Chapter 1
Functional Genomics Research in Aquaculture: Principles and General Approaches

Shikai Liu, Yu Zhang, Fanyue Sun, Yanliang Jiang, Ruijia Wang, Chao Li, Jiaren Zhang, and Zhanjiang (John) Liu

Abstract: Functional analysis has always been more difficult, but it is nothing compared with structural analysis. This is especially true when the number of genes under study is increased to cover various systems and pathways on a genome scale. In this chapter, we provide an overview of functional genomics focusing on general approaches for functional genomics to include (1) Functional inference based on expression profiling such as analysis of expressed sequence tag (EST) analysis, microarray analysis, and RNA-Seq; (2) Functional inference of gene functions based on positional analysis such as genome-wide association studies (GWAS), quantitative trait loci (QTL) mapping, and expression quantitative trait loci (eQTL) mapping; (3) Functional inference of gene functions by comparative genome analysis; (4) Gene pathway analysis; and (5) Experimental determination of gene functions using novel technologies, such as the zinc finger nuclease (ZFN) technology. We also provide a section on epigenetics and analysis of protein–DNA interactions. At the end of the chapter, we offer our assessment of the potential of various technologies for functional genomics in aquaculture.

Introduction

Genomics as a branch of science started to make headway during the early to mid-1980s.

At the beginning, it started with a major research project "The Human Genome Project." As the genomic information was accumulated, and data was analyzed, a series of specific genomic methodologies were developed. With the rapid advances in technology, particularly the advances in PCR and sequencing technologies, a series of highly efficient approaches for genomic studies were developed. As a result of scientific demand and technological advances, a very specific branch of science evolved that is now called Genomics.

To gain better understanding of genomics, we must examine its roots. The term "genome" itself is more than 75 years old and refers to the entire genetic material of an organism, or its complete set of genes located on chromosomes (Hieter and Boguski, 1997). In 1986, "genomics" was coined by Thomas Roderick to describe the scientific discipline of mapping, sequencing, and analyzing genomes (Mckusick, 1989). The term of genomics has become universally accepted over the past two

Functional Genomics in Aquaculture, First Edition. Edited by Marco Saroglia and Zhanjiang (John) Liu.

decades. However, genomics is now undergoing a transition or expansion from the mapping and sequencing of genomes to an emphasis on genome functions. To reflect this shift, genome analysis may be generally divided into "structural genomics" and "functional genomics." Structural genomics represents an initial phase of genome analysis, studies the structure, organization, and evolution of genomes, while functional genomics, which studies expression and functions of the genomes. Structural genomics has a clear end point—the construction of high-resolution genetic, physical, and sequence maps of an organism. The ultimate map of an organism is its complete DNA sequence with a resolution of every single base pair (Hieter and Boguski, 1997). Although, genomics in its major research objectives can be divided into structural genomics and functional genomics, there is no clear separation of these subdisciplines. Furthermore, structural genomics is the basis for functional genomics.

Functional genomics represents a new phase of genome analysis (Hieter and Boguski, 1997). It requires the development of innovative technologies that make use of the vast resource of structural genomics information. Specifically, functional genomics refers to the development and application of genome-wide experimental approaches to assess gene functions by making use of the information and reagents provided by structural genomics. It is characterized by high throughput or large-scale experimental methodologies combined with statistical and computational analysis of the results. The fundamental strategy in a functional genomics approach is to expand the scope of biological investigation from studying single gene or protein to studying all genes or proteins at once on a genome-wide scale. Such operations allow the generation of tremendously large data sets that demand additional capacities of data analysis to draw information relevant to biology. Assistance is needed from all areas of biology, and more so from disciplines outside biology that can handle large data sets. Computer sciences and mathematics are among the first disciplines genomics has demanded cooperation from. Computational biology will play a critical and expanding role in this area: structural genomics has been characterized by data generation and management, whereas functional genomics will be characterized by mining the data sets for valuable biological information. Functional genomics promises to rapidly narrow the gap between sequence and function and to yield new insights into the behavior of biological systems (Hieter and Boguski, 1997).

The goal of this chapter is to provide some basic concepts of functional genomics, and provide a general description of approaches for functional genomics research in aquaculture.

The Concept of Functional Genomics

Genes and Gene Functions

Most eukaryotic organisms harbor tens of thousands of genes, and in most cases, perhaps more than 20,000 but fewer than 40,000 genes. Each gene has its own functions. Historically, gene functions were determined by observations of a phenotypic mutation followed by genetic mapping of the mutated phenotype and eventually trace to the gene controlling the trait. Such an approach is usually considered to be a forward genetics approach. This approach is highly straightforward; however, mutations, whether as a result of spontaneous mutation, or induced mutation, are rare and mutated phenotypes are oftentimes difficult to be observed in the first place. With the rapid progress of DNA sequencing technologies, scientists started to know much quicker about a specific gene sequence and its protein structures than the functions of the gene. As a result, a new set of reverse genetics approaches was developed. In reverse genetics, the simple concept is to inactivate the gene, and then determine the changes of the phenotypes; alternatively, to add more of the gene products, and then determine the

changes of the phenotypes. In the former, genes can be specifically targeted to "knockout" the gene. This worked really well with some model species where embryonic stem cell technology is available. For instance, with rat or mouse, a specific gene can be knocked out in an embryonic cell line, and then individuals can be developed from the cells, and phenotypes can be observed in the "knocked out" animals. However, this approach is to date not applicable to aquaculture species because embryonic stem cell technologies have not been developed for aquaculture species.

In recent years, the discovery of RNA interference (RNAi) has lent research tools for the study of gene functions. RNAi is an RNA-dependent gene silencing process that has been applied to knockdown gene expression (Hannon, 2002). This is particularly useful for some aquaculture species, as RNAi technologies have been applied to investigate gene function as well as to develop antiviral agents to combat various infections in some fish and crustaceans species (Acosta et al., 2005; Copf et al., 2006; Liu et al., 2006; Wargelius et al., 1999; Kelly and Hurlstone, 2011). For instance, the highly efficient gene knockdown was observed to result in similar embryonic defects to the known mutant phenotypes in zebrafish (Wargelius et al., 1999). In a study of genes expressed differentially in freshwater crayfish infected with the white spot syndrome virus (WSSV), the anti-lipopolysaccharide factor gene (AF) was shown to protect against WSSV infection, and knockdown of AF by RNAi specifically resulted in higher rates of viral propagation (Liu et al., 2006).

In addition to gene knockout and knockdown technologies, transgenic technology has been widely used to demonstrate the functions of genes. The basic principle is that if a gene has certain functions, its over-expression should cause changes in phenotypes. Such an approach was best demonstrated by transgenic fish harboring the growth hormone gene that grow much faster and bigger than their non-transgenic controls (Du et al., 1992; Gross et al., 1992; Devlin et al., 1994; Rahman et al., 1998; Rahman et al., 2001).

All the traditional approaches for the study of gene functions are effective, but they have fatal problems that include the following:

1. Not all genes can cause a visible phenotype.
2. Knockout of one gene may cause numerous other changes in genome expression; many of these are compensatory or consequential, making analysis of gene functions very difficult.
3. Functional study of gene function by "one gene at a time" is too laborious, too expensive, and too slow.

As a result, scientists have explored to study gene functions on the genomic scale that leads to the emergence of functional genomics.

Concept of Functional Genomics

Functional genomics can be defined as a discipline for the understanding of gene functions and regulation on a genome-wide scale. It is a field of molecular biology that attempts to make use of the vast wealth of data produced by genomic projects (such as genome sequencing projects) to describe gene (and protein) functions and their interactions. Unlike structural genomics and proteomics, functional genomics focuses on the dynamic aspects such as gene transcription, translation, and protein–protein interactions and interactions between proteins, DNA, and RNA. As opposed to the static aspects of the genomic information such as nucleotide and amino acid sequences or structures, functional genomics attempts to answer questions about the functions of DNA at the levels of genes, RNA transcripts, and protein products (Hackett and Clark, 2007). The ultimate goal for functional genomics is to bridge the gap between the blueprint (genome sequence or genotype) and the living organism (trait or phenotype) under various environmental conditions (Cogburn et al., 2007). Spawned by the technological revolution in genome

sequencing (Venter et al., 1996; Rowen et al., 1997), functional genomics studies currently are greatly stimulated by the high-throughput sequencing and screening technologies. Therefore, a key characteristic of functional genomics studies is the genome-wide approach generally involving high-throughput technologies rather than the traditional "gene-by-gene" approach.

Goals of Functional Genomics

The goals of functional genomics are to gain better understanding of the roles of functional elements in the genome that directly or indirectly affect the development, growth, metabolism, immunity, behavior, reproduction, and various other processes of an organism. One of the primary tasks of functional genomics is assigning specific functions to genes, noncoding RNAs, and *cis*-Acting DNA elements involved in the processes explained earlier (Hackett and Clark, 2007). The goals of functional genomics vary depending on organism and project. In the broad sense, functional genomics studies are conducted with the following objectives:

1. To discover functional elements from genomic sequence of organisms, including protein coding genes and regulatory noncoding regions;
2. To obtain the global assessment of how the expression of all genes in the genome varies under changing conditions;
3. To generate resources and develop methodologies for genome-wide mutagenesis deducing the functions of novel genes by mutating them and studying the mutant phenotype;
4. To genetically manipulate organisms for specific purposes;
5. To understand the evolution of genomes in relation to biology of organisms across a spectrum of evolutionarily-related species.

Approaches to Functional Genomics

Numerous approaches with sophisticated experimental techniques have been developed for functional genomics studies. The huge numbers of genes, and even larger numbers of transcripts contribute to the complexity of functional genomics analysis. Such complexities are amplified again by an even greater diversity of polypeptides and their post-translational modifications, the variation in expression from different tissues under different developmental stages, and under various environmental conditions (Hackett and Clark, 2007). High throughput methodologies are required to simultaneously examine the expression and functions of all genes of an organism. Clearly, high-speed computation is an essential feature of most functional genomics techniques. In the later sections, we describe some of the general approaches currently used in functional genomics studies, with the understanding that more powerful methods are constantly emerging, especially as computational power continues to improve. The improvements include tools and instruments for digital recording of information and newer algorithms for sorting and analyzing information developed as our understanding of gene function increases. Because of the large quantity of data generated by these techniques and the desire to discover biologically meaningful patterns, bioinformatics is crucial to functional genomics data analysis (Hackett and Clark, 2007).

While whole genome sequencing is now a routine, functional analysis continues to be a great challenge. There are several general strategies for elucidating the functions and regulation of gene expression. In terms of strategies, functional genomics approaches can be divided into four general categories: (1) Functional correlation of gene expression and thereby inference of gene functions; (2) Functional correlation of gene positions and thereby inference of candidate gene functions; (3) Functional assignment by association

with gene pathways; and (4) Direct testing of gene functions.

Functional Correlation of Gene Expression Profiling

Short of direct testing that is often very difficult, gene functions can be inferred from their correlations of expression with function, or their correlations of position with function. The basis for functional inference from expression correlation is expression profiling. The idea is that if a gene is involved in a specific trait, e.g., defense response, its expression would respond to infection. Gene expression profiling is the measurement of expression levels for thousands of genes to paint a general global picture of gene expression under a specific developmental stage, environmental condition, or treatment. In practice, gene expression profiling experiments often involve measuring the relative amount of mRNA expressed in two or more experimental conditions ("treatment"). The correlation of "treatment" and gene expression profile could provide inference on gene functions.

Various technologies have been developed to quantify gene expression, including hybridization-based and sequence-based approaches. Hybridization-based approaches typically involve incubating fluorescently-labeled complementary DNA (cDNA) with custom-made microarrays or commercial high-density oligo microarrays. In contrast to microarray methods, sequence-based approaches determine gene expression levels by directly sequencing cDNAs. The relative abundance of cDNAs reflect gene expression levels. Initially, Sanger sequencing of cDNA or expressed sequence tag (EST) libraries were used (Gerhard et al., 2004), and then tag-based methods were developed to improve the throughput, including serial analysis of gene expression (SAGE) (Velculescu et al., 1995; Harbers and Carninci, 2005) and massively parallel signature sequencing (MPSS) (Brenner et al., 2000). These tag-based sequencing approaches are high throughput and can provide precise, digital gene expression (DGE) levels. Recently, the development of high-throughput DNA sequencing technologies provides additional strengths for gene expression profiling. This method, termed as RNA-Seq, has clear advantages over previous sequence-based approaches and is expected to revolutionize the manner in which transcriptomes are analyzed for both gene discovery and global gene expression profiling.

In terms of functional genomics, the largest portion of published literature to date involves gene expression profiling through sequencing-based approaches such as EST analysis, SAGE, MPSS, and more recently developed RNA-Seq, and hybridization-based microarray analysis. We will limit our discussion to EST analysis, microarrays, and RNA-Seq as these are the most commonly used approaches.

Analysis of Expressed Sequence Tags (ESTs)

ESTs are single-pass sequences of random cDNA clones from cDNA libraries. They are traditionally generated using Sanger sequencing and therefore the resultant sequences are approximately 500 to 800 base pairs in length. Several years ago, because sequencing was relatively cheap, large numbers of ESTs can be generated at a reasonably low cost from either the 5′ or 3′ end of a cDNA clone to get an insight into transcriptionally active regions. ESTs were used as a primary resource for human gene discovery (Adams et al., 1991). Thereafter, there has been an exponential growth in the generation and accumulation of EST data in public databases for various organisms, with approximately 71 million ESTs now available in public databases (http://www.ncbi.nlm.nih.gov/dbEST/, September 2011, all species). Readers can refer to Chapter 2 of this volume for availability of ESTs among various aquaculture species.

EST analysis is an effective genomic approach for rapid identification of expressed genes, and has been widely used in genome-wide gene expression studies in various tissues, developmental stages or under different environmental conditions (Adams et al., 1995; Ronning et al., 2003). In addition, the availability of cDNA sequences has accelerated further molecular characterization of genes of interest and provided sequence information for microarray construction and genome annotation (Bailey et al., 1998; Lo et al., 2003; Kim et al., 2006).

Gene expression analysis plays an important role in identifying differentially expressed genes under different environmental conditions and gene expression regulation, shedding light on gene functions. EST analysis has been demonstrated effective for detection of differential expression and regulation of certain genes. Without normalization or subtraction in library construction, the number of the sequenced ESTs for a given gene reflected the abundance of the gene expression at the corresponding scenario (e.g., environmental conditions, developmental stages, treatments, etc.).

Direct EST sequencing is inefficient in discovery of rarely expressed genes. To solve this problem, the method to construct normalized cDNA libraries was developed (Soares et al., 1994; Bonaldo et al., 1996). The basic principle is using hybridization to reduce redundant genes and increase the representation of rarely expressed genes.

Initial annotation of ESTs can be conducted by simple sequence similarity comparisons. Further annotation analysis can be carried out after obtaining the consensus sequences (putative unigenes), such as determination of gene identity based on homology search, open reading frame (ORF) identification, Gene Ontology (GO) annotation and gene-enrichment analysis (e.g., Nakaya et al., 2007).

In order to assign gene identity to contigs and singletons, homology search is widely used. Such an approach is especially helpful for newly-studied species. BLAST is the most widely used program to obtain high throughput EST analysis and annotation results. BLAST package provides different flavors of algorithm for sequence similarity searching. BLASTX is used to search against protein database by translated consensus EST sequences while BLASTN is used to search against nucleotide sequence databases. NCBI, ENSEMBL and Swiss-Prot are three important databases for BLAST search. For instance, Swiss-Prot database have fully manually curated and annotated unigene database, Uniprot, which can be used for identifying putative function for unigene by BALSTX. NCBI provide dbEST database that can be used to search novel transcript by BLASTN. dbEST is a main ESTs resource database including ESTs for over 200 aquaculture species. ENSEMBL database can provide chromosome location information of genes, which is a useful tool for comparative genome analysis. However, BLAST sequence similarity comparison provides only sequence homology information, and one cannot purely rely on BLAST for gene identification. Detailed phylogenetic analysis and/or orthology analysis is needed to determine the identities of genes.

For a greater level of annotation, ORF is identified to determine the full or portion of coding region in the unigene. The unigene with a full ORF usually represent a full-length cDNA. There are some useful tools for ORF detection. For example, ESTScan (Iseli et al., 1999) can extract coding regions from low-quality ESTs and correct frame shift errors. OrfPredictor (Min et al., 2005) is another program for identification of protein-coding sequences from ESTs through predicting most probable coding regions from all the six translation frames.

GO annotation can provide description of gene products behaving in a cellular context. Gene functions are placed into three categories: biological processes, cellular components, and molecular functions. Consensus sequences can be linked to GO terms

and assigned a possible function by Blast2GO (Conesa et al., 2005).

GO enrichment analysis is to cluster most relevant GO terms associated with certain biological pathway. GOEAST (Zheng and Wang, 2008), Ontologizer (Bauer et al., 2008), GeneTrail (Backes et al., 2007), and DAVID functional annotation tool (Huang et al., 2009) are useful tools for these analysis.

EST analysis is an efficient approach for gene discovery and gene identification. For instance, during 2001 to 2007, catfish ESTs increased from 10,000 to 44,000 and the putative genes number increased from 5905 to 25,000 (Li et al., 2007). In Pacific oyster (*Crassostrea gigas*), 40,845 high-quality ESTs represented 29,745 unique transcribed sequences (Fleury et al., 2009). In gilthead sea bream (*Sparus auratus*), 30,000 ESTs represented 18,196 putative unigenes (Louro et al., 2010). Currently, there are over 180 aquaculture species having more than 100 ESTs in dbEST (see Chapter 2 of this volume on existing genome resources for details).

EST analysis can provide comparisons of gene expression profiling in different tissues and conditions. For instance, in a recent study with rainbow trout (*Oncorhynchus mykiss*), Kondo et al. (2011) sequenced over 30,000 ESTs from rainbow trout adipose tissue. These ESTs were used to search adipokine-related genes. The result showed that none of them encoded adipokine and PPAR-γ gene, which play important roles in mammalian adipocytes. Further qRT-PCR result confirmed EST analysis results, that is, rainbow trout adiponectin transcripts were weakly detected in adipose tissue but strongly detected in muscle, suggesting the difference of energy metabolism between fish and mammal (Kondo et al., 2011). Chini et al. (2008), constructed normalized cDNA libraries from liver, ovary and testis in blue fin tuna (*Thunnus thynnus*), identifying several sequences with known function in other organisms, but not previously described in this species. Also, sequences were described being expressed in one, two, or more tissue libraries. Similarly, Zou et al. (2011) constructed normalized cDNA libraries from testis, ovary, and mixed organs of mud crab (*Scylla paramamosain*). Through EST analysis, sex-specific transcripts were identified.

