

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

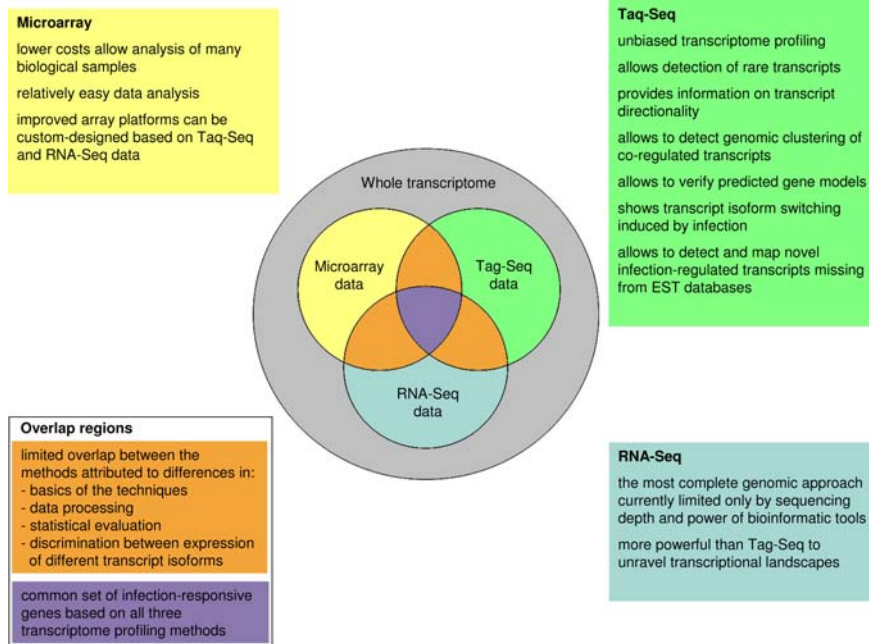


Figure 1. Complementary use of different transcriptome profiling techniques. Microarray, Tag-Seq and RNA-Seq methods as used in this thesis to study the response to bacterial infections in zebrafish can be considered as complementary techniques to analyze the transcriptome. The specific advantages of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

