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).

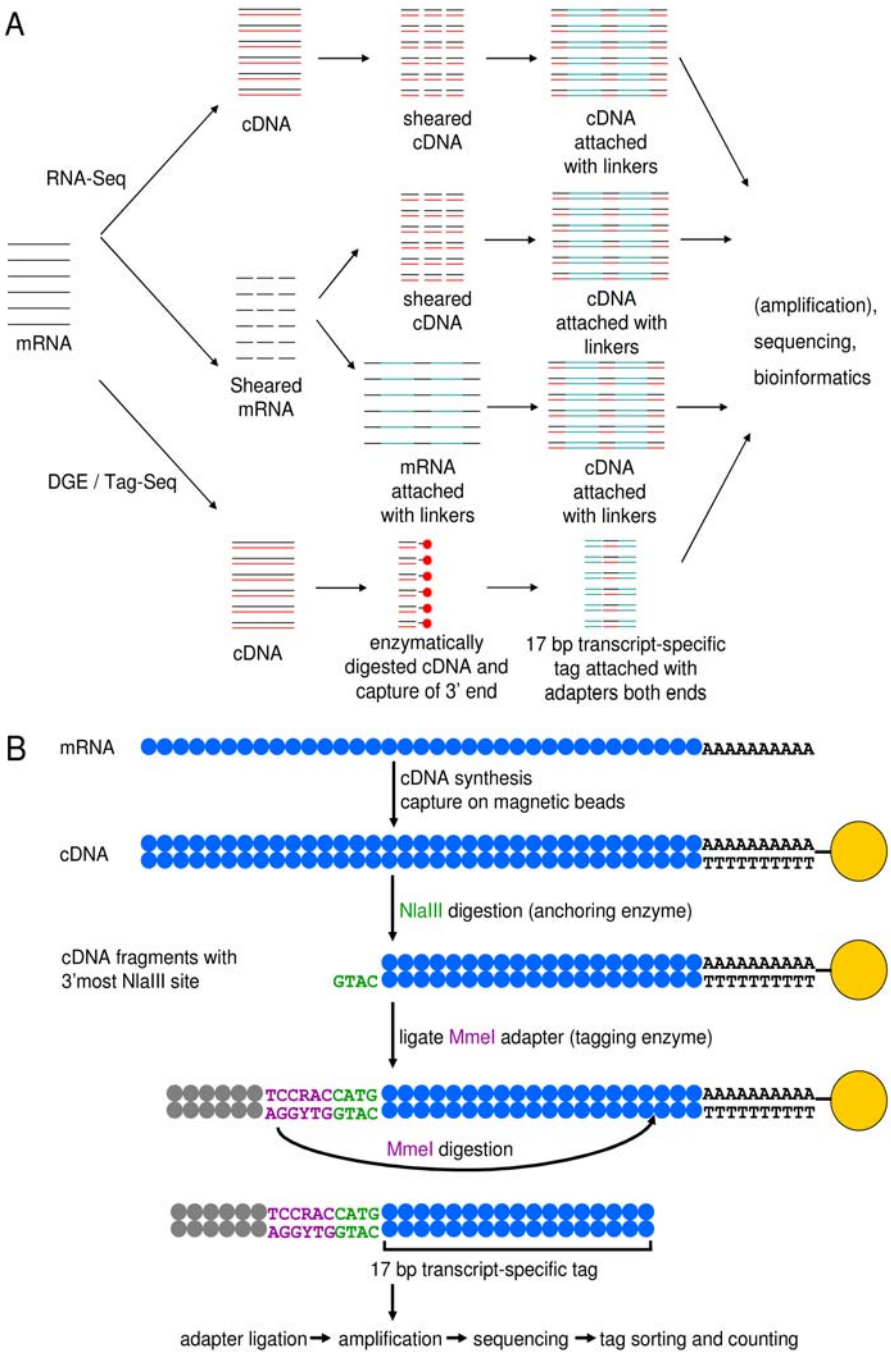
Another approach to quantify and annotate transcriptomes is the sequencing-based approach which overcomes some limitations of array-based methods (Irizarry et al., 2005; Pedotti et al., 2008; t Hoen et al., 2008; Wilhelm and Landry, 2009). Sequencing-based transcriptome profiling methods directly determine 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
Sheared mRNA
DGE / Tag-Seq
cDNA
sheared cDNA
cDNA attached with linkers
sheared cDNA
cDNA attached with linkers
mRNA attached with linkers
cDNA attached with linkers
cDNA
enzymatically digested cDNA and capture of 3' end
17 bp transcript-specific tag attached with adapters both ends
(amplification), sequencing, bioinformatics
B
mRNA
AAAAAAAAAA
cDNA synthesis
capture on magnetic beads
cDNA
AAAAAAAAAA
TTTTTTTTTT
NlaIII digestion (anchoring enzyme)
cDNA fragments with 3'most NlaIII site
GTAC
ligate MmeI adapter (tagging enzyme)
TCCRACCATG
AGGYTGGTAC
MmeI digestion
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