EST resources provide sequence information for microarray development. For instance, in a recent study, Booman et al. (2011) developed a large-scale oligonucleotide microarray platform containing 20,000 features (20K), which was used to study immune response of the Atlantic cod spleen with stimulation of formalin-killed, atypical *Aeromonas salmonicida* (Booman et al., 2011). Similarly, oligo microarray for gilthead sea bream (*Sparus aurata*) was developed based on ESTs, and the microarray was used to identify 1050 differentially expressed genes between two developmental stages (Ferraresso et al., 2008).

Although EST analysis has been important for transcriptome characterization, it is now becoming expensive, relative to several of the most recently developed approaches, as described in the following text. However, EST resources still have a great value to serve as reference for RNA-Seq analysis. We found that ESTs are essential for high-quality reference-guided assembly of next-generation sequencer-generated short reads (Liu et al., 2011).

Microarrays

Microarray is a powerful tool that allows analysis of global gene expression in individual cells or tissues under different conditions (Schena et al., 1995). The core principle of microarray is dense placement of gene target sequences in a small area and hybridization. Tens of thousands of DNA sequences termed probes are anchored or spotted onto the solid surface of a chip. Fluorescence-labeled probes are used to hybridize with the features on the microarray. Microarray combines simple nucleic acid hybridization with high-density spotting robots, fluorescence-based signal detection and high-resolution laser scanners (Peatman and Liu, 2007). High-density spotting

robots and photolithography allow each feature to be placed accurately on the slide in high densities. Fluorescence-labeled probe provides much clearer signal than the traditional radiation labeling. Moreover, the high-resolution laser scanner allows accurate fluorescence-signal quantification.

Based on the construction and sample labeling, there are two primary approaches to DNA microarray used in aquaculture species: the spotted arrays (or printed arrays) and the in situ arrays. Spotted arrays are constructed by spotting cDNA, small fragments of PCR products or long oligos using robot. This technique is adapted by most researchers to produce "in-house" printed microarray, because it is relatively low-cost and flexible. The researchers can decide the probes, generate their own probes, spot the array, hybridize the samples to the array, and scan the arrays with their own machine. However, it is labor-consuming. The number of spots (features) is limited to avoid cross-contamination. Two-color fluorescence, such as Cy3 and Cy5, are usually used for sample labeling for spotted array (Schena et al., 1996).

In situ arrays are constructed by synthesizing short oligos directly onto the slide surface by photolithography instead of depositing intact sequences. The oligo probes may be longer, like 65-mer (Mathavan et al., 2005), or shorter, like 24-mer (Peatman et al., 2007). Longer probes are more specific, whereas the shorter ones are cheaper and can be spotted in higher densities. In order to overcome the short probes to improve the specificity and sensitivity, in situ arrays contain high-density features, usually multiple probes per target (Miller and Tang, 2009). A perfect match (PM) and mismatch (MM) system is used to further improve the specificity (Irizarry et al., 2003; Han et al., 2004). MM probes contain one or more mismatched nucleotides within the PM probe sequences and act as a negative control to detect the false-positive signal resulted from the nonspecific cross hybridization.

Affymetrix (http://www.affymetrix.com/) is one of the most widely known industries for in situ array. Semiconductor-based photochemical synthesis and photolithographic masks are used to synthesize oligo probes for Affymetrix GeneChip. The photolithographic masks either block or allow light to reach the microarray surface. In the area the mask covers, the addition of the nucleotides will be prevented since the UV light has been blocked, whereas in the area exposed to the UV light, the specific nucleotide can be added. After many cycles of unmasked, addition of nucleotide and masked, the sequences of every oligo probes are fully constructed. Another commercial microarray manufacturer Roche NimbleGen (http://www.nimblegen.com/) has developed a maskless array synthesis, which uses digital mirrors instead of the photolithographic masks (Nuwaysir et al., 2002) at a significantly low start-up cost. Agilent Technologies construct the oligo probes for in situ array use glass slides and inkjet printing, neither photolithographic masks nor digital mirrors. Instead of Cy3/Cy5 labeling system, the sample using in situ array is usually Biotin–Streptavidin labeled. No matter which platform is used, spotted array or in situ array, the basic procedure for gene expression experiment is similar, starting with RNA. As shown in Figure 1.1, the RNA is extracted from the sample that we are interested and reverse transcribed to cDNA after quantification and quality check. The cDNA is fluorescently labeled and hybridized to the probes on the microarray. The hybridization will result in fluorescence signal, which can be measured by a fluorescence scanner and then be analyzed by using software, e.g., R/Bioconductor (http://www.bioconductor.org). Background correction and data normalization are conducted then to minimize variation caused by nonbiological effects (Xiang and Chen, 2000), followed by cluster analysis (Eisen et al., 1998), which establish gene expression patterns and define the relationships between gene expression profiles across different samples (e.g., treatment sample vs. control sample).

When starting the microarray experiment, researchers need to keep in mind that all

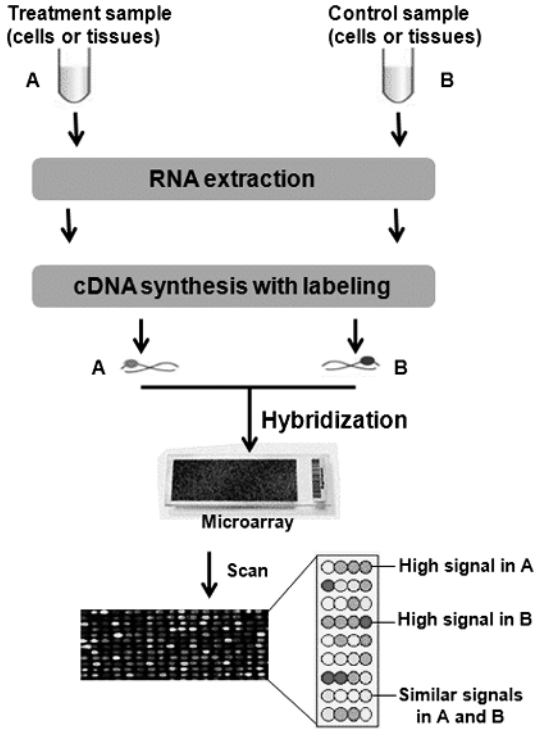

Figure 1.1 A schematic presentation of microarray experiment where expression of two samples are compared with A being the sample under treatment, and B being the sample for control. RNA is separately isolated from the samples, and fluorescently-labeled probes are made separately with different labels, e.g., cy3 for treatment (green), and cy5 for control (red). The probes are simultaneously used to hybridize an array containing the features representing the transcripts of the organism. After hybridization, signals are scanned and analyzed. If the signals are high in the treatment, green will be detected, and if the signals are high in control, red will be detected; if the signals are equal, yellow will be detected. Based on the relative signals of red and green, expression levels are determined. See color insert.

designs should meet the standards of the Microarray Gene Expression Data Society (MGED) and be compliant with the Minimum Information About a Microarray Experiment (MIAME) guidelines (http://www.mged.org/Workgroups/MIAME/miame.html). Depending on biological questions they are interested and availability of the financial and genetic resources, the researchers need to make appropriate decisions to construct the microarray experiment.

In recent years, along with more and more genetic resources available, such as ESTs, transcriptome sequences, whole genome sequences and so on, microarray technologies have been dramatically advanced and broadly applied. A variety of microarrays including low-density or high-density cDNA arrays and oligo arrays have been developed in aquaculture species, such as zebrafish (Ton et al., 2002; Mathavan et al., 2005), Salmonidae (von Schalburg et al., 2005; Koop et al., 2008), catfish (Li and Waldbieser, 2006; Peatman et al., 2007), Atalantic cod (Booman et al., 2011; Edvardsen et al., 2011), shrimp (Wongsurawat et al., 2010; Aoki et al., 2011; Leelatanawit et al., 2011), oyster (Wang et al., 2010; Dheilly et al., 2011), and so on. See Chapter 2 for information on the detail microarray resources. The applications of microarray in aquaculture species are mainly focused on the following aspects: (1) *Development*: Determining how genes interact and change the expression level during developmental process has been the main goal for development biology. Microarray, an efficient technology to globally identify the patterns of gene expression, is well utilized in aquaculture for development especially embryogenesis. Ton et al. (2002) constructed a zebrafish cDNA array to reveal dynamic change in levels of gene expression involved in development (Ton et al., 2002). An Atlantic cod cDNA microarray representing 7000 genes were used to analyze the temporal activity of the transcriptome during early cod embryogenesis (Drivenes et al., 2011). (2) *Immunity or disease resistance*: Many studies have been done by using microarray to screen and identifying genes that are involved in the immune system and disease resistance. For instance, Meijer et al. (2005) use a zebrafish oligo array containing 16K features to analyze a host transcriptome response to mycobacterium *Mycobacterium marinum* infection at the

organismal level (Meijer et al., 2005). 28K oligo arrays have been developed in catfish to identify immune-related genes in catfish (Peatman et al., 2007). (3) *Response to environmental variation or stress*: Environmental variations, such as hypoxia, water temperature, and salinity, will cause changes in physiology, genomics, and gene expression for aquaculture species. A microarray containing 8046 medaka unigenes was developed to measure gene expression profiling in the brain, gill, and liver of medaka after exposed to hypoxia (Ju et al., 2007). Microarray analyzes also have been conducted on gene expression change in water temperature (Kassahn et al., 2007; Hirayama et al., 2008). (4) *Reproduction*: Reproduction is an important trait in aquaculture industry. Karoonuthaisiri et al. (2009) utilized a cDNA microarray to screen reproduction-related genes in giant tiger shrimp, and several transcripts were identified that play important roles during shrimp ovarian development.

Currently, microarrays are mainly used to accelerate gene expression analysis under various experimental conditions. In the future, it looks promising to use microarray for single nucleotide polymorphisms (SNP) analysis, quantitative trait loci (QTL) mapping, and disease diagnosis. However, microarray study in aquaculture species is in its infancy, mostly because of the incomplete whole genome sequences in most aquaculture species. It is an essential task for the aquaculture community to exploit and adapt the advances for their respective species.

High Throughput Sequencing of mRNA (RNA-Seq)

RNA-Seq takes advantage of high-throughput DNA sequencing technology to capture the complete set of mRNA transcripts in a cell of an organism (Nagalakshmi et al., 2010). In this approach, mRNA is reverse transcribed into cDNA and fragmented, then sequenced using a next-generation technology to generate reads that can be assembled to cover a good portion of the transcripts, if not the full length of transcripts (see illustrations in Figure 1.2). Based upon different choices of sequencing technology, the sequencing yields and read lengths vary.

Currently, three main next-generation sequencing platforms are widely used in the RNA-Seq, the 454, Illumina and ABI SOLiD. Among these platforms, the throughput varies from hundreds of thousands of reads for the 454 system to hundreds of millions of reads for the Illumina and ABI SOLiD systems (Marguerat and Bahler, 2010). The read lengths typically range from 30–100 bp for Illumina and SOLiD to 200–500 bp for 454. In general, Illumina and SOLiD platforms are relatively inexpensive, while the 454 technology offers longer reads, but is more expensive per run. Illumina, SOLiD and 454 technologies can be combined in a "hybrid assembly" strategy: short reads that are sequenced at a greater depth are assembled into contigs, and long reads are subsequently used to scaffold the contigs and resolve variants (Martin and Wang, 2011).

Two main approaches can be used for RNA-Seq data analysis. One way is to map the resulting reads to a reference genome or reference transcriptome. This is usually taken in well-studied species with sequenced genome. The other way is to do the de novo assembly for species without reference genome or transcriptome. Consequently, a genome-scale map that is composed of both the transcriptional structure and/or level of expression for each gene can be generated (Wang et al., 2009).

RNA-Seq, as a way of high-throughput sequencing method, is being widely used in functional genomics studies in aquaculture species and their related model fish species such as zebrafish (Hegedus et al., 2009; Aanes et al., 2011; Bontems et al., 2011; Ordas et al., 2011; Rosel et al., 2011; Vesterlund et al., 2011), catfish (Liu et al., 2011), Japanese sea bass (Xiang et al., 2010), Atlantic cod (Johansen et al., 2011), large yellow croaker (Mu et al., 2010), rainbow trout (Lewis et al., 2010; Salem

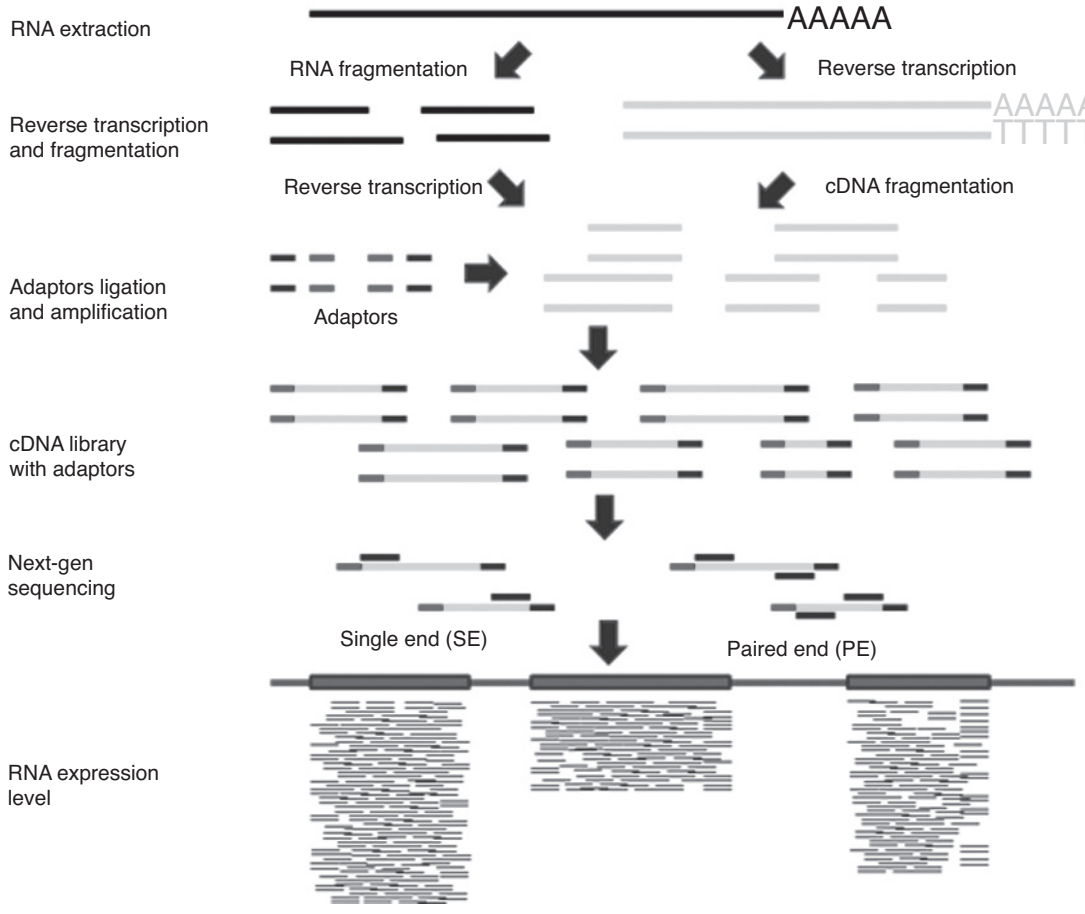

Figure 1.2 A schematic presentation of RNA-Seq. The extracted RNA is first converted into a library of complementary DNA (cDNA fragments through either RNA fragmentation (left) or DNA fragmentation (right). Sequencing adaptors (depicted by short red bars and short purple bars) are subsequently ligated to each cDNA fragment (green lines) and short sequence reads (single end or paired ends) from each cDNA are generated using high-throughput sequencing technology. The resulting sequence reads [short lines beneath the genome sequence with three genes shown (fat blue bars)] are aligned with the reference genome to evaluate gene expression by counting mapped reads. In the given example, the gene on the very left are expressed at a very high level, and the gene in the middle is expressed at a relatively lower level. See color insert.

et al., 2010; Purcell et al., 2011), European eel (Coppe et al., 2010), and spotted gar (Amores et al., 2011). The applications of RNA-Seq in aquaculture species are focused on these aspects: (1) Gene expression profiling, (2) Transcriptome characterization and gene annotation, and (3) Identification of gene-associated markers.

RNA-Seq can be used to identify differentially expressed genes under different treatments by measuring the expression level. For instance, in the study of transcriptome changes in zebrafish with mycobacterium infection (Hegedus et al., 2009) and zebrafish embryos with Salmonella infection (Ordas et al., 2011), Illumina's DGE system revealed the

high degree of transcriptional complexity of the host response to both infections and resulted in the discovery of a common set of infection-responsive genes with induced expression in infected individuals. Stockhammer et al. (2010) used the combination of microarray analysis and whole transcriptome deep sequencing to analyze the response to bacterial infection with emphasis on identification of a gene set whose responsiveness during infection is highly dependent on Traf6. In a study with large yellow croaker infected with *Aeromonas hydrophila*, changes of multiple signaling pathways involved in immunity were revealed, which will facilitate the comprehensive understanding of the mechanisms involved in the immune response to bacterial infection (Mu et al., 2010). Deep sequencing-based transcriptome profiling analysis of bacteria-challenged Japanese sea bass provided insight into the immune-relevant genes in marine fish. In the study, over 1000 strong infection-responsive transcripts were identified as significantly up- or down-regulated genes, suggesting the considerable alteration of the host transcriptome profile after the *Vibrio harveyi* infection (Xiang et al., 2010). In a study focusing on gene expression changes during development, RNA-Seq is used to compare the transcription profiles of four early developmental stages in zebrafish on a global scale. An enrichment of gene transcripts with molecular functions of DNA binding, protein folding and processing as well as metal ion binding was observed with progression of development (Vesterlund et al., 2011).

