicate of 7604 probes of 60-oligonucleotide length. The probes consisted of 2x22 nucleotide sequences antisense to mature miRNAs separated by a spacer of 8 nucleotides (CGATCTTT) and with a second spacer with the same sequence at the end. From 7604 probes 546 were designed for left (5') and right (3') arms of the hairpins of zebrafish miRNAs that are currently known in miRBase. The remainder 7058 probes were classified into 3 categories: 62 probes into C1 category (novel confident miRNAs for which there is compelling experimental evidence), 54 probes into C2 category (candidate miRNAs with clear hairpin structure but less experimental evidence), and 6942 into C3 category (additional hairpins identified in the zebrafish genome sequence that are less likely to be miRNAs).

RNA isolation, labeling and hybridization

Adult fish and embryos for RNA isolation were snap frozen in liquid nitrogen and subsequently stored at −80 °C. Adult fish were homogenized in liquid nitrogen as described (van der Sar et al., 2009) and portions of 50–100 µg of powdered tissue were used for RNA extraction. Total RNA from adult or embryo sample was isolated using the miRNeasy Mini kit to preserve small RNA species. RNA labeling was carried out with miRCURY™ LNA microRNA, Hy3™/Hy5™ Power Labeling kit (Exiqon) using 1 µg of total RNA according to the manufacturer's instructions. RNA samples from mycobacterium-infected adult fish were labeled with Hy3 and hybridized against Hy5-labeled RNA samples from PBS-injected controls. For the yolk infection study of 2-day-old embryos, the RNA samples from mycobacterium-injected, L. casei Shirota injected, and PVP-injected embryos were labeled with Hy3 and hybridized against Hy5-labeled samples from uninjected controls. RNA samples from salmonella-infected 1-day-old embryos were labeled with Hy3 and hybridized

against Hy5-labeled RNA samples from PBS-injected controls. The dual colour hybridization of the microarray chips was performed according to Agilent protocol GE2-v5_95_Feb07 and GE2_105_Jan09 (www.Agilent.com) for two-color microarray-based gene expression analysis except that hybridization and washing was performed at 37 °C.

Data analysis

Microarray data were processed from raw data image files with Feature Extraction Software 10.1.1 and 10.5.1 (Agilent Technologies). Processed data were subsequently imported into Rosetta Resolver 7.2 (Rosetta Biosoftware) and subjected to default ratio error modeling. Ratio results from control vs. infected replicates were combined using the default ratio experiment builder. Significance cut-offs for the ratios of infected versus control were set at 1.5-fold change at $P \leq 10-4$. Two-dimensional hierarchical cluster analyses were performed with MultiExperiment Viewer version 4.4.1 (TM4 Software Development Team) settings for average linkage method with Euclidean distance.

Target prediction

To predict miR-146 family member target sites in TLR pathway genes *traf6*, *irak1* and *myd88* and rank these predictions we employed three different miRNA target site prediction methods: miRanda (www.microrna.org/) (John et al., 2004), RNAhybrid *(bibiserv.techfak.uni-bielefeld.de/rnahybrid/) (Rehmsmeier et al., 2004),* and TargetScan (www.targetscan.org/) (Lewis et al., 2003), and an integration algorithm for these three methods developed by Yanju Zhang and Fons Verbeek, Leiden Insititute for Advanced Computer Science. To predict further conserved targets of miR-146 family members in zebrafish, human, mouse, rat and chicken we used MicroCosm Targets (formerly known as miRBase Targets) (http://www.ebi.ac.uk/enright-srv/microcosm/htdocs/targets/v5/). The predicted targets from zebrafish, mouse, rat and chicken were converted to human Ensembl Gene ID homologues using the Biomart program of the Ensembl database. Targets which appeared across zebrafish and human or multiple other species were selected to study their involvement in biological pathways using the Reactome pathway database (www.reactome.org/), GeneALaCart (www.genecards.org) and AmiGO pathways (amigo.geneontology.org/cgi-bin/amigo/go.cgi).

Supplementary data

Supplementary data associated with this chapter can be found at http://apo.szbk.u-szeged.hu/transfer/A_ORDAS/.

Acknowledgments

We are grateful to Edwin Cuppen and Eugene Berezikov (Hubrecht Institute, Utrecht, The Netherlands) for the Agilent miRNA array design.

References

Androulidaki A., Iliopoulos D., Arranz A., Doxaki C., Schworer S., Zacharioudaki V., Margioris A. N., Tsichlis P. N. and Tsatsanis C. (2009) The kinase Akt1 controls macrophage response to lipopolysaccharide by regulating microRNAs. *Immunity* **31**, 220-31.

Balkhi M. Y., Fitzgerald K. A. and Pitha P. M. IKKalpha negatively regulates IRF-5 function in a MyD88-TRAF6 pathway. *Cell Signal* **22**, 117-27.

Baltimore D., Boldin M. P., O'Connell R. M., Rao D. S. and Taganov K. D. (2008) MicroRNAs: new regulators of immune cell development and function. *Nat Immunol* **9**, 839-45.

Belair C., Darfeuille F. and Staedel C. (2009) Helicobacter pylori and gastric cancer: possible role of microRNAs in this intimate relationship. *Clin Microbiol Infect* **15**, 806-12.

Bhaumik D., Scott G. K., Schokrpur S., Patil C. K., Campisi J. and Benz C. C. (2008) Expression of microRNA-146 suppresses NF-kappaB activity with reduction of metastatic potential in breast cancer cells. *Oncogene*.

Carissimi C., Fulci V. and Macino G. (2009) MicroRNAs: novel regulators of immunity. *Autoimmun Rev* **8**, 520-4.

Chen C. Z., Li L., Lodish H. F. and Bartel D. P. (2004) MicroRNAs modulate hematopoietic lineage differentiation. *Science* **303**, 83-6.

Chen C. Z. and Lodish H. F. (2005) MicroRNAs as regulators of mammalian hematopoiesis. *Semin Immunol* **17**, 155-65.

Choi W. Y., Giraldez A. J. and Schier A. F. (2007) Target protectors reveal dampening and balancing of Nodal agonist and antagonist by miR-430. *Science* **318**, 271-4.

Costinean S., Zanesi N., Pekarsky Y., Tili E., Volinia S., Heerema N. and Croce C. M. (2006) Pre-B cell proliferation and lymphoblastic leukemia/high-grade lymphoma in E(mu)-miR155 transgenic mice. *Proc Natl Acad Sci U S A* **103**, 7024-9.

Dai R., Phillips R. A., Zhang Y., Khan D., Crasta O. and Ahmed S. A. (2008) Suppression of LPS-induced Interferon-gamma and nitric oxide in splenic lymphocytes by select estrogen-regulated microRNAs: a novel mechanism of immune modulation. *Blood* **112**, 4591-7.

Davis J. M., Clay H., Lewis J. L., Ghori N., Herbomel P. and Ramakrishnan L. (2002) Real-time visualization of mycobacterium-macrophage interactions leading to initiation of granuloma formation in zebrafish embryos. *Immunity* **17**, 693-702.

Debernardi S., Skoulakis S., Molloy G., Chaplin T., Dixon-McIver A. and Young B. D. (2007) MicroRNA miR-181a correlates with morphological sub-class of acute myeloid leukaemia and the expression of its target genes in global genome-wide analysis. *Leukemia* **21**, 912-6.

Esquela-Kerscher A. and Slack F. J. (2006) Oncomirs - microRNAs with a role in cancer. *Nat Rev Cancer* **6**, 259-69.

Evangelisti C., Florian M. C., Massimi I., Dominici C., Giannini G., Galardi S., Bue M. C., Massalini S., McDowell H. P., Messi E., Gulino A., Farace M. G. and Ciafre S. A. (2009) MiR-128 up-regulation inhibits Reelin and DCX expression and reduces neuroblastoma cell motility and invasiveness. *Faseb J* **23**, 4276-87.

Garzon R., Calin G. A. and Croce C. M. (2009) MicroRNAs in Cancer. *Annu Rev Med* **60**, 167-79.

Hiroki E., Akahira J. I., Suzuki F., Nagase S., Ito K., Suzuki T., Sasano H. and Yaegashi N. (2009) Changes in microRNA expression levels correlate with clinicopathological features and prognoses in endometrial serous adenocarcinomas. *Cancer Sci*.

Hou J., Wang P., Lin L., Liu X., Ma F., An H., Wang Z. and Cao X. (2009) MicroRNA-146a feedback inhibits RIG-I-dependent Type I IFN production in macrophages by targeting TRAF6, IRAK1, and IRAK2. *J Immunol* **183**, 2150-8.

Hu G. and Barnes B. J. (2009) IRF-5 is a mediator of the death receptor-induced apoptotic signaling pathway. *J Biol Chem* **284**, 2767-77.

Hurst D. R., Edmonds M. D., Scott G. K., Benz C. C., Vaidya K. S. and Welch D. R. (2009) Breast cancer metastasis suppressor 1 up-regulates miR-146, which suppresses breast cancer metastasis. *Cancer Res* **69**, 1279-83.

John B., Enright A. J., Aravin A., Tuschl T., Sander C. and Marks D. S. (2004) Human MicroRNA targets. *PLoS Biol* **2**, e363.

Kannan M., Mohan K. V., Kulkarni S. and Atreya C. (2009) Membrane array-based differential profiling of platelets during storage for 52 miRNAs associated with apoptosis. *Transfusion* **49**, 1443-50.

Kim D. H., Saetrom P., Snove O., Jr. and Rossi J. J. (2008) MicroRNA-directed transcriptional gene silencing in mammalian cells. *Proc Natl Acad Sci U S A* **105**, 16230-5.

Kimmel C. B., Ballard W. W., Kimmel S. R., Ullmann B. and Schilling T. F. (1995) Stages of embryonic development of the zebrafish. *Dev Dyn* **203**, 253-310.

Kloosterman W. P., Lagendijk A. K., Ketting R. F., Moulton J. D. and Plasterk R. H. (2007) Targeted inhibition of miRNA maturation with morpholinos reveals a role for miR-375 in pancreatic islet development. *PLoS Biol* **5**, e203.

Kloosterman W. P., Steiner F. A., Berezikov E., de Bruijn E., van de Belt J., Verheul M., Cuppen E. and Plasterk R. H. (2006) Cloning and expression of new microRNAs from zebrafish. *Nucleic Acids Res* **34**, 2558-69.

Larson R. A. (2009) Micro-RNAs and copy number changes: New levels of gene regulation in acute myeloid leukemia. *Chem Biol Interact*.

Lehmann U., Hasemeier B., Christgen M., Muller M., Romermann D., Langer F. and Kreipe H. (2008) Epigenetic inactivation of microRNA gene hsa-mir-9-1 in human breast cancer. *J Pathol* **214**, 17-24.

Lei Z., Li B., Yang Z., Fang H., Zhang G. M., Feng Z. H. and Huang B. (2009) Regulation of HIF-1alpha and VEGF by miR-20b tunes tumor cells to adapt to the alteration of oxygen concentration. *PLoS One* **4**, e7629.

Lewis B. P., Shih I. H., Jones-Rhoades M. W., Bartel D. P. and Burge C. B. (2003) Prediction of mammalian microRNA targets. *Cell* **115**, 787-98.

Li Q. J., Chau J., Ebert P. J., Sylvester G., Min H., Liu G., Braich R., Manoharan M., Soutschek J., Skare P., Klein L. O., Davis M. M. and Chen C. Z. (2007) miR-181a is an intrinsic modulator of T cell sensitivity and selection. *Cell* **129**, 147-61.

Lindsay M. A. (2008) microRNAs and the immune response. *Trends Immunol* **29**, 343-51.

Loffler D., Brocke-Heidrich K., Pfeifer G., Stocsits C., Hackermuller J., Kretzschmar A. K., Burger R., Gramatzki M., Blumert C., Bauer K., Cvijic H., Ullmann A. K., Stadler P. F. and Horn F. (2007) Interleukin-6 dependent survival of multiple myeloma cells involves the Stat3-mediated induction of microRNA-21 through a highly conserved enhancer. *Blood* **110**, 1330-3.

Lu J., Getz G., Miska E. A., Alvarez-Saavedra E., Lamb J., Peck D., Sweet-Cordero A., Ebert B. L., Mak R. H., Ferrando A. A., Downing J. R., Jacks T., Horvitz H. R. and Golub T. R. (2005) MicroRNA expression profiles classify human cancers. *Nature* **435**, 834-8.

Lu L. F. and Liston A. (2009) MicroRNA in the immune system, microRNA as an immune system. *Immunology* **127**, 291-8.

Lu T. X., Munitz A. and Rothenberg M. E. (2009) MicroRNA-21 is up-regulated in allergic airway inflammation and regulates IL-12p35 expression. *J Immunol* **182**, 4994-5002.

Marcucci G., Radmacher M. D., Maharry K., Mrozek K., Ruppert A. S., Paschka P., Vukosavljevic T., Whitman S. P., Baldus C. D., Langer C., Liu C. G., Carroll A. J., Powell B. L., Garzon R., Croce C. M., Kolitz J. E., Caligiuri M. A., Larson R. A. and Bloomfield C. D. (2008) MicroRNA expression in cytogenetically normal acute myeloid leukemia. *N Engl J Med* **358**, 1919-28.

Medina P. P. and Slack F. J. (2008) microRNAs and cancer: an overview. *Cell Cycle* **7**, 2485-92.

Meijer A. H., Gabby Krens S. F., Medina Rodriguez I. A., He S., Bitter W., Ewa Snaar-Jagalska B. and Spaink H. P. (2004) Expression analysis of the Toll-like receptor and TIR domain adaptor families of zebrafish. *Mol Immunol* **40**, 773-83.

Meijer A. H., Verbeek F. J., Salas-Vidal E., Corredor-Adamez M., Bussman J., van der Sar A. M., Otto G. W., Geisler R. and Spaink H. P. (2005) Transcriptome profiling of adult zebrafish at the late stage of chronic tuberculosis due to Mycobacterium marinum infection. *Mol Immunol* **42**, 1185-203.

Motsch N., Pfuhl T., Mrazek J., Barth S. and Grasser F. A. (2007) Epstein-Barr virus-encoded latent membrane protein 1 (LMP1) induces the expression of the cellular microRNA miR-146a. *RNA Biol* **4**, 131-7.

Nakasa T., Miyaki S., Okubo A., Hashimoto M., Nishida K., Ochi M. and Asahara H. (2008) Expression of microRNA-146 in rheumatoid arthritis synovial tissue. *Arthritis Rheum* **58**, 1284-92.

O'Connell R. M., Taganov K. D., Boldin M. P., Cheng G. and Baltimore D. (2007) MicroRNA-155 is induced during the macrophage inflammatory response. *Proc Natl Acad Sci U S A* **104**, 1604-9.

Pandey A. K., Yang Y., Jiang Z., Fortune S. M., Coulombe F., Behr M. A., Fitzgerald K. A., Sassetti C. M. and Kelliher M. A. (2009) NOD2, RIP2 and IRF5 play a critical role in the type I interferon response to Mycobacterium tuberculosis. *PLoS Pathog* **5**, e1000500.

Pedersen I. and David M. (2008) MicroRNAs in the immune response. *Cytokine* **43**, 391-4.

Pekarsky Y., Santanam U., Cimmino A., Palamarchuk A., Efanov A., Maximov V., Volinia S., Alder H., Liu C. G., Rassenti L., Calin G. A., Hagan J. P., Kipps T. and Croce C. M. (2006) Tcl1 expression in chronic lymphocytic leukemia is regulated by miR-29 and miR-181. *Cancer Res* **66**, 11590-3.

Perry M. M., Williams A. E., Tsitsiou E., Larner-Svensson H. M. and Lindsay M. A. (2009) Divergent intracellular pathways regulate interleukin-1beta-induced miR-146a and miR-146b expression and chemokine release in human alveolar epithelial cells. *FEBS Lett* **583**, 3349-55.

Rehmsmeier M., Steffen P., Hochsmann M. and Giegerich R. (2004) Fast and effective prediction of microRNA/target duplexes. *Rna* **10**, 1507-17.

Schier A. F. and Giraldez A. J. (2006) MicroRNA function and mechanism: insights from zebra fish. *Cold Spring Harb Symp Quant Biol* **71**, 195-203.

Schmidt W. M., Spiel A. O., Jilma B., Wolzt M. and Muller M. (2009) In vivo profile of the human leukocyte microRNA response to endotoxemia. *Biochem Biophys Res Commun* **380**, 437-41.

Shi L., Cheng Z., Zhang J., Li R., Zhao P., Fu Z. and You Y. (2008) hsa-mir-181a and hsa-mir-181b function as tumor suppressors in human glioma cells. *Brain Res* **1236**, 185-93.

Sonkoly E. and Pivarcsi A. (2009) Advances in microRNAs: implications for immunity and inflammatory diseases. *J Cell Mol Med* **13**, 24-38.

Sonkoly E., Stahle M. and Pivarcsi A. (2008) MicroRNAs and immunity: novel players in the regulation of normal immune function and inflammation. *Semin Cancer Biol* **18**, 131-40.

Stanczyk J., Pedrioli D. M., Brentano F., Sanchez-Pernaute O., Kolling C., Gay R. E., Detmar M., Gay S. and Kyburz D. (2008) Altered expression of MicroRNA in synovial fibroblasts and synovial tissue in rheumatoid arthritis.

Arthritis Rheum **58**, 1001-9.
Stockhammer O. W., Zakrzewska A., Hegedus Z., Spaink H. P. and Meijer A. H. (2009) Transcriptome profiling and functional analyses of the zebrafish embryonic innate immune response to Salmonella infection. *J Immunol* **182**, 5641-53.
Taganov K. D., Boldin M. P. and Baltimore D. (2007) MicroRNAs and immunity: tiny players in a big field. *Immunity* **26**, 133-7.
Taganov K. D., Boldin M. P., Chang K. J. and Baltimore D. (2006) NF-kappaB-dependent induction of microRNA miR-146, an inhibitor targeted to signaling proteins of innate immune responses. *Proc Natl Acad Sci U S A* **103**, 12481-6.
Tang Y., Luo X., Cui H., Ni X., Yuan M., Guo Y., Huang X., Zhou H., de Vries N., Tak P. P., Chen S. and Shen N. (2009) MicroRNA-146A contributes to abnormal activation of the type I interferon pathway in human lupus by targeting the key signaling proteins. *Arthritis Rheum* **60**, 1065-75.
Tili E., Michaille J. J., Cimino A., Costinean S., Dumitru C. D., Adair B., Fabbri M., Alder H., Liu C. G., Calin G. A. and Croce C. M. (2007) Modulation of miR-155 and miR-125b levels following lipopolysaccharide/TNF-alpha stimulation and their possible roles in regulating the response to endotoxin shock. *J Immunol* **179**, 5082-9.
Traver D., Herbomel P., Patton E. E., Murphey R. D., Yoder J. A., Litman G. W., Catic A., Amemiya C. T., Zon L. I. and Trede N. S. (2003) The zebrafish as a model organism to study development of the immune system. *Adv Immunol* **81**, 253-330.
van der Sar A. M., Abdallah A. M., Sparrius M., Reinders E., Vandenbroucke-Grauls C. M. and Bitter W. (2004a) Mycobacterium marinum strains can be divided into two distinct types based on genetic diversity and virulence. *Infect Immun* **72**, 6306-12.
van der Sar A. M., Appelmelk B. J., Vandenbroucke-Grauls C. M. and Bitter W. (2004b) A star with stripes: zebrafish as an infection model. *Trends Microbiol* **12**, 451-7.
van der Sar A. M., Musters R. J., van Eeden F. J., Appelmelk B. J., Vandenbroucke-Grauls C. M. and Bitter W. (2003) Zebrafish embryos as a model host for the real time analysis of Salmonella typhimurium infections. *Cell Microbiol* **5**, 601-11.
van der Sar A. M., Spaink H. P., Zakrzewska A., Bitter W. and Meijer A. H. (2009) Specificity of the zebrafish host transcriptome response to acute and chronic mycobacterial infection and the role of innate and adaptive immune components. *Mol Immunol* **46**, 2317-32.
Vasudevan S., Tong Y. and Steitz J. A. (2007) Switching from repression to activation: microRNAs can up-regulate translation. *Science* **318**, 1931-4.
Wienholds E., Kloosterman W. P., Miska E., Alvarez-Saavedra E., Berezikov E., de Bruijn E., Horvitz H. R., Kauppinen S. and Plasterk R. H. (2005) MicroRNA expression in zebrafish embryonic development. *Science* **309**, 310-1.
Williams A. E., Perry M. M., Moschos S. A., Larner-Svensson H. M. and Lindsay M. A. (2008) Role of miRNA-146a in the regulation of the innate immune response and cancer. *Biochem Soc Trans* **36**, 1211-5.
Xia H., Qi Y., Ng S. S., Chen X., Li D., Chen S., Ge R., Jiang S., Li G., Chen Y., He M. L., Kung H. F., Lai L. and Lin M. C. (2009) microRNA-146b inhibits glioma cell migration and invasion by targeting MMPs. *Brain Res* **1269**, 158-65.
Xiao C., Calado D. P., Galler G., Thai T. H., Patterson H. C., Wang J., Rajewsky N., Bender T. P. and Rajewsky K. (2007) MiR-150 controls B cell differentiation by targeting the transcription factor c-Myb. *Cell* **131**, 146-59.
Yang L., Belaguli N. and Berger D. H. (2009) MicroRNA and colorectal cancer. *World J Surg* **33**, 638-46.
Zhang Z., Li Z., Gao C., Chen P., Chen J., Liu W., Xiao S. and Lu H. (2008) miR-21 plays a pivotal role in gastric cancer pathogenesis and progression. *Lab Invest* **88**, 1358-66.
Zhou R., Hu G., Liu J., Gong A. Y., Drescher K. M. and Chen X. M. (2009) NF-kappaB p65-dependent transactivation of miRNA genes following Cryptosporidium parvum infection stimulates epithelial cell immune responses. *PLoS Pathog* **5**, e1000681.
Zhu S., Wu H., Wu F., Nie D., Sheng S. and Mo Y. Y. (2008) MicroRNA-21 targets tumor suppressor genes in invasion and metastasis. *Cell Res* **18**, 350-9.