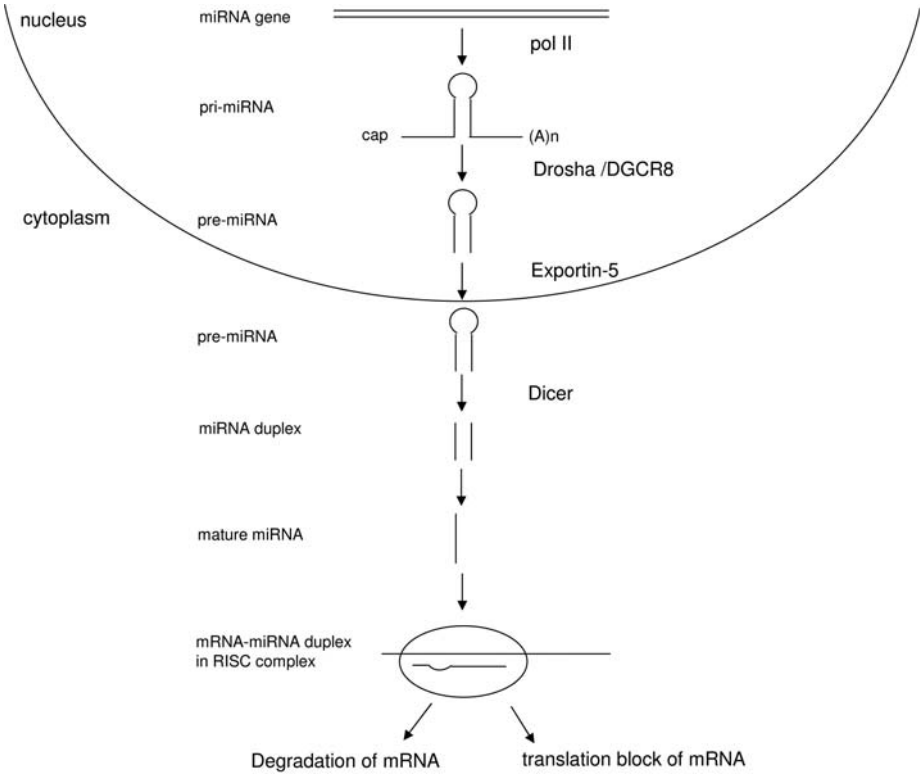


Figure 2. MicroRNA biogenesis. In the nucleus miRNA genes are first transcribed as primary miRNA (pri-miRNA) transcripts with 5′ cap and 3′ poly(A) tail by RNA polymerase II (pol II) which are recognized and cleaved by Drosha/DGRC8. The generated hairpin precursor miRNA (pre-miRNA) transcript is exported 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-

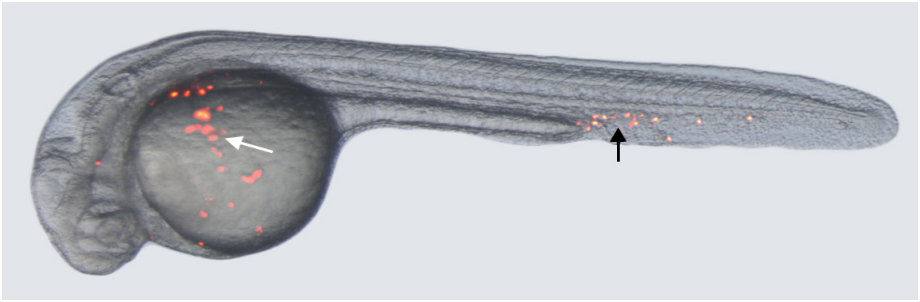


Figure 3. Real-time visualization of bacterial infection in the zebrafish embryo. The embryo was injected at 27 hours post fertilization with approximately 250 colony forming units of DsRed-labelled Salmonella typhimurium bacteria into the caudal vein and imaged at 2 hours post infection. An overlay 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.