Transcriptome characterization and gene annotation is another area RNA-Seq can be applied. Transcriptome data generated through RNA-Seq can provide accurate and effective reagents for annotating the protein-coding genes. Despite the availability of complete genome sequences, a complete genome annotation would require knowledge of all transcription start and polyadenylation sites, exon–intron boundaries, splice variants, and regulatory sequences (Morozova et al., 2009). Because of its longer read lengths compared with other new sequencing technologies, 454 has been effectively used for de novo assembly of the transcriptome in several aquaculture species including lake sturgeon (Hale et al., 2009), rainbow trout (Salem et al., 2010), Atlantic cod (Johansen et al., 2011), and Yesso scallop (Hou et al., 2011). In an early study to characterize gene expression of gonad transcriptome in polyploid lake sturgeon using 454 sequencing, thousands of contigs were assembled and characterized from 454 reads providing an overview of transcription in lake sturgeon gonads, including the discovery of the genes and SNPs (Hale et al., 2009). High-throughput sequencing of the rainbow trout transcriptome using 454 sequencing technology significantly increased the suite of ESTs available for rainbow trout, allowing improved assembly and annotation of the transcriptome (Salem et al., 2010). A more recent study in the guppy, de novo assembly of the guppy (*Poecilia reticulata*) transcriptome using 454 sequence reads were conducted to detect sex-specific transcripts and provide a reference for gene expression analysis (Fraser et al., 2011).

Although shorter reads produced by Illumina or SOLiD compared with the 454 technology may be more challenging for de novo sequence assembly, the preexisting ESTs produced by Sanger sequencing can be used to facilitate the assembly (Liu et al., 2011), and the algorithms for short reads de novo assembly are being developed (e.g., Grabherr et al., 2011). Xiang et al. assembled the short reads from Illumina RNA-Seq deep sequencing to generate the nonredundant consensus which is subsequently used as references for DGE profile analysis (Xiang et al., 2010). RNA deep sequencing of the Atlantic cod transcriptome was conducted using the combination of 454, Illumina and ABI SOLiD platforms to increase the efficiency of assembly (Johansen et al., 2011).

RNA-Seq has been extensively used for the identification of gene-associated markers. In catfish, hundreds of thousands of gene-associated SNPs have been identified by deep

sequencing of RNA from many individuals of both channel catfish and blue catfish, which will be used in development of high-density catfish SNP chips for genome-wide association studies (GWAS) (Liu et al., 2011). In the study to understand the adaptive divergence between dwarf and normal lake whitefish species, 454 sequencing was used with the aim to generate a set of SNP markers, 89 SNPs showed pronounced allele frequency differences between sympatric normal and dwarf whitefish (Renaut et al., 2010).

Comparisons of Gene Expression Profiling Techniques

Hybridization-based approaches represented by microarrays are currently most popular for gene expression profiling and are readily affordable for many laboratories. Various commercial and academic microarray platforms have been developed that vary in genome coverage, availability, specificity, and sensitivity (e.g., Affymetrix, Agilent, and NimbleGen). Microarray approaches are high throughput and relatively inexpensive. However, these methods have several limitations, including relying on prior knowledge about genome sequence; high background levels owing to cross-hybridization (Royce et al., 2007); and a limited dynamic range of detection because of both background and saturation of signals. Moreover, comparing gene expression levels across different microarray experiments is often difficult and can require complicated normalization methods (Wang et al., 2009), although metagenomic analysis is possible.

In contrast to microarrays, direct sequencing of cDNAs was a digital method for gene expression measurement by counting mRNA molecules in the sample. Sequencing of cDNA or EST libraries was initially conducted using Sanger sequencing technology (Gerhard et al., 2004), but Sanger sequencing is relatively low throughput, expensive, and laborious. Therefore, EST analysis using Sanger sequencing for gene expression profiling is no longer a good choice.

More recently, RNA-Seq has become an alternative to microarrays (Wang et al., 2009). RNA-Seq provides many advantages over the traditional tag-based transcriptome analysis or microarray analysis including (1) Similarity between traditional SAGE analysis and MPSS, where RNA-Seq does not require any prior knowledge of genome sequence information; (2) Its extremely high throughput greatly improves the coverage of the transcriptome. Currently, Illumina HiSeq 2000 can generate over 200 million reads per lane, and the reads can be increased three times more using the newest chemistry, i.e., over 600 million reads per lane, allowing capture of the vast majority of transcriptome including many of the rarely expressed transcripts; (3) It overcomes many of the shortcomings of microarrays such as biases introduced during hybridization of microarrays; (4) Its cost is relatively low. One lane of RNA-Seq costs approximately $4000 while the construction of a comprehensive microarray often costs more; (5) The high throughput of reads allow technical assembly of sequences into contigs for additional studies such as gene structures. However, such a strength is also a weakness as assembly of RNA-Seq reads without reference genome or transcriptome poses challenges. De novo assembly, in particular, requires a greater level of bioinformatic expertise.

While microarrays have limitations for in-depth gene expression analyzes, they have the advantage being very useful for the high throughput analysis of multiple samples (Reinartz et al., 2002). Therefore, it may be helpful to consider the microarray and RNA-Seq as being complementary in nature which can be used as different tools for different types of experiments. For instance, to generate in-depth and quantitative gene expression data for species lacking genome information such as most of aquaculture species, RNA-Seq would be the best technology of choice. After the generation of expressed sequence data, it may be

necessary to examine whether sets of genes are differentially expressed in a large number of samples (e.g., individual gene expression variation) or under different conditions (e.g., different treatments), microarray analysis may be a better choice in terms of results and costs.

Functional Correlation of Gene Positions

In addition to expression correlation, gene functions can also be inferred from correlations of gene positions with the traits that are genetically also mapped to the same genomic location. In order to locate the positions of genes that are responsible for a certain trait, GWAS can be conducted. GWAS is a quantitative approach to analyze the association of whole genome DNA polymorphisms and a phenotypic trait, thereby localizing the genes underlining the trait.

Genome-Wide Association Studies (GWAS)

GWAS is a holistic whole-genome approach to robustly determine the association of DNA polymorphisms with correlated phenotypic traits. Most often, GWAS requires use of genome-wide polymorphic markers such as SNPs and at least hundreds of individuals with the phenotype information. We must stress the number of markers used because the more the markers, the better the markers over the entire genome; we also must stress the number of individuals because use of fewer than 200 individuals does not provide the confidence for the association. Many scientists feel that SNP association studies that are conducted on fewer than 200 animals that do not have either a confirmation population or functional data to support the association data are not reliable (e.g., James Reecy, Iowa State University, personal communication). The basis for GWAS analysis is that on a genome scale, most SNPs are distributed randomly in relation to the trait of interest, and only those SNPs tightly linked with the genes underlining the traits are in linkage disequilibrium (LD) in relation to the trait. GWAS usually involve tens or hundreds of thousands of SNPs that are tested on hundreds or thousands of individuals. These studies normally compare the DNA of two groups of participants: individuals with the trait, e.g., resistant fish, and individuals without the trait, e.g., susceptible fish.

GWAS has been extensively used for human disease research. For instance, in 2005, an association was found between age-related macular degeneration (ARMD) and a variation in the gene for complement factor H (CFH). Complement is a protein that regulates inflammation. Use of variation in the CFH gene, along with four other variants, can predict half the risk of ARMD between siblings, and this work was regarded as among the most successful examples of GWAS (Klein et al., 2005). Similarly, a GWAS involving genotyping of around 400K SNPs in a French case–control cohort allowed detection of an association between Type 2 diabetes and a variation in several SNPs in the genes TCF7L2, SLC30A8 and others (Sladek et al., 2007), which can explain a substantial portion of disease risk. In 2007, the Wellcome Trust Case Control Consortium carried out a GWAS of 14,000 cases of seven common diseases including coronary heart disease, Type 1 diabetes, Type 2 diabetes, rheumatoid arthritis, Crohn's disease, bipolar disorder, and hypertension. This study was successful in uncovering many new disease genes underlying these diseases (The Wellcome Trust Case Control Consortium, 2007).

In spite of being very powerful, GWAS has not been applied to aquaculture species. The primary reason was the lack of genome-wide polymorphic markers until recently. Now that a large number of SNPs are available for a number of aquaculture species, future application of GWAS in aquaculture species is clearly technically feasible. However, challenges related to low funding with aquaculture species are still

paramount as genotyping of a large number of polymorphic markers with a large number of individuals is expensive.

The problems associated with functional inference based on genomic positions come from the inaccuracy of GWAS analysis. On the one hand, the genomic location that are "in suspicion" to be involved in the trait can still involve large genomic segments, e.g., millions of base pairs that include many genes within the segment. On the other hand, GWAS may point to several or even many genomic locations for the trait of interest, complicating further functional analysis.

Analysis of Quantitative Trait Loci (QTL)

QTL analysis reveals statistically significant linkage between phenotypes and genotypes, thereby providing explanation for the genetic basis of variation in complex traits (Falconer and Mackay, 1996; Lynch and Walsh, 1998). In a sense, QTL analysis can be viewed as "incomplete" GWAS analysis with limited number of markers that does not cover the entire genome. As such, if one or few QTLs are found, there may be more QTLs in the genome to be discovered. More importantly, in the absence of closely linked markers in the genomic regions containing significant QTLs for the trait, the most significant genes responsible for the trait can be missed. However, because of historical reasons such as the lack of genome-wide markers, or the lack of funding, QTL analysis is still very important for aquaculture.

Many QTL studies have been conducted with aquaculture species. Most of these studies were conducted in salmonids. Table 1.1 lists some examples of QTL studies of aquaculture species.

For functional inference, QTL analysis provides limited power, as the resolution is low. However, in some cases, fine mapping of QTLs has allowed identification of causation genes or candidate genes for important aquaculture traits (see Table 1.1 for references).

Expression Quantitative Trait Loci Analysis

Studies have shown that mRNA levels for many genes are inheritable, thus amenable to genetic analysis (Brem et al., 2002; Cheung et al., 2003; Schadt et al., 2003). In the past few years, genetic and gene expression approaches have been brought together, in what has been coined "genetical genomics," to study the genetic basis of gene expression (Jansen and Nap, 2001), or expression quantitative trait loci (eQTL).

In the context of eQTL, gene expression, as measured by transcript abundance, is considered as a quantitative trait or phenotype, and in combination with genetic markers spaced across the genome, QTL are deciphered that account for variation in gene expression (Jansen and Nap, 2001). In other words, eQTL are genomic loci that regulate expression levels of mRNAs or proteins (http://en.wikipedia.org/wiki/Expression_quantitative_trait_loci-cite_note-0; Consoli et al., 2002).

Expression traits differ from most other classical complex traits in one important respect that the measured mRNA or protein trait almost always is the product of a single gene with a specific chromosomal location. eQTLs can be in either *cis* or *trans* with respect to the gene of interest. *cis*-Acting QTLs that map to the approximate location of their gene-of-origin are sequences flanking the gene (e.g., the promoter region) that regulate gene expression or transcript stability. On the other hand, *trans*-Acting QTLs are thought to involve transcription factors or other modulators, which map far from the location of their gene-of-origin, and can be on different chromosomes (Hansen et al., 2008).

eQTL analysis is the straightforward integration of traditional linkage analysis and global gene expression profiling. Both gene expression data and DNA marker data are collected in tissue samples from genetically related individuals. The expression level of each of the thousands of genes is treated as a separate quantitative trait, just like traits such as body weight

Table 1.1 Some examples of QTL studies of aquaculture species.

Species	Traits	References
Atlantic salmon	Body weight and condition factor	Reid et al., 2005
Atlantic salmon	Adaptive traits	Boulding et al., 2008
Atlantic salmon	Growth	Baranski et al., 2010
Atlantic salmon	Resistance against IPV	Houston et al., 2008; Moen et al., 2009; Gheyas et al., 2010; Houston et al., 2010
Atlantic salmon	Resistance against ISA	Moen et al., 2007
Atlantic salmon	Flesh colour	Baranski et al., 2010
Atlantic salmon	Life history	Vasemagi et al., 2010
Coho salmon	Hatch timing, weight, length and growth	McClelland and Naish, 2010
Rainbow trout	Upper thermal tolerance	Jackson et al., 1998; Danzmann et al., 1999; Perry et al., 2001; Perry et al., 2005
Rainbow trout	Life history	Leder et al., 2006
Rainbow trout	Spawning time	O'Malley et al., 2003; Colihueque et al., 2010
Rainbow trout	Osmoregulation capacities	Le Bras et al., 2011
Rainbow trout	Development rate	Robison et al., 2001; Easton et al., 2011
Rainbow trout	Whirling disease resistance	Baerwald et al., 2011
Rainbow trout	Growth	Wringe et al., 2010
Rainbow trout	Smoltification	Nichols et al., 2008
Arctic charr	Body weight, condition factor and age of sexual maturation	Moghadam et al., 2007; Kuttner et al., 2011
Arctic charr	Salinity tolerance	Norman et al., 2011
Asian seabass	Growth	Wang et al., 2006, 2011
European seabass	Body weight, morphometric traits and stress response	Massault et al., 2010
Gilthead Sea bream	Sex determination and body growth	Loukovitis et al., 2011
Tilapia	Sex determination	Shirak et al., 2006; Cnaani et al., 2007
Common carp	Muscle fiber-related QTL	Zhang et al., 2011
Eastern oyster	Disease resistance	Yu and Guo, 2006
Pacific Oyster	Growth	Guo et al., 2011
Pacific oyster	Resistance against summer mortality	Sauvage et al., 2010
Zhikong scallop	Size-related traits	Zhan et al., 2009
Pacific abalone	Growth-related traits	Liu et al., 2007
Blacklip abalone	Growth	Baranski et al., 2008

or disease resistance. The genomic loci that affect steady-state levels of each transcript are then determined by conventional QTL analysis. In this case, a significant QTL means that different genotypes at a polymorphic marker locus are associated with different expression levels. With the availability of an assembled genome sequence and genome-wide DNA microarray, this approach could be more powerful, simpler to implement, and used for any quantitative trait (Wayne and McIntyre, 2002). The advent of high-throughput array-based

methods to measure mRNA abundance in the early 2000s catalyzed an impressive number of expression QTL studies in plants, animals, and humans (Schadt et al., 2003; Morley et al., 2004; Bystrykh et al., 2005; Chesler et al., 2005; Hubner et al., 2005; Lan et al., 2006). Because RNA-Seq can provide the more accurate assessment of expression, a recent suggestion is to extend this technology to studies in eQTL analysis (Majewski and Pastinen, 2011). RNA-Seq could provide a platform independent and objective standard compared with the microarray approach. Recent eQTL studies using RNA-Seq technology have both confirmed and further clarified previous microarray results (Montgomery et al., 2010; Pickrell et al., 2010).

Mapping eQTLs is conducted using standard QTL mapping methods that examine the linkage or association between variation in expression and genetic polymorphisms. The only special consideration is that eQTL studies can involve a million or more expression microtraits. Genomic regions with a high proportion of eQTL could represent areas with genes that have common transcriptional regulators, which control important biological pathways (Cogburn et al., 2007). The ability to identify *trans*-Acting loci is particularly attractive because it is difficult to identify expression regulators even with a complete genome sequence. Standard gene mapping software packages can be used, although it is often faster to use custom code such as QTL Reaper or the web-based eQTL mapping system GeneNetwork (Wang et al., 2003; Wu et al., 2004). GeneNetwork hosts many large eQTL mapping data sets and provide access to fast algorithms to map single loci and epistatic interactions. As is true in all QTL mapping studies, the final step in defining DNA variants that cause variation in traits is usually difficult and requires additional rounds of experimentation. This is especially true for *trans* eQTL that do not benefit from the strong prior probability that relevant variants are in the immediate vicinity of the impacted gene. Statistical, graphical, and bioinformatic methods are used to evaluate positional candidate genes and entire systems of interactions (Lee et al., 2009).

Despite successes in other species, eQTL studies have not been conducted with aquaculture species due largely to the lack of technology in the past. However, it is important to point out that eQTL studies can be very expensive, and therefore, may not be suited well for aquaculture species. Even for the future, some caution should be considered before implementing eQTL analysis for aquaculture, similarly as had been pointed out for poultry (Cogburn et al., 2007). First, only one or a few tissues and time points are monitored in an individual, which provides a limited snapshot of the complete transcriptome. This must be kept in mind when results are interpreted (Cogburn et al., 2007). Second, how to properly analyze the data can be a significant challenge (Gibson and Weir, 2005). Current methods are limited in their ability to handle nonadditive, epistatic, or other complex gene effects. Third, experimental designs that do not have sufficient statistical power can compound the analysis. Only genes with large expression variation can be mapped because of the limitation of statistical power (Cogburn et al., 2007). It is widely recognized that gene expression measurements are subject to "noise," but without biological and technical replicates, one cannot determine the source of the problem. With regard to this issue, readers are referred to an excellent review by Rosa et al. (2006) for experimental design strategies for using microarrays in genetical genomics studies. Related to this limited statistical power is the ability to determine whether *cis*-Acting QTL are truly *cis*, where gene expression is regulated by the gene sequence itself, or the inability to separate out *trans*-Acting eQTL, which are closely linked (Cogburn et al., 2007). Finally, a large eQTL analysis project with a large number of genotyping and microarray assays represents an expensive venture that is not affordable for most laboratories working with aquaculture species. Although new and improved technologies with lower costs per data

point are available, the cost of genotyping and genome-wide transcriptional profiling needs to be reduced for implementation in aquaculture.

Functional Inference by Comparative Genome Analysis

A variation of position-based functional inference is functional inference through comparative genome analysis. With this approach, gene orthologies and conserved genome syntenic regions are first identified between the species of interest with well-studied model species. If the functions of a specific gene were well characterized in the model species, by inference, the orthologous gene would have the same or similar functions in the species under study. Given the large numbers of genes and the difficulties involved in determining the functions of genes, this approach could be the most popular approach for aquaculture species for the vast majority of genes. Genome-wide comparative analysis enables the transfer of genome annotation from better-annotated species to a newly-sequenced, related species. However, caution need to be exercised because: (1) Gene functions may have evolved through evolution such that the orthologous gene now may have different functions; (2) for important aquaculture traits, there would be no known "functional equivalent" in the model species. For instance, the enteric septicemia disease of catfish exists only in catfish, and it would become obscure as to what genes we should look after in zebrafish for potential candidate genes.