Chapter 5

Liver tumor-related microRNA expression is conserved between zebrafish and human

This chapter is based on:
Ordas A, He S, Gong Z, Zhang Y, Verbeek FJ, Mink M, Spaink HP, Snaar-Jagalska B.E., Meijer AH. Liver tumor-related microRNA expression is conserved between zebrafish and human, in preparation

Abstract

The use of zebrafish models for liver cancer has been supported by a strong conservation of gene expression signatures between zebrafish and human liver tumors. The expression levels of genes involved in many cellular processes are modulated by the activity of microRNAs. Aberrant expression of these small non-coding RNAs has been associated with many diseases, including cancer. In this study we compared the miRNA transcriptomes of zebrafish liver tumors and human hepatocellular carcinoma (HCC). Microarray analysis of five individual zebrafish with carcinogen-induced liver tumors indicated a consistent tumor-specific expression signature for a large set of known and predicted miRNAs. The most notable similarities between zebrafish and human liver cancer were the common up-regulation of miR-21, miR-23a, miR-146a/b, miR-221 and miR-222, and the common down-regulation of miR-1 and miR-122. In addition, several miRNAs not previously linked to HCC but linked to several other types of human cancer showed differential expression in zebrafish liver tumors. A subset of the liver tumor-related miRNAs, including members of the miR-21 and miR-146 families, was also responsive to *Mycobacterium marinum* infection in zebrafish, suggesting a common role in cancer and the immune response. Finally, we investigated the possible conservation of miRNA target genes and found that several putative target genes of the miR-1, miR-146 and miR-221/miR-222 families are conserved between human and zebrafish. The conserved expression signatures of these tumor-related miRNAs and the involvement of their predicted target genes in cancer-related processes and signal transduction pathways corroborate the similarity between zebrafish and human liver cancer.

Introduction

Zebrafish is increasingly used as a model organism for cancer research (Amatruda et al., 2002; Berghmans et al., 2005; Feitsma and Cuppen, 2008; Lam and Gong, 2006). Several types of cancers have been studied in zebrafish as they can spontaneously develop almost any type of tumors with notable similarities to human and mammalian tumors. The ease of performing tumor transplantations and visualization of tumor invasiveness stimulated the development of zebrafish cancer models using forward and reverse genetic screens and transgenesis methods.

Recently, zebrafish has been used as a model species to study liver cancer. It has been shown that zebrafish and human liver tumors share molecular, biochemical and cellular similarities (Chu and Sadler, 2009; Lam and Gong, 2006; Lam et al., 2006; Ung et al., 2009). They have largely overlapping expression profiles of genes involved in cell cycle/proliferation, apoptosis, DNA replication and repair, metastasis and cell adhesion, cytoskeletal organization and cell motility, and RNA processing and protein synthesis. Differential expression of genes in these processes is consistent with the characteristics of human liver cancer such as defective apoptosis, uncon-

trolled proliferation, limitless replication and loss of tissue-specific functions (Chu and Sadler, 2009; Lam and Gong, 2006; Lam et al., 2006).

MicroRNAs (miRNA) are highly conserved small non-coding RNAs that control gene expression by regulating mRNA stability and translation. They play crucial roles in essential physiological functions including cell cycle, differentiation, development, and metabolism (Schmidt et al., 2009; Sonkoly and Pivarcsi, 2009). A large number of recent studies support that miRNAs are important regulators of the immune system and that specific alterations of their expression contribute to pathogenesis of cancer and other diseases (Calin et al., 2002; Garofalo et al., 2008; Garzon et al., 2009; Lu et al., 2005; Lu and Liston, 2009; Medina and Slack, 2008). More than 50% of miRNA genes were found to reside in cancer-associated genomic regions (Calin et al., 2004).

Differentially expressed miRNAs in cancer can act either as direct causatives of tumorigenesis (oncogenic and tumor suppressor oncomirs) or can be indirectly involved in genomic, epigenomic or physiological changes associated with cancer (Esquela-Kerscher and Slack, 2006). Recently, a significant group of miRNAs have been found to be expressed in the liver and to modulate several liver functions (Chen, 2009). Their deregulation may play a crucial role in liver diseases including hepatocellular carcinoma (HCC). HCC is the most common malignancy in the liver and one of the most common causes of death from cancer triggered either by viral (hepatitis B and C) or nonviral (alcohol and aflatoxin B1) agents. Revealing alterations in miRNA expression and uncovering their mRNA targets is important to our understanding of carcinogenesis and could lead to novel therapeutic strategies based on the modulation of miRNA activity. Using a murine liver cancer model, it was recently shown that systemic administration of miR-26a, a miRNA showing reduced expression in HCC cells, could suppress tumorigenesis (Kota et al., 2009). In addition to this miRNA replacement strategy, the application of synthetic inhibitors (antagomirs) of miRNAs up-regulated in HCC may prove to be a promising approach to liver cancer treatment (Krutzfeldt et al., 2005; Pineau et al.; Tsai et al., 2009).

While numerous studies have uncovered miRNA profiles of different cancer types, their exact functions are not well understood. MiRNAs that have been linked to the progression of HCC and metastasis among others include miR-10, miR-15a, miR-21, miR-26a, miR-29a, miR-34a, miR-101, miR-122, miR-146, miR-199, miR-221/222, and miR-224 (Budhu et al., 2008; Chen, 2009; Huang et al., 2008; Ladeiro et al., 2008; Murakami et al., 2006; Su et al., 2009; Varnholt et al., 2008). Several miRNAs that are differentially expressed in HCC show similar alterations in other cancers, suggesting that there are common miRNAs regulating basic processes of carcinogenesis. Each miRNA is predicted to regulate several hundreds of target mRNAs. The identification of true *in vivo* targets is crucial for understanding of miRNA function. A number of targets of deregulated miRNAs in HCC and other types of cancer have now been experimentally validated, including several oncogenes and tumor suppressor genes, inhibitors of apoptosis, cell cycle regulators and genes involved in cell

proliferation, migration and invasion (Bueno et al., 2008; Mann et al., 2007; Medina and Slack, 2008; Mott, 2009; Qin and Tang, 2002).

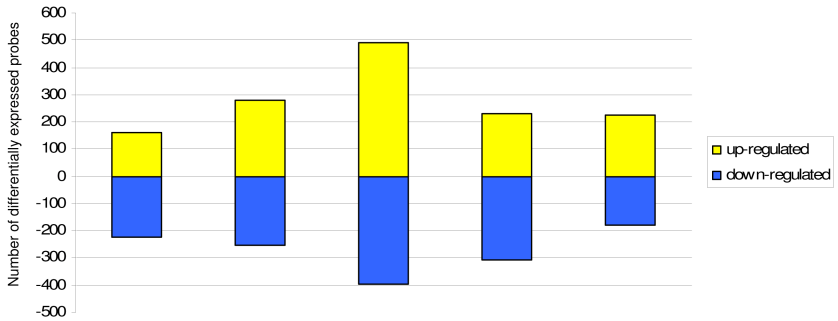
Here, we analyzed the miRNA transcriptome of zebrafish liver tumors and compared our data set with miRNAs previously detected in human liver cancer. Furthermore, we determined the overlap between the liver cancer-related miRNAs and miRNAs identified in previous infection studies of zebrafish (chapter 4). Finally, we compiled experimentally validated targets of differentially expressed tumor-related miRNAs to shed more light on their role in cancer mechanisms. With these studies we have defined tumor and immune-related miRNA marker sets and predicted targets conserved between human and zebrafish.

Results

Differentially expressed miRNAs in liver tumor samples from zebrafish

Previously, it has been shown that gene expression profiles between zebrafish and human liver tumors are strongly overlapping (Lam and Gong, 2006). In order to determine if miRNA expression is also conserved between zebrafish and human liver tumors, we obtained samples from these previous studies and analyzed them using a custom-designed Agilent 8x15k microarray for relative quantification of miRNA expression. We compared liver tissues from 5 fish with liver tumors induced by carcinogen (DMBA) treatment with normal liver tissues from 5 healthy control fish. In the five biological replicates the numbers of differentially expressed probes showing more than 1.5-fold induction or repression ($P < 1.00E{-}5$) ranged between 385 and 888 (Fig. 1). The percentage of up-regulated probes (45%-60%) was similar to that of down-regulated probes in all replicates (Fig.1). The custom-designed zebrafish microarray platform contained probes for both known and predicted miRNAs, the latter classified into three categories: C1 - novel confident miRNAs for which there is compelling experimental evidence, C2 - candidate miRNAs with clear hairpin structure but less experimental evidence, and C3 - additional hairpins identified in the zebrafish genome sequence that are less likely to be miRNAs. Approximately one third of the differentially expressed probes corresponded to known miRNAs, even though these represent less than one tenth of the total number of probes on the microarray (546 out of 7604 probes). The largest group of probes on the microarray belongs to the C3 category (6942 probes) and between 4-9% of these probes showed differential expression in the microarray analysis. In the smaller C1 (62 probes) and C2 (54 probes) categories only a few probes showed altered expression in the replicate liver tumor samples (C1: 4 to 8; C2: 1 to 5). Next, we determined the overlap of differentially expressed probes between the biological replicates (Fig. 2). Taking together the known miRNA probes and the C1/C2 probes for experimentally supported miRNA candidates, we found that 72 probes were consistently changed in at least 4 out of 5 biological replicates, of which 26 probes showed differential expression in all 5 replicates. A good overlap was also observed for the probes in the C3 category of putative

A



B

	#1	#2	#3	#4	#5
known up	59	104	113	106	130
known down	40	32	90	36	17
C1 up	2	2	5	4	4
C1 down	3	2	3	2	2
C2 up	0	1	1	2	1
C2 down	4	3	2	3	0
C3 up	101	174	373	120	90
C3 down	176	214	301	267	161

Figure 1. Differentially expressed miRNA probes in zebrafish liver tumor samples. (A) Total number of up-regulated and down-regulated probes in five biological replicates of healthy liver versus liver tumor. (B) Number of up-regulated and down-regulated probes in four annotation categories: known miRNAs, and predicted miRNAs, C1, C2, and C3, with decreasing level of confidence as further detailed in the text. The total number of probes on the array per category is: 546 probes in known, 62 probes in C1, 54 probes in C2 and 6942 probes in C3 category.

miRNAs without experimental support, where 140 probes were changed in at least 4 out of 5 biological replicates. In conclusion, our microarray analysis of five individual fish with carcinogen-induced liver tumors indicated a consistent tumor-specific expression signature for a large set of known and predicted miRNAs.

Comparison of miRNA regulation between zebrafish and human

To study the miRNA transcriptome similarities between zebrafish and human liver tumors we compiled in Table 1 those miRNAs that were differentially expressed in at least 4 replicates of the zebrafish liver tumor study and investigated which of these miRNAs have previously been related to human hepatocellular carcinoma (HCC) or to other types of human cancer. Furthermore, we also included in Table 1 all other human miRNAs that were previously linked to HCC in the literature and indicated the number of zebrafish liver tumor replicates in which these miRNAs were differentially expressed. The results of the comparative analysis in Table 1 are summarized in Figure 3, showing that several miRNAs were commonly regulated in zebrafish and human liver cancer. MiRNAs previously reported to show increased expression in HCC and that we also found up-regulated in zebrafish liver tumors included miR-

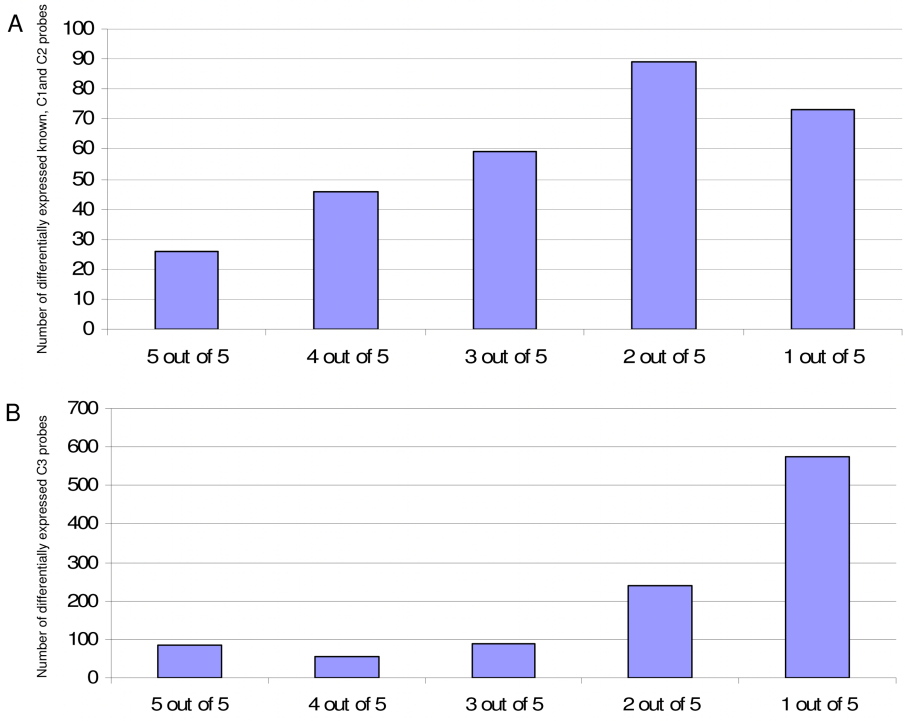


Figure 2. Overlap of differentially expressed miRNA probes between biological replicates of healthy liver versus liver tumor. (A) Overlap of probes in the combined categories of known miRNAs and C1 and C2 miRNAs that include candidate miRNAs supported by experimental evidence. (B) Overlap of probes in the C3 category of putative miRNAs not supported by experimental evidence. Probes were sorted according to the number of replicates in which they were differentially expressed.

21, miR-23a, miR-146a, miR-146b, miR-221 and miR-222, which were consistently changed in at least 4 replicates of our microarray study. Furthermore, miR-1 and miR-122, for which decreased expression was reported in HCC, were consistently down-regulated in zebrafish tumors. Other miRNAs increased (miR-10a/b, miR-18c, miR-93, miR-125b and miR-181a/b, miR-338) or decreased (let-7a, miR-101b, miR-148 and miR-199) in HCC, were also changed in the same direction in zebrafish liver tumors, but only in 2 or 3 out of the 5 replicates. We also found miRNAs (including miR-9, miR-15a, miR-29a/b and miR-34 changed in 4-5 replicates, and miR-125a and miR-223 changed in 2-3 replicates) for which the observed direction of change in zebrafish liver tumors was opposite to changes reported for HCC (Figure 3). In addition, several miRNAs not previously linked to HCC showed differential expression in at least 4 replicates of the zebrafish liver tumor study, including miR-15b, miR-16-a/b, miR-27a/b, miR-132, miR-153a/b/c, miR-193a/b, miR-210, miR-212, miR-217, miR-

Table 1. Liver cancer-related miRNA probes.

Zebrafish miRNA	Expression in zf liver cancer	Human miRNA	Expression in HCC	Other associations to cancer	Functional links	References
dre-let-7a (let-7a-1_l, let-7a-2_l, let-7a-3_l, let-7a-4_l, let-7a-6_l)	down (2)	hsa-let-7a	down	down in lung squamous carcinoma	tumor suppression, proliferation, cell death	Mott JL 2009 Yang J et al 2008 Kutay H et al 2006
dre-miR-1 (miR-1-1_r, miR-1-2_r)	down (5)	hsa-miR-1	down	down in lung cancer, colon cancer	apoptosis	Datta J et al 2008 Nasser MW et al 2008 Melkamu T et al 2009 Tang Y et al 2009 Sarver AL et al 2009
dre-miR-7a* (miR-7a-2_r)	down (4)	hsa-miR-7-1*, hsa-miR-7-2*		hsa-mir-7: down in glioblastoma, breast cancer, urothelial carcinomas, lung cancer	hsa-mir-7: tumor suppression, aggressiveness	Kefas B et al 2008 Foekens JA et al 2008 Veerla S et al 2009 Webster RJ et al 2009 Reddy SD et al 2008
dre-miR-7b (miR-7b_l)	up (3) / down (1)	hsa-miR-7	-			
dre-miR-9 (miR-9-1_l, miR-9-3_l, miR-9-6_l)	down (4)	hsa-miR-9	up	down in gastric carcinoma	metastasis, invasion	Zhao Y et al 2009 Hao-Xiang T et al 2009 Luo H et al 2009 Budhu et al 2008
dre-mir-10a (mir-10a_l)	up (2)	hsa-miR-10a	up	up in pancreatic cancer, urothelial carcinoma, down in CML, CLL, AML,	metastasis	Varnholt H et al 2008 Weiss FU et al 2009 Veerla S et al 2009 Lund AH et al 2009
dre-mir-10b (mir-10b-1_l, mir-10b-2_l)	up (2) / down (1)	hsa-miR-10b	up	up in breast cancer, glioma	cell migration, metastasis, invasion	Ladiero et al 2008 Sasayama T et al 2009 Ma L et al 2007 Lund AH et al 2009
dre-miR-10d (miR-10d-1 _l)	up (3) / down (1)	-	-	-	-	
dre-miR-15a (miR-15a-1_l)	up (5)	hsa-miR-15a	down,deletion	up in APL, down in CLL, prostate cancer, adenomas	dedifferentiation, chromosomal instability, apoptosis, tumor suppression	Budhu et al 2008 Skawran B et al 2008 Aqeilan RI et al 2009
dre-miR-15b (miR-15b_l)	up (4)	hsa-miR-15b		up in melanomas, colorectal cancer, cervical cancer, APL, pancreatic cancer, down in gastric cancer	proliferation, apoptosis	Satzger I etal 2009 Xia L et al 2008 Wang X et al 2008 Zhang Y et al 2009
dre-miR-15c (miR-15c_l)	up (4)	-	-	-	-	
dre-miR-16a (miR-16a_l)	up (5)	hsa-mir-16	-	up in APL, cervical cancer, down in CLL	apoptosis, tumor suppression	Aqeilan RI et al 2009 Calin GA et al 2002 Xia L et al 2008 Wang X et al 2008
dre-miR-16b (miR-16b_l)	up (5)					

Zebrafish miRNA	Expression in zf liver cancer	Human miRNA	Expression in HCC	Other associations to cancer	Functional links	References
-	-	hsa-miR-17-92 cluster	up	up in lung cancer, colorectal adenoma, ALL	proliferation, angiogenesis	Li Y et al 2009 Liu WH 2009 Connolly E et al 2008 Kutay H et al 2006 Ebi H et al 2009 Zanette DL et al 2007
dre-miR-18c (miR-18c_l)	up (2) / down (3)	hss-mir-18b	up	-	tumor suppression	Li Q et al 2009 Li Y et al 2009
dre-miR-21 (miR-21-1_l, miR-21-2_l)	up (5)	hsa-miR-21	hsa-miR-21: up	hsa-mir-21: up in glioma, colon cancer, breast cancer, gastric cancer	hsa-mir-21: tumor growth, invasion, metastasis, cell cycle progression, cell death, angiogenesis	Gramantieri L et al 2008 Connolly E et al 2008 Ladeiro et al 2008 Meng F et al 2007 Kutay H et al 2006 Zhang et al 2008
dre-miR-21* (miR-21-1_r)	up (5)	hsa-miR-21*				
dre-miR-23a (miR-23a-1_r, miR-23a-3_r)	up (4)	hsa-miR-23a	up	up in bladder cancer	cell growth	Kutay H et al 2006 Kulshreshtha R et al 2007 Huang S et al 2008 Gottardo F et al 2007
dre-miR-24* (miR-24-3_l)	up (4)	hsa-miR-24	-	up in kaposi sarcoma, hodgkin lymphoma, deregulated in colorectal cancer	erythropoiesis, cellular transformation, cellular proliferation	O'Hara AJ et al 2009 Gibcus JH et al 2009 Mishra PJ et al 2009
-	-	hsa-mir-25	up	up in gastric cancer, esophageal adenocarcinoma, prostate cancer	proliferation, tumorigenicity	Li Y et al 2009 Kim YK et al 2009 Kan T et al 2009 Ambs S et al 2009
dre-mir-26a (mir-26a-1_l, mir-26a-2_l, mir-26a-3_l	up (1) / down (1)	hsa-miR-26a	down/up	up in glioma, colorectal cancer, down in thyroid anaplastic carcinomas, Burkitt lymphoma	cell cycle regulation, tumor suppression	Ji J et al 2009 Kota J et al 2009 Huse JT et al 2009 Xi Y et al 2006
dre-miR-27a (miR-27a_r)	up (5)	hsa-miR-27a	-	up in renal cell carcinoma, down in breast cancer	differentiation	Gottardo F et al 2007 Feng J et al 2009 Tsuchiya Y et al 2006 Lin Q et al 2009
dre-miR-27b (miR-27b_r)	up (5)	hsa-miR-27b				
dre-miR-27d (miR-27d_r)	up (4)	-				
dre-miR-27e (miR-27e_r)	up (5)	-				
dre-miR-29a (miR-29a-1_r, miR-29a-2_r)	up (4)	hsa-miR-29a	down	up in mesenchymal, metastatic mouse breast cells, AML, down in CLL, lung cancer,	apoptosis	Su H et al 2009 Ura S et al 2009 Xiong Y et al 2009 Mott JL et al 2007 Park SY et al 2009 Mraz M et al 2009 Calin GA et al 2007 Fabbri M et al 2007 Garzon R et al 2008 Li W et al 2008
dre-miR-29b (miR-29b-1_r, miR-29b-2_r, miR-29b-3_r)	up (4)	hsa-miR-29b				
-	-	hsa-miR-29c				

Zebrafish miRNA	Expression in zf liver cancer	Human miRNA	Expression in HCC	Other associations to cancer	Functional links	References
dre-miR-34 (miR-34_l)	up (4)	hsa-mir-34a	down	down in melanomas, lung cancer, osteosarcoma	maintenance and survival of cancer stem cells, tumour suppression, apoptosis, cell-cycle arrest, senescence	Satzger I etal 2009 Ji Q et al 2009 Christoffersen NR et al 2009 Hermeking H. et al 2009 Yamakuchi M, Lowenstein CJ 2009 Pogribny IP et al 2007 Gallardo E et al 2009 Li N et al 2009 Pineau P et al 2009
dre-mir-93 (mir-93_l)	up (2)	hsa-miR-93	up	up in gastric cancer, dult T-cell leukemia, ovarian carcinoma, esophageal adenocarcinoma, prostate cancer	tumor suppression	Li W et al 2008 Kim YK et al 2009 Yeung ML et al 2008 Nam EJ et al 2008 Kan T et al 2009 Ambs S et al 2008
dre-mir-101a (mir-101a_r) dre-miR-101b (miR-101b_r)	up (1) / down (1) down (2)	hsa-miR-101	down	down in endometrial serous adenocarcinomas, colon cancer	apoptosis, tumorigenicity	Mott JL 2009 Su H et al 2009 Li W et al 2008 Hiroki E et al 2009 Strillacci A et al 2009
dre-miR-103 (miR-103_r)	up (3) / down (1)	hsa-miR-103		up in esophageal squamous cell carcinoma	-	Guo Y et al 2008
-	-	hsa-miR-106b	up	up in esophageal adenocarcinoma, gastric cancer, prostate cancer	progression, proliferation	Li W et al 2008 Kan T et al 2009 Kim YK et al 2009 Ambs S et al 2008
dre-miR-107 (miR-107_r)	up (3) / down (1)	hsa-miR-107		up in esophageal squamous cell carcinoma, pancreatic cancer	-	Guo Y et al 2008 Lee KH et al 2009
dre-miR-122 (miR-122_l)	down (4)	hsa-miR-122	down	-	metastasis, tumor suppression	Bai S et al 2009 Coulouarn C et al 2009 Tsai WC et al 2009 Lin CJ et al 2008 Gramantieri L et al 2007 Kutay H et al 2006
dre-mir-125a (mir-125a-1_l, mir-125a-2_l)	up (2)	hsa-miR-125a	down	breast cancer, down in ovarian cancer	tumor suppression, apoptosis	Mott JL 2009 Meng F et al 2007 Guo X et al 2009 Cowden Dahl KD et al 2009
dre-mir-125b (mir-125b-1_l, mir-125b-2_l, mir-125b-3_l) dre-mir-125b* (mir-125b-2_r, mir-125b-3_r)	up (2) up (3)	hsa-miR-125b hsa-miR-125b-2*	hsa-miR-125b: up	hsa-miR-125b: up in leukemia, down in oral squamous cell carcinoma, bladder cancer, breast cancer	hsa-miR-125b: proliferation, apoptosis	Varnholt H et al 2008 Meng F et al 2007 Xia HF et al 2009 Henson BJ et al 2009 Ichimi T et al 2009 Hui AB et al 2009 Gefen N et al 2010

Zebrafish miRNA	Expression in zf liver cancer	Human miRNA	Expression in HCC	Other associations to cancer	Functional links	References
dre-mir-126* (mir-126_l)	up (1)	hsa-miR-126	down	down in breast cancer, colon cancer, B-lineage ALL, lung cancer	invasion, cell cycle	Wong QW et al 2008 Liu B et al 2009 Fulci V et al 2009 Zhang J et al 2008 Guo C et al 2008 Crawford M et al 2008
dre-mir-126 (mir-126_r)	up (1)	hsa-miR-126*	down	down in lung cancer	invasion	Ladiero et al 2008
dre-miR-128* (miR-128-2_l)	up (4)	-	-	hsa-mir-128: prostate cancer, up in acute lymphoblastic leukemia, down in glioma	hsa-mir-128: invasion	Khan AP et al 2009 Fulci V et al 2009 Godlewski J et al 2009, Zhang Y et al 2009, Ciafrè SA et al 2005 Mi S, et al 2007
dre-miR-132 (miR-132-1_r, miR--132-2_r)	up (5)	hsa-miR-132	-	up in squamous cell carcinoma of tongue, down in adult T-cell leukemia	innate immunity	Bellon et al 2009 Wong et al 2008
dre-miR-135c (miR-135c-1_l, miR-135c-3_l)	up (3) / down (1)	hsa-135a	-	up in colorectal adenomas, megakaryoblastic leukemia		Nagel R et al 2008 Garzon R et al 2006
dre-miR-143 (miR-143-1_l)	up (3) / down (1)	hsa-miR-143*	hsa-mir-143: up	hsa-miR-143: colorectal cancer, T-cell leukemia, down in B-cell malignancies, gastric cancers	hsa-miR-143: metastasis, apoptosis	Zhang X et al 2009 Ng EK et al 2009 Akao Y et al 2009, 2007 Takagi T et al 2009 Clapé C et al 2009
dre-miR-146a (miR-146a_l)	up (5)	hsa-miR-146a	hsa-miR-146a: up, SNP association	hsa-miR-146a: up in breast cancer, SNP in thyroid carcinoma, breast/ovarian cancer, down in pancreatic cancer	innate immunity, invasiveness	Xu T et al 2008 Belair C et al 2009 Shen J et al 2008 Huang YS et al 2008 Li Y et al 2010 Pogribny IP et al 2010 Jazdzewski K et al 2008
dre-miR-146a* (miR-146a_r)	up (4)	hsa-miR-146a*				
dre-miR-146b (miR-146b-1_l)	up (5)	hsa-miR-146b-5p	up	up in lung cancer, leukemia, down in adult T-cell leukemia, breast cancer		Bellon et al 2009 Hurst DR et al 2009 Huang YS et al 2008 Patnaik SK et al 2010 Fulci V et al 2009
dre-mir-148* (mir-148_l)	down (1)	hsa-miR-148a*/b*	hsa-miR-148a, b: down	hsa-miR-148a, b: CLL	hsa-miR-148a, b: metastasis	Budhu et al 2008 Visone R et al 2009 Lujambio A et al 2008 Zhao Y et al 2009
dre-mir-148 (mir-148_r)	down (2)	hsa-miR-148a, b				
dre-miR-153a (miR-153a_r)	up (4)	hsa-miR-153	-	down in glioblastoma	-	Xu J et al 2009
dre-miR-153b (miR-153b_r)	up (4)	hsa-miR-153				
dre-miR-153c (miR-153c_r)	up (4)	hsa-miR-153				

Zebrafish miRNA	Expression in zf liver cancer	Human miRNA	Expression in HCC	Other associations to cancer	Functional links	References
dre-miR-181a (miR-181a_l, mir-181a-1_l, mir-181a-2_l,	up (2) / down (1)	hsa-miR-181a	up	up in oral squamous cell carcinoma, down in AML, CLL, glioma		Ji J et al 2009 Larson RA et al 2009 Visone R et al 2009
dre-mir-181b (mir-181b-1_l, mir-181b-2_l)	up (2) / down (1)	hsa-miR-181b	up	down in AML, CLL	aggressiveness	Shi L et al 2008 Cervigne NK et al 2009 Schaefer A et al 2010
dre-mir-181c (mir-181c_l)	up (1) / down (1)	hsa-miR-181c	up	down in AML, CLL		
dre-miR-192 (miR-192_l)	up (1) / down (3)	hsa-miR-192	polymorphism	colon cancer	tumor suppression	Yang J et al 2008 Song B et al 2008 Georges SA et al 2008 Braun CJ et al 2008 Yantiss RK et al 2009
dre-miR-193a (miR-193a_r, miR-193a-1_r)	up (4)	hsa-miR-193a-3p	-	epigenetically silenced during oral carcinogenesis	tumor suppression	Kozaki K et al 2008
dre-miR-193b (miR-193b_r)	up (5)	hsa-miR-193b	-	down in breast cancer, adenocarcinoma	metastasis, tumor progression, invasion	Li XF et al 2009 Leivonen SK et al 2009 Wu W et al 2009
-	-	hsa-miR-195	down	up in breast cancer, down in adrenocortical carcinoma, bladder cancer, prostate carcinoma	cell cycle control	Xu T et al 2009 Soon PS et al 2009 Zhang H et al 2009 Ichimi T et al 2009 Zanette DL et al 2007
dre-miR-199 (miR-199_l, mir-199-1_l, mir-199-3_l)	down (2)	hsa-miR-199a-5p	down	up in cervical carcinomas, AML, CLL, down in bladder cancer,ovarian cancer	metastasis, apoptosis	Chen XM 2009 Mott JL 2009 Li W et al 2008 Ichimi T et al 2009 Kim S et al 2008 Lee JW et al 2008 Garzon R et al 2008 Iorio MV et al 2007
dre-miR-199* (miR-199_r, mir-199-1_r, mir-199-3_r)	up (1) / down (1)	hsa-miR-199a-3p				
dre-miR-210 (miR-210_r)	up (4)	hsa-miR-210	-	up in renal cell carcinoma, pancreatic tumor, breast cancer, B-cell lymphoma	metastasis, cell cycle regulation, differentiation, DNA repair, regulation of angiogenesis, neuronal cell survival	Juan D et al 2009 Greither T et al 2010 Foekens JA et al 2008 Crosby ME et al 2009 Fasanaro P et al 2009
dre-miR-212 (miR-212_r)	up (5)	hsa-miR-212	-	down in gastric cancer	-	Wada R et al 2009
dre-miR-217 (miR-217-1_l)	down (4)	hsa-miR-217	-	down in pancreatic cancer	-	Szafranska AE et al 2007
dre-miR-221 (miR-221_r)	up (4)	hsa-miR-221	up	-	apoptosis, proliferation, aggressiveness	Pineau P et al 2009 Gramantieri L et al 2009 Fornari F et al 2008
dre-miR-222* (miR-222_l)	up (4)	hsa-miR-222*				Wong QW et al 2008 Ladeiro Y et al 2008 Sun T et al 2009 Greither T et al 2010 Ueda R et al 2009 Liu X et al 2009
dre-miR-222 (miR-222_r)	up (4)	hsa-miR-222	hsa-mir-222: up	hsa-mir-222: up in pancreatic tumor, prostate cancer	hsa-mir-222: invasion	

Zebrafish miRNA	Expression in zf liver cancer	Human miRNA	Expression in HCC	Other associations to cancer	Functional links	References
dre-mir-223* (mir-223_l) dre-mir-223 (mir-223_r)	up (1) up (2)	hsa-miR-223* hsa-miR-223	hsa-miR-223: down	hsa-miR-223: gastric cancer, small cell lung cancer, down in CLL, ovarian cancer	-	Wong QW et al 2008 L X et al 2009 Miko E et al 2009 Visone R et al 2009 Laios A et al 2008
-	-	hsa-miR-224	up	up in pancreatic ductal adenocarcinoma, AML, prostate cancer	proliferation, migration, invasion	Chen XM 2009 Su H et al 2009 Li W et al 2008 Wang Y et al 2008 Li Z et al 2008 Mees ST et al 2009 Prueitt RL et al 2008
dre-miR-338 (miR-338_r, mir-338-1_r, mir-338-3_r)	up (2) / down (1)	hsa-miR-338-3p	down	gastric cancer, up in squamous cell carcinoma of tongue	aggressiveness	Huang XH et al 2009 Zhao Y et al 2009 L X et al 2009 Wong TS et al 2008
dre-miR-363 (miR-363_r)	up (2) / down (2)	hsa-miR-363	-	up in T-cell lymphoma, down in malignant bronchial epithelial cells	-	Lum AM et al 2007 Shen YL et al 2009
dre-miR-365 (miR-365-1_r, miR-365-2_r, miR-365-3_r)	up (5)	hsa-miR-365	-	up-regulated in breast cancer	-	Yan LX et al 2008
dre-miR-455 (miR-455_l) dre-miR-455* (miR-455_r)	up (5) up (5)	hsa-miR-455-5p hsa-miR-455-3p	-	down in endometrial serous adenocarcinoma	differentiation, invasion	Walden TB et al 2009 Hiroki E et al 2009
dre-miR-457a (miR-457a_l)	up (4)	-	-	-	-	
dre-miR-457b (miR-457b_l)	up (4)	-	-	-	-	
dre-miR-489* (miR-489_l)	up (1) / down (4)	-	-	hsa-miR-489: down in breast cancer	-	Miller TE et al 2008
dre-miR-499 (miR-499_l)	up (1) / down (3)	hsa-miR-499-5p	-	breast cancer	senescence	Catucci I et al 2009 Hu Z et al 2009 Lafferty-Whyte K et al 2009
dre-miR-722 (miR-722_r)	up (2) / down (3)	-	-	-	-	
dre-miR-733 (miR-733_l)	up (2) / down (2)	-	-	-	-	
dre-miR-735 (miR-735_r)	up (5)	-	-	-	-	
(block361064_C1_r)	up (4)	-	-	-	-	
(sblock169393_C1_r)	down (5)	-	-	-	-	

The miRNAs that showed more than 1.5-fold differential expression (P<10-5) in four or more zebrafish liver tumor experiments and a set of miRNAs previously linked to human hepatocellular carcinoma (HCC) in the literature are listed in the table. MiRNA names were derived from miRBase and star sequences (not in all cases present in miRBase) were named accordingly. Probe names on the custom microarray are indicated in brackets behind the zebrafish (dre) miRNA names, and the corresponding human (hsa) miRNA names are given. For some human miRNAs there are no probes for corresponding zebrafish miRNA on the microarray, which is due to absence of the miRNA genes in zebrafish (miR-106, miR-195, miR-224) or an omission/error in the array design (miR-25, miR-29c, miR-92a). The direction of change of miRNA expression is indicated both for the zebrafish liver tumor samples and for the data on HCC derived from the literature. Numbers in brackets indicate the number of biological replicates of zebrafish healthy liver versus liver tumor where the miRNA probes were differentially expressed. Data on association to other types of cancer in human and on biological processes affected by miRNA function were collected from the literature. In some cases we observed differential expression of the miRNA star sequences in addition to or instead of expression of the mature miRNAs reported in miRBase. As indicated in the table, the literature data are for primary miRNAs. Abbreviations: ALL - acute lymphocytic leukemias, AML - acute myeloid leukemia, APL - acute promyelocytic leukemia, CLL - chronic lymphocytic leukemia, CML - chronic myeloid leukemia, HCC-hepatocellular carcinoma.

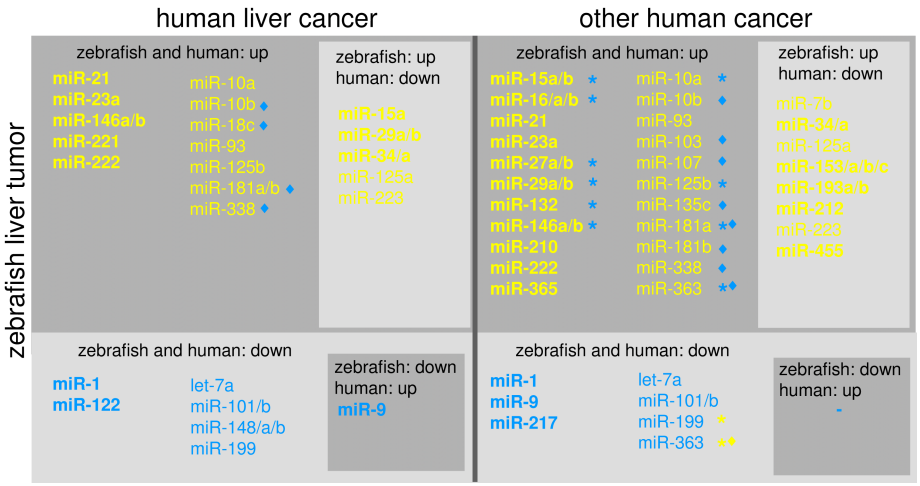


Figure 3. Common regulation of miRNA expression in zebrafish liver tumors and human cancer. Differential expression of miRNAs in zebrafish liver tumors is compared with literature data on the expression of the homologous human miRNAs in human liver cancer (hepatocellular carcinoma) and in other types of cancer as further detailed in Table 1. Yellow indicates miRNAs up-regulated in zebrafish liver tumor tissue compared to normal liver and blue indicates down-regulated miRNAs in zebrafish liver tumors. MiRNAs showing differential expression in 4 out of 5 or 5 out of 5 biological replicates are in bold and miRNAs differentially expressed in 2 or 3 biological replicates are in normal script. MiRNAs for which both up- and down-regulation has been reported in human cancer are indicated with asterisks. MiRNAs for which both up- and down-regulation was observed in zebrafish liver tumors are indicated with diamonds.

365 and miR-455. These miRNAs have been linked to other types of cancer in human and like those linked to HCC, these miRNAs have also been suggested to play roles in tumorigenicity, tumor suppression, apoptosis, cell proliferation, invasion, and metastasis (Table 1, Figure 3).

Expression of miRNA star sequences

Besides regulation of the described mature miRNAs, in several cases we also detected differential expression of probes matching miRNA star sequences (Table 1). Star sequences (miRNA*) are derived from the other arm of the pre-miRNA hairpin structure during miRNA biogenesis, but are not loaded into the RISC complex and are degraded. However, sometimes miRNA* can have similar 5' end stability as the mature miRNA strands and similar incorporation efficiency into the RISC complex, and therefore can also be functional (Soares AR et al 2009). For several miRNAs (miR-21, miR-125b, miR-126, miR-146a, miR-148, miR-199, miR-222, miR-223, miR-455) we found both the primary miRNA and the miRNA* to be differentially expressed in liver tumors, while in other cases (miR-7a*, miR-24*, miR-128*, miR-

Table 2. Common tumor- and immune-related miRNAs.

miRNA	Liver cancer	Mycobacterium infection			
		chronic strain (E11)		acute strain (Mma20)	
		1 dpi	6 dpi	1 dpi	6 dpi
dre-miR-15c (miR-15c_l)	up (4)				up
dre-miR-16a (miR-16a_l)	up (5)			up	
dre-miR-21 (**miR-21-1_l, miR-21-2_l**)	up (5)				up
dre-miR-21* (**miR-21-1_r**)	up (5)	down			up
dre-miR-34 (miR-34_l)	up (4)		up	down	
dre-miR-128* (miR-128-2_l)	up (4)	down	down	down	down
dre-miR-132 (**miR-132-1_r**, miR-132-2_r)	up (5)			down	
dre-miR-146a (miR-146a_l)	up (5)				up
dre-miR-146a* (miR-146a_r)	up (4)				up
dre-miR-146b (miR-146b-1_l)	up (5)				up
dre-miR-212 (miR-212_r)	up (5)			down	
dre-miR-217 (miR-217-1_l)	down (4)	down		down	down
dre-miR-455 (miR-455_l)	up (5)			down	
dre-miR-455* (miR-455_r)	up (5)			down	

Differentially regulated miRNAs in the zebrafish liver tumor study were compared with data from our previous infection study where we used *Mycobacterium marinum* strains Mma20 and E11 as pathogens causing acute disease or chronic granulomatous tuberculosis in zebrafish adults, respectively (chapter 4). Microarray analysis was carried out using RNA samples collected from two time points, 1 and 6 days post infection (dpi). Probe names on the custom microarray are indicated in brackets behind the zebrafish (dre) miRNA names. If there are more probes for one miRNA, those that were expressed in the infection study are indicated in bold. The direction of change of miRNA expressions is indicated both for the zebrafish liver tumor samples and the samples from the infection study. Only the probes that showed consistent up- or down-regulation in at least four biological replicates of the liver tumor study were considered for comparison with the infection data. Significance cut-offs were set at $P<10\text{-}5$ and more 1.5-fold change for analysis of both the liver tumor and infection data sets.