Theoretical Basis for Functional Inference Based on Orthologies

The evolution of life on earth, which appears to have begun billions of years ago, is recorded in the genome of modern organisms. Since the organisms are more likely to be negatively affected by mutations in functional regions, the genes and other functional elements undergo mutation at a slower rate than the rest of the genome. Therefore, the genomes of current organisms inherited from their ancestral counterparts preserve conservation because of the evolutionary pressure. Based on this principle, comparative genomics approaches have been applied to not only study the evolution of genome, but also to expand our knowledge of biological processes across multiple species by projecting high-quality annotations from one genome to another. This is a promising direction in current functional genomics research as all the genomes of many different species are being sequenced. For instance, researchers have learned a great deal about the function of human genes by examining their counterparts in simpler model organisms such as the mouse (Lee et al., 2004; Crozat et al., 2010). Because the research resources invested in model species are far greater than those invested in aquaculture species, transferring the genome information from well-characterized species can significantly advance our knowledge in aquaculture species. Comparative genomics approaches have been proven to be very effective in identifying homologous genes and functional elements in complex genomes (Cogburn et al., 2007). With the aid of actively developed bioinformatics tools, genome sequence similarity, chromosomal location and organization, and conserved sequences are integrated to effectively detect and annotate novel genes in newly-studied organisms, such as aquaculture species.

Conserved Synteny and Orthology

An important concept of comparative genomics is conserved syntenies, i.e., conserved gene sequences and their chromosomal organization. During the course of evolution, chromosome rearrangement, gene duplication, gene divergence, and gene loss have occurred. However, in vertebrates, gene arrangement in chromosomes is usually conserved, and can be classified into conserved syntenies. Strictly speaking, conserved syntenies are conserved segments that are uninterrupted by other chromosomal

segments, and conserved order of genes in the same linear orientation on the chromosome (Andersson et al., 1996). However, in many cases, as the genomic region of conserved segments extends, the exact gene order and orientation of genes may vary among related species (Eichler and Sankoff, 2003; Hurst et al., 2004).

Orthology refers to a gene in two or more species that has evolved from a common ancestor. As evolutionary studies are often from inference rather than tested by experiments, the evolutionary origin of a gene is mostly inferred from conserved syntenies.

For functional inference, if orthologies can be established, and assuming the orthologous genes have similar functions in related species, then functions of a gene in aquaculture species can be inferred from its closely-related model species. In this regard, the fish model zebrafish is highly useful because many genes can be tested on their functions.

Another application of comparative genomics is to identify regulatory sequences that control gene expression (Cogburn et al., 2007). Regulatory sequences are *cis*-Acting modules (i.e., promoters and enhancers) that govern the spatiotemporal expression of genes. Identification of regulatory sequences in the genome is far more difficult than prediction of genes. In the absence of systematic biochemical analysis, the prediction of functional elements in the genome of a new organism depends heavily on computational and comparative analysis (Jones and Pevzner, 2006). Genome sequences harboring these functional elements seem to be conserved among related species. The alignment of genome sequences of several related species allows the identification of evolutionary conserved sequences containing transcription regulatory elements. Comparative analysis of the human with mouse, rat, and dog genomes sequences have identified many common regulatory motifs in human promoters and 3′-untranslated regions (Xie et al., 2005). Various computational tools have been developed and assessed for the efficiency in predicting transcription factor-binding site (TFBS) motifs in genome sequence (Tompa et al., 2005). However, because of the nature of the regulatory sequences being short, species-specific and variable, not all TFBS can be identified by comparative or computational approaches unless accompanied by robust statistical confirmation and experimental validation (Elnitski et al., 2006; GuhaThakurta, 2006). The most reliable approach to detect regulatory regions with transcription factor binding is ChIP-chip analysis (Horak and Snyder, 2002; Valouev et al., 2008), and more recent ChIP-Seq which allowed genome-wide detection of TFBS with unprecedented sensitivity and specificity (Hu et al., 2010; Schmidt et al., 2010).

Gene Pathway Analysis

In order to effectively mine EST, microarray, or RNA-Seq data for meaningful biological information, gene set enrichment analysis (GSEA) has been recently adopted to provide clues as to what gene pathway may be turned on or turned off under a specific treatment (e.g., He et al., 2011; Kim et al., 2011b; Zhang et al., 2011). A number of pathway analysis software packages are available such as Pathway Studio (http://www.ariadnegenomics.com/), and MetaCore™ (http://www.genego.com/metacore.php). In such software packages, the algorithms calculate the statistical significance of the expression changes across every group or pathway in the database, thus, allowing identification of groups or pathways most strongly affected by the observed expression changes (http://www.ariadnegenomics.com/technology-research/pathway-analysis/). In simple terms, such software finds the genes mostly up- or down-regulated, and then places them into the gene pathways these genes are involved in, and then pins down the gene pathways that are operating under the specific condition of the "treatment."

Although gene pathway analysis has been fruitful for studies involving mammals, it has

not yet been extensively used in aquaculture species. Recently, a few studies were conducted with fish (Olsvik et al., 2008; Wang et al., 2010; Sanchez et al., 2011; Thomas et al., 2011). These studies demonstrated the usefulness of GSEA for functional genomics analysis in aquaculture species.

Profiling of DNA–Protein Interactions and Epigenetic Modifications

The control of gene expression mostly occurs at the level of transcription, so information on genome-wide chromatin profiles and DNA–protein interactions is essential to decipher the inherent logic of transcriptional regulation (Schones and Zhao, 2008). The study of epigenetics focuses on heritable changes in gene expression that does not involve the underlying DNA sequence but the epigenetic modifications such as DNA methylation and histone modification. Recent research has implicated the importance of epigenetic modifications in development, cell differentiation, and oncogenesis, setting the grounds for the Human Epigenome Project (HEP) initiative, which aims to catalog DNA methylation patterns on a genome-wide scale (Esteller, 2006). Genome-wide mapping of epigenetic modifications and protein–DNA interactions is essential for a full understanding of transcriptional regulation, and there is a growing recognition that systematic profiling of the epigenomes in multiple cell types and stages may be needed for understanding developmental processes and disease states (Bernstein et al., 2007).

The main approach for investigating epigenetic mechanisms is chromatin immunoprecipitation (ChIP), which is a technique for assaying protein–DNA binding in vivo (Solomon et al., 1988). In ChIP, antibodies are used to select specific proteins or nucleosomes, which enriches for DNA fragments that are bound to these proteins or nucleosomes (see illustrations in Figure 1.3). The introduction of microarrays allowed the fragments obtained from ChIP to be identified by hybridization to a microarray (ChIP–chip), therefore enabling a genome-scale view of DNA–protein interactions (Ren et al., 2000). Briefly, chromatin fragments are isolated using antibodies that are specific to a particular histone modification and the isolated fragments are amplified to generate micrograms of fluorescently labeled DNA, which is subsequently hybridized to DNA microarrays.

The rapid technological developments in next-generation sequencing enabled the combination of ChIP with massively parallel sequencing (ChIP–Seq), which allows researchers to survey more of the genome in less time and promises to unveil new aspects of biology (Johnson et al., 2007; Robertson et al., 2007). In ChIP–Seq, the DNA fragments of interest are sequenced directly instead of being hybridized on an array. Briefly, chromatin fragments are isolated using antibodies that are specific to a particular histone modification, and the isolated ChIP DNA is ligated to a pair of adaptors and subjected to high throughput sequencing. The number of sequenced reads mapped to a genomic locus is directly proportional to its modification level. In contrast to ChIP-chip, ChIP-Seq requires less PCR amplification and does not rely on the efficiency of probe hybridization, it is probably more quantitative and the modification levels at different genomic regions can be directly compared (Schones and Zhao, 2008).

Application of these techniques has led to great advances in our understanding of how epigenetic phenomena are regulated and how they affect gene expression. Epigenetic studies are currently solely limited to human and some other model species such as mouse, fruit fly, and yeast. However, the advances of high throughput microarray and next-generation sequencing can make these techniques feasible to decipher epigenetic mechanism in aquaculture species. In the following text, we give a summary of the applications of these techniques in model species. Interested readers are referred to the excellent reviews published in the last

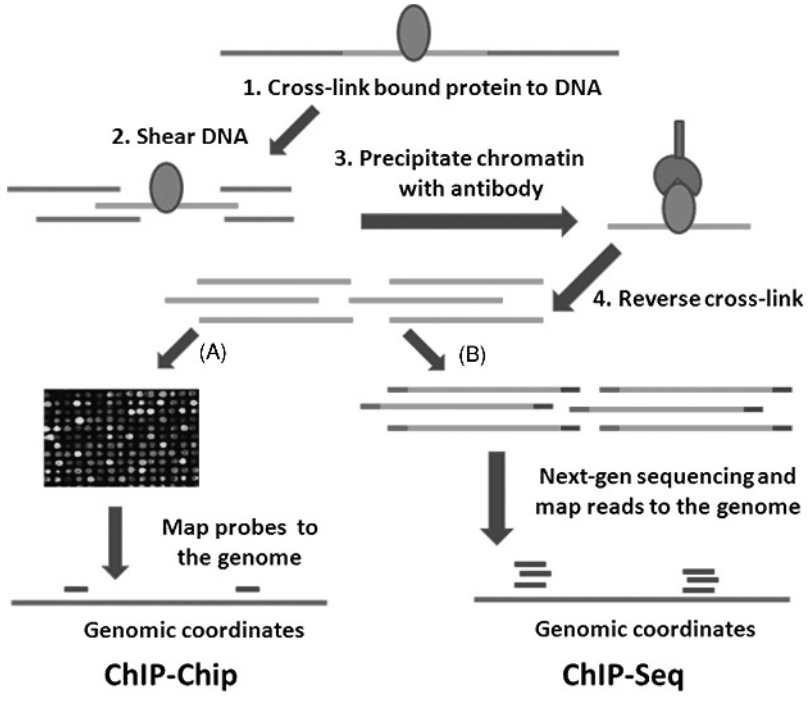

Figure 1.3 A schematic presentation of Chromatin immunoprecipitation (ChIP) combined with microarray (ChIP-Chip) and high-throughput sequencing techniques (ChIP–Seq). The first step of ChIP is to cross-link bound proteins to DNA, then isolate the chromatin and shear DNA, precipitate chromatin with protein-specific antibody, and lastly, reverse cross-link and digest protein to release the bound DNA. In ChIP-Chip (left lower panel), the ChIP DNA is then amplified to obtain sufficient DNA followed by hybridization to a DNA microarray. The microarray probes can then be mapped to the genome to yield genomic coordinates; in ChIP–Seq (right lower panel), the ChIP DNA ends are repaired and ligated to a pair of adaptors, followed by high throughput sequencing. The resulting sequence reads are mapped to a reference genome to obtain genomic coordinates that correspond to the immunoprecipitated fragments. See color insert.

several years (Ren et al., 2000; Holliday, 2006; Goldberg et al., 2007; Henikoff, 2008; Schones and Zhao, 2008; Suzuki and Bird, 2008; Park, 2009; Laird, 2010).

DNA Methylation

DNA methylation is defined as the addition of a methyl group to the cytosine in a CpG dinucleotide. This covalent modification is known to have an important role in genome function, such as gene regulation, chromosomal stability, and parental imprinting (Bird, 2002). CpG islands, GC-rich regions within mammalian genomes, are resistant to methylation and are associated with most human genes (Ng and Bird, 1999). Gene inactivation can be observed by de novo methylation of promoters that contain CpG islands (Bird, 2002), and DNA methylation of TBFS can influence their binding (Tate and Bird, 1993; Bell et al., 1999).

Studies of DNA methylation on a genome scale primarily rely on "local" techniques combined with global approaches such as DNA microarrays or high-throughput sequencing. The most important local techniques include genomic mapping of cleavage sites by restriction enzymes that differentiate between methylated and unmethylated CpG DNA sequences. Sequencing of DNA after treatment with sodium

bisulphite, which converts all unmethylated cytosines to uracils, allows differentiation of methylated and unmethylated sites. Affinity purification can be used to isolate protein-bound DNA with methylcytosine DNA-binding domain (MBD) proteins, and immunoprecipitation of DNA with an antibody that recognizes 5-methyl cytosine (known as MeDIP, mDIP or mCIP) (Schones and Zhao, 2008). In the past few years, immunoprecipitation of methylated DNA has been combined with microarrays to study DNA methylation at a genome scale in human cells (Weber et al., 2005; Keshet et al., 2006). Recently, MeDIP has been combined with tiling microarrays to provide a DNA methylation map of the *Arabidopsis thaliana* genome (Zhang et al., 2006; Zilberman et al., 2007), and with high-density promoter arrays to map the human promoter methylome (Shen and Waterland, 2007; Weber et al., 2007).

Histone Modifications

Histone proteins play important roles in gene regulation process. In general, genes those are active have less bound histones, while inactive genes are highly associated with histones during interphase. Histone proteins are subject to a number of covalent post-translational modifications, primarily at their N-terminal tails, including methylation, acetylation, phosphorylation, ubiquitylation, and ADP-ribosylation. Histones undergo post-translational modifications that alter their interaction with DNA and nuclear proteins, which affect their function of gene regulation (Schones and Zhao, 2008).

The most prevalent technique used to map histone modifications at a genomic scale has been the combination of ChIP with DNA microarrays (ChIP-chip). The first ChIP-chip studies of histone modifications in *Saccharomyces cerevisiae* (Bernstein et al., 2002; Robyr et al., 2002) and *Drosophila melanogaster* (Schubeler et al., 2004) suggested that histone modifications are associated with distinct genomic regions and with distinct transcription states. Subsequent studies with higher resolution tiling arrays in yeast (Liu et al., 2005; Pokholok et al., 2005) reinforced the concept of redundancy in histone-modification maps. ChIP-chip has also been used to profile histone modifications in mammalian genomes (Bernstein et al., 2005; Heintzman et al., 2007).

More recently, ChIP-Seq is used to profile histone modifications. The first applications of ChIP-Seq were done in $CD4^+$ T cells (Barski et al., 2007) and mouse embryonic stem (ES) cells (Mikkelsen et al., 2007).

Numerous genome-wide profiles of histone modifications have been conducted in various model organisms. In studies using yeast, the histone H3 lysine 4 methylation (H3K4me) and histone acetylation were found positively correlated with transcription levels, and are highly enriched in promoter regions and extend into the transcribed regions significantly (Liu et al., 2005; Pokholok et al., 2005). A general concept is now emerging in which distinct genomic regions (e.g., enhancers, promoters and genes) have distinct histone-modification patterns (Heintzman et al., 2007). Genome-wide distribution patterns of histone acetylation have allowed successful identification of functional enhancer elements (Roh et al., 2007). Novel transcription units and functional transcription start sites have also been determined based on modification patterns (Barski et al., 2007). It has become apparent that histone modification patterns can be a useful means to facilitate a more precise annotation of the human and other genomes.

DNA–Proteins Interactions

Gene expression is often regulated by proteins that activate or repress transcription by binding to short, specific DNA sequences. Such *cis*-Acting sites are usually located close to the promoter (TFBS) for the regulated gene. In a ChIP-Seq experiment for DNA-binding proteins, DNA fragments associated with a specific protein are enriched and sequenced.

The earliest reports using ChIP-Seq focused on identification of DNA-binding sites for the transcription factors NRSF (neuron-restrictive silencer factor) (Johnson et al., 2007) and STAT1 (Signal transducer and activator of transcription 1) (Robertson et al., 2007). Genome-wide mapping of many transcription factors has now been reported (Jothi et al., 2008; Valouev et al., 2008; Wederell et al., 2008). Many of these studies show that TBFS are found upstream of the transcription start site, downstream of the termination signal, within introns and exons of genes and in genomic regions far away from any annotated genes. Such diversity in DNA-binding site locations highlights the power of genome-wide analysis and our need to better understand mechanisms of gene regulation. As thousands of DNA-binding sites are normally found for each transcription factor, determining the functions of protein–DNA interactions can be difficult. Gene expression profiling can be used to determine whether the associated gene is expressed in the cells or tissues. In addition, patterns of specific histone modification around TFBS can be used to evaluate whether the site is acting to activate or repress transcription (Heintzman et al., 2007; Koch et al., 2007). By combining transcription factor binding analysis with gene expression profiling and histone modification, a detailed picture of binding site activity can emerge (Cullum et al., 2011).

Direct Test of Gene Functions

Traditional Approaches

Gene functions can be directly tested with forward genetics or reverse genetics approaches. In forward genetics, the phenotype is already known, usually from a mutant with visible or measurable phenotype. The genes responsible for the mutation can be mapped and characterized. The basic concept of the forward genetics approaches is that if a mutant phenotype is mapped to a locus where a gene mutation is found, then the gene harboring the mutation is responsible for the mutant phenotype. In this way, functions of many genes have been determined. In zebrafish, mutant libraries have been created to study the genes responsible for the mutation (e.g., Ellett and Lieschke, 2010; Lessman, 2011; Peal et al., 2011). However, its limitation is the inability for generation of mutants with a visible phenotype, particularly for aquaculture traits. Functional genomics using forward genetics approaches are well reviewed by Hackett and Clark (2007), and interested readers are referred to his review.

Reverse genetics approaches have also been used for the study of gene functions. Transgenic technology and knockout/knockdown technologies are among the most used approaches. With transgenic technology, enhanced growth rate was clearly demonstrated with the transfer of growth hormone genes (Du et al., 1992; Gross et al., 1992; Devlin et al., 1994; Rahman et al., 1998; Rahman et al., 2001). The basic concept of functional analysis through transgenesis is that if a gene has a specific function, then the introduction of the gene or an increase in its copy numbers in the genome should generate a noticeable phenotypic change. While transgenic technology is very effective for the analysis of gene functions, genome-wide test of functions of many genes may be difficult because the inefficiency of transgenic technology, screening for transgenic animals, and the associated time, efforts, and costs.