489*) altered expression of only the miRNA* was detected. Most of the zebrafish miRNA* that we found to be differentially expressed in liver tumors have not been previously reported in miRBase. However, miRBase provides evidence for expression of several of these miRNA* in human, such as miR-7-1/2*, miR-21*, miR-24*, miR-125b*, miR-126*, miR-143*, miR-146a*, miR-148a*/b*, miR-222*, and miR-223*. Our microarray data support that these miRNAs* are also expressed in zebrafish and suggest that they play a role in liver cancer similar as the primary miRNAs.

Comparison of miRNA expression in liver cancer and tuberculosis infection

Besides being involved in the pathogenesis of cancer, miRNAs are also known to play important regulatory roles in the development and function of the immune system (Lu and Liston, 2009). Previously we used the same microarray platform to investigate miRNA expression during the immune response of adult zebrafish to infection with *Mycobacterium marinum*, the causative agent of fish tuberculosis (chapter 4). Here we compared the data from this infection study with the liver cancer study in order to define common tumor and immune-related miRNAs (Table 2).

In our infection study we used *Mycobacterium marinum* strains Mma20 and E11, which respectively cause acute disease or chronic granulomatous tuberculosis in adult zebrafish (van der Sar et al., 2004; van der Sar et al., 2009), and determined miRNA profiles at 1 and 6 days post infection (dpi). Comparing the datasets of liver tumor-related miRNAs to *M. marinum* infection-related miRNAs we found that 8 miRNAs, miR-15c, miR-16a, miR-21, miR-21*, miR-34, miR-146a, miR-146a* and miR-146b, were commonly up-regulated in liver tumors and at either the 1 or 6 dpi time point of infection with the Mma20 or E11 strains (Table 2). Two of these miRNAs (miR-21* and miR-34) were up-regulated by only one of the *M. marinum* strains, but down-regulated by the other. Expression of miR-217 was down-regulated in liver tumors and at different time points of *M. marinum* infection with both strains. Five miRNAs (miR-128*, miR-132, miR-212, miR-455, miR-455*) were up-regulated in liver tumors, but down-regulated at one or more time points of infection with the *M. marinum* strains. Other miRNAs that were up-regulated in liver tumors (miR-15a, miR-15b, miR-16b, miR-23a, miR-24*, miR-27a/b/c/d, miR-29a/b, miR-153a/b/c, miR-193a/b, miR-210, miR-221, miR-222, miR-222*, miR-365, miR-457a/b and miR-735) or down-regulated in liver tumors (miR-1, miR-7a*, miR-9, miR-122) were not differentially expressed in the samples of the *M. marinum* infection study. Among the miRNAs commonly regulated in the liver cancer and tuberculosis studies, in particular the members of the miR-21 and miR-146 family can be strongly linked to the immune response, as they were also found to be induced by different bacterial infections in zebrafish embryos (chapter 4).

Experimentally validated targets of liver tumor-related miRNAs

To shed further light on the role of differentially expressed liver tumor-related miRNAs in the mechanism of cancer we searched the TarBase database for putative

Table 3. Pathways and processes in which the experimentally validated target genes of liver tumor-related miRNAs are involved.

Pathway/process	Gene symbol	miRNA	Additional pathway
Apoptosis signaling	*ACVR1B*	miR-24	cell cycle, immune system process, MAPK signaling pathway
	API5	miR-224	
	BCL2	miR-15A, miR-15B	cell proliferation, cell cycle, immune system process
	CDKN1A	miR-106B	cell proliferation, cell cycle, immune system process, ErbB signaling pathway, p53 signaling pathway
	CDKN1B	miR-221, miR-222	cell proliferation, cell cycle, cell motility, immune system process, cytoskeleton organization and biogenesis, ErbB signaling pathway
	CDKN2A	miR-24	cell proliferation, cell cycle, cell adhesion, p53 signaling
	CXCR4	miR-146A	cell proliferation, cell motility, immune system process,
	FADD	miR-146A	I-kappaB kinase/NF-kappaB cascade,
	GCH1	miR-1	
	GJA1	miR-1	cell proliferation, I-kappaB kinase/NF-kappaB cascade
	IHPK2	miR-1	
	KRAS	miR-146B-5P	cell proliferation, cell cycle, cytoskeleton organization and biogenesis, small GTPase mediated signal transduction, ErbB signaling pathway, MAPK signaling pathway, VEGF signaling pathway, Fc epsilon RI signaling pathway
	MCL1	miR-29B	
	PDCD4	miR-1, miR-21	cell cycle
	POGK	miR-1	
	PTEN	miR-21	cell proliferation, cell cycle, cell motility, cell adhesion, p53 signaling pathway
	ROCK1	miR-146A	cytoskeleton organization and biogenesis, small GTPase mediated signal transduction, Wnt signaling
	TIMP3	miR-1	
Cell proliferation	***CDKN1C***	miR-221, miR-222	cell cycle
	ERBB2	miR-125A	immune system process, ErbB signaling pathway
	HDAC4	miR-1	cell cycle, immune system process
	KIT	miR-221, miR-222	immune system process
	IGF1	miR-1	cell motility, small GTPase mediated signal transduction, p53 signaling pathway
	LAMC1	miR-29C	cell motility, cell adhesion
	MXD4	miR-1	
	MYCN	miR-101	
	NF2	LET-7A	cell cycle, cell motility, cell adhesion, cytoskeleton organization and biogenesis
	NOTCH1	miR-24, miR-27B	immune system process
	SMAD1	miR-26A	
	CDK6	miR-34a	cell cycle, cell adhesion, immune system process, p53 signaling
Cell cycle	*AXL*	miR-1	
	DMTF1	miR-15A	
	EML4	miR-1	cytoskeleton organization and biogenesis
	MAPK7	miR-143*	MAPK signaling pathway
	E2F3	miR-34a	
	CCND1	miR-34a	p53 signaling, wnt signaling

Pathway/process	Gene symbol	miRNA	Additional pathway
Cell motility	*CAP1*	miR-1	cytoskeleton organization and biogenesis
	MAPK14	miR-24	MAPK signaling pathway, VEGF signaling pathway, Fc epsilon RI signaling pathway
	SERPINB5	miR-1, miR-21	p53 signaling pathway
	TPM1	miR-21	
	TPM4	miR-1	
Cell adhesion	*COL15A1*	miR-29C	
	CXCL12	miR-23A	cytoskeleton organization and biogenesis, immune system process
	VCAM1	miR-126	
Cytoskeleton organization and biogenesis	*ARHGEF18*	miR-1	small GTPase mediated signal transduction
	HCN4	miR-1	
	LASP1	miR-1	p53 signaling pathway
	PREX1	miR-1	small GTPase mediated signal transduction, immune system process
	TMSB4X	miR-1	
I-kappaB kinase/NF-kappaB cascade	***IRAK1***	miR-146A, miR-146B-5P	
	IRAK2	miR-146A	
	TRAF6	miR-146A, miR-146B-5P	apoptosis, immune system process, MAPK signaling pathway
	STAT1	miR-146A	apoptosis, cell cycle, immune system processs,
	IRF5	miR-146A, miR-146B-5P	
ErbB signaling pathway	*AKT2*	miR-146B-5P	MAPK signaling pathway, VEGF signaling pathway, Fc epsilon RI signaling pathway
	ERBB3	miR-125A	
Small GTPase mediated signal transduction	*ARF3*	miR-1	
	ARF4	miR-1	
	RABGAP1L	miR-1	
	RABL2A	miR-1	
	RABL2B	miR-1	

Experimentally validated targets were collected from the TarBase v5 database. Gene ontology analysis using DAVID tools classified these targets into various cancer-related pathways and processes listed in the table. Target genes predicted to be conserved in zebrafish are indicated in bold.

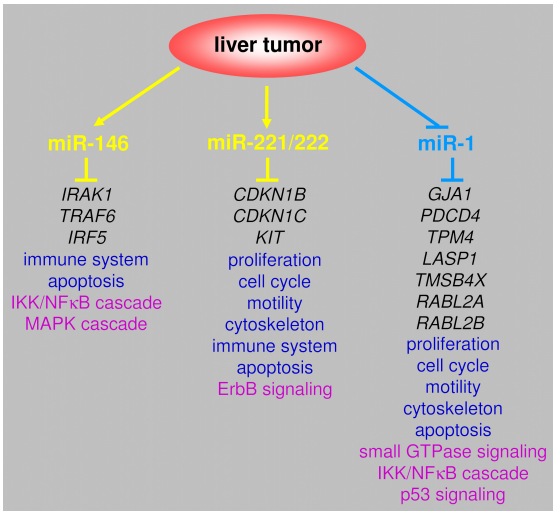


Figure 4. Common predicted target genes of liver tumor-related miRNAs in zebrafish and human. Up-regulation (yellow) of members of the miR-146 and miR-221/miR-222 families and down-regulation (blue) of miR-1 was observed in at least 4 out of 5 biological replicates of zebrafish liver tumor tissue compared to normal liver and changes in the same direction have been reported in the literature on human liver cancer (hepatocellular carcinoma) as detailed in Table 1. Members of the miR-146, miR-221/miR-222 and miR-1 families are predicted to target several genes whose human homologs have been reported as experimentally validated target genes in the TarBase database. The figure shows the common predicted target genes and the biological pathways and signaling pathways that can be associated with these genes by gene ontology analysis.

target genes that have been experimentally validated targets in human cell-based assays (Table 3). Gene Ontology analysis using DAVID tools linked these target genes to biological processes such as apoptosis signaling, cell proliferation, cell cycle, cell motility, cell adhesion, cytoskeleton organization and biogenesis, and immune system process, as well as to signaling pathways, including the I-kappaB kinase/NF-kappaB cascade, small GTPase mediated signal transduction, ErbB signaling, MAPK signaling, VEGF signaling, p53 signaling, and several KEGG cancer pathways (Table 3). These processes and pathways are often deregulated in HCC suggesting that these miRNAs can be part of signaling pathways that significantly contribute to cancer (Chen et al., 2002; Kim et al., 2004; Neo et al., 2004; Okabe et al., 2001; Xu et al., 2001). Next we used different miRNA target prediction programs to investigate if the homologous zebrafish genes might contain miRNA target sites for the same miRNAs as those involved in HCC. This analysis showed that several target genes of the miR-146 and the miR-221/miR-222 families, which we found to be commonly up-regulated in zebrafish liver tumors and in HCC, might be conserved between zebrafish and human (Table 3, Figure 4). For the miR-146 family these included two pivotal signaling intermediates, *IRAK1* and *TRAF6*, of the NF-kappaB pathway and the *IRF5* transcription factor gene that is a target of this pathway. Predicted common targets of the human and zebrafish miR-221/miR-222 family included cyclin-dependent kinase inhibitor genes (*CDKN1B* and *CDKN1C*), linked with epidermal growth factor receptor (ErbB) signaling, and the proto-oncogene tyrosine-protein kinase

KIT. Furthermore, we could predict 7 conserved target genes for miR-1, which we identified as a commonly down-regulated miRNA in zebrafish liver tumors and in reports on HCC. One of these was a member of the connexin family (*GJA1*, gap junction protein, alpha 1, 43kDa) related to cell-cell signaling and positive regulation of the NF-kappaB pathway. Another predicted conserved miR-1 target was the programmed cell death 4 (*PDC4*) gene, also known as neoplastic transformation inhibitor, which has a still unknown function in apoptosis regulation. Three other genes encoded cytoskeleton-related proteins, including the stress-responsive, actin-binding tropomyosin 4 (TPM4) protein, the actin-binding protein LASP1 (LIM and SH3 protein 1), which has been shown to play an essential role in tumor cell growth and migration and was reported as p53 transcriptional target involved in HCC (Wang et al., 2009a), and the actin-sequestering protein TMSB4X (thymosin beta 4, X-linked), which has been reported to positively regulate the expression of hepatocyte growth factor (Barnaeva et al., 2007) and has been implicated in cell proliferation, migration, and differentiation as well as tumor metastasis and angiogenesis (Wang et al., 2004; Zhang et al., 2008a). Finally, two members of the RAS oncogene family-like 2 (RABL2A and RABL2B) were also predicted to be conserved targets of miR-1 in human and zebrafish. Taken together, our results show that several miRNAs are commonly regulated in human and zebrafish liver cancer and predict conservation of a number of target genes strongly linked to cancer-related biological processes and signaling pathways.

Discussion

Recently, it has been shown that there are remarkable molecular similarities between zebrafish and human liver cancer not only with respect to general molecular hallmarks (Lam et al., 2006) but also at the level of biological modules related to tumorigenesis and transcription factor target modules (Ung et al., 2009). In this study we show that there is conservation at the level of miRNA expression as well. Using a custom microarray platform we observed highly reproducible induction and repression of miRNAs in zebrafish liver tumors, many of which were previously linked to human hepatocellular carcinoma (HCC) or to other types of human cancer. We also show that a subset of these liver tumor-related miRNAs overlap with miRNAs responsive to *Mycobacterium marinum* infection, suggesting a common role in cancer and the immune response. Finally, based on target predictions for zebrafish miRNAs and experimental target validation data reported for human miRNAs, we suggest that several target genes of the miR-1, miR-146 and miR-221/miR-222 families, which are commonly deregulated in liver tumors, are conserved between human and zebrafish.

To compare miRNA expression signatures between human and zebrafish liver tumors we took two approaches. First, using a custom Agilent microarray platform we analyzed differential expression in carcinogen-induced zebrafish liver tumors as

compared to healthy liver tissue. Second, by PubMed abstract searching we compiled human miRNAs previously reported to show altered expression in HCC or associated to HCC by polymorphisms. Together this resulted in the identification of 83 liver tumor-related miRNAs, including 70 primary miRNAs reported in miRBase and 13 miRNA star sequences. The majority of the miRNAs previously linked to HCC showed altered expression in zebrafish liver tumors. The most notable similarities were the common up-regulation of miR-21, miR-23a, miR-146a/b, miR-221 and miR-222, and the common down-regulation of miR-1 and miR-122. These miRNAs showed consistent up- or down-regulation in four or five out of five biological replicates of zebrafish liver tumors and have been reported to show the same direction of expression changes in HCC. A similar number of HCC-linked miRNAs was also changed in the same direction in zebrafish liver tumors, but in less than four replicates. Some other miRNAs, most notably miR-9, miR-15a, miR-29a/b and miR-34, showed opposite changes in zebrafish liver tumors compared to literature reports on HCC. Approximately half of the zebrafish miRNAs with consistent differential expression in liver tumors were not previously linked to HCC. Some of these miRNAs lacked human counterparts, but in most cases the human counterparts of these miRNAs were already associated with various other types of cancer and therefore it is conceivable that associations with human liver cancer may also be found in future studies. Most of the miRNAs associated with HCC or other human tumors, including those commonly regulated in zebrafish liver tumors, have been found to play roles in different mechanisms of cancer such as tumorigenicity, tumor suppression, apoptosis, cell proliferation, differentiation, invasion, and metastasis. These data further support that zebrafish liver tumors are appropriate for modeling human liver tumors and that the molecular similarities between the species extend from conserved tumor-associated gene expression signatures to shared expression alterations of a set of tumor-associated miRNAs.

The custom microarray platform used in this study not only contained probes for almost all known miRNAs reported in miRBase but also for the putative star sequences of all primary miRNAs and for predicted hairpin structures with or without experimental support for miRNA identity. Our microarray data support the expression of several miRNA star sequences that were not yet reported in miRBase and show that in most cases altered expression of these miRNA star sequences in liver tumors occurs concomitantly with expression of the corresponding primary miRNAs. We also noted that 86 probes for predicted hairpin structures without experimental support for miRNA identity were differentially expressed in all five biological replicates of the liver tumor study. Similarly, we previously observed differential expression of many probes in this category during bacterial infection of zebrafish (chapter 4). The possibility that these probes represent an additional group of non-coding small RNAs with putative regulatory functions during cancer and infection is of great interest for further investigation.

Differential expression of miRNAs has been implicated not only in different

mechanisms of many types of cancer but also in the immune response. In agreement, our previous work suggests that many miRNAs implicated in cancer processes are also connected to regulation of the immune response in zebrafish (chapter 4). In the previous study we used the same microarray platform to analyze miRNAs induced by *Mycobacterium marinum* infection in zebrafish (chapter 4). *M. marinum* is closely related to *Mycobacterium tuberculosis* and the pathology of *M. marinum* infection in zebrafish strongly resembles that of human tuberculosis. In particular, the formation of granulomas, a hallmark of tuberculosis, is well conserved between zebrafish and human. Granuloma formation involves the aggregation of infected and non-infected macrophages and other immune cells into tight structures. In this respect granuloma formation shares interesting similarities with tumor formation, which generally is also associated with the attraction of macrophages and other immune cells. Therefore, it was of interest to determine the miRNAs commonly regulated in liver tumors and in response to *M. marinum* infection. We identified several miRNAs that were commonly up-regulated in liver tumors and at one or more stages of infection with two *M. marinum* strains, including miRNAs of the miR-15, miR-16, miR-21, miR-34, and miR-146 families. Three of the most interesting commonly regulated miRNAs (miR-146a, miR-146a* and miR-146b) are members of the miR-146 family, which has previously been implicated in regulation of innate immunity signaling pathways and recently suggested also to modulate adaptive immunity (Hou et al., 2009; Perry et al., 2009; Schmidt et al., 2009; Taganov et al., 2006; Williams et al., 2008). Altered expression of human miR-146 miRNAs was observed in viral and bacterial infections, under inflammatory conditions as well as in different types of cancer, including HCC, leukemia and breast cancer (Bellon et al., 2009; Nakasa et al., 2008; Perry et al., 2009; Shen et al., 2008; Williams et al., 2008; Xia et al., 2009; Xu et al., 2008). Similarly, elevated expressed members of the human miR-21 family, which were also commonly up-regulated in zebrafish liver tumors and mycobacterium infection, was previously reported in studies of infection, allergic inflammation and cancer, including HCC, gastric cancer, glioma, colon cancer and breast cancer (Cameron et al., 2008; Connolly et al., 2008; Gramantieri et al., 2007; Lu et al., 2009; Wang et al., 2009b; Zhang et al., 2008b; Zhou et al., 2009).). It has been suggested that the connection between processes of cancer and the immune response might be due to changes in expression of miRNAs regulating shared signaling pathways (Williams et al., 2008). The common and conserved regulation of several miRNAs in tumors and during infection that we observed in this study further supports this connection.

Finally, we investigated the possible conservation of miRNA target genes. Although there are numerous predicted miRNA targets detected by bioinformatics approaches only a few have been validated. In this study we compiled experimentally validated targets of human liver tumor-related miRNAs. Functional classification of these targets by gene ontology analysis revealed their involvement in signaling pathways and biological processes that formerly have been identified as common

molecular hallmarks of zebrafish and human liver cancers (Lam and Gong, 2006; Lam et al., 2006; Ung et al., 2009). Next we checked the 3'UTR regions of the homologous zebrafish genes for the presence of target sites for the same miRNAs. For several miRNAs no conserved targets were predicted by MiRanda, TargetScan and RNA-hybrid algorithms; however, we cannot exclude the possible presence of target sites in introns or the presence of target sites not recognized due to limitations of the target prediction algorithms or incomplete annotation of the 3'UTR regions in the Ensembl database. Interestingly, target predictions indicated several conserved target genes for miRNAs of three families, including the miR-146 and miR-221/miR-222 families that were commonly up-regulated in human and zebrafish liver tumors, and miR-1, a commonly down-regulated miRNA family in human and zebrafish liver tumors. Shared target genes of the miR-146 family encoded signaling intermediates (IRAK1, TRAF6) and a transcription factor (IRF5) involved in the immune response. As discussed above, the miR-146 family is also induced by infection in human and zebrafish; therefore, the predicted conservation of these target genes further supports that this miRNA family functions at the crossroads of cancer and the immune response. Proteins encoded by the predicted conserved target genes of the miR-221/222 family (CDKN1B, CDKN1C, KIT) and the miR-1 family (GJA1, PDCD4, TPM4, LASP1, TMSB4X, RABL2A, RABL2B) are involved in several cancer-related biological processes such as proliferation, cell cycle regulation, apoptosis, and cytoskeletal organization and motility. Furthermore, these genes have been implicated in cancer-related signaling pathways, including ErbB, NF-κB and p53 signaling, and small GTPase-mediated signal transduction. The ErbB signaling pathway has been implicated in the development and malignancy of numerous types of human cancers, including liver cancer (Altimari et al., 2003; Ito et al., 2001; Olayioye et al., 2000). NF-κB signaling is involved in the immune response, cell adhesion, differentiation, proliferation, angiogenesis and apoptosis and imbalanced regulation of this pathway has been associated with development of multiple types of tumors (Steele and Lane, 2005; Sun and Zhang, 2007), similar as for deregulation of the p53 tumor-suppression pathway (Steele and Lane, 2005). Finally, predicted target genes involved in small GTPase-mediated signal transduction are potential markers for tissue invasiveness of liver tumor cells (Lam and Gong, 2006). The predicted conservation of target genes in these pathways and processes further supports the similarity between zebrafish and human liver cancer and that miRNAs might play a pivotal role in regulating these cancer-related mechanisms in both species.