There are still no effective knockout systems in aquaculture species, but knockdown using RNAi and morpholinos have been conducted in many studies (e.g., Huang et al., 2008; Anastasaki et al., 2011). For instance, RNAi was employed to knockdown three subunits of the gonadotropin alpha (GtHalpha), follicle-stimulating hormone beta (FSHbeta), and luteinizing hormone beta (LHbeta) genes. Expression of GtH mRNA was obviously and more efficiently depressed by GtHalpha RNAi expression. A GtHalpha morpholino analysis showed that the GtHalpha morpholino led to

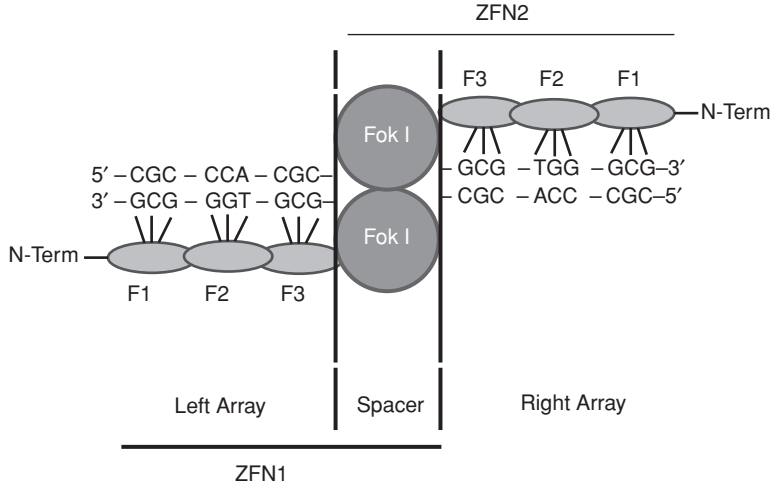

Figure 1.4 Presentation of engineered zinc finger nuclease (ZFN). The sequence specificity is determined by individual ZFN1 and ZFN2 fingers. The Fok I endonuclease domain introduces double-stranded breaks after sequence-specific DNA binding. (This figure was adopted from Dr. Stephen C. Ekker (2008) with his permission.) See color insert.

suppression of embryonic development and the production of embryonic mutants as a result of an injection of GtHalpha–shRNA (Huang et al., 2008). However, these approaches are by no means ideal for functional genomics studies in aquaculture species. The main problems of knockdown technologies are their inability to provide sufficient knockdown. The level of suppression of gene expression by these technologies may vary depending on the systems.

Zinc Finger Nuclease (ZFN) Technology

A recently developed technology, the zinc finger nuclease (ZFN) technology, enables targeted editing of the genome by introducing double-strand breaks in DNA at specific locations by taking advantage of ZFNs. ZFN technology shows a great promise for functional genomics research, including aquaculture species.

ZFN technology is based on engineered ZFNs. Engineered ZFNs are artificial restriction enzymes composed of zinc finger (ZF) DNA-binding domain and a DNA-cleavage domain (Kim et al., 1996). As depicted in Figure 1.4, the DNA-binding domain of ZFN determines the specificity, while the endonuclease Fok I cleave the DNA.

The ZF domains can be engineered to bind target DNA sequences of interest (Segal and Barbas, 2001; Beerli and Barbas, 2002), enabling the creation of double-strand DNA breaks (DSBs) at targeted genome locations. Repair of ZFN-induced DSBs can be a nonhomologous end-joining (NHEJ) event that is error-prone and leads to efficient introduction of insertion or deletion (indels) mutations at the site of DSB (Figure 1.5).

Alternatively, repair of a DSB by homologous recombination (HR) machinery with supplied DNA fragment as a template can promote efficient introduction of alterations or insertions at or near the break site (Sander et al., 2011)

With the popularity of ZFN technology, a ZFN Consortium has been established: The Zinc Finger (ZF) Consortium (http://www.zincfingers.org/scientific-background.htm). This Consortium is committed to developing

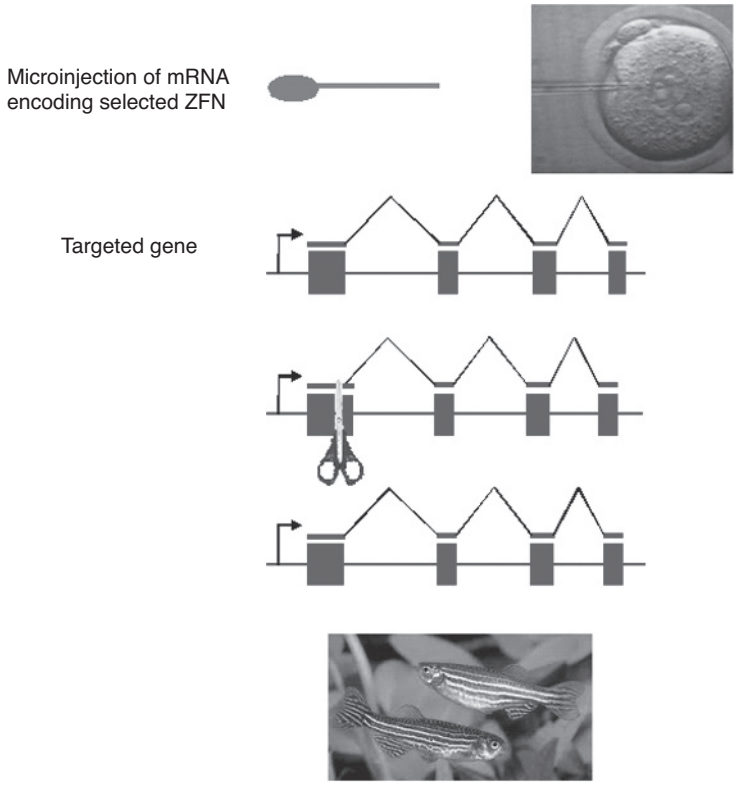

Figure 1.5 Schematic presentation of the processes of making a targeted gene knockout in zebrafish using ZFNs. mRNA encoding the selected ZFN is injected into one-cell zebrafish embryos. This custom ZFN binds and makes a double stranded break at the specified locus. Cell repairs the double stranded DNA break imprecisely to introduce mutations to the selected gene. The specific molecular lesion is clonally selected and determined after genotyping the offspring from these mosaic founders. (This figure is adopted from Dr. Stephen C. Ekker (2008) with his permission.) See color insert.

resources, software, and other tools for engineering ZFs and for performing genome engineering that are robust, user-friendly, and publicly available to the academic scientific community.

Approaches for engineering zinc-finger arrays were developed to allow widespread adoption and large-scale use of ZFN technology. As summarized by Maeder et al. (2008), zinc-finger engineering methods can be grouped into two general categories: The first is "Modular assembly," which involves joining together single fingers with pre-characterized specificities (Liu et al., 2001; Beerli and Barbas, 2002; Bae et al., 2003; Segal et al., 2003; Mandell and Barbas, 2006; Kim et al., 2011a). Modular assembly is easy to perform, but the success rate is relatively low (6%) (Ramirez et al., 2008), and can yield ZFNs with low activities and/or high toxicities (Cornu et al., 2008; Pruett-Miller et al., 2008). The second approach involves combinatorial selection-based methods that yield multi-finger domains possessing high DNA-binding affinities and specificities (Greisman and Pabo, 1997; Isalan et al., 1997, 2001; Hurt et al., 2003) and high activities and low toxicities when expressed as ZFNs in human cells (Cornu et al., 2008; Pruett-Miller et al., 2008). The weakness

of selection-based methods, as summarized by Maeder et al. (2008) is that it requires construction and interrogation of large numbers of randomized libraries (typically >108 in size). Recently, Maeder et al. (2008) developed a method known as OPEN (Oligomerized Pool ENgineering). The efficiencies of OPEN are high, ranging from 1%–50%. Apparently, the publicly available OPEN platform will enable routine practice and further development of ZFN technology.

Several software tools have been developed including ZiFiT, ZiFDB, and ZFNGenome. The ZiFiT software package identifies potential target sites in DNA sequences for which ZF proteins may be engineered (Sander et al., 2010a; 2010b; Sander et al., 2007). ZiFDB is a web-accessible database of ZFs and engineered ZF arrays, which organizes information on both individual ZF modules and engineered ZF arrays (Fu et al., 2009). ZFNGenome is a GBrowse-based tool for identifying and visualizing potential target sites of OPEN-generated ZFNs (Reyon et al., 2011).

ZFNGenome currently includes a total of more than 11 million potential ZFN target sites mapped within the fully sequenced genomes of seven model species.

ZFN technology can be used for gene knockouts. For instance, ZFN-induced knockouts have been created in the kdr gene in zebrafish (Meng et al., 2008). In another study, NHEJ-based repair of ZFN-induced DSBs were exploited to disrupt gol and ntl genes in zebrafish to generate mutants (Doyon et al., 2008). These results demonstrated the potential of this technology for reverse genetics applications. To date, no published data are available for use of ZFN technology in aquaculture species. However, heritable knockout lines of aquaculture fish species have been produced (Dr. Qingshun Zhao, Nanjing University, China, Personal communications).

Because homologous recombination relies on homologous DNA to repair the DSB, gene "editing" can be achieved by providing an exogenous "donor" template (Reyon et al., 2011). This results in replication of the "donor" DNA sequence at the specified locus. Many studies have taken advantage of this process to introduce small mutations or large insertions in plants and animals such as maize (Shukla et al., 2009), fruit fly (Beumer et al., 2006) and human (Zou et al., 2009).

Although most of previous ZFN works focus on model species, we can envision that ZFN technology will be applied to aquaculture species in the near future. The results derived from the applications of ZFN technology in zebrafish will absolutely offer prior knowledge for similar work in aquaculture fish species. At the same time, more and more genome sequencing projects of aquaculture species will provide the fundamental genomic resources for the identification of potential target site and the design of engineered ZF arrays.

Functional Genomics Approaches Suitable for Aquaculture

Currently, the whole genome projects have been "completed" in six fish species including Zebrafish (*Danio rerio*), medaka (*Oryzias latipes*), fugu (*Takifugu rubripes*), *Tetraodon* (*Tetraodon nigroviridis*) Stickleback (*Gasterosteus aculeatus*), and Atlantic cod (*Gadus morhua*). With the advances in next-generation sequencing technology, more than 30 whole genome sequencing projects are underway for aquaculture species such as Atlantic salmon (*Salmo salar*), rainbow trout (*O. mykiss*), nile tilapia (*Oreochromis niloticus*), Half-smooth tongue sole (*Cynoglossus semilaevis*), channel catfish (*Ictalurus punctatus*), and many more. For shellfish and crustaceans, a number of species are also being sequenced such as Pacific oyster (*C. gigas*), zhikong scallop (*Chlamys farreri*), and Pacific white shrimp (*Litopenaeus vannamei*). Although none of these genomes has been completely finished to a stage that sequences are assembled together into chromosome, the huge-scale sequence resources

generated during this process are valuable for genome-scale functional genomics studies. For species lack of genome projects, a wealth of sequence data exist which are largely from cDNA sequencing projects (see Chapter 2 for EST and RNA-Seq resources generated in aquaculture species). Although all the technologies will continue to provide functional genomics information, in the following text, we attempt to make some assessment of the existing technology.

With the genome or transcriptome resources, microarrays can be readily developed for functional genomics research in aquaculture species. While situation may vary, we anticipate that microarrays will still have important roles to play for functional genomics analysis, especially for species with relatively larger research communities.

In spite of its historical contributions, EST analysis will become a limited approach because of its high cost. In contrast, the next-generation sequencing based gene expression analysis is increasingly feasible with the sequencing cost becoming lower and lower. We see that RNA-Seq can be a very rapid and technologically ready approach for aquaculture functional genomics, but large-scale application with many individuals may still be cost-prohibitive presently.

With the increasingly more availability of genome-scale SNP markers, the GWAS perhaps will become the most important approach among all for the determination of genes involved in performance and production traits important for aquaculture. Most aquaculture species have high fecundities with the ability to produce a large number of progenies in a single spawn, allowing genetic associations to be analyzed with minimal "noise" in the system. On the other hand, because of the aquatic living environment, phenotypes of aquaculture species are more difficult to measure, making GWAS more challenging as well. Traditional QTL analysis will continue to have an important role in providing basic information of genomic locations underlying various traits.

Although many eQTL studies have been conducted in plants and human as discussed in the previous text, few studies were conducted in aquaculture species. Recently, an eQTL analysis conducted on the dwarf and normal whitefish species pairs identified 253 genes differentially expressed in white muscle between these two coregonid ecotypes, 33 of which were associated with 53 eQTL (Derome et al., 2008). However, as discussed in the previous text, application of eQTL in aquaculture species will prove to be difficult, not because of technical reasons, but because of limitations of funding.

As the whole genome sequences are being assembled with aquaculture species, comparative genomics will perhaps be a very powerful approach for functional inference, simply because it is not possible to determine functions of genes in each species, related functions can be inferred one another based on evolutionary conservation of genes and their functions. In this regard, work in model species will continue to be useful for aquaculture species.

Epigenetics is a rapidly developing field of functional genomics because of its recognized role in determination of the phenotype. At present, the worm, fruit fly, and mouse are the most frequently used models in epigenetics. Epigenetic study in aquaculture is still in its infancy. Marbled crayfish (*Marmorkrebs*) was recently proposed as a model organism for research in epigenetics (Vogt, 2008). DNA methylation studied for marbled crayfish revealed that a global DNA methylation in the hepatopancreas and abdominal musculature varied from 1.52–1.94% and 1.77–2.09%, which corresponds to roughly half of the values in humans (Vogt, 2008). In spite of its importance, epigenetic studies in aquaculture species are generally limited by funding at present.

Among many of the direct experimental approaches for the determination of gene functions, the ZFN technology shows the greatest promise for use in aquaculture species. Clearly, investment should be made to develop ZFN technology and its application with aquaculture species.

Acknowledgments

Research in our laboratory is supported by grants from the United States Department of Agriculture (USDA) National Research Initiative (NRI) Animal Genome and Genetic Mechanisms Program, the USDA NRI Basic Genome Reagents and Tools Program, the Mississippi–Alabama Sea Grant Consortium, the Alabama Department of Conservation, the United States Agency for International Department (USAID), and the Binational Agriculture Research and Development Fund (BARD).

References

Aanes H, Winata CL, Lin CH, Chen JP, Srinivasan KG, Lee SG, Lim AY, Hajan HS, Collas P, Bourque G, Gong Z, Korzh V, Alestrom P, Mathavan S (2011) Zebrafish mRNA sequencing deciphers novelties in transcriptome dynamics during maternal to zygotic transition. *Genome Research*, 21: 1328–1338.

Acosta J, Carpio Y, Borroto I, Gonzalez O, Estrada MP (2005) Myostatin gene silenced by RNAi show a zebrafish giant phenotype. *Journal of Biotechnology*, 119: 324–331.

Adams MD, et al. (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. *Science*, 252: 1651–1656.

Adams MD, et al. (1995) Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. *Nature*, 377: 3–174.

Amores A, Catchen J, Ferrara A, Fontenot Q, Postlethwait JH (2011) Genome evolution and meiotic maps by massively parallel DNA sequencing: spotted gar, an outgroup for the teleost genome duplication. *Genetics*, 188: 799–808.

Anastasaki C, Longman D, Capper A, Patton EE, Caceres JF (2011) Dhx34 and Nbas function in the NMD pathway and are required for embryonic development in zebrafish. *Nucleic Acids Research*, 39: 3686–3694.

Andersson L, et al. (1996) Comparative genome organization of vertebrates. The First International Workshop on Comparative Genome Organization. *Mammalian Genome*, 7: 717–734.

Aoki T, Wang HC, Unajak S, Santos MD, Kondo H, Hirono I (2011) Microarray analyses of shrimp immune responses. *Marine Biotechnology (NY)*, 13: 629–638.

Backes C, Keller A, Kuentzer J, Kneissl B, Comtesse N, Elnakady YA, Muller R, Meese E, Lenhof HP (2007) GeneTrail–advanced gene set enrichment analysis. *Nucleic Acids Research*, 35: W186–W192.

Bae KH, et al. (2003) Human zinc fingers as building blocks in the construction of artificial transcription factors. *Nature Biotechnology*, 21: 275–280.

Baerwald MR, Petersen JL, Hedrick RP, Schisler GJ, May B (2011) A major effect quantitative trait locus for whirling disease resistance identified in rainbow trout (*Oncorhynchus mykiss*). *Heredity*, 106: 920–926.

Bailey LC, Searls DB, Overton GC (1998) Analysis of EST-driven gene annotation in human genomic sequence. *Genome Research*, 8: 362–376.

Baranski M, Moen T, Vage DI (2010) Mapping of quantitative trait loci for flesh colour and growth traits in Atlantic salmon (*Salmo salar*). *Genetics Selection Evolution*, 42: 17.

Baranski M, Rourke M, Loughnan S, Hayes B, Austin C, Robinson N (2008) Detection of QTL for growth rate in the blacklip abalone (*Haliotis rubra* Leach) using selective DNA pooling. *Animal Genetics*, 39: 606–614.

Barski A, Cuddapah S, Cui KR, Roh TY, Schones DE, Wang ZB, Wei G, Chepelev I, Zhao KJ (2007) High-resolution profiling of histone methylations in the human genome. *Cell*, 129: 823–837.

Bauer S, Grossmann S, Vingron M, Robinson PN (2008) Ontologizer 2.0–a multifunctional tool for GO term enrichment analysis and data exploration. *Bioinformatics*, 24: 1650–1651.

Beerli RR, Barbas CF 3rd (2002) Engineering polydactyl zinc-finger transcription factors. *Nature Biotechnology*, 20: 135–141.

Bell AC, West AG, Felsenfeld G (1999) The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell*, 98: 387–396.

Bernstein BE, Humphrey EL, Erlich RL, Schneider R, Bouman P, Liu JS, Kouzarides T, Schreiber SL (2002) Methylation of histone H3 Lys 4 in coding regions of active genes. *Proceedings of the National Academy of Sciences of the United States of America*, 99: 8695–8700.

Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, Huebert DJ, McMahon S, Karlsson EK, Kulbokas EJ 3rd, Gingeras TR, Schreiber SL,

Lander ES (2005) Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell*, 120: 169–181.

Bernstein BE, Meissner A, Lander ES (2007) The mammalian epigenome. *Cell*, 128: 669–681.

Beumer K, Bhattacharyya G, Bibikova M, Trautman JK, Carroll D (2006) Efficient gene targeting in Drosophila with zinc-finger nucleases. *Genetics*, 172: 2391–2403.

Bird A (2002) DNA methylation patterns and epigenetic memory. *Genes & Development*, 16: 6–21.

Bonaldo MDF, Lennon G, Soares MB (1996) Normalization and subtraction: Two approaches to facilitate gene discovery. *Genome Research*, 6: 791–806.

Bontems F, Baerlocher L, Mehenni S, Bahechar I, Farinelli L, Dosch R (2011) Efficient mutation identification in zebrafish by microarray capturing and next generation sequencing. *Biochemical and Biophysical Research Communications*, 405: 373–376.

Booman M, Borza T, Feng CY, Hori TS, Higgins B, Culf A, Leger D, Chute IC, Belkaid A, Rise M, Gamperl AK, Hubert S, Kimball J, Ouellette RJ, Johnson SC, Bowman S, Rise ML (2011) Development and experimental validation of a 20K Atlantic Cod (*Gadus morhua*) oligonucleotide microarray based on a collection of over 150,000 ESTs. *Marine Biotechnology*, 13: 733–750.

Boulding EG, Culling M, Glebe B, Berg PR, Lien S, Moen T (2008) Conservation genomics of Atlantic salmon: SNPs associated with QTLs for adaptive traits in parr from four trans-Atlantic backcrosses. *Heredity*, 101: 381–391.

Brem RB, Yvert G, Clinton R, Kruglyak L (2002) Genetic dissection of transcriptional regulation in budding yeast. *Science*, 296: 752–755.

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

Bystrykh L, Weersing E, Dontje B, Sutton S, Pletcher MT, Wiltshire T, Su AI, Vellenga E, Wang JT, Manly KF, Lu L, Chesler EJ, Alberts R, Jansen RC, Williams RW, Cooke MP, de Haan G (2005) Uncovering regulatory pathways that affect hematopoietic stem cell function using 'genetical genomics'. *Nature Genetics*, 37: 225–232.

Chesler EJ, Lu L, Shou SM, Qu YH, Gu J, Wang JT, Hsu HC, Mountz JD, Baldwin NE, Langston MA, Threadgill DW, Manly KF, Williams RW (2005) Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. *Nature Genetics*, 37: 233–242.

Cheung VG, Conlin LK, Weber TM, Arcaro M, Jen KY, Morley M, Spielman RS (2003) Natural variation in human gene expression assessed in lymphoblastoid cells. *Nature Genetics*, 33: 422–425.

Chini V, Cattaneo AG, Rossi F, Bernardini G, Terova G, Saroglia M, Gornati R (2008) Genes expressed in Blue Fin Tuna (*Thunnus thynnus*) liver and gonads. *Gene*, 410: 207–213.

Cnaani A, Lee BY, Ozouf-Costaz C, Bonillo C, Baroiller JF, D'Cotta H, Kocher T (2007) Mapping of sox2 and sox14 in tilapia (*Oreochromis* spp.). *Sex Development*, 1: 207–210.

Cogburn LA, Porter TE, Duclos MJ, Simon J, Burgess SC, Zhu JJ, Cheng HH, Dodgson JB, Burnside J (2007) Functional genomics of the chicken—A model organism. *Poultry Science*, 86: 2059–2094.

Colihueque N, Cardenas R, Ramirez L, Estay F, Araneda C (2010) Analysis of the association between spawning time QTL markers and the biannual spawning behavior in rainbow trout (*Oncorhynchus mykiss*). *Genetics and Molecular Biology*, 33: 578–582.

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 21: 3674–3676.

Consoli L, Lefevre A, Zivy M, de Vienne D, Damerval C (2002) QTL analysis of proteome and transcriptome variations for dissecting the genetic architecture of complex traits in maize. *Plant Molecular Biology*, 48: 575–581.

Copf T, Rabet N, Averof M (2006) Knockdown of spalt function by RNAi causes de-repression of Hox genes and homeotic transformations in the crustacean *Artemia franciscana*. *Developmental Biology*, 298: 87–94.

Coppe A, Pujolar JM, Maes GE, Larsen PF, Hansen MM, Bernatchez L, Zane L, Bortoluzzi S (2010) Sequencing, de novo annotation and analysis of the first *Anguilla anguilla* transcriptome: EeelBase

opens new perspectives for the study of the critically endangered European eel. *BMC Genomics*, 11: 635.

Cornu TI, Thibodeau-Beganny S, Guhl E, Alwin S, Eichtinger M, Joung J, Cathomen T (2008) DNA-binding specificity is a major determinant of the activity and toxicity of zinc-finger nucleases. *Molecular Therapy*, 16: 352–358.

Crozat K, Guiton R, Guilliams M, Henri S, Baranek T, Schwartz-Cornil I, Malissen B, Dalod M (2010) Comparative genomics as a tool to reveal functional equivalences between human and mouse dendritic cell subsets. *Immunological Reviews*, 234: 177–198.

Cullum R, Alder O, Hoodless PA (2011) The next generation: using new sequencing technologies to analyse gene regulation. *Respirology*, 16: 210–222.

Danzmann RG, Jackson TR, Ferguson MM (1999) Epistasis in allelic expression at upper temperature tolerance QTL in rainbow trout. *Aquaculture*, 173: 45–58.

Derome N, Bougas B, Rogers SM, Whiteley AR, Labbe A, Laroche J, Bernatchez L (2008) Pervasive sex-linked effects on transcription regulation as revealed by expression quantitative trait loci mapping in lake whitefish species pairs (*Coregonus* sp., *sahnonidae*). *Genetics*, 179: 1903–1917.

Devlin RH, Yesaki TY, Biagi CA, Donaldson EM, Swanson P, Chan WK (1994) Extraordinary salmon growth. *Nature*, 371: 209–210.

Dheilly NM, Lelong C, Huvet A, Favrel P (2011) Development of a Pacific oyster (*Crassostrea gigas*) 31,918-feature microarray: identification of reference genes and tissue-enriched expression patterns. *BMC Genomics*, 12: 468.

Doyon Y, McCammon JM, Miller JC, Faraji F, Ngo C, Katibah GE, Amora R, Hocking TD, Zhang L, Rebar EJ, Gregory PD, Urnov FD, Amacher SL (2008) Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nature Biotechnology*, 26: 702–708.

Drivenes O, Taranger GL, Edvardsen RB (2011) Gene expression profiling of Atlantic Cod (*Gadus morhua*) embryogenesis using microarray. *Marine Biotechnology*, 14: 167–176.

Du SJ, Gong ZY, Fletcher GL, Shears MA, King MJ, Idler DR, Hew CL (1992) Growth enhancement in transgenic Atlantic salmon by the use of an "all fish" chimeric growth hormone gene construct. *Bio/Technology*, 10: 176–181.

Easton AA, Moghadam HK, Danzmann RG, Ferguson MM (2011) The genetic architecture of embryonic developmental rate and genetic covariation with age at maturation in rainbow trout *Oncorhynchus mykiss*. *Journal of Fish Biology*, 78: 602–623.

Edvardsen RB, Malde K, Mittelholzer C, Taranger GL, Nilsen F (2011) EST resources and establishment and validation of a 16 k cDNA microarray from Atlantic cod (*Gadus morhua*). *Comprative Biochemistry and Physiology Part D Genomics&Proteomics*, 6: 23–30.

Eichler EE, Sankoff D (2003) Structural dynamics of eukaryotic chromosome evolution. *Science*, 301: 793–797.

Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the United States of America*, 95: 14863–14868.

Ekker (2008) SC. Zinc finger-based knockout punches for zebrafish genes. *Zebrafish*, 5: 121–123.

Ellett F, Lieschke GJ (2010) Zebrafish as a model for vertebrate hematopoiesis. *Current Opinion in Pharmacology*, 10: 563–570.

Elnitski L, Jin VX, Farnham PJ, Jones SJ (2006) Locating mammalian transcription factor binding sites: a survey of computational and experimental techniques. *Genome Research*, 16: 1455–1464.

Esteller M (2006) The necessity of a human epigenome project. *Carcinogenesis*, 27: 1121–1125.

Falconer DS, Mackay TF (1996) *Introduction to Quantitative Genetics*, 4th edn. Longman, Harlow.

Ferraresso S, Vitulo N, Mininni AN, Romualdi C, Cardazzo B, Negrisolo E, Reinhardt R, Canario AVM, Patarnello T, Bargelloni L (2008) Development and validation of a gene expression oligo microarray for the gilthead sea bream (*Sparus aurata*). *BMC Genomics*, 9: 580.

Fleury E, Huvet A, Lelong C, de Lorgeril J, Boulo V, Gueguen Y, Bachere E, Tanguy A, Moraga D, Fabioux C, Lindeque P, Shaw J, Reinhardt R, Prunet P, Davey G, Lapegue S, Sauvage C, Corporeau C, Moal J, Gavory F, Wincker P, Moreews F, Klopp C, Mathieu M, Boudry P, Favrel P (2009) Generation and analysis of a 29,745 unique expressed sequence tags from the Pacific oyster (*Crassostrea gigas*) assembled into a publicly accessible database: the GigasDatabase. *BMC Genomics*, 10: 341.

Fraser BA, Weadick CJ, Janowitz I, Rodd FH, Hughes KA (2011) Sequencing and characterization of the guppy (*Poecilia reticulata*) transcriptome. *BMC Genomics*, 10: 341.

Fu F, Sander JD, Maeder M, Thibodeau-Beganny S, Joung JK, Dobbs D, Miller L, Voytas DF (2009) Zinc finger database (ZiFDB): a repository for information on C2H2 zinc fingers and engineered zinc-finger arrays. *Nucleic Acids Research*, 37: D279–D283.

Gerhard DS, Wagner L, Feingold EA, Shenmen CM, Grouse LH, Schuler G, Klein SL, Old S, Rasooly R, Good P, Guyer M, Peck AM, Derge JG, Lipman D, Collins FS, Team MP (2004) The status, quality, and expansion of the NIH full-length cDNA project: The mammalian gene collection (MGC). *Genome Research*, 14: 2121–2127.

Gheyas AA, Haley CS, Guy DR, Hamilton A, Tinch AE, Mota-Velasco JC, Woolliams JA (2010) Effect of a major QTL affecting IPN resistance on production traits in Atlantic salmon. *Animal Genetics*, 41: 666–668.

Gibson G, Weir B (2005) The quantitative genetics of transcription. *Trends in Genetics*, 21: 616–623.

Goldberg AD, Allis CD, Bernstein E (2007) Epigenetics: A landscape takes shape. *Cell*, 128: 635–638.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng QD, Chen ZH, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29: 644–652.

Greisman HA, Pabo CO (1997) A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. *Science*, 275: 657–661.

Gross ML, Schneider JF, Moav N, Moav B, Alvarez C, Myster SH, Liu ZJ, Hallerman EM, Hackett PB, Guise KS, Faras AJ, Kapuscinski AR (1992) Molecular analysis and growth evaluation of Northern Pike (Esox-Lucius) microinjected with growth-hormone genes. *Aquaculture*, 103: 253–273.

GuhaThakurta D (2006) Computational identification of transcriptional regulatory elements in DNA sequence. *Nucleic Acids Research*, 34: 3585–3598.

Guo X, Li Q, Wang QZ, and Kong LF (2011) Genetic mapping and QTL analysis of growth-related traits in the Pacific oyster. *Marine Biotechnology*, 14: 218–226.

Hackett PB, Clark KJ (2007) Functional genomics. In: Liu Z (ed) *Aquaculture Genome Technologies*. Blackwell Publishing Ltd, Oxford. 427–460. DOI: 10.1002/9780470277560.ch24.

Hale MC, McCormick CR, Jackson JR, Dewoody JA (2009) Next-generation pyrosequencing of gonad transcriptomes in the polyploid lake sturgeon (Acipenser fulvescens): the relative merits of normalization and rarefaction in gene discovery. *BMC Genomics*, 10: 203.

Han ES, Wu YM, McCarter R, Nelson JF, Richardson A, Hilsenbeck SG (2004) Reproducibility, sources of variability, pooling, and sample size: Important considerations for the design of high-density oligonucleotide array experiments. *Journals of Gerontology Series a-Biological Sciences and Medical Sciences*, 59: 306–315.

Hannon GJ (2002) RNA interference. *Nature*, 418: 244–251.

Hansen BG, Halkier BA, Kliebenstein DJ (2008) Identifying the molecular basis of QTLs: eQTLs add a new dimension. *Trends in Plant Science*, 13: 72–77.

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

He K, Chen Z, Ma Y, Pan Y (2011) Identification of high-copper-responsive target pathways in Atp7b knockout mouse liver by GSEA on microarray data sets. *Mammalian Genome*, 22: 703–713.

Hegedus Z, Zakrzewska A, Agoston VC, Ordas A, Racz P, Mink M, Spaink HP, Meijer AH (2009) Deep sequencing of the zebrafish transcriptome response to mycobacterium infection. *Molecular Immunology*, 46: 2918–2930.

Heintzman ND, Stuart RK, Hon G, Fu YT, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu CX, Ching KA, Wang W, Weng ZP, Green RD, Crawford GE, Ren B (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genetics*, 39: 311–318.

Henikoff S (2008) Nucleosome destabilization in the epigenetic regulation of gene expression. *Nature Reviews Genetics*, 9: 15–26.

Hieter P, Boguski M (1997) Functional genomics: It's all how you read it. *Science*, 278: 601–602.

Hirayama M, Ahsan MN, Mitani H, Watabe S (2008) CYR61 is a novel gene associated

with temperature-dependent changes in fish metabolism as revealed by cDNA microarray analysis on a medaka *Oryzias latipes* cell line. *Journal of Cellular Biochemistry*, 104: 1297–1310.

Holliday R (2006) Epigenetics: a historical overview. *Epigenetics*, 1: 76–80.

Horak CE, Snyder M (2002) ChIP-chip: a genomic approach for identifying transcription factor binding sites. *Methods in Enzymology*, 350: 469–483.

Hou R, Bao Z, Wang S, Su H, Li Y, Du H, Hu J, Hu X (2011) Transcriptome sequencing and de novo analysis for Yesso scallop (*Patinopecten yessoensis*) using 454 GS FLX. *PLoS ONE*, 6: e21560.

Houston RD, Haley CS, Hamilton A, Guy DR, Mota-Velasco JC, Gheyas AA, Tinch AE, Taggart JB, Bron JE, Starkey WG, McAndrew BJ, Verner-Jeffreys DW, Paley RK, Rimmer GSE, Tew IJ, Bishop SC (2010) The susceptibility of Atlantic salmon fry to freshwater infectious pancreatic necrosis is largely explained by a major QTL. *Heredity*, 105: 318–327.

Houston RD, Haley CS, Hamilton A, Guyt DR, Tinch AE, Taggart JB, McAndrew BJ, Bishop SC (2008) Major quantitative trait loci affect resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*). *Genetics*, 178: 1109–1115.

Hu M, Yu JD, Taylor JMG, Chinnaiyan AM, Qin ZHS (2010) On the detection and refinement of transcription factor binding sites using ChIP-Seq data. *Nucleic Acids Research*, 38: 2154–2167.

Huang DW, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 4: 44–57.

Huang WT, Hsieh JC, Chiou MJ, Chen JY, Wu JL, Kuo CM (2008) Application of RNAi technology to the inhibition of zebrafish GtH alpha, FSH beta, and LH beta expression and to functional analyses. *Zoological Science*, 25: 614–621.

Hubner N, Wallace CA, Zimdahl H, Petretto E, Schulz H, Maciver F, Mueller M, Hummel O, Monti J, Zidek V, Musilova A, Kren V, Causton H, Game L, Born G, Schmidt S, Muller A, Cook SA, Kurtz TW, Whittaker J, Pravenec M, Aitman TJ (2005) Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. *Nature Genetics*, 37: 243–253.

Hurst LD, Pal C, Lercher MJ (2004) The evolutionary dynamics of eukaryotic gene order. *Nature Reviews Genetics*, 5: 299–310.

Hurt JA, Thibodeau SA, Hirsh AS, Pabo CO, Joung JK (2003) Highly specific zinc finger proteins obtained by directed domain shuffling and cell-based selection. *Proceedings of the National Academy of Sciences USA*, 100: 12271–12276.

Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, 4: 249–264.

Isalan M, Choo Y, Klug A (1997) Synergy between adjacent zinc fingers in sequence-specific DNA recognition. *Proceedings of the National Academy of Sciences USA*, 94: 5617–5621.

Isalan M, Klug A, Choo Y (2001) A rapid, generally applicable method to engineer zinc fingers illustrated by targeting the HIV-1 promoter. *Nature Biotechnology*, 19: 656–660.

Iseli C, Jongeneel CV, Bucher P (1999) ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. *Proceedings of International Conference on Intelligent Systems for Molecular Biology*, pp. 138–148.

Jackson TR, Ferguson MM, Danzmann RG, Fishback AG, Ihssen PE, O'Connell M, Crease TJ (1998) Identification of two QTL influencing upper temperature tolerance in three rainbow trout (*Oncorhynchus mykiss*) half-sib families. *Heredity*, 80: 143–151.

Jansen RC, Nap JP (2001) Genetical genomics: the added value from segregation. *Trends in Genetics*, 17: 388–391.

Johansen SD, Karlsen BO, Furmanek T, Andreassen M, Jorgensen TE, Bizuayehu TT, Breines R, Emblem A, Kettunen P, Luukko K, Edvardsen RB, Nordeide JT, Coucheron DH, Moum T (2011) RNA deep sequencing of the Atlantic cod transcriptome. *Comparative biochemistry and physiology. Part D, Genomics & proteomics*, 6: 18–22.

Johnson DS, Mortazavi A, Myers RM, Wold B (2007) Genome-wide mapping of in vivo protein-DNA interactions. *Science*, 316: 1497–1502.

Jones NC, Pevzner PA (2006) Comparative genomics reveals unusually long motifs in mammalian genomes. *Bioinformatics*, 22: E236–E242.

Jothi R, Cuddapah S, Barski A, Cui K, Zhao K (2008) Genome-wide identification of in vivo protein-DNA binding sites from ChIP-Seq data. *Nucleic Acids Research*, 36: 5221–5231.

Ju ZL, Wells MC, Heater SJ, Walter RB (2007) Multiple tissue gene expression analyses in Japanese medaka (*Oryzias latipes*) exposed to hypoxia. *Comparative Biochemistry and Physiology C-Toxicology & Pharmacology*, 145: 134–144.