Taken together, our data suggest that there is high conservation between zebrafish and human liver tumors not only at the level of gene expression signatures and transcriptional targets, but also at the level of the regulation by miRNAs. These results further support the usefulness of zebrafish as a useful model for studying liver cancer and functions of miRNAs in the mechanisms of liver cancer.

Materials and methods

Generation of zebrafish liver tumors

Liver tumor samples were obtained from the group of Zhiyuan Gong (National University of Singapore) and generated by carcinogen treatment as described in Lam et al., 2006. In brief, three-week-old juvenile zebrafish were treated with 0.75 ppm 7,12-dimethylbenz[α]anthracene (DMBA) or dimethyl sulfoxide (DMSO, vehicle) for 24 h and the treatment was repeated once at five weeks of age for another 24 hours with 1.25 ppm DMBA or DMSO. Treated fish were rinsed three times in fresh water and transferred into new tanks for maintenance. Fish were sampled during 6-10 months after the onset of DMBA exposure. Tumor samples used for the microarray study were all bigger than 3 mm in diameter. Partial liver tumors were used for RNA extraction and the rest of the tissues were used for histopathological diagnosis. The normal livers were also sampled at 6 months to 10 months after the control treatment (DMSO, vehicle). Zebrafish husbandry was as described (Lam et al., 2006).

Microarray design

Custom-designed 8x15k microarray slides were ordered from Agilent Technologies. The 15k custom design was obtained from Edwin Cuppen and Eugene Berezikov, Hubrecht Institute, Utrecht, The Netherlands and will be submitted into the Gene Expression Omnibus database. The 15k design contained a duplicate of 7604 probes of 60-oligonucleotide length. The probes consisted of 2x22 nucleotide sequences antisense to mature miRNAs separated by a spacer of 8 nucleotides (CGATCTTT) and with a second spacer with the same sequence at the end. From 7604 probes 546 were designed for left (5') and right (3') arms of the hairpins of zebrafish miRNAs that are currently known in miRBase. The remainder 7058 probes were classified into 3 categories: 62 probes into C1 category (novel confident miRNAs for which there is compelling experimental evidence), 54 probes into C2 category (candidate miRNAs with clear hairpin structure but less experimental evidence), and 6942 into C3 category (additional hairpins identified in the zebrafish genome sequence that are less likely to be miRNAs).

RNA isolation, labeling and hybridization

Total RNA from five liver tumors and five normal livers was isolated using the miRNeasy Mini kit to preserve small RNA species. Three of the liver tumor samples were from male zebrafish and two from female zebrafish and each was compared with normal liver samples from fish with the same sex. RNA labeling was carried out with miRCURY™ LNA microRNA, Hy3™/Hy5™ Power Labeling kit (Exiqon) using 1 μg of total RNA according to the manufacturer's instructions. RNA samples from liver tumors were labeled with Hy3 and hybridized against Hy5-labeled RNA samples from normal livers. The dual colour hybridization of the microarray chips was performed according to Agilent protocol GE2-v5_95_Feb07 and GE2_105_Jan09

(www.Agilent.com) for two-color microarray-based gene expression analysis except that hybridization and washing was performed at 37 °C.

Data analysis

Microarray data were processed from raw data image files with Feature Extraction Software 10.5.1 (Agilent Technologies). Processed data were subsequently imported into Rosetta Resolver 7.2 (Rosetta Biosoftware) and subjected to default ratio error modeling. Significance cut-offs for the ratios of infected versus control were set at 1.5-fold change at $P \leq 10-5$. Experimentally validated target genes of human miRNAs were obtained from TarBase V.5 (http://diana.cslab.ece.ntua.gr/tarbase/) (Papadopoulos et al., 2009). Gene ontology and pathway analysis was carried out using Database for Annotation, Visualization and Integrated Discovery (DAVID) software tools (http://david.abcc.ncifcrf.gov/) (Huang da et al., 2009). MiRNA target site predictions for the homologous zebrafish genes were derived from the MicroCosm database and were performed by locally running miRanda, TargetScan and RNAhybrid prediction algorithms using the 3'UTR regions derived from the Ensembl Zv8 zebrafish database. For the zebrafish homolog of *IRAK1* the 3'UTR region experimentally determined in chapter 4 was included in the predictions.

Supplementary data

Supplementary data associated with this chapter can be found at http://apo.szbk.u-szeged.hu/transfer/A_ORDAS/.

Acknowledgments

We are grateful to Edwin Cuppen and Eugene Berezikov (Hubrecht Institute, Utrecht , The Netherlands) for the Agilent miRNA array design.

References

Altimari A., Fiorentino M., Gabusi E., Gruppioni E., Corti B., D'Errico A. and Grigioni W. F. (2003) Investigation of ErbB1 and ErbB2 expression for therapeutic targeting in primary liver tumours. *Dig Liver Dis* **35**, 332-8.

Amatruda J. F., Shepard J. L., Stern H. M. and Zon L. I. (2002) Zebrafish as a cancer model system. *Cancer Cell* **1**, 229-31.

Barnaeva E., Nadezhda A., Hannappel E., Sjogren M. H. and Rojkind M. (2007) Thymosin beta4 upregulates the expression of hepatocyte growth factor and downregulates the expression of PDGF-beta receptor in human hepatic stellate cells. *Ann N Y Acad Sci* **1112**, 154-60.

Bellon M., Lepelletier Y., Hermine O. and Nicot C. (2009) Deregulation of microRNA involved in hematopoiesis and the immune response in HTLV-I adult T-cell leukemia. *Blood* **113**, 4914-7.

Berghmans S., Jette C., Langenau D., Hsu K., Stewart R., Look T. and Kanki J. P. (2005) Making waves in cancer research: new models in the zebrafish. *Biotechniques* **39**, 227-37.

Budhu A., Jia H. L., Forgues M., Liu C. G., Goldstein D., Lam A., Zanetti K. A., Ye Q. H., Qin L. X., Croce C. M., Tang Z. Y. and Wang X. W. (2008) Identification of metastasis-related microRNAs in hepatocellular carcinoma. *Hepatology* **47**, 897-907.

Bueno M. J., de Castro I. P. and Malumbres M. (2008) Control of cell proliferation pathways by microRNAs. *Cell Cycle* **7**, 3143-8.

Calin G. A., Dumitru C. D., Shimizu M., Bichi R., Zupo S., Noch E., Aldler H., Rattan S., Keating M., Rai K., Rassenti L., Kipps T., Negrini M., Bullrich F. and Croce C. M. (2002) Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A* **99**, 15524-9.

Calin G. A., Liu C. G., Sevignani C., Ferracin M., Felli N., Dumitru C. D., Shimizu M., Cimmino A., Zupo S., Dono M., Dell'Aquila M. L., Alder H., Rassenti L., Kipps T. J., Bullrich F., Negrini M. and Croce C. M. (2004) MicroRNA profiling reveals distinct signatures in B cell chronic lymphocytic leukemias. *Proc Natl Acad Sci U S A* **101**, 11755-60.

Cameron J. E., Fewell C., Yin Q., McBride J., Wang X., Lin Z. and Flemington E. K. (2008) Epstein-Barr virus growth/latency III program alters cellular microRNA expression. *Virology* **382**, 257-66.

Chen X., Cheung S. T., So S., Fan S. T., Barry C., Higgins J., Lai K. M., Ji J., Dudoit S., Ng I. O., Van De Rijn M., Botstein D. and Brown P. O. (2002) Gene expression patterns in human liver cancers. *Mol Biol Cell* **13**, 1929-39.

Chen X. M. (2009) MicroRNA signatures in liver diseases. *World J Gastroenterol* **15**, 1665-72.

Chu J. and Sadler K. C. (2009) New school in liver development: lessons from zebrafish. *Hepatology* **50**, 1656-63.

Connolly E., Melegari M., Landgraf P., Tchaikovskaya T., Tennant B. C., Slagle B. L., Rogler L. E., Zavolan M., Tuschl T. and Rogler C. E. (2008) Elevated expression of the miR-17-92 polycistron and miR-21 in hepadnavirus-associated hepatocellular carcinoma contributes to the malignant phenotype. *Am J Pathol* **173**, 856-64.

Esquela-Kerscher A. and Slack F. J. (2006) Oncomirs - microRNAs with a role in cancer. *Nat Rev Cancer* **6**, 259-69.

Feitsma H. and Cuppen E. (2008) Zebrafish as a cancer model. *Mol Cancer Res* **6**, 685-94.

Garofalo M., Condorelli G. and Croce C. M. (2008) MicroRNAs in diseases and drug response. *Curr Opin Pharmacol* **8**, 661-7.

Garzon R., Calin G. A. and Croce C. M. (2009) MicroRNAs in Cancer. *Annu Rev Med* **60**, 167-79.

Gramantieri L., Ferracin M., Fornari F., Veronese A., Sabbioni S., Liu C. G., Calin G. A., Giovannini C., Ferrazzi E., Grazi G. L., Croce C. M., Bolondi L. and Negrini M. (2007) Cyclin G1 is a target of miR-122a, a microRNA frequently down-regulated in human hepatocellular carcinoma. *Cancer Res* **67**, 6092-9.

Hou J., Wang P., Lin L., Liu X., Ma F., An H., Wang Z. and Cao X. (2009) MicroRNA-146a feedback inhibits RIG-I-dependent Type I IFN production in macrophages by targeting TRAF6, IRAK1, and IRAK2. *J Immunol* **183**, 2150-8.

Huang da W., Sherman B. T. and Lempicki R. A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57.

Huang Y. S., Dai Y., Yu X. F., Bao S. Y., Yin Y. B., Tang M. and Hu C. X. (2008) Microarray analysis of microRNA expression in hepatocellular carcinoma and non-tumorous tissues without viral hepatitis. *J Gastroenterol Hepatol* **23**, 87-94.

Ito Y., Takeda T., Sakon M., Tsujimoto M., Higashiyama S., Noda K., Miyoshi E., Monden M. and Matsuura N. (2001) Expression and clinical significance of erb-B receptor family in hepatocellular carcinoma. *Br J Cancer* **84**, 1377-83.

Kim J. W., Ye Q., Forgues M., Chen Y., Budhu A., Sime J., Hofseth L. J., Kaul R. and Wang X. W. (2004) Cancer-associated molecular signature in the tissue samples of patients with cirrhosis. *Hepatology* **39**, 518-27.

Kota J., Chivukula R. R., O'Donnell K. A., Wentzel E. A., Montgomery C. L., Hwang H. W., Chang T. C., Vivekanandan P., Torbenson M., Clark K. R., Mendell J. R. and Mendell J. T. (2009) Therapeutic microRNA delivery suppresses tumorigenesis in a murine liver cancer model. *Cell* **137**, 1005-17.

Krutzfeldt J., Rajewsky N., Braich R., Rajeev K. G., Tuschl T., Manoharan M. and Stoffel M. (2005) Silencing of microRNAs in vivo with 'antagomirs'. *Nature* **438**, 685-9.

Ladeiro Y., Couchy G., Balabaud C., Bioulac-Sage P., Pelletier L., Rebouissou S. and Zucman-Rossi J. (2008) MicroRNA profiling in hepatocellular tumors is associated with clinical features and oncogene/tumor suppres-

sor gene mutations. *Hepatology* **47**, 1955-63.
Lam S. H. and Gong Z. (2006) Modeling liver cancer using zebrafish: a comparative oncogenomics approach. *Cell Cycle* **5**, 573-7.
Lam S. H., Wu Y. L., Vega V. B., Miller L. D., Spitsbergen J., Tong Y., Zhan H., Govindarajan K. R., Lee S., Mathavan S., Murthy K. R., Buhler D. R., Liu E. T. and Gong Z. (2006) Conservation of gene expression signatures between zebrafish and human liver tumors and tumor progression. *Nat Biotechnol* **24**, 73-5.
Lu J., Getz G., Miska E. A., Alvarez-Saavedra E., Lamb J., Peck D., Sweet-Cordero A., Ebert B. L., Mak R. H., Ferrando A. A., Downing J. R., Jacks T., Horvitz H. R. and Golub T. R. (2005) MicroRNA expression profiles classify human cancers. *Nature* **435**, 834-8.
Lu L. F. and Liston A. (2009) MicroRNA in the immune system, microRNA as an immune system. *Immunology* **127**, 291-8.
Lu T. X., Munitz A. and Rothenberg M. E. (2009) MicroRNA-21 is up-regulated in allergic airway inflammation and regulates IL-12p35 expression. *J Immunol* **182**, 4994-5002.
Mann C. D., Neal C. P., Garcea G., Manson M. M., Dennison A. R. and Berry D. P. (2007) Prognostic molecular markers in hepatocellular carcinoma: a systematic review. *Eur J Cancer* **43**, 979-92.
Medina P. P. and Slack F. J. (2008) microRNAs and cancer: an overview. *Cell Cycle* **7**, 2485-92.
Mott J. L. (2009) MicroRNAs involved in tumor suppressor and oncogene pathways: implications for hepatobiliary neoplasia. *Hepatology* **50**, 630-7.
Murakami Y., Yasuda T., Saigo K., Urashima T., Toyoda H., Okanoue T. and Shimotohno K. (2006) Comprehensive analysis of microRNA expression patterns in hepatocellular carcinoma and non-tumorous tissues. *Oncogene* **25**, 2537-45.
Nakasa T., Miyaki S., Okubo A., Hashimoto M., Nishida K., Ochi M. and Asahara H. (2008) Expression of microRNA-146 in rheumatoid arthritis synovial tissue. *Arthritis Rheum* **58**, 1284-92.
Neo S. Y., Leow C. K., Vega V. B., Long P. M., Islam A. F., Lai P. B., Liu E. T. and Ren E. C. (2004) Identification of discriminators of hepatoma by gene expression profiling using a minimal dataset approach. *Hepatology* **39**, 944-53.
Okabe H., Satoh S., Kato T., Kitahara O., Yanagawa R., Yamaoka Y., Tsunoda T., Furukawa Y. and Nakamura Y. (2001) Genome-wide analysis of gene expression in human hepatocellular carcinomas using cDNA microarray: identification of genes involved in viral carcinogenesis and tumor progression. *Cancer Res* **61**, 2129-37.
Olayioye M. A., Neve R. M., Lane H. A. and Hynes N. E. (2000) The ErbB signaling network: receptor heterodimerization in development and cancer. *Embo J* **19**, 3159-67.
Papadopoulos G. L., Alexiou P., Maragkakis M., Reczko M. and Hatzigeorgiou A. G. (2009) DIANA-mirPath: Integrating human and mouse microRNAs in pathways. *Bioinformatics* **25**, 1991-3.
Perry M. M., Williams A. E., Tsitsiou E., Larner-Svensson H. M. and Lindsay M. A. (2009) Divergent intracellular pathways regulate interleukin-1beta-induced miR-146a and miR-146b expression and chemokine release in human alveolar epithelial cells. *FEBS Lett* **583**, 3349-55.
Pineau P., Volinia S., McJunkin K., Marchio A., Battiston C., Terris B., Mazzaferro V., Lowe S. W., Croce C. M. and Dejean A. miR-221 overexpression contributes to liver tumorigenesis. *Proc Natl Acad Sci U S A* **107**, 264-9.
Qin L. X. and Tang Z. Y. (2002) The prognostic molecular markers in hepatocellular carcinoma. *World J Gastroenterol* **8**, 385-92.
Schmidt W. M., Spiel A. O., Jilma B., Wolzt M. and Muller M. (2009) In vivo profile of the human leukocyte microRNA response to endotoxemia. *Biochem Biophys Res Commun* **380**, 437-41.
Shen J., Ambrosone C. B., DiCioccio R. A., Odunsi K., Lele S. B. and Zhao H. (2008) A functional polymorphism in the miR-146a gene and age of familial breast/ovarian cancer diagnosis. *Carcinogenesis* **29**, 1963-6.
Sonkoly E. and Pivarcsi A. (2009) Advances in microRNAs: implications for immunity and inflammatory diseases. *J Cell Mol Med* **13**, 24-38.
Steele R. J. and Lane D. P. (2005) P53 in cancer: a paradigm for modern management of cancer. *Surgeon* **3**, 197-205.
Su H., Yang J. R., Xu T., Huang J., Xu L., Yuan Y. and Zhuang S. M. (2009) MicroRNA-101, down-regulated in hepatocellular carcinoma, promotes apoptosis and suppresses tumorigenicity. *Cancer Res* **69**, 1135-42.
Sun X. F. and Zhang H. (2007) NFKB and NFKBI polymorphisms in relation to susceptibility of tumour and other diseases. *Histol Histopathol* **22**, 1387-98.
Taganov K. D., Boldin M. P., Chang K. J. and Baltimore D. (2006) NF-kappaB-dependent induction of microRNA miR-146, an inhibitor targeted to signaling proteins of innate immune responses. *Proc Natl Acad Sci U S A* **103**, 12481-6.
Tsai W. C., Hsu P. W., Lai T. C., Chau G. Y., Lin C. W., Chen C. M., Lin C. D., Liao Y. L., Wang J. L., Chau Y. P., Hsu M. T., Hsiao M., Huang H. D. and Tsou A. P. (2009) MicroRNA-122, a tumor suppressor microRNA that regulates intrahepatic metastasis of hepatocellular carcinoma. *Hepatology* **49**, 1571-82.
Ung C. Y., Lam S. H. and Gong Z. (2009) Comparative transcriptome analyses revealed conserved biological and transcription factor target modules between the zebrafish and human tumors. *Zebrafish* **6**, 425-31.
van der Sar A. M., Abdallah A. M., Sparrius M., Reinders E., Vandenbroucke-Grauls C. M. and Bitter W. (2004) Mycobacterium marinum strains can be divided into two distinct types based on genetic diversity and virulence. *Infect Immun* **72**, 6306-12.

van der Sar A. M., Spaink H. P., Zakrzewska A., Bitter W. and Meijer A. H. (2009) Specificity of the zebrafish host transcriptome response to acute and chronic mycobacterial infection and the role of innate and adaptive immune components. *Mol Immunol* **46**, 2317-32.

Varnholt H., Drebber U., Schulze F., Wedemeyer I., Schirmacher P., Dienes H. P. and Odenthal M. (2008) MicroRNA gene expression profile of hepatitis C virus-associated hepatocellular carcinoma. *Hepatology* **47**, 1223-32.

Wang B., Feng P., Xiao Z. and Ren E. C. (2009a) LIM and SH3 protein 1 (Lasp1) is a novel p53 transcriptional target involved in hepatocellular carcinoma. *J Hepatol* **50**, 528-37.

Wang P., Zou F., Zhang X., Li H., Dulak A., Tomko R. J., Jr., Lazo J. S., Wang Z., Zhang L. and Yu J. (2009b) microRNA-21 negatively regulates Cdc25A and cell cycle progression in colon cancer cells. *Cancer Res* **69**, 8157-65.

Wang W. S., Chen P. M., Hsiao H. L., Wang H. S., Liang W. Y. and Su Y. (2004) Overexpression of the thymosin beta-4 gene is associated with increased invasion of SW480 colon carcinoma cells and the distant metastasis of human colorectal carcinoma. *Oncogene* **23**, 6666-71.

Williams A. E., Perry M. M., Moschos S. A., Larner-Svensson H. M. and Lindsay M. A. (2008) Role of miRNA-146a in the regulation of the innate immune response and cancer. *Biochem Soc Trans* **36**, 1211-5.

Xia H., Qi Y., Ng S. S., Chen X., Li D., Chen S., Ge R., Jiang S., Li G., Chen Y., He M. L., Kung H. F., Lai L. and Lin M. C. (2009) microRNA-146b inhibits glioma cell migration and invasion by targeting MMPs. *Brain Res* **1269**, 158-65.

Xu T., Zhu Y., Wei Q. K., Yuan Y., Zhou F., Ge Y. Y., Yang J. R., Su H. and Zhuang S. M. (2008) A functional polymorphism in the miR-146a gene is associated with the risk for hepatocellular carcinoma. *Carcinogenesis* **29**, 2126-31.

Xu X. R., Huang J., Xu Z. G., Qian B. Z., Zhu Z. D., Yan Q., Cai T., Zhang X., Xiao H. S., Qu J., Liu F., Huang Q. H., Cheng Z. H., Li N. G., Du J. J., Hu W., Shen K. T., Lu G., Fu G., Zhong M., Xu S. H., Gu W. Y., Huang W., Zhao X. T., Hu G. X., Gu J. R., Chen Z. and Han Z. G. (2001) Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver. *Proc Natl Acad Sci U S A* **98**, 15089-94.