Karoonuthaisiri N, Sittikankeaw K, Preechaphol R, Kalachikov S, Wongsurawat T, Uawisetwathana U, Russo JJ, Ju J, Klinbunga S, Kirtikara K (2009) ReproArray(GTS): a cDNA microarray for identification of reproduction-related genes in the giant tiger shrimp Penaeus monodon and characterization of a novel nuclear autoantigenic sperm protein (NASP) gene. *Comp Biochem Physiol Part D Genomics Proteomics*, 4: 90–99.

Kassahn KS, Caley MJ, Ward AC, Connolly AR, Stone G, Crozier RH (2007) Heterologous microarray experiments used to identify the early gene response to heat stress in a coral reef fish. *Molecular Ecology*, 16: 1749–1763.

Kelly A, Hurlstone AF (2011) The use of RNAi technologies for gene knockdown in zebrafish. *Brief Funct Genomics*, 10: 189–196.

Keshet I, Schlesinger Y, Farkash S, Rand E, Hecht M, Segal E, Pikarski E, Young RA, Niveleau A, Cedar H, Simon I (2006) Evidence for an instructive mechanism of de novo methylation in cancer cells. *Nature Genetics*, 38: 149–153.

Kim JY, Cha DY, Park JH, Chang JB, Lee SH, Lee WJ (2006) Genome wide identification of parturition associated genes using cDNA microarray in the placenta of pregnant women between with labor and without labor. *American Journal of Obstetrics and Gynecology*, 195: S101–S101.

Kim S, Lee MJ, Kim H, Kang M, Kim JS (2011a) Preassembled zinc-finger arrays for rapid construction of ZFNs. *Nature Methods*, 8: 7.

Kim SW, Rai D, Aguiar R (2011b) Gene-set enrichment analysis unveils the mechanism for the phosphodiesterase 4B control of glucocorticoid response in B-cell lymphoma. *Clinical Cancer Research*, 17: 6723–6732.

Kim YG, Cha J, Chandrasegaran S (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences of the United States of America*, 93: 1156–1160.

Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh J (2005) Complement factor H polymorphism in age-related macular degeneration. *Science*, 308: 385–389.

Koch A, Hrychyk A, Hartmann W, Waha A, Mikeska T, Schuller U, Sorensen N, Berthold F, Goodyer CG, Wiestler OD, Birchmeier W, Behrens J, Pietsch T (2007) Mutations of the Wnt antagonist AXIN2 (Conductin) result in TCF-dependent transcription in medulloblastomas. *International Journal of Cancer*, 121: 284–291.

Kondo H, Suga R, Suda S, Nozaki R, Hirono I, Nagasaka R, Kaneko G, Ushio H, Watabe S (2011) EST analysis on adipose tissue of rainbow trout *Oncorhynchus mykiss* and tissue distribution of adiponectin. *Gene*, 485: 40–45.

Koop BF, von Schalburg KR, Leong J, Walker N, Lieph R, Cooper GA, Robb A, Beetz-Sargent M, Holt RA, Moore R, Brahmbhatt S, Rosner J, Rexroad CE 3rd, McGowan CR, Davidson WS. 2008. A salmonid EST genomic study: genes, duplications, phylogeny and microarrays. *BMC Genomics*, 9: 545.

Kuttner E, Moghadam HK, Skulason S, Danzmann RG, Ferguson MM (2011) Genetic architecture of body weight, condition factor and age of sexual maturation in Icelandic Arctic charr (*Salvelinus alpinus*). *Molecular Genetics and Genomics*, 286: 67–79.

Laird PW (2010) Principles and challenges of genomewide DNA methylation analysis. *Nature Reviews Genetics*, 11: 191–203.

Lan H, Chen M, Flowers JB, Yandell BS, Stapleton DS, Mata CM, Mui ETK, Flowers MT, Schueler KL, Manly KF, Williams RW, Kendziorski C, Attie AD (2006) Combined expression trait correlations and expression quantitative trait locus mapping. *PLoS Genetics*, 2: 51–61.

Le Bras Y, Dechamp N, Krieg F, Filangi O, Guyomard R, Boussaha M, Bovenhuis H, Pottinger TG, Prunet P, Le Roy P, Quillet E (2011) Detection of QTL with effects on osmoregulation capacities in the rainbow trout (*Oncorhynchus mykiss*). *BMC Genetics*, 12: 46.

Leder EH, Danzmann RG, Ferguson MM (2006) The candidate gene, Clock, localizes to a strong spawning time quantitative trait locus region in rainbow trout. *Journal of Heredity*, 97: 74–80.

Lee JS, Chu IS, Mikaelyan A, Calvisi DF, Heo J, Reddy JK, Thorgeirsson SS (2004) Application of comparative functional genomics to identify best-fit mouse models to study human cancer. *Nature Genetics*, 36: 1306–1311.

Lee SI, Dudley AM, Drubin D, Silver PA, Krogan NJ, Pe'er D, Koller D (2009) Learning a Prior on Regulatory Potential from eQTL Data. *PLoS Genetics*, 5: e1000358.

Leelatanawit R, Uawisetwathana U, Klinbunga S, Karoonuthaisiri N (2011) A cDNA microarray, UniShrimpChip, for identification of genes relevant to testicular development in the black tiger shrimp (*Penaeus monodon*). *BMC Molecular Biology*, 12: 15.

Lessman CA (2011) The developing zebrafish (*Danio rerio*): A vertebrate model for high-throughput screening of chemical libraries. *Birth Defects Res C Embryo Today*, 93: 268–280.

Lewis JM, Hori TS, Rise ML, Walsh PJ, Currie S (2010) Transcriptome responses to heat stress in the nucleated red blood cells of the rainbow trout (*Oncorhynchus mykiss*). *Physiological Genomics*, 42: 361–373.

Li P, Peatman E, Wang SL, Feng JN, He CB, Baoprasertkul P, Xu P, Kucuktas H, Nandi S, Somridhivej B, Serapion J, Simmons M, Turan C, Liu L, Muir W, Dunham R, Brady Y, Grizzle J, Liu ZJ (2007) Towards the ictalurid catfish transcriptome: generation and analysis of 31,215 catfish ESTs. *BMC Genomics*, 8: 177.

Li RW, Waldbieser GC (2006) Production and utilization of a high-density oligonucleotide microarray in channel catfish, *Ictalurus punctatus*. *BMC Genomics*, 7: 134.

Liu CL, Kaplan T, Kim M, Buratowski S, Schreiber SL, Friedman N, Rando OJ (2005) Single-nucleosome mapping of histone modifications in S-cerevisiae. *PLoS Biology*, 3: 1753–1769.

Liu H, Jiravanichpaisal P, Soderhall I, Cerenius L, Soderhall K (2006) Antilipopolysaccharide factor interferes with white spot syndrome virus replication in vitro and in vivo in the crayfish *Pacifastacus leniusculus*. *Journal of Virology*, 80: 10365–10371.

Liu N, Caderas G, Deillon C, Hoffmann S, Klauser S, Cui T, Gutte B (2001) Fusion proteins from artificial and natural structural modules. *Current Protein and Peptide Science*, 2: 107–121.

Liu SK, Zhou ZC, Lu JG, Sun FY, Wang SL, Liu H, Jiang YL, Kucuktas H, Kaltenboeck L, Peatman E, Liu ZJ (2011) Generation of genome-scale gene-associated SNPs in catfish for the construction of a high-density SNP array. *BMC Genomics*, 12: 53.

Liu XD, Liu X, Zhang GF (2007) Identification of quantitative trait loci for growth-related traits in the Pacific abalone Haliotis discus hannai Ino. *Aquaculture Research*, 38: 789–797.

Lo J, Lee SC, Xu M, Liu F, Ruan H, Eun A, He YW, Ma WP, Wang WF, Wen ZL, Peng JR (2003) 15,000 unique zebrafish EST clusters and their future use in microarray for profiling gene expression patterns during embryogenesis. *Genome Research*, 13: 455–466.

Loukovitis D, Sarropoulou E, Tsigenopoulos CS, Batargias C, Magoulas A, Apostolidis AP, Chatziplis D, Kotoulas G (2011) Quantitative trait loci involved in sex determination and body growth in the Gilthead sea bream (*Sparus aurata* L.) through targeted genome scan. *PLoS ONE*, 6: e16599.

Louro B, Passos ALS, Souche EL, Tsigenopoulos C, Beck A, Lagnel J, Bonhomme F, Cancela L, Cerda J, Clark MS, Lubzens E, Magoulas A, Planas JV, Volckaert FAM, Reinhardt R, Canario AVM (2010) Gilthead sea bream (*Sparus auratus*) and European sea bass (*Dicentrarchus labrax*) expressed sequence tags: Characterization, tissue-specific expression and gene markers. *Marine Genomics*, 3: 179–191.

Lynch M, Walsh BJ (1998) *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, MA.

Maeder ML, Thibodeau-Beganny S, Osiak A, Wright DA, Anthony RM, Eichtinger M, Jiang T, Foley JE, Winfrey RJ, Townsend JA, Unger-Wallace E, Sander JD, Müller-Lerch F, Fu F, Pearlberg J, Göbel C, Dassie JP, Pruett-Miller SM, Porteus MH, Sgroi DC, Iafrate AJ, Dobbs D, McCray PB Jr, Cathomen T, Voytas DF, Joung JK (2008) Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. *Molecular Cell*, 31: 294–301.

Majewski J, Pastinen T (2011) The study of eQTL variations by RNA-seq: from SNPs to phenotypes. *Trends in Genetics*, 27: 72–79.

Mandell JG, Barbas CF 3rd (2006) Zinc Finger Tools: custom DNA-binding domains for transcription factors and nucleases. *Nucleic Acids Research*, 34: W516–523.

Marguerat S, Bahler J (2010) RNA-seq: from technology to biology. *Cellular and Molecular Life Sciences*, 67: 569–579.

Martin JA, Wang Z (2011) Next-generation transcriptome assembly. *Nature reviews Genetics*, 12: 671–682.

Massault C, Hellemans B, Louro B, Batargias C, Van Houdt JKJ, Canario A, Volckaert FAM,

Bovenhuis H, Haley C, de Koning DJ (2010) QTL for body weight, morphometric traits and stress response in European sea bass *Dicentrarchus labrax*. *Animal Genetics*, 41: 337–345.

Mathavan S, Lee SG, Mak A, Miller LD, Murthy KR, Govindarajan KR, Tong Y, Wu YL, Lam SH, Yang H, Ruan Y, Korzh V, Gong Z, Liu ET, Lufkin T (2005) Transcriptome analysis of zebrafish embryogenesis using microarrays. *PLoS Genetics*, 1: 260–276.

McClelland EK, Naish KA (2010) Quantitative trait locus analysis of hatch timing, weight, length and growth rate in coho salmon, *Oncorhynchus kisutch*. *Heredity*, 105: 562–573.

Mckusick VA (1989) The human-genome-organization—history, purposes, and membership. *Genomics*, 5: 385–387.

Meijer AH, Verbeek FJ, Salas-Vidal E, Corredor-Adamez M, Bussman J, van der Sar AM, Otto GW, Geisler R, Spaink HP (2005) Transcriptome profiling of adult zebrafish at the late stage of chronic tuberculosis due to *Mycobacterium marinum* infection. *Molecular Immunology*, 42: 1185–1203.

Meng X, Noyes MB, Zhu LJ, Lawson ND, Wolfe SA (2008) Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nature Biotechnology*, 26: 695–701.

Mikkelsen TS, Ku MC, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, Lee W, Mendenhall E, O'Donovan A, Presser A, Russ C, Xie XH, Meissner A, Wernig M, Jaenisch R, Nusbaum C, Lander ES, Bernstein BE (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, 448: 553–560.

Miller MB, Tang YW (2009) Basic concepts of microarrays and potential applications in clinical microbiology. *Clinical Microbiology Reviews*, 22: 611–633.

Min XJ, Butler G, Storms R, Tsang A (2005) OrfPredictor: predicting protein-coding regions in EST-derived sequences. *Nucleic Acids Research*, 33: W677–W680.

Moen T, Baranski M, Sonesson AK, Kjoglum S (2009) Confirmation and fine-mapping of a major QTL for resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*): population-level associations between markers and trait. *BMC Genomics*, 10: 368.

Moen T, Sonesson AK, Hayes B, Lien S, Munck H, Meuwissen THE (2007) Mapping of a quantitative trait locus for resistance against infectious salmon anaemia in Atlantic salmon (*Salmo Salar*): comparing survival analysis with analysis on affected/resistant data. *BMC Genetics*, 8: 53.

Moghadam HK, Poissant J, Fotherby H, Haidle L, Ferguson MM, Danzmann RG (2007) Quantitative trait loci for body weight, condition factor and age at sexual maturation in Arctic charr (*Salvelinus alpinus*): comparative analysis with rainbow trout (*Oncorhynchus mykiss*) and Atlantic salmon (*Salmo salar*). *Molecular Genetics and Genomics*, 277: 647–661.

Montgomery SB, Sammeth M, Gutierrez-Arcelus M, Lach RP, Ingle C, Nisbett J, Guigo R, Dermitzakis ET (2010) Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature*, 464: 773-777.

Morley M, Molony CM, Weber TM, Devlin JL, Ewens KG, Spielman RS, Cheung VG (2004) Genetic analysis of genome-wide variation in human gene expression. *Nature*, 430: 743–747.

Morozova O, Hirst M, Marra MA (2009) Applications of new sequencing technologies for transcriptome analysis. *Annual Review Genomics Human Genetics*, 10: 135–151.

Mu Y, Ding F, Cui P, Ao J, Hu S, Chen X (2010) Transcriptome and expression profiling analysis revealed changes of multiple signaling pathways involved in immunity in the large yellow croaker during *Aeromonas hydrophila* infection. *BMC Genomics*, 11: 506.

Nagalakshmi U, Waern K, Snyder M (2010) RNA-Seq: a method for comprehensive transcriptome analysis. *Current Protocol of Molecular Biology*, 11: 11–13.

Nakaya HI, Amaral PP, Louro R, Lopes A, Fachel AA, Moreira YB, El-Jundi TA, da Silva AM, Reis EM, Verjovski-Almeida S (2007) Genome mapping and expression analyses of human intronic noncoding RNAs reveal tissue-specific patterns and enrichment in genes related to regulation of transcription. *Genome Biology*, 8: R43.

Ng HH, Bird A (1999) DNA methylation and chromatin modification. *Current Opinion in Genetics & Development*, 9: 158–163.

Nichols KM, Edo AF, Wheeler PA, Thorgaard GH (2008) The genetic basis of smoltification-related traits in *Oncorhynchus mykissi*. *Genetics*, 179: 1559–1575.

Norman JD, Danzmann RG, Glebe B, Ferguson MM (2011) The genetic basis of salinity tolerance traits in Arctic charr (*Salvelinus alpinus*). *BMC Genetics*, 12: 81.

Nuwaysir EF, Huang W, Albert TJ, Singh J, Nuwaysir K, Pitas A, Richmond T, Gorski T, Berg JP, Ballin J, McCormick M, Norton J, Pollock T, Sumwalt T, Butcher L, Porter D, Molla M, Hall C, Blattner F, Sussman MR, Wallace RL, Cerrina F, Green RD (2002) Gene expression analysis using oligonucleotide arrays produced by maskless photolithography. *Genome Research*, 12: 1749–1755.

Olsvik PA, Lie KK, Mykkeltvedt E, Samuelsen OB, Petersen K, Stavrum AK, Lunestad BT (2008) Pharmacokinetics and transcriptional effects of the anti-salmon lice drug emamectin benzoate in Atlantic salmon (*Salmo salar* L.). *BMC Pharmacology*, 8: 16.

O'Malley KG, Sakamoto T, Danzmann RG, Ferguson MM (2003) Quantitative trait loci for spawning date and body weight in rainbow trout: Testing for conserved effects across ancestrally duplicated chromosomes. *Journal of Heredity*, 94: 273–284.

Ordas A, Hegedus Z, Henkel CV, Stockhammer OW, Butler D, Jansen HJ, Racz P, Mink M, Spaink HP, Meijer AH (2011) Deep sequencing of the innate immune transcriptomic response of zebrafish embryos to Salmonella infection. *Fish & Shellfish Immunology*, 31: 716–724.

Park PJ (2009) ChIP-seq: advantages and challenges of a maturing technology. *Nature Reviews Genetics*, 10: 669–680.

Peal DS, Mills RW, Lynch SN, Mosley JM, Lim E, Ellinor PT, January CT, Peterson RT, Milan DJ (2011) Novel chemical suppressors of long QT syndrome identified by an in vivo functional screen. *Circulation*, 123: 23–30.

Peatman E, Baoprasertkul P, Terhune J, Xu P, Nandi S, Kucuktas H, Li P, Wang SL, Somridhivej B, Dunham R, Liu ZJ (2007) Expression analysis of the acute phase response in channel catfish (*Ictalurus punctatus*) after infection with a Gram-negative bacterium. *Developmental and Comparative Immunology*, 31: 1183–1196.

Peatman E, Liu Z (2007) Microarray fundamentals: Basic principles and application in aquaculture. In: Liu Z (ed) *Aquaculture Genome Technologies*. Blackwell Publishing Ltd., Oxford. DOI: 10.1002/9780470277560.ch21.

Perry GML, Danzmann RG, Ferguson MM, Gibson JP (2001) Quantitative trait loci for upper thermal tolerance in outbred strains of rainbow trout (*Oncorhynchus mykiss*). *Heredity*, 86: 333–341.

Perry GML, Ferguson MM, Sakamoto T, Danzmann RG (2005) Sex-linked quantitative trait loci for thermotolerance and length in the rainbow trout. *Journal of Heredity*, 96: 97–107.

Pickrell JK, Marioni JC, Pai AA, Degner JF, Engelhardt BE, Nkadori E, Veyrieras JB, Stephens M, Gilad Y, Pritchard JK (2010) Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature*, 464: 768–772.

Pokholok DK, Harbison CT, Levine S, Cole M, Hannett NM, Lee TI, Bell GW, Walker K, Rolfe PA, Herbolsheimer E, Zeitlinger J, Lewitter F, Gifford DK, Young RA (2005) Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell*, 122: 517–527.