Zhang Y., Feurino L. W., Zhai Q., Wang H., Fisher W. E., Chen C., Yao Q. and Li M. (2008a) Thymosin Beta 4 is overexpressed in human pancreatic cancer cells and stimulates proinflammatory cytokine secretion and JNK activation. *Cancer Biol Ther* **7**, 419-23.

Zhang Z., Li Z., Gao C., Chen P., Chen J., Liu W., Xiao S. and Lu H. (2008b) miR-21 plays a pivotal role in gastric cancer pathogenesis and progression. *Lab Invest* **88**, 1358-66.

Zhou R., Hu G., Liu J., Gong A. Y., Drescher K. M. and Chen X. M. (2009) NF-kappaB p65-dependent transactivation of miRNA genes following Cryptosporidium parvum infection stimulates epithelial cell immune responses. *PLoS Pathog* **5**, e1000681.

Chapter 6

Summary and General discussion

Recently, new developments in transcriptome profiling technologies have significantly advanced genomic research. The use of these technologies to analyze the expressed transcripts in several prokaryotic and eukaryotic genomes has revealed the high complexity of transcriptomes. Furthermore, quantification of expressed transcripts using these technologies has facilitated our understanding of the molecular background of specific developmental processes and pathological conditions. Several high-throughput methods for transcriptome profiling have been developed with two basic approaches, microarray-based and sequencing-based methods, both offering great opportunities for large-scale analysis. Microarray analysis is a widely used technique able to simultaneously interrogate thousands of transcripts and has led to crucial advances in a variety of biological problems. However, microarrays have several limitations inherent to hybridization-based analysis and are biased by a predefined array content. The discovery of next generation sequencing methods allows the detection of millions of transcripts simultaneously and overcomes the limitations of microarray technology. The two approaches can be considered as complementary techniques to scrutinize the genome, producing not only overlapping results but also disclosing technology-dependent findings. Transcriptome analysis is invaluable to reveal the regulation of gene expression. Gene expression is regulated on different levels, from transcription to post-translational modification of the protein. Recently, it has become evident that small non-coding RNAs such as miRNAs are important components of this regulation. MiRNAs negatively regulate mRNA translation or stability and have been implicated in a variety of biological processes, for instance the development and function of the immune system. Aberrant expression of miRNAs has been associated with different disorders such as infectious diseases and cancer. Revealing groups of differentially expressed miRNAs under disease conditions and demonstrating their biological functions is therefore important to our understanding of pathogenesis. To study mRNA and miRNA profiles affected by disease processes, both *in vitro* and *in vivo* models are applied. Zebrafish is a commonly applied vertebrate model organism used for developmental biology, embryogenesis and more recently for immunological research due to the remarkable similarities of its immune system to that of human. In the recent years zebrafish models for several infectious diseases and types of cancer have proved useful additions to mammalian models.

Taking advantage of the novel technological opportunities for transcriptome analysis and the possibilities of applying zebrafish as a model organism, this thesis focuses on the analysis of transcriptome complexity during infectious disease and cancer and aims to gain insight into regulatory functions of microRNAs.

In **Chapter 2** we report for the first time on profiling of the transcriptional response to infectious disease using Solexa/Illumina's digital gene expression (DGE or Tag-Seq) technology, a tag-based transcriptome sequencing method. In this study we investigated the zebrafish host response to *Mycobacterium marinum* infection modeling human tuberculosis. We generated two Tag-Seq libraries of control fish

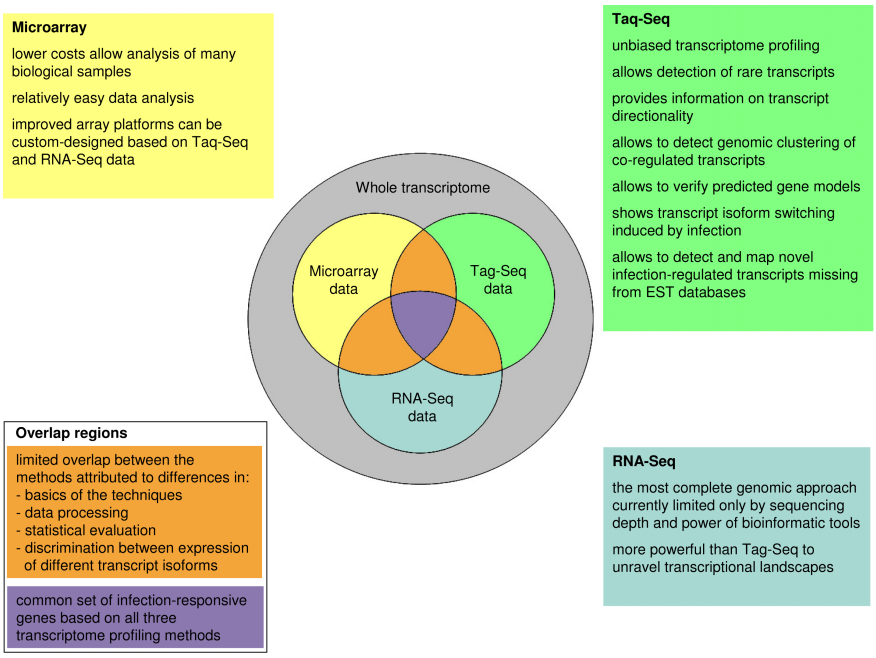


Figure 1. Complementary use of different transcriptome profiling techniques. Microarray, Tag-Seq and RNA-Seq methods as used in this thesis to study the response to bacterial infections in zebrafish can be considered as complementary techniques to analyze the transcriptome. The specific advantages of the different methods are summarized in the figure.

and two Tag-Seq libraries of mycobacterium-infected fish and confirmed the technical reproducibility of the method. With this approach we detected 5-8 million tags per library and mapping analysis showed that the libraries contained tag sequences for over 70% of all genes represented in transcript databases. The tag sequences that were significantly changed upon mycobacterium infection mapped to approximately 2% of the database transcripts. In addition to quantitative information on gene expression, Tag-Seq analysis also provided information on the directionality of the transcripts. There is increasing evidence for the occurrence of antisense transcription; however, its biological relevance is still unclear. In our study, approximately 40% of the mapping events were detected on the antisense strand of database transcripts consistently with other recent deep sequencing studies. However, when taking only the significantly changed tag entities into account we found that the transcriptional regulation in the mycobacterium-induced immune response acts most strongly on the sense strand. Thus, our work focused on the regulation of the sense strand transcripts, although the regulation of antisense expression is of great interest and warrants further research.

To validate our results we compared our Tag-Seq data to a previously compiled

reference set of mycobacterium-regulated genes in zebrafish generated by multiplatform microarray analysis. This comparison revealed strong correlation between the two methods. Furthermore, we performed microarray analysis using a new custom-designed Agilent platform, also demonstrating a considerable overlap with the Tag-Seq data. Moreover, gene ontology analysis showed that both profiling methods detected similar functional groups of genes, further supporting comparability of the two approaches. On the other hand, comparisons uncovered some differences due to limitations of microarray (limited sensitivity, cross-hybridization problems and inadequate probe design) and Tag-Seq (unable to identify some transcripts that lack a unique tag sequence or cutting site for the DGE anchoring enzyme). Therefore, microarray analysis and Tag-Seq can be considered complementary to each other (Fig. 1). In addition to the comparison with microarray analysis, we verified a subset of differentially expressed tag sequences by qPCR analysis, which supported the Tag-Seq results even in the cases of conflicting results between Tag-Seq and microarray. Subsequently, we demonstrated with several examples how additional valuable information can be obtained from Tag-Seq transcriptome profiling. Overall, we showed that Tag-Seq data can be used to investigate genomic clustering of co-regulated transcripts, to verify predicted gene models, to detect transcript isoform switching induced by infection, and to detect and map novel mycobacterium-regulated transcripts missing from EST databases. Taken together, with these results we proved here that Solexa/Illumina's digital gene expression system is a remarkably useful technology for transcriptome quantification and revealing the complexity of the transcriptome.

In **Chapter 3** we used tag-based (Tag-Seq) and full transcript sequencing (RNA-Seq) to extend our knowledge on the transcriptome of zebrafish embryos during the innate host response to *Salmonella enterica* serovar Typhimurium (*S. typhimurium*) infection. We analyzed *S. typhimurium* infected 1-day-old zebrafish embryos at 8 hours post infection (hpi) based on the strong induction of inflammatory genes detected previously by microarray. We determined a set of 165 transcripts that were commonly responsive in Tag-Seq and microarray analysis. The comparison of Tag-Seq with RNA-Seq revealed an overlap of 241 transcripts differentially expressed, thus we showed, for the first time, the comparable performance of both methods in quantifying the transcriptome response to infection (Fig. 1). Combining the sequencing-based and microarray-based transcriptome data we presented 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 anti-microbial activities, as well as many known or novel proteins not previously linked to the immune response. Furthermore, we compared our Tag-Seq data of *S. typhimurium* infection performed in zebrafish embryos to the Tag-Seq data of *M. marinum* infection in adult zebrafish reported in Chapter 2. The end stage of *M. marinum* infection in adult fish is associated with a strong inflammatory response,

similar to what we observed during *S. typhimurium* infection of zebrafish embryos. The comparison revealed a common set of 206 up-regulated infection-responsive genes, including transcription factors and signaling components involved in the innate host defense, as well as genes not previously linked to the immune response. With this transcriptome analysis we provided a valuable reference for further studies of host-pathogen interactions using zebrafish infection models.

In **Chapter 4** we aimed to reveal miRNA functions in the vertebrate immune system during infectious diseases. As in the previous chapters, we used adult zebrafish and zebrafish larvae infected with the bacterial pathogens *M. marinum* and *S. typhimurium* as disease models. By using embryonic and adult zebrafish in our experiments we could determine miRNA expression distinctive to the innate and adaptive immune response, due to the fact that in embryos only the innate immune system is developed. Using a custom-designed Agilent microarray platform we identified a set of common infection-responsive zebrafish miRNAs, including members of the miR-21, miR-146 and miR-181 families. These results were consistent with previous studies showing that these highly conserved miRNAs are strongly associated to the immune system and also to development of cancer in human and mammalian models. However, our study is the first that links these miRNAs with mycobacterium and salmonella infections. Moreover, our results support previous suggestions that common miRNAs might participate in the regulation of the immune response and processes of cancer.

The miR-146 family has previously been implicated in the regulation of innate immunity signaling pathways in human and animal models. Our zebrafish data are consistent with these findings, as miR-146a/b were commonly induced by infection in adult zebrafish and in embryos that rely solely on innate defense mechanisms. The role of miR-146a/b in innate immunity is believed to be a fine-tuning function of the TLR and IL1R signaling involved in pathogen recognition and activation of the innate immune response. To investigate if the function of miR-146 might be conserved in zebrafish, we subsequently identified predicted target genes of this family. We showed that *irak1* and *traf6*, two genes involved in TLR and IL1R signaling and targeted by mammalian miR-146a/b, also contain target sites for these miRNAs in zebrafish. Additionally, we found that the transcript encoding MyD88, the common adaptor in TLR and IL1R signaling pathways, contained putative target sites for zebrafish miR-146a/b. The induction of miR-146 by bacterial pathogens is in line with the mRNA expression data in chapters 2 and 3, showing that several TLR pathway genes as well as downstream effector genes and negative regulatory genes of the TLR pathway are induced during infections in zebrafish. Thus, our data support that miRNA regulation plays an important role as a feed back mechanism to properly control the immune response. We also used target prediction software to identify other possible targets of the miR-146 family. Most of the predicted miR-146 target genes conserved between human and zebrafish were previously linked to immune response processes related to apoptosis, regulation of NF-κB transcription factor

activity, haemostasis, infections, T-cell development and function, TLR signalling, and also tumor progression. Thus, these target predictions further support the importance of miR-146 in the regulation of the immune response. With this study we demonstrated that the zebrafish is a useful model to scrutinize miRNA functions in the vertebrate immune system based on the highly conserved infection-responsive miRNAs that we detected during bacterial infections and the considerable overlap in predicted miRNA target genes between human and zebrafish.

In **Chapter 5** we assessed miRNA expression patterns in zebrafish liver tumors in order to increase understanding of the contribution of miRNAs to the mechanisms of cancer. Zebrafish has been demonstrated to be a valuable disease model for liver cancer due to remarkable molecular similarities to human hepatocellular carcinoma (HCC), including conserved expression signatures and transcriptional regulation of several cancer-related gene groups. To determine if there is also conservation at the level of miRNA expression in zebrafish and human liver tumors, we performed miRNA transcriptome profiling of carcinogen-induced zebrafish liver tumor samples. We compared the significantly differentially expressed miRNAs in zebrafish liver tumors with the miRNAs that had previously been associated with human HCC and identified a set of miRNAs that were commonly deregulated in liver tumors of both organisms. The most notable similarities were the common up-regulation of miR-21, miR-23a, miR-146a/b, miR-221 and miR-222, and the common down-regulation of miR-1 and miR-122. Furthermore, we identified another set of miRNAs in zebrafish liver tumors that have not been previously linked to human HCC but have been connected to other types of human cancer with roles in tumorigenicity, tumor suppression, apoptosis, cell proliferation, differentiation, invasion, and metastasis. Besides the deregulation of primary miRNAs, in several cases we also detected deregulation of putative star sequences of the miRNAs. Furthermore, we observed differential expression of a large group of probe sequences corresponding to hairpin structures in the zebrafish genome that might represent additional miRNAs or other types of non-coding RNAs that remain to be further studied. Our results therefore further support that zebrafish liver tumors are suitable models of human liver tumors, as they share high molecular similarities, and demonstrate that custom Agilent microarray technology is useful to detect expression of novel predicted transcripts representing putative miRNAs.

Subsequently, we compared the liver tumor-related miRNA set to the miRNA set induced by Mycobacterium infection that we established in chapter 4. The comparison revealed a common group of tumor and immune-related miRNAs including members of the miR-15, miR-16, miR-21, miR-34, and miR-146 families, supporting the view that many miRNAs implicated in regulation of the immune response are also connected to cancer processes.

Finally, we investigated the possible conservation of target genes of cancer-related miRNAs. Functional annotation of experimentally validated targets of human liver tumor-related miRNAs revealed that they are involved in various cancer-relat-

ed processes and pathways. Prediction of miRNA recognition sites in the homologous zebrafish genes showed that several target genes of the miR-146 family (*IRAK1, TRAF6, IRF5*), the miR-221/miR222 families (*CDKN1B, CDKN1C, KIT*), and the miR-1 family (*GJA1, PDCD4, TPM4, LASP1, TMSB4X, RABL2A, RABL2B*) may be conserved between human and zebrafish. These genes are involved in the immune system as well as in cancer-related processes, such as cell cycle regulation, proliferation, apoptosis, cell adhesion, cell motility, and cytoskeletal organization, and they function in cancer- and immune-related pathways, such as MAPK signaling, NF-κB signaling, p53 signaling, small GTPase-mediated signal transduction, and the ErbB signaling pathway. These results further support the similarity between zebrafish and human liver cancer and the hypothesis that miRNAs might play an important role in regulating these cancer-related mechanisms in both species. Therefore, our miRNA results present a new level of conservation between zebrafish and human liver tumors in addition to the high conservation at the level of gene expression and transcriptional regulation by transcription factors. Hence, zebrafish as a liver cancer model can offer further valuable information of miRNA functions in cancer and contribute to studying the interaction between cancer and the immune system.

In conclusion, the studies described in this thesis further support that zebrafish is a valuable model system for infectious diseases and cancer as well as for scrutinizing the regulatory functions of miRNAs in the vertebrate immune system and disease. Using next-generation sequencing methods to analyze zebrafish mycobacterium and salmonella infection models, we have demonstrated for the first time that Solexa/Illumina's digital gene expression (DGE or Tag-Seq) system is a useful technology for transcriptome analysis of the vertebrate host response to infection. Our results demonstrated that this technology provides valuable complementary information to array-based methods, such as insight into infection-induced expression of alternative transcript isoforms. Furthermore, we have shown that another deep-sequencing technology, RNA-Seq, is also useful for quantification of the transcriptome response to infection. Since RNA-Seq is even more powerful than Tag-seq for studying transcriptional landscapes, future in depth analysis of RNA-Seq data can be expected to provide further insights into alternative isoform expression during infection. By combining sequencing-based and microarray-based transcriptome data we have provided a valuable reference set of infection-responsive genes in zebrafish infection models. These infection-responsive genes include not only well-known vertebrate immune response genes but also genes not previously linked to infectious disease and that are of great interest for functional analysis of their possible role in host-pathogen interactions. In addition, we have evaluated a set of infection-responsive and liver cancer-related miRNAs conserved between human and zebrafish, supporting that common miRNAs might play regulatory roles in the immune response and processes underlying the development of cancer. Our combined mRNA and miRNA transcriptome data provide a strong basis for future applications of zebrafish as an infection and cancer model.

Samenvatting

In het wetenschappelijk onderzoek naar de werking van alle genen (genomica) is de laatste jaren een enorme vooruitgang geboekt door nieuwe technologische ontwikkelingen waarmee de expressie van genen (het transcriptoom) op grote schaal in kaart gebracht kan worden. De toepassing van deze nieuwe technologieën heeft duidelijk gemaakt dat de transcriptomen van zowel prokaryote als eukaryote organismen veel complexer zijn dan voorheen werd gedacht. Bovendien heeft het gebruik van deze technologieën voor transcriptoomanalyse nieuwe inzichten verschaft in de moleculaire basis van ontwikkelings- en ziekteprocessen. De technologieën voor grootschalige transcriptoomanalyse berusten op twee benaderingen, microarrayanalyse en sequentieanalyse. Microarrayanalyse is een veelgebruikte techniek waarmee de expressieniveaus van duizenden transcripten gelijktijdig bepaald kunnen worden. Deze techniek heeft bijgedragen aan de oplossing van een groot aantal biologische vraagstellingen. Echter, microarrayanalyse heeft verschillende beperkingen die inherent zijn aan alle analysemethoden die gebaseerd zijn op hybridisatie van DNA of RNA. Bovendien is onderzoek waarbij microarrays gebruikt worden bevooroordeeld in die zin dat de samenstelling van de genen op de array voorafgaand aan het experiment is bepaald. Met de opkomst van de nieuwe generatie van technologieën voor sequentieanalyse is het mogelijk geworden om miljoenen transcripten tegelijk te detecteren en zijn de beperkingen van microarrayanalyse overwonnen. Transcriptoomanalyse is onmisbaar om de regulering van genexpressie aan het licht te brengen. Genexpressie wordt gereguleerd op verschillende niveaus, van transcriptie tot post-translationele modificatie van het gecodeerde eiwit. Recentelijk is het duidelijk geworden dat kleine, niet-coderende RNA-moleculen, zoals microRNA's (miRNA's) belangrijke componenten van het reguleringsmechanisme van genexpressie zijn. MiRNA's werken als negatieve regulatoren van mRNA-translatie of mRNA-stabiliteit en spelen een rol bij een groot aantal biologische processen, bijvoorbeeld de ontwikkeling en de functie van het immuunsysteem. Afwijkende expressie van miRNA's is waargenomen bij verschillende ziekten, waaronder infectieziekten en kanker. Het in kaart brengen van de verschillen in miRNA-expressie tijdens ziekteprocessen en het ontrafelen van de biologische functies van miRNA's is daarom van groot belang voor het begrip van de moleculaire mechanismen van ziekten. Om te onderzoeken hoe ziekteprocessen de expressiepatronen van mRNA's en miRNA's beïnvloeden, worden zowel *in vitro*-systemen (celculturen) als *in vivo*-systemen (proefdiermodellen) toegepast. De zebravis is een veelgebruikt modelorganisme voor ontwikkelingsbiologie en recentelijk ook voor immunologisch onderzoek, vanwege de opvallende overeenkomsten tussen het immuunsysteem van de zebravis en dat van de mens. In de laatste jaren hebben zebravismodellen voor verschillende infectieziekten en vormen van kanker zich bewezen als nuttige aanvullingen op zoogdiermodellen.

Het doel van het in dit proefschrift beschreven onderzoek was om inzicht te krij-

gen in de complexiteit van het transcriptoom tijdens infectieziekten en kanker en in de regulerende functies van miRNA's bij deze ziekteprocessen, door gebruik te maken van de nieuwe technologische mogelijkheden voor transcriptoomanalyse en de mogelijkheden van de zebravis als modelorganisme.