Pruett-Miller SM, Connelly JP, Maeder ML, Joung JK, Porteus MH. 2008. Comparison of zinc finger nucleases for use in gene targeting in mammalian cells. *Molecular Therapy*, 16: 707–717.

Purcell MK, Marjara IS, Batts W, Kurath G, Hansen JD (2011) Transcriptome analysis of rainbow trout infected with high and low virulence strains of infectious hematopoietic necrosis virus. *Fish & Shellfish Immunology*, 30: 84–93.

Rahman MA, Mak R, Ayad H, Smith A, Maclean N (1998) Expression of a novel piscine growth hormone gene results in growth enhancement in transgenic tilapia (*Oreochromis niloticus*). *Transgenic Research*, 7: 357–369.

Rahman MA, Ronyai A, Engidaw BZ, Jauncey K, Hwang GL, Smith A, Roderick E, Penman D, Varadi L, Maclean N (2001) Growth and nutritional trials on transgenic Nile tilapia containing an exogenous fish growth hormone gene. *Journal of Fish Biology*, 59: 62–78.

Ramirez CL, et al. (2008) Unexpected failure rates for modular assembly of engineered zinc-fingers. *Nature Methods*, 5: 374–375.

Reid DP, Szanto A, Glebe B, Danzmann RG, Ferguson MM (2005) QTL for body weight and condition factor in Atlantic salmon (*Salmo salar*): comparative analysis with rainbow trout (*Oncorhynchus mykiss*) and Arctic charr (*Salvelinus alpinus*). *Heredity*, 94: 166–172.

Reinartz J, Bruyns E, Lin JZ, Burcham T, Brenner S, Bowen B, Kramer M, Woychik R (2002) Massively parallel signature sequencing (MPSS) as a tool for in-depth quantitative gene expression

profiling in all organisms. *Brief Functional Genomics & Proteomics*, 1: 95–104.

Ren J, Jenkins TC, Chaires JB (2000) Energetics of DNA intercalation reactions. *Biochemistry*, 39: 8439–8447.

Renaut S, Nolte AW, Bernatchez L (2010) Mining transcriptome sequences towards identifying adaptive single nucleotide polymorphisms in lake whitefish species pairs (*Coregonus* spp. *Salmonidae*). *Molecular Ecology*, 19(Suppl 1): 115–131.

Reyon D, Kirkpatrick JR, Sander JD, Zhang F, Voytas DF, Joung JK, Dobbs D, Coffman CR (2011) ZFNGenome: a comprehensive resource for locating zinc finger nuclease target sites in model organisms. *BMC Genomics*, 12: 83.

Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao YJ, Zeng T, Euskirchen G, Bernier B, Varhol R, Delaney A, Thiessen N, Griffith OL, He A, Marra M, Snyder M, Jones S (2007) Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nature Methods*, 4: 651–657.

Robison BD, Wheeler PA, Sundin K, Sikka P, Thorgaard GH (2001) Composite interval mapping reveals a major locus influencing embryonic development rate in rainbow trout (*Oncorhynchus mykiss*). *Journal of Heredity*, 92: 16–22.

Robyr D, Suka Y, Xenarios I, Kurdistani SK, Wang A, Suka N, Grunstein M (2002) Microarray deacetylation maps determine genome-wide functions for yeast histone deacetylases. *Cell*, 109: 437–446.

Roh TY, Wei G, Farrell CM, Zhao K (2007) Genome-wide prediction of conserved and non-conserved enhancers by histone acetylation patterns. *Genome Research*, 17: 74–81.

Ronning CM, Stegalkina SS, Ascenzi RA, Bougri O, Hart AL, Utterbach TR, Vanaken SE, Riedmuller SB, White JA, Cho J, Pertea GM, Lee Y, Karamycheva S, Sultana R, Tsai J, Quackenbush J, Griffiths HM, Restrepo S, Smart CD, Fry WE, Van Der Hoeven R, Tanksley S, Zhang P, Jin H, Yamamoto ML, Baker BJ, Buell CR (2003) Comparative analyses of potato expressed sequence tag libraries. *Plant Physiology*, 131: 419–429.

Rosa GJ, de Leon N, Rosa AJ (2006) Review of microarray experimental design strategies for genetical genomics studies. *Physiological Genomics*, 28: 15–23.

Rosel TD, Hung LH, Medenbach J, Donde K, Starke S, Benes V, Ratsch G, Bindereif A (2011) RNA-Seq analysis in mutant zebrafish reveals role of U1C protein in alternative splicing regulation. *The EMBO journal*, 30: 1965–1976.

Rowen L, Mahairas G, Hood L (1997) Sequencing the human genome. *Science*, 278: 605–607.

Royce TE, Rozowsky JS, Gerstein MB (2007) Assessing the need for sequence-based normalization in tiling microarray experiments. *Bioinformatics*, 23: 988–997.

Salem M, Rexroad CE 3rd, Wang J, Thorgaard GH, Yao J (2010) Characterization of the rainbow trout transcriptome using Sanger and 454-pyrosequencing approaches. *BMC Genomics*, 11: 564.

Sanchez BC, Carter B, Hammers HR, Sepulveda MS (2011) Transcriptional response of hepatic largemouth bass (*Micropterus salmoides*) mRNA upon exposure to environmental contaminants. *Journal of Applied Toxicology*, 31: 108–116.

Sander JD, Dahlborg EJ, Goodwin MJ, Cade L, Zhang F, Cifuentes D, Curtin SJ, Blackburn JS, Thibodeau-Beganny S, Qi Y, Pierick CJ, Hoffman E, Maeder ML, Khayter C, Reyon D, Dobbs D, Langenau DM, Stupar RM, Giraldez AJ, Voytas DF, Peterson RT, Yeh JR, Joung JK (2011). Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nature Methods*, 8: 67–69.

Sander JD, Reyon D, Maeder ML, Foley JE, Thibodeau-Beganny S, Li X, Regan MR, Dahlborg EJ, Goodwin MJ, Fu F, Voytas DF, Joung JK, Dobbs D (2010a). Predicting success of oligomerized pool engineering (OPEN) for zinc finger target site sequences. *BMC Bioinformatics*, 11: 543.

Sander JD, Maeder ML, Reyon D, Voytas DF, Joung JK, Dobbs D (2010b). ZiFiT (zinc finger targeter): an updated zinc finger engineering tool. *Nucleic Acids Research*, 38: W462–W468.

Sander JD, Zaback P, Joung JK, Voytas DF, Dobbs D (2007) Zinc finger targeter (ZiFiT): an engineered zinc finger/target site design tool. *Nucleic Acids Research*, 35: W599–W605.

Sauvage C, Boudry P, de Koning DJ, Haley CS, Heurtebise S, Lapegue S (2010) QTL for resistance to summer mortality and OsHV-1 load in the Pacific oyster (*Crassostrea gigas*). *Animal Genetics*, 41: 390–399.

Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, Ruff TG, Milligan SB, Lamb JR, Cavet G, Linsley PS, Mao M, Stoughton RB, Friend SH (2003) Genetics of gene expression surveyed in maize, mouse and man. *Nature*, 422: 297–302.

Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, 270: 467–470.

Schena M, Shalon D, Heller R, Chai A, Brown PO, Davis RW (1996) Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proceedings of the National Academy of Sciences of the United States of America*, 93: 10614–10619.

Schmidt D, Wilson MD, Ballester B, Schwalie PC, Brown GD, Marshall A, Kutter C, Watt S, Martinez-Jimenez CP, Mackay S, Talianidis I, Flicek P, Odom DT (2010) Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science*, 328: 1036–1040.

Schones DE, Zhao K (2008) Genome-wide approaches to studying chromatin modifications. *Nature Reviews Genetics*, 9: 179–191.

Schubeler D, MacAlpine DM, Scalzo D, Wirbelauer C, Kooperberg C, van Leeuwen F, Gottschling DE, O'Neill LP, Turner BM, Delrow J, Bell SP, Groudine M (2004) The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. *Genes & Development*, 18: 1263–1271.

Segal DJ, Beerli RR, Blancafort P, Dreier B, Effertz K, Huber A, Koksch B, Lund CV, Magnenat L, Valente D, Barbas CF 3rd (2003) Evaluation of a modular strategy for the construction of novel polydactyl zinc finger DNA-binding proteins. *Biochemistry*, 42: 2137–2148.

Segal DJ, Barbas CF 3rd (2001) Custom DNA-binding proteins come of age: polydactyl zinc-finger proteins. *Current Opinion in Biotechnology*, 12: 632–637.

Shen LL, Waterland RA (2007) Methods of DNA methylation analysis. *Current Opinion in Clinical Nutrition and Metabolic Care*, 10: 576–581.

Shirak A, Seroussi E, Cnaani A, Howe AE, Domokhovsky R, Zilberman N, Kocher TD, Hulata G, Ron M (2006) Amh and Dmrta2 genes map to tilapia (*Oreochromis* spp.) linkage group 23 within quantitative trait locus regions for sex determination. *Genetics*, 174: 1573–1581.

Shukla VK, Doyon Y, Miller JC, DeKelver RC, Moehle EA, Worden SE, Mitchell JC, Arnold NL, Gopalan S, Meng X, Choi VM, Rock JM, Wu YY, Katibah GE, Zhifang G, McCaskill D, Simpson MA, Blakeslee B, Greenwalt SA, Butler HJ, Hinkley SJ, Zhang L, Rebar EJ, Gregory PD, Urnov FD (2009) Precise genome modification in the crop species Zea mays using zinc-finger nucleases. *Nature*, 459: 437–441.

Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S, Balkau B, Heude B, Charpentier G, Hudson TJ, Montpetit A, Pshezhetsky AV, Prentki M, Posner BI, Balding DJ, Meyre D, Polychronakos C, Froguel P (2007) A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature*, 445: 881–885.

Soares MB, Bonaldo MD, Jelene P, Su L, Lawton L, Efstratiadis A (1994) Construction and characterization of a normalized Cdna library. *Proceedings of the National Academy of Sciences of the United States of America*, 91: 9228–9232.

Solomon MJ, Larsen PL, Varshavsky A (1988) Mapping protein-DNA interactions in vivo with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene. *Cell*, 53: 937–947.

Stockhammer OW, Rauwerda H, Wittink FR, Breit TM, Meijer AH, Spaink HP (2010) Transcriptome analysis of Traf6 function in the innate immune response of zebrafish embryos. *Molecular Immunology*, 48: 179–190.

Suzuki MM, Bird A (2008) DNA methylation landscapes: provocative insights from epigenomics. *Nature Reviews Genetics*, 9: 465–476.

Tate PH, Bird AP (1993) Effects of DNA methylation on DNA-binding proteins and gene expression. *Current Opinion in Genetics & Development*, 3: 226–231.

The Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3000 shared controls. *Nature*, 447: 661–678.

Thomas MA, Yang L, Carter BJ, Klaper RD (2011) Gene set enrichment analysis of microarray data from *Pimephales promelas* (Rafinesque), a non-mammalian model organism. *BMC Genomics*, 12: 66.

Tompa M, Li N, Bailey TL, Church GM, De Moor B, Eskin E, Favorov AV, Frith MC, Fu Y, Kent WJ, Makeev VJ, Mironov AA, Noble WS, Pavesi

G, Pesole G, Regnier M, Simonis N, Sinha S, Thijs G, van Helden J, Vandenbogaert M, Weng Z, Workman C, Ye C, Zhu Z (2005) Assessing computational tools for the discovery of transcription factor binding sites. *Nature Biotechnology*, 23: 137–144.

Ton C, Stamatiou D, Dzau VJ, Liew CC (2002) Construction of a zebrafish cDNA microarray: gene expression profiling of the zebrafish during development. *Biochemical and Biophysical Research Communications*, 296: 1134–1142.

Valouev A, Johnson DS, Sundquist A, Medina C, Anton E, Batzoglou S, Myers RM, Sidow A (2008) Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. *Nature Methods*, 5: 829–834.

Vasemagi A, Gross R, Palm D, Paaver T, Primmer CR (2010) Discovery and application of insertion-deletion (INDEL) polymorphisms for QTL mapping of early life-history traits in Atlantic salmon. *BMC Genomics*, 11: 156.

Velculescu VE, Zhang L, Vogelstein B, Kinzler KW (1995) Serial analysis of gene expression. *Science*, 270: 484–487.

Venter JC, Smith HO, Hood L (1996) A new strategy for genome sequencing. *Nature*, 381: 364–366.

Vesterlund L, Jiao H, Unneberg P, Hovatta O, Kere J (2011) The zebrafish transcriptome during early development. *BMC Developmental Biology*, 11: 30.

Vogt G (2008) The marbled crayfish: a new model organism for research on development, epigenetics and evolutionary biology. *Journal of Zoology*, 276: 1–13.

von Schalburg KR, Rise ML, Cooper GA, Brown GD, Gibbs AR, Nelson CC, Davidson WS, Koop BF (2005) Fish and chips: Various methodologies demonstrate utility of a 16,006-gene salmonid microarray. *BMC Genomics*, 6: 126.

Wang CM, Bai ZY, He XP, Lin G, Xia JH, Sun F, Lo LC, Feng F, Zhu ZY, Yue GH (2011) A high-resolution linkage map for comparative genome analysis and QTL fine mapping in Asian seabass, *Lates calcarifer*. *BMC Genomics*, 12: 174.

Wang CM, Lo LC, Zhu ZY, Yue GH (2006) A genome scan for quantitative trait loci affecting growth-related traits in an F1 family of Asian seabass (*Lates calcarifer*). *BMC Genomics*, 7: 174.

Wang J, Williams RW, Manly KF (2003) WebQTL: web-based complex trait analysis. *Neuroinformatics*, 1: 299–308.

Wang S, Peatman E, Liu H, Bushek D, Ford SE, Kucuktas H, Quilang J, Li P, Wallace R, Wang Y, Guo X, Liu Z (2010) Microarray analysis of gene expression in eastern oyster (*Crassostrea virginica*) reveals a novel combination of antimicrobial and oxidative stress host responses after dermo (*Perkinsus marinus*) challenge. *Fish & Shellfish Immunology*, 29: 921–929.

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

Wargelius A, Ellingsen S, Fjose A (1999) Double-stranded RNA induces specific developmental defects in zebrafish embryos. *Biochemical and Biophysical Research Communications*, 263: 156–161.

Wayne ML, McIntyre LM (2002) Combining mapping and arraying: An approach to candidate gene identification. *Proceedings of the National Academy of Sciences of the United States of America*, 99: 14903–14906.

Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL, Schubeler D (2005) Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nature Genetics*, 37: 853–862.

Weber M, Hellmann I, Stadler MB, Ramos L, Paabo S, Rebhan M, Schubeler D (2007) Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature Genetics*, 39: 457–466.

Wederell ED, Bilenky M, Cullum R, Thiessen N, Dagpinar M, Delaney A, Varhol R, Zhao YJ, Zeng T, Bernier B, Ingham M, Hirst M, Robertson G, Marra MA, Jones S, Hoodless PA (2008) Global analysis of in vivo Foxa2-binding sites in mouse adult liver using massively parallel sequencing. *Nucleic Acids Research*, 36: 4549–4564.

Wongsurawat T, Leelatanawit R, Thamniemdee N, Uawisetwathana U, Karoonuthaisiri N, Menasveta P, Klinbunga S (2010) Identification of testis-relevant genes using in silico analysis from testis ESTs and cDNA microarray in the black tiger shrimp (*Penaeus monodon*). *BMC Molecular Biology*, 11: 55.

Wringe BF, Devlin RH, Ferguson MM, Moghadam HK, Sakhrani D, Danzmann RG (2010) Growth-related quantitative trait loci in domestic and wild rainbow trout (*Oncorhynchus mykiss*). *BMC Genetics*, 11:63.

Wu CC, Huang HC, Juan HF, Chen ST (2004) GeneNetwork: an interactive tool for reconstruction of genetic networks using microarray data. *Bioinformatics*, 20: 3691–3693.

Xiang CC, Chen YD (2000) cDNA microarray technology and its applications. *Biotechnology Advances*, 18: 35–46.

Xiang LX, He D, Dong WR, Zhang YW, Shao JZ (2010) Deep sequencing-based transcriptome profiling analysis of bacteria-challenged *Lateolabrax japonicus* reveals insight into the immune-relevant genes in marine fish. *BMC Genomics*, 11: 472.

Xie XH, Lu J, Kulbokas EJ, Golub TR, Mootha V, Lindblad-Toh K, Lander ES, Kellis M (2005) Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. *Nature*, 434: 338–345.

Yu ZN, Guo XM (2006) Identification and mapping of disease-resistance QTLs in the eastern oyster, *Crassostrea virginica* Gmelin. *Aquaculture*, 254: 160–170.

Zhan A, Hu J, Hu X, Hui M, Wang M, Peng W, Huang X, Wang S, Lu W, Sun C, Bao Z (2009) Construction of microsatellite-based linkage maps and identification of size-related quantitative trait loci for Zhikong scallop (*Chlamys farreri*). *Animal Genetics*, 40: 821–831.

Zhang XY, Yazaki J, Sundaresan A, Cokus S, Chan SWL, Chen HM, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE , Ecker JR (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in Arabidopsis. *Cell*, 126: 1189–1201.

Zhang Y, Xu P, Lu C, Kuang Y, Zhang X, Cao D, Li C, Chang Y, Hou N, Li H, Wang S, Sun X (2011) Genetic linkage mapping and analysis of muscle fiber-related QTLs in common carp (*Cyprinus carpio* L.). *Marine Biotechnology (NY)*, 13: 376–392.

Zheng Q, Wang XJ (2008) GOEAST: a web-based software toolkit for Gene Ontology enrichment analysis. *Nucleic Acids Research*, 36: W358–363.

Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S (2007) Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nature Genetics*, 39: 61–69.

Zou J, Maeder ML, Mali P, Pruett-Miller SM, Thibodeau-Beganny S, Chou BK, Chen G, Ye Z, Park IH, Daley GQ, Porteus MH, Joung JK, Cheng L (2009) Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. *Cell Stem Cell*, 5: 97–110.

Zou ZH, Zhang ZP, Wang YL, Han KH, Fu MJ, Lin P, Jia XW (2011) EST analysis on the gonad development related organs and microarray screen for differentially expressed genes in mature ovary and testis of *Scylla paramamosain*. *Comparative Biochemistry and Physiology D-Genomics & Proteomics*, 6: 150–157.