In **Hoofdstuk 2** rapporteren wij voor het allereerst over de toepassing van grootschalige sequentieanalyse voor onderzoek naar de reactie van het transcriptoom van een gastheerorganisme op een bacteriële infectieziekte. Hierbij hebben wij gebruik gemaakt van het Solexa/Illumina-systeem voor digitale genexpressie (DGE of Tag-Seq), een methode die is gebaseerd op het detecteren van korte transcript-specifieke sequenties ("tags"). In deze studie onderzochten wij de gastheerreactie van de zebravis op een infectie met *Mycobacterium marinum*, een model voor humane tuberculose. Door twee Tag-Seq-banken te maken van controlevissen en twee Taq-seq-banken van mycobacterium-geïnfecteerde vissen konden wij de technische reproduceerbaarheid van de methode bevestigen. Met deze benadering konden wij 5-8 miljoen sequenties per bank detecteren en de kartering daarvan toonde aan deze representatief waren voor meer dan 70% van alle genen die aanwezig zijn in de publieke transcriptdatabanken. De Tag-Seq-sequenties die een significant verschillende expressie vertoonden na mycobacterium-infectie omvatten ongeveer 2% van de transcripten in de databases. Behalve kwantitatieve informatie over genexpressie, kon uit de Tag-Seq-analyse ook worden afgeleid of de transcriptsequenties van de coderende ("sense") of van de niet-coderende ("antisense") streng van het DNA afkomstig waren. Er is toenemend bewijs voor het bestaan van antisense-transcripten, hoewel de biologische relevantie hiervan nog veel vragen oproept. In onze studie vonden wij ongeveer 40% antisense-sequenties, in overeenstemming met andere recente grootschalige sequentiestudies in de muis en in menselijke cellen. Echter, het grootste deel van de sequenties die significant verschillend waren tussen geïnfecteerde vissen en de controles betrof sense-transcripten. Dit geeft aan dat de transcriptionele regulering van de door mycobacterium-geïnduceerde immuunreactie vooral op het niveau van de sense-transcripten plaatsvindt. Daarom hebben wij onze verdere studie gericht op de analyse van de sense-transcripten, hoewel de regulering van antisense-expressie ook zeer interessant is en verder onderzoek behoeft.

Om onze resultaten te valideren hebben wij onze Tag-Seq-data vergeleken met een eerder samengestelde referentieset van mycobacterium-gereguleerde genen van de zebravis, die bepaald was door middel van een multiplatform-microarrayanalyse. Deze vergelijking liet een sterke correlatie tussen de twee methoden zien. Daarnaast hebben wij een nieuwe microarrayanalyse uitgevoerd met een zelfontworpen Agilent-platform, waaruit ook een aanzienlijke overlap met de Tag-Seq-data bleek. Het feit dat met Agilent-microarrayanalyse en met Tag-Seq-analyse dezelfde functionele groepen van genen gedetecteerd werden, toonde aan dat deze twee benaderingen voor transcriptoomanalyse goed vergelijkbaar zijn. Aan de andere kant bracht de vergelijking ook verschillen aan het licht die verklaard kunnen worden door de beperkingen van microarrayanalyse (zoals beperkte gevoeligheid, hybridisatie-artefac-

ten en onvolkomenheden in het array-ontwerp) of van de Tag-Seq-methode (waarbij sommige transcripten gemist worden omdat ze geen unieke sequentie bevatten of een knipplaats voor het enzym dat gebruikt wordt voor de vervaardiging van Tag-Seq-banken). Daarom kunnen microarrayanalyse en Tag-Seq-analyse beschouwd worden als complementaire technieken. Naast de vergelijking met microarrayanalyse, hebben wij ook enkele van de differentieel tot expressie komende transcriptsequenties geverifieerd met behulp van kwantitatieve PCR-analyse. Dit ondersteunde de Tag-Seq-resultaten zelfs in gevallen waar er een discrepantie was tussen de Taq-Seq- en microarray-resultaten. Vervolgens hebben wij aan de hand van verschillende voorbeelden laten zien hoe transcriptoomanalyse met behulp van Tag-Seq additionele waardevolle informatie over genexpressie kan opleveren. Op basis van de Taq-Seq-data konden wij concluderen dat er groepen zijn van gecoördineerd-gereguleerde genen op bepaalde chromosomen. De Tag-Seq-data waren tevens bruikbaar om voorspellingen van genmodellen (intron-exon-structuren) te verifiëren. Bovendien toonden de Tag-Seq-data aan dat infectie kan leiden tot expressie van andere transcript-isovormen en konden wij nieuwe mycobacterium-gereguleerde transcripten in kaart brengen die ontbraken in de beschikbare transcript-databanken. Bij elkaar genomen hebben wij met deze resultaten aangetoond dat het Solexa/Illumina-systeem voor digitale genexpressie een uiterst waardevolle technologie is om expressie te kwantificeren en om de complexiteit van het transcriptoom te ontrafelen.

In **Hoofdstuk 3** hebben wij sequentieanalyse gebruikt om meer kennis te vergaren over het transcriptoom van zebravisembryo's tijdens de reactie van het aangeboren ("innate") immuunsysteem op een infectie met *Salmonella enterica* serovar Typhimurium (*S. typhimurium*). Hierbij hebben wij twee sequentiemethoden toegepast, de Tag-Seq-methode, die is gebaseerd op het bepalen van korte transcript-specifieke sequenties verkregen door enzymatische digestie, en de RNA-Seq-methode, waarbij de gehele transcripten worden gefragmenteerd en de sequentie bepaald. Wij analyseerden 1 dag oude zebravisembryo's op 8 uur na infectie met *S. typhimurium*, welk tijdspunt was gebaseerd op eerder microarraydata waarbij een sterke inflammatoire reactie was waargenomen. Onze analyse resulteerde in een set van 165 transcripten die een overeenkomstige reactie vertoonden in zowel de Tag-Seq-analyse als in de microarrayanalyse. Bij de vergelijking van Tag-Seq met RNA-Seq vonden wij een overlap van 241 transcripten die differentieel tot expressie kwamen, waarmee wij hebben laten zien dat deze twee sequentiebepalingsmethoden een vergelijkbare prestatie leveren bij het kwantificeren van de transcriptoomreactie op infectie. De gecombineerde transcriptoomdata van de sequentie- en microarrayanalyses hebben wij vervolgens gebruikt om een geannoteerde referentieset samen te stellen van infectie-responsieve genen in zebravisembryo's. De genen in deze referentieset coderen voor transcriptiefactoren, signaaltransductie-eiwtten, cytokinen en chemokinen, complementfactoren, eiwitten betrokken bij apoptose en proteolyse alsmede eiwitten met anti-microbiële activiteiten. Daarnaast omvatte de referentieset genen voor zowel bekende als nieuwe eiwitten die niet eerder met de immuunreactie in verband

waren gebracht. Wij hebben de Tag-Seq-data van de *S. typhimurium*-infectie van zebravisembryo's ook vergeleken met de Tag-Seq-data van *M. marinum*-infectie in volwassen zebravissen uit Hoofdstuk 2. Het eindstadium van de *M. marinum*-infectie in volwassen vissen gaat gepaard met een sterke inflammatoire reactie, vergelijkbaar met de reactie van zebravisembryo's op *S. typhimurium*-infectie. De vergelijking resulteerde in een gemeenschappelijke set van 206 genen die een verhoogd expressieniveau vertoonden in beide infectiemodellen. Deze set omvatte transcriptiefactorgenen en signaaltransductiegenen betrokken bij de aangeboren immuunreactie, evenals genen waarvoor een rol bij de immuunreactie niet eerder bekend was. Deze transcriptoomdata vormen een waardevolle referentie voor toekomstig onderzoek naar gastheer-pathogeen-interacties met gebruik van zebravis-infectiemodellen.

Het onderzoek beschreven in **Hoofdstuk 4** was gericht op functies van miRNA's bij het immuunsysteem van gewervelde dieren (vertebraten) tijdens infectieziekten. Net als in de voorgaande hoofdstukken, gebruikten wij als ziektemodellen zowel volwassen zebravissen als zebravisembryo's die geïnjecteerd waren met de bacteriele pathogenen *M. marinum* and *S. typhimurium*. Door embryonale en volwassen zebravissen in onze experimenten te gebruiken konden wij onderscheid maken in miRNA-expressiepatronen die kenmerkend waren voor het aangeboren ("innate") immuunsysteem en voor het verworven ("adaptive") immuunsysteem, aangezien in embryo's alleen nog het aangeboren systeem is ontwikkeld. Door gebruik te maken van een nieuw Agilent-microarrayplatform voor miRNA-detectie konden we een set van miRNA's identificeren die algemeen reageren op verschillende typen infecties, waaronder leden van de miR-21-, miR-146- en miR-181-families. De resultaten waren in overeenstemming met eerdere studies die een verband hadden aangetoond van deze miRNA's met het immuunsysteem en met kanker in humane patiëntenmaterialen en celcultuursystemen en in proefdiermodellen van zoogdieren. Echter, onze studie is de eerste die deze miRNA's in verband brengt met mycobacterium- en salmonella-infecties. Bovendien ondersteunen onze resultaten eerdere hypothesen dat regulering van de immuunreactie en kankerprocessen afhankelijk is van gemeenschappelijke miRNA's.

De miR-146-familie is eerder in verband gebracht met de regulering van signaalroutes van het aangeboren immuunsysteem in de mens en in diermodellen. Onze data zijn hiermee in overeenstemming, aangezien miR-146a en miR-146b niet alleen geïnduceerd werden in volwassen zebravissen, maar ook in embryo's die volledig afhankelijk zijn van het aangeboren immuunsysteem. Er wordt gedacht dat miR-146a en -b een rol spelen bij de fijnafstelling van de Toll-receptor- en interleukine-1-receptor-signaalroutes (TLR- en IL1R-signaalroutes) die een sleutelrol spelen bij de herkenning van pathogenen en bij de activering van de aangeboren immuunreactie. Om te onderzoeken of de functie van miR-146 geconserveerd is in de zebravis hebben wij een voorspelling gemaakt van de genen die door deze miRNA's gereguleerd zouden kunnen worden (doelwitgenen of "target genes"). Wij vonden herkenningsplaatsen voor miR-146-a en -b van de zebravis in *irak1* en *traf6*, twee genen die een rol spelen

in de TLR- en IL1R-signaalroutes en die beide herkend ook worden door de miR-146 -familie van zoogdieren. Daarnaast vonden wij dat het transcript dat codeert voor MyD88, een essentieel adaptereiwit in de TLR- en IL1R-signaalroutes, ook mogelijke herkenningsplaatsen voor miR-146a en -b van de zebravis bevat. De inductie van miR-146 door bacteriële pathogenen is in overeenstemming met de genexpressiedata uit hoofdstukken 2 en 3, die laten zien dat tijdens infecties in de zebravis er verschillende genen van de TLR-signaalroute geïnduceerd worden, evenals doelwitgenen en negatieve regulatoren van deze route. Daarom ondersteunen onze data de hypothese dat miRNA-regulering een belangrijke rol speelt als een negatief terugkoppelingsmechanisme om de immuunreactie op de juiste manier onder controle te houden. Om andere mogelijke doelwitgenen van de miR-146-familie te identificeren hebben wij gebruik gemaakt van softwareprogramma's waarmee herkenningsplaatsen voor miRNA's voorspeld kunnen worden. De meeste van de voorspelde doelwitgenen van miR-146 die geconserveerd waren tussen de zebravis en de mens, waren genen die eerder in verband gebracht waren met infecties, tumorprogressie en processen die onderdeel uitmaken van de immuunreactie, zoals TLR-signaaltransductie, regulering van de activiteit van de centrale transcriptiefactor NF-κB, apoptose, hemostase en T-cell-ontwikkeling en -functie. Onze voorspelling van geconserveerde miR-146-doelwitgenen ondersteunt daarom de veronderstelde functie van miR-146 bij de regulering van de immuunreactie. Met deze studie hebben wij laten zien dat de zebravis een bruikbaar model is om miRNA-functies bij het immuunsysteem van vertebraten te onderzoeken, aangezien de miRNA's die tijdens bacteriële infecties tot expressie komen sterk geconserveerd zijn en er een aanzienlijk overlap is in de voorspelde doelwitgenen van miRNA's in de mens en in de zebravis.

Om meer inzicht te krijgen in de rol van miRNA's in mechanismen van kanker hebben wij in **Hoofdstuk 5** miRNA-expressiepatronen in levertumoren van de zebravis onderzocht. Eerder was aangetoond dat de zebravis een waardevol ziektemodel is voor leverkanker, aangezien er opvallende moleculaire overeenkomsten zijn met humane hepatocellulaire carcinoma (HCC), zoals geconserveerde patronen van genexpressie en transcriptionele regulering van verschillende bij kanker betrokken gengroepen. Op te bepalen of er ook conservering is op het niveau van miRNA-expressie tussen levertumoren van de zebravis en van de mens, hebben wij een miRNA-transcriptoomanalyse uitgevoerd van carcinogeen-geïnduceerde levertumoren van zebravissen. Wij hebben miRNA's die een significant verschillende expressie vertoonden tussen levertumoren en gezonde lever van de zebravis vergeleken met de miRNA's die eerder in verband waren gebracht met humane HCC. Deze vergelijking resulteerde in een set van miRNA's waarvan de expressie ontregeld was in levertumoren van zowel de zebravis als de mens. De meest opvallende overeenkomsten waren de inductie van miR-21, miR-23a, miR-146a/b, miR-221 en miR-222, en de repressie van miR-1 en miR-122. Bovendien konden wij diverse andere miRNA's aantonen in levertumoren van de zebravis die niet eerder in verband gebracht waren met humane HCC, maar die wel eerder geassocieerd waren met verschillende andere typen van

kanker. Voor deze miRNA's zijn functies gepostuleerd bij het ontstaan van tumoren, bij de suppressie van tumoren, bij apoptose, proliferatie, differentiatie, invasiviteit en metastase. Naast de ontregeling van primaire miRNA's, vonden wij in verschillende gevallen ook ontregeling van de zogenoemde ster-sequenties. Dit zijn miRNA's die afkomstig zijn van de tegenovergestelde arm van de haarspeldstructuur van precursor-miRNA's. Bovendien konden wij een differentiële expressie waarnemen van een grote groep van sequenties die overeenkomen met voorspelde haarspeldstructuren in het zebravisgenoom. Van deze structuren is het nog niet bekend of het mogelijk ook miRNA-genen zijn of dat er andere typen niet-coderende RNA's van worden afgeschreven, die nog verdere studie behoeven. Samen met de eerder gepubliceerde genexpressiestudies tonen onze miRNA-expressiedata aan dat zebravis-levertumoren een geschikt model zijn voor het bestuderen van humane levertumoren, gezien de aanzienlijke moleculaire overeenkomsten. Tevens tonen onze resultaten aan dat de Agilent-microarraytechnology zeer bruikbaar is om nieuw voorspelde miRNA's of andere niet-coderende, regulerende RNA-moleculen te detecteren.

Vervolgens hebben wij de set van levertumor-gerelateerde miRNA's vergeleken met de set van miRNA's die geïnduceerd waren door mycobacterium-infectie, zoals bepaald in hoofdstuk 4. Deze vergelijking resulteerde in een gemeenschappelijke groep van tumor- en immuun-gerelateerde miRNA's, waaronder leden van de miR-15-, miR-16-, miR-21-, miR-34- en miR-146-families. Dit ondersteunt de opvatting dat veel miRNA's die betrokken zijn bij regulering van de immuunreactie, ook een rol spelen bij processen van kanker.Tenslotte hebben wij onderzocht of de doelwitgenen van kanker-gerelateerde miRNA's mogelijk ook geconserveerd zijn. Functionele annotatie van experimenteel-gevalideerde doelwitgenen van miRNA's die betrokken zijn bij humane leverkanker, toont aan dat deze genen een rol spelen in diverse kanker-gerelateerde processen en signaalroutes. Voorspelling van de miRNA-herkenningsplaatsen in de overeenkomstige genen van de zebravis liet zien dat verschillende doelwitgenen van de miR-146-familie (*IRAK1, TRAF6, IRF5*), de miR-221/miR222-familie (*CDKN1B, CDKN1C, KIT*) en de miR-1-familie (*GJA1, PDCD4, TPM4, LASP1, TMSB4X, RABL2A, RABL2B*) mogelijk geconserveerd zijn tussen de mens en de zebravis. Deze genen zijn betrokken bij het immuunsysteem zowel als bij kanker-gerelateerde processen, zoals regulering van de celcyclus, proliferatie, apoptose, celadhesie, celmotiliteit en organisatie van het cytoskelet. Bovendien functioneren deze genen in kanker- en immuun-gerelateerde signaalroutes, zoals MAPK-, NF-κB-, p53- en ErbB-signaaltransductie, alsmede signaaltransductie door kleine G-eiwitten. Door deze resultaten wordt de overeenkomst tussen leverkanker van de zebravis en de mens verder versterkt, evenals de hypothese dat miRNA's een belangrijke rol spelen bij deze kanker-gerelateerde processen in beide soorten. Onze miRNA-resultaten hebben een nieuw niveau van conservering tussen levertumoren van de zebravis en de mens aan het licht gebracht naast de al aangetoonde sterke conservering van genexpressiepatronen en regulering door transcriptiefactoren. Daarom kan de zebravis als leverkankermodel bijdragen om miRNA-functies

bij kanker te ontrafelen en om de interactie tussen kanker en het immuunsysteem verder te onderzoeken.

Concluderend heeft het in dit proefschrift beschreven onderzoek ondersteunend bewijs geleverd dat de zebravis een waardevol modelsysteem is voor infectieziekten en kanker, evenals voor het bestuderen van de regulerende functies van miRNA's bij het immuunsysteem en ziekten van vertebraten. Met het gebruik van de nieuwe generatie van technologieën voor sequentieanalyse voor de studie van mycobacterium- and salmonella-infectiemodellen van de zebravis, hebben wij voor het eerst aangetoond dat het Solexa/Illumina-systeem voor digitale genexpressie (DGE of Tag-Seq) een zeer bruikbare technologie is voor transcriptoomanalyse van de gastheerreactie op een infectie. Onze resultaten hebben aangetoond dat deze technologie waardevolle additionele informatie oplevert ten opzichte van microarrayanalyse, zoals inzicht in infectie-geïnduceerde expressie van alternatieve transcript-isovormen. Bovendien hebben wij laten zien dat een andere grootschalige sequentietechnologie, RNA-Seq, evenzeer geschikt is voor het kwantificeren van de transcriptoomreactie op infectie. Aangezien RNA-Seq een nog krachtiger techniek is dan Tag-Seq om transcriptionele landschappen in kaart te brengen, is het te verwachten dat een toekomstige diepte-studie van RNA-Seq-data verdere inzichten zal opleveren in de expressie van alternatieve isovormen tijdens infectie. Door combinatie van de transcriptoomdata van de sequentie- en microarrayanalyses hebben wij een waardevolle referentieset samengesteld van infectie-responsieve genen in zebravis-infectiemodellen. Deze infectie-responsieve genen omvatten niet alleen welbekende immuunreactiegenen van vertebraten, maar ook genen die niet eerder met infectieziekten in verband zijn gebracht en die daarom zeer interessant zijn voor functioneel onderzoek naar hun mogelijke rol bij gastheer-pathogeen-interacties. Daarnaast hebben wij een set van infectie-responsieve en leverkanker-gerelateerde miRNA's in kaart gebracht die geconserveerd zijn tussen de mens en de zebravis en hiermee ondersteunend bewijs geleverd dat gemeenschappelijke miRNA's een regulerende rol spelen bij de immuunreactie en bij processen die ten grondslag liggen aan de ontwikkeling van kanker. Onze gecombineerde mRNA- en miRNA-transcriptoomdata vormen een sterke basis voor de toepassing van de zebravis als infectie- en kankermodel in toekomstig onderzoek.

Összefoglalás

Az utóbbi években a transzkriptómikai vizsgálatokban alkalmazott technológiák fejlődésének köszönhetően a genom kutatás területén jelentős előrehaladás figyelhető meg. A disszertáció **első fejezete** ezen technológiákról, a génexpresszió szabályozásáról, és a szabályozás vizsgálatához használt modellekről ad átfogó képet. A transzkriptómikában alkalmazott technológiák fényt derítettek néhány prokarióta és eukarióta szervezet transzkriptómájának komplexitására, továbbá bizonyos fejlődési és patológiás folyamatok molekuláris hátterére. Számos high-throughput technikát fejlesztettek már ki a transzkritómikában, melyek alapvetően két csoportba sorolhatóak: a mikrochip alapú és a szekvenálás alapú eljárások, melyek mindegyike lehetőséget biztosít az átfogó vizsgálatokra. A mikrochip analízis széles körben használt technológia, mely több ezer transzkript egyidejű vizsgálatára képes. Bár a mikrochipek sok biológiai probléma megoldásában hasznosnak bizonyultak, a hibridizáció és a chipre felvitt, előre meghatározott szekvenciájú próbák hátrányt jelentenek a transzkriptómikai vizsgálatok során. Az új-generációs szekvenálási technikák kifejlesztésével ezen hátrányokat legyőzve több millió transzkript szimultán detektálására van lehetőség. A két technika egymást kiegészítő eljárásként használható a genom vizsgálatára, melyek nem csak átfedő, de az egyes módszerek tulajdonságából következő specifikus eredményeket is hoznak. Transzkriptómikai elemzések elengedhetetlenül fontosak a génexpresszió szabályozásának vizsgálatánál. A génexpresszió több szinten szabályzott, egészen a transzkripciótól a fehérjék poszt-transzlációs módosításáig. Az utóbbi években fény derült arra, hogy ebben a szabályozásban nagy szerepet játszanak a kisméretű nem-kódoló RNS-ek, mint például a mikroRNS-ek (miRNS). A miRNS-ek negatívan szabályozzák a mRNS transzlációját vagy stabilitását. Számos biológia folyamatban bizonyították már szerepüket, így például a fejlődés valamint az immunrendszer működése során. A miRNS-ek megváltozott expresszióját mutatták ki különböző betegségek, mint például a fertőző betegségek és a rák esetében. Annak érdekében, hogy jobban megismerjük ezen megbetegedések patogenezisét, fontos tudnunk, mely miRNS-ek expressziója változott meg az egyes esetekben. Az egyes betegségek következtében megváltozott mRNS és miRNS expressziós vizsgálatokhoz *in vitro,* és *in vivo* modelleket is használnak. A zebrahal egy gyakran alkalmazott gerinces modellszervezet, melyet fejlődésbiológiai, embriogenenezist érintő, valamint újabban immunológiai kísérletekben használnak köszönhetően annak, hogy immunrendszere jelentős hasonlóságot mutat a humánéval. Az utóbbi években a zebrahal, az emlős állatmodellek mellett, egy kiváló modellszervezetnek bizonyult számos fertőző betegség és a tumoros megbetegedés vizsgálatában.

Kihasználva a transzkriptómikában megjelenő újfajta technológiai lehetőségeket, valamint alkalmazva a zebrahalat, mint modellszervezetet ez a doktori disszertáció a fertőző betegségek és tumoros megbetegedések során bekövetkező transzkriptómikai változások vizsgálatát és annak eredményeit foglalja össze, valamint betekintést nyújt a miRNS-ek szabályzó funkciójába.

A **második fejezetben** elsőként mutatjuk be a fertőző megbetegedések során bekövetkező transzkriptómikai változásokat, melyet a Solexa/Illumina rendszer úgynevezett digitális génexpresszió (DGE vagy Tag-Seq) technológiájával vizsgáltunk, mely egy szekvencia mintavételezésen alapuló, ún. tag-alapú transzkriptóm szekvenálási technika. Ebben a tanulmányban a humán tuberkulózist modellező *Mycobacterium marinum* fertőzés során fellépő változásokat vizsgáltuk zebrahalban. Kontrol és fertőzött halak esetében is két-két könyvtárat hoztunk létre, melyek bizonyították a technikai reprodukálhatóságot. Az eljárással könyvtáranként 5-8 millió tag-et detektáltunk, melyek lefedték a transzkript adatbázisokban található összes gén több mint 70%-át. A szignifikánsan megváltozott tag szekvenciák az adatbázis transzkriptjeinek 2%-ára térképeződtek. A Tag-Seq analízis nem csak kvantitatív információt nyújt a génexpresszióról, hanem a transzkriptek irányultságáról is tájékoztat. Vizsgálatunk során a tag-ek közel 40%-a térképeződött az antiszensz szálra, mely megegyezik más tanulmányokkal. Azonban, mivel a szignifikánsan megváltozott tag-ek nagy része a szensz szálra térképeződött, és bár az antiszensz jelenségre egyre több bizonyíték létezik, de biológiai jelentősége még nehezen értelmezhető, a továbbiakban csak a szensz transzkripteket tanulmányoztuk.

Annak érdekében, hogy bizonyítsuk a Tag-Seq technológiával nyert eredményeink helyességét, egy korábban, különböző típusú mikrochip kísérletek során felállított, mikobaktérium által regulált gének referencialistájával hasonlítottuk össze Tag-Seq eredményünket. Az összevetés erős korrelációt mutatott alátámasztva a kétféle metódus összehasonlíthatóságát, másrészről azonban felfedte az egyes eljárások hátrányait is. Ezért a két technika egymást kiegészítő eljárásnak tekinthető. A mikrochip mellett qPCR segítségével is ellenőriztünk egyes megváltozott tag szekvenciák expresszióját. Ezek a Tag-Seq eredményeket igazolták olyan esetekben is, ahol ellentétes expressziót mutatott a Tag-Seq és mikrochip technika. Ezenkívül bemutattuk, hogy a Tag-Seq technológia milyen más értékes információt nyújthat a kutató számára. Használható együttesen szabályzott transzkriptek vizsgálatára, prediktált génmodellek megerősítésére, fertőzés által indukált izoforma-változás vizsgálatához, újabb, mikobaktérium által regulált transzkriptek detektálására és térképezésére. Összefoglalva tehát bemutattuk, hogy a Solexa/Illumina digitális génexpressziós rendszere egy igen alkalmas technológia a transzkriptómikai analízisek során, kvantifikációs és komplexitási vizsgálatokhoz egyaránt.

A **harmadik fejezetben** tag-alapú szekvenálást (Tag-Seq) és teljes transzkript szekvenálást (RNA-Seq) alkalmazva végeztünk transzkriptómikai analízist olyan 1 napos zebrahal embriók esetében, amelyeket *Salmonella enterica* serovar Typhimuriummal (*S. typhimurium*) fertőztünk 8 órával a vizsgálat előtt. Egy korábbi mikrochip kísérletben a *S. typhimurium*-mal fertőzött embriók transzkriptómikai vizsgálatakor számos gyulladást okozó gén indukcióját mutattuk ki. A Tag-Seq és a mikrochip kísérletek eredményeit összehasonlítva 165 olyan transzkriptet találtunk, mely mindkét esetben változást mutatott. A Tag-Seq és a RNA-Seq összehasonlításával 241 mindkét esetben egyaránt megváltozott expressziójú transzkriptet azonosí-

tottunk, így elsőként demonstráltuk a két módszer hasonló teljesítményét fertőzés során fellépő transzkriptómikai változások vizsgálatakor. Egyesítve a szekvenálás alapú és a mikrochip alapú transzkriptóm vizsgálatok eredményeit összeállítottunk egy annotált referencia génlistát, mely zebrahal embriókban fertőzés következtében megváltozik. Ide tartoznak a transzkripciós faktorokat, szignál transzdukciós fehérjéket, citokineket és kemokineket, komplement faktorokat, apoptózisban és proteolízisben részt vevő fehérjéket, anti-mikrobiális aktivitással bíró fehérjéket kódoló gének, valamint számos olyan ismert és nem ismert gén, amelyet eddig nem hoztak összefüggésbe az immunválasszal. Ezenkívül összehasonlítottuk a *S. typhimurium*-mal fertőzött zebrahal embriókkal végzett kísérlet Tag-Seq adatait a második fejezetben leírt *M. marinum*-mal fertőzött kifejlett zebrahalakkal végzett kísérlet Tag-Seq adataival, mivel mindkét esetben erős gyulladásos immunválaszt figyeltünk meg. Összesen 206 fertőzés következtében megnövekedett expressziójú gént azonosítottunk, melyek a természetes immunválaszban részvevő transzkripciós faktorok és jelátviteli komponensek, valamint olyan gének, melyeket eddig nem kötöttek az immunválasz folyamatához. Ezzel a transzkriptómikai analízissel egy értékes referencialistát állítottunk össze, mely a további, zebrahal és patogén kapcsolatát tanulmányozó vizsgálatban hasznos lehet.

A **negyedik fejezetben** célul tűztük ki, hogy megismerjük a miRNS-ek fertőzéses megbetegedések esetén betöltött szerepét a gerincesek immunrendszerében. *M. marinum* és *S. typhimurium* bakteriális patogénekkel fertőzött kifejlett zebrahalakat, valamint embriókat alkalmaztunk, hogy modellezzük az egyes humán betegségeket. Azzal, hogy mind embrionális, mind pedig kifejlett egyedeket használtunk, meghatározhattuk, hogy az egyes miRNS-ek a természetes vagy a szerzett immunválaszban játszhatnak szerepet, hiszen az embriókban csak a természetes immunitás van kifejlődve. Egyedi tervezésű mikrochipet alkalmazva számos közös, fertőzésre reagáló miRNS-t sikerült azonosítanunk zebrahalban, beleértve a miR-21, miR-146 és a miR-181 családot. Ezek konzerváltságát, immunrendszerben való közreműködését, illetve tumorok kifejlődésében való szerepét humánban és más emlősben több egyéb kísérlet is jól bizonyítja. Azonban ez a tanulmány hozza kapcsolatba elsőként ezen miRNS-eket mikobaktérium és szalmonella fertőzésekkel. Továbbá eredményeink azt a korábbi feltevést is igazolják, miszerint azonos miRNS-ek vehetnek részt az immunrendszer és a tumoros folyamatok szabályozásában.

A miR-146 család természetes immunitásban betöltött szerepét már korábban bizonyították emberben és állat modellekben. Eredményeink megegyeznek ezekkel az adatokkal, hiszen a miR-146 a/b expresszió növekedést mutatott fertőzés következtében a kifejlett halakban és az embriókban egyaránt, melyeknél csak a természetes immunrendszer van kifejlődve. Feltételezések szerint a miR-146 a/b természetes immunválaszban betöltött szerepe a TLR és az IL1R szignalizáció szabályozásában lehet, amely a patogének felismerésében és a természetes immunválasz aktiválásában jelentős. Annak érdekében, hogy megtudjuk, hogy a miR-146 funkciója konzervált-e zebrahalban, azonosítottuk a család prediktált célgénjeit. Ezek között találtuk a TLR

és IL1R szignalizációs útvonalakban részt vevő *irak1* és *traf6* gént, mely emlősökben is a miR-146 a/b célgénje. Ezenkívül találtunk vélt miR-146 a/b kötődési pontokat a zebrahal MyD88 génjén, mely a két szignalizációs útvonal közös adaptere. A miR-146 bakteriális fertőzés hatására bekövetkező indukciója összhangban áll a második és harmadik fejezetben leírt mRNS expressziós adatokkal, ezzel demonstrálva néhány TLR szignalizációs gén, effektor gén és azok szabályzó génjeinek indukálódását fertőzés hatására zebrahalban. Tehát eredményeink alátámasztják azt a feltevést, hogy a miRNS-ek fontos feed-back mechanizmusként szerepet játszanak az immunválasz pontos szabályozásában. Ezenkívül bioinformatikai szoftverek segítségével azonosítottunk további prediktált miR-146 célgént. A legtöbb gén, mely konzervált a humán és a zebrahal között, közreműködik az immunrendszer egyes folyamataiban, így például az apoptózisban, az NF-KB transzkripciós faktor aktivitását szabályozásában, hemosztázisban, fertőzés során, T-sejtek fejlődésében és funkciójában, TLR szignalizációban, valamint tumor fejlődésében. Összességében tehát ez a tanulmány megerősítést ad arról, hogy a zebrahal egy megfelelő modellszervezet a gerincesek immunrendszerében részt vevő miRNS-ek funkciójának vizsgálatára, hiszen számos bakteriális fertőzésben résztvevő konzervált miRNS-t, valamint számos konzervált miRNS célgént azonosítottunk.

Az **ötödik fejezetben** célunk az volt, hogy megtudjuk, hogy a miRNS-ek milyen szerepet töltenek be a rák mechanizmusában, ezért azonosítottuk azokat a miRNS-eket, amelyek szerepet játszhatnak a májtumor kialakulásában zebrahalak esetében. A zebrahal ígéretes modellnek bizonyult a májtumor modellezésében, hiszen molekulárisan nagyon hasonlít a humán hepatocelluláris karcinómára (HCC). Annak érdekében, hogy kiderítsük, hogy ez a hasonlóság, konzerváltság megtalálható-e a miRNS expresszió szintjén is, miRNS transzkriptómikai analízist végeztünk zebrahalból nyert, karcinogénnel indukált májtumor mintákkal. Összehasonlítottuk a zebrahalban általunk azonosított és a humán HCC-vel korábban kapcsolatba hozott szignifikánsan megváltozott expressziójú miRNS-eket, és összeállítottunk azoknak a listáját, melyek mindkét szervezetben deregulációt mutattak. Ide sorolható a megnövekedett expressziójú miR-21, miR-23a, miR-146a/b, miR-221 és miR-222, valamint a csökkent expressziójú miR-1 és miR-122. Ezeken felül kimutattunk olyan miRNS-eket is, amelyeket eddig még csak másfajta humán tumoros megbetegedéssel kapcsolatban publikáltak, és amelyek részt vesznek a tumor képződésében, tumor szuppresszióban, apoptózisban, sejtosztódásban, differenciációban, invázióban és metasztázisban. Az érett miRNS-ek mellett bizonyos esetekben megváltozott expressziót figyeltünk meg a miRNS hajtű-struktúra másik karjánál is, illetve olyan szekvenciák esetében is, amelyek újabb miRNS-eket vagy másfajta nem-kódoló RNS struktúrákat reprezentálhatnak. Eredményeink tehát igazolják, hogy a zebrahal májtumor megfelelően modellezi a humán májtumort, valamint hogy az Agilent mikrochip technológiával lehetőség van újabb, vélt miRNS-eket expressziójának kimutatására.

Ezt követően összehasonlítottuk a májtumorral összefüggésbe hozott miRNS-ek listáját azon miRNS-ekkel, melyek mikobaktérium fertőzés hatására indukálódtak (4.

fejezet). A közösen megjelenő miRNS-ek, a miR-15, miR-16, miR-21, miR-34, és miR-146 család, megerősítik azt a nézetet, miszerint sok miRNS, amely szerepet játszik az immunválaszban, annak szerepe van a tumoros folyamatokban is.

Végül, vizsgáltuk a tumorban résztvevő miRNS-ek célgénjeinek lehetséges konzerváltságát. Elvégeztük a humán májtumorban közreműködő miRNS-ek kísérletesen is bizonyított célgénjeinek funkcionális annotációját, amely azt mutatta, hogy ezek a gének részt vesznek különböző tumorral kapcsolatos folyamatokban, szignalizációs útvonalakban. A homológ zebrahal géneknél végzett miRNS kötő hely prediktáció azt mutatta, hogy néhány miR-146 család célgén (*IRAK1, TRAF6, IRF5*), miR-221/222 család célgén (*CDKN1B, CDKN1C, KIT*) és miR-1 család célgén (*GJA1, PDCD4, TPM4, LASP1, TMSB4X, RABL2A, RABL2B*) konzervált lehet humán és zebrahal között. Ezek a gének mind az immunrendszerben, mind pedig a tumoros folyamatokban szerepet játszanak, így például a sejtciklus szabályozásban, sejtosztódásban, apoptózisban, sejtadhezióban, sejtmozgásban, citoszkeletális organizációban, valamint immunválasszal és tumorral kapcsolatos szignalizációs útvonalakban, a MAPK, NF-κB, p53, ErbB szignalizációban, és a kis GTPáz-hoz kapcsolt szignál transzdukciós útvonalakban. Ezek az eredmények alátámasztják a zebrahal és humán májtumor hasonlóságát, valamint azt a hipotézist, miszerint a miRNS-ek fontos szerepet játszhatnak mindkét organizmusban a tumorral kapcsolatos mechanizmusok szabályzásában. MiRNS tanulmányunk tehát a zebrahal és humán májtumor közti hasonlóság újabb szintjét mutatja be a génexpresszió és a transzkripciós faktorok általi transzkripciós szabályzáson túl.

Összefoglalva, a tézisben leírt tanulmányaink alátámasztják azt, hogy a zebrahal alkalmas a fertőző és a tumoros megbetegedések modellezésére, valamint alkalmasak a miRNS-eknek a gerincesek immunrendszerében és egyes megbetegedésekben betöltött szabályzó funkciójának vizsgálatára. Bemutattuk, hogy az újgenerációs szekvenálási eljárás, a Solexa/Illumina digitális génexpressziós rendszere (DGE vagy Tag-Seq) jól alkalmazható a gerincesek fertőzésre adott immunválaszának transzkriptómikai vizsgálatára, és mint egy kiegészítő technológiaként használható a mikrochip alapú eljárások mellett. Továbbá bemutattuk, hogy egy másik újgenerációs szekvenálási technika, az RNA-Seq ugyancsak megfelelő a fertőzésekre adott transzkriptómikai változások elemzésére. A szekvenálás alapú és a mikrochip alapú transzkriptómikai adatok egyesítésével egy értékes referencialistát állítottunk össze azokból a génekből, melyek expressziója a fertőzés következtében változott meg zebrahalban. Ez a génlista nemcsak jól ismert, a gerinces immunválaszban részvevő géneket tartalmaz, hanem olyanokat is, amelyeket eddig még nem figyeltek meg fertőző betegségeknél, viszont értékes információkat rejthetnek a gazda-patogén kölcsönhatásban betöltött szerepüket illetően. Ezenkívül összeállítottuk azon miRNS listáját, amelyek fertőzésekben és a májtumor folyamatában egyaránt közreműködnek, és konzerváltságot mutatnak humán és zebrahal közt, mindezzel megerősítve azt, hogy közös miRNS-ek vehetnek részt az immunválasz és a tumor kialakulásának szabályzásában. Tanulmányainkban végzett mRNS és miRNS transzkriptómikai adatok ala-

pul szolgálhatnak a jövőben olyan vizsgálatokhoz, melyek a zebrahalat a fertőző és rákos megbetegedések modellezésére kívánják alkalmazni.

.

Curriculum vitae

Anita Ordas was born on 21 October 1980 in Szeged, Hungary where she attended secondary education at the Miklos Radnoti Gymnasium. She studied biology at University of Szeged, Hungary from 1999 and received her MSc in 2004. She also received a diploma as an English Translator of Biology at the same university (2000-2004). From 2002 she worked in the lab of Dr. Edit Hajdu at the Institute of Clinical Microbiology of University of Szeged, Hungary where her project was the analysis of antibiotic resistance of *Streptococcus pneumoniae*. During her studies she received three times the Scholarship of the City of Szeged for her prominent scholastic record and scientific work (2001, 2002, 2003). From 2005 she worked as a research assistant in the lab of Prof. Dr. Katalin Csiszar in the Cardiovascular Research Center of University of Hawaii, USA and studied molecular aspects of a novel human fibrotic disorder. From 2007 she joined the group of Dr. Matyas Mink in the Department of Genetics of University of Szeged, Hungary and started working on the EU-funded project 'High-throughput Tools for Biomedical Screens in Zebrafish' (ZF-TOOLS) in collaboration with the Institute of Biology, University of Leiden. In the context of this project she started her PhD work in the lab of Prof. Dr. Herman Spaink under supervision of Dr. Annemarie H. Meijer in the Institute of Biology at University of Leiden. The focus of her research and the main work for her thesis has been the analysis of gene and miRNA expression in infectious diseases and cancer in zebrafish.

List of publications

Ordas A, Szauter KM, Laxer RM, Pope E, Wherrett D, Alman B, Mink M, Csiszar K and Hinek A. A novel fibrotic disorder associated with increased dermal fibroblast proliferation and downregulation of genes of the microfibrillar network (submitted to *Br J Dermatol*, accepted with minor revisions)

Hegedus Z, Zakrzewska A, Agoston VC, **Ordas A**, Racz P, Mink M, Spaink HP, Meijer AH. Deep sequencing of the zebrafish transcriptome response to mycobacterium infection. *Mol Immunol*. 2009 Sep;46(15):2918-30.

Racz P, Mink M, **Ordas A**, Cao T, Szalma S, Szauter KM, Csiszar K. The human orthologue of murine Mpzl3 with predicted adhesive and immune functions is a potential candidate gene for immune-related hereditary hair loss. *Exp Dermatol*. 2009 Mar;18(3):261-3.

Hajdu E, Matuz M, Benko R, **Ordas A**, Nagy E. An 8-Year Evaluation of Antibiotic Consumption and Antibiotic Resistance Among *Streptococcus pneumoniae* from In- and Out-Patients in Szeged, Hungary. *J Chemother*. 2007 Oct;19(5):519-27.

Hajdú E Matuz M, Benkő R, **Ordas A**, Blazsovszky M, Jeszenszky M, Szurdi M, Nagy E. *S. pneumoniae* törzsek antibiotikum érzékenységének retrospektív értékelése 1998-2005 között az antibiotikum fogyasztás tükrében. *Infektológia és Klinikai Mikrobiológia* 2006; XIII. 127-136

In preparation:
Ordas A, Zakrzewska A, Stockhammer OW, Carvalho R, van der Sar AM, Zhang Y, Verbeek FJ, Mink M, Spaink HP, Meijer AH. MicroRNA expression during bacterial infections in zebrafish

Ordas A, He S, Gong Z, Zhang Y, Verbeek FJ, Mink M, Spaink HP, Snaar-Jagalska B.E., Meijer AH. Liver tumor-related microRNA expression is conserved between zebrafish and human

Ordas A, Hegedus Z, Henkel C, Stockhammer OW, Butler D, Jansen HJ, Racz P, Mink M, Spaink HP, Meijer AH. Deep sequencing of the innate immune transcriptomic response of zebrafish embryos to Salmonella infection

Racz P, Zakrzewska A, Csukonyi E, **Ordas A**, Spaink HP, Mink M, Meijer AH. Functional analysis of the *mpzl3* gene in zebrafish