Spatio-temporal gene expression analysis from 3D in situ hybridization images

Welten, M.C.M.

Chapter 1

General introduction

1. Introduction

Understanding developmental processes requires a broad range of approaches and tools. These comprise biological (molecular-biological and genetic) as well as computational tools.

In this thesis, several developmental processes in both zebrafish and chicken embryos are investigated in two groups of case studies. The focus is on imaging spatial as well as temporal gene expression patterns during embryonic development, exploring *in situ* hybridization methods and computer assisted tools. Based on the use of these tools, a workflow has been developed in order to generate a large amount of biological data in a straightforward manner, thus allowing statistical analysis and pattern recognition.

Developmental biologists use a wide range of vertebrate model species such as zebrafish, clawed toad, mouse and chicken to elucidate developmental processes and as models to study human disorders. In recent years, a large amount of molecular data from these model systems has become available from developmental genetics and functional studies of genes involved in human development and disorders. This abundance of molecular data is generated with the mere goal to provide insight in developmental processes and - in some cases- evolution, and supplements to the existing knowledge. Now that all these molecular data are available, genes involved in developmental processes can be analyzed and compared across model species, to extrapolate experimental findings to other model systems and human.

2. Zebrafish as a model system

During the last decades, zebrafish has become an increasingly popular model system. In the late 1960s of the past century, George Streisinger (University of Oregon) started working with zebrafish as a model system. He had experienced its many advantages such as high fecundity, small size, short generation time, external fertilization, and numerous transparent embryos (Grunwald and Eisen, 2002). In the mid 1970s, Streisinger developed a technique to produce recessive mutations present in the maternal germ line, facilitating the analysis of mutants in zebrafish, a vertebrate model system (Streisinger et al.1981). In the mid-1980s, a research community was founded in Oregon by Streisinger's colleagues Kimmel, Westerfield and others, focusing on genetic and developmental studies in zebrafish (Grunwald and Eisen, 2002). In 1993, large screenings of embryonic zebrafish mutants were initiated in Tübingen, Germany (Haffter and Nüsslein-Volhardt, 1996) and Boston (Driever et al. 1996). These genetic screenings provided the possibility to analyze the molecular mechanisms underlying specific developmental processes. In recent years, large amounts of data from developmental marker genes have become available from functional genomics, clinical studies and molecular developmental research (Grunwald and Eisen 2002; Stern and Zon, 2003). These data are available from internet resources (e.g. www.zfin.org, http://cegs.stanford.edu/search.jsp) and include both micro array and spatial, i.e. *in situ* gene expression data.

In this thesis we will not further exploit the genetics of zebrafish, but rather analyze the spatial gene expression profiles in developmental time series of wild-type zebrafish.

3. Tools to visualize gene expression

Temporal gene expression patterns can be analyzed in several ways, such as RT-PCR, Northern blot, or immunolocalization of the proteins, i.e. the product of transcription. Recently, micro-RNAs (miRNAs) have been described as non-coding small RNAs that regulate expression of target mRNAs (reviewed by Ruvkun, 2001).

Nowadays, microarrays are a popular instrument to study gene expression (Lipshutz et al. 1995; Schena, 1996). These methods reveal gene expression profiles for a large number of genes at different time-points. In general, however, they are limited in providing spatial information of gene expression patterns during development. Tetko et al. (2006) describe spatio-temporal gene expression patterns in *Arabidopsis thaliana*. In this study, genes expressed in specific regions of the plant, e.g. root, inflorescence and leaves, are analyzed using microarrays. Though gene expression is extensively studied in a wide variety of plant organs as well as developmental stages, no true spatial – i.e. in situ - gene expression patterns are shown. Also, isolating specific organs from plant embryos is probably easier than from zebrafish embryos and early embryos of other animal models. In order to investigate spatial patterns of gene expression*, in situ* hybridization (ISH) is the most suitable tool. This method facilitates visualization of spatial characteristics of cell and tissue specific gene expression patterns (Wilkinson, 1998; Darnell et al. 2006). Moreover, large numbers of embryos can be hybridized simultaneously (Wilkinson, 1998) – though the amount of genes that are analyzed in microarrays rises to ten thousands. In zebrafish research ISH is in most cases applied to whole mount embryos, using digoxigenin-labeled antisense RNA probes and the alkaline phosphatase (AP) detection method (Wilkinson, 1998). A high resolution ISH protocol with the AP detection method has been developed by Thisse et al. (Thisse et al, 1993, 2004), and has shown to be suitable for high throughput genomic screens. Spatial patterns of gene expression are compared by microscope images of the whole-mount ISH, providing two-dimensional information (Thisse et al.1993, 2004; www.ZFIN.org). However, ISH in itself is not the most suitable tool to provide true spatial information. Accurate visualization of the gene expression domains and their spatial relationship requires additional serial sectioning of the hybridized embryos. But even though these techniques provide complementary information, reconstruction of the gene expression patterns is still required since the information is only in 2D.

For mouse as well as for *Xenopus*, a digital 3Datlas of embryonic development and a gene expression database have been developed (mouse: http://genex.hgu.mrc.ac.uk; Davidson et al. 1997; *Xenopus*: http://www.xenbase.org3DModels, http://3dexpress.org, http://xlaevis.cpsc.ucalgary.ca; Gerth and Vize, 2004). The 3D atlas of mouse development is based on serial sections of mouse embryos. 2D gene expression patterns from ISH as well as 3D gene expression patterns obtained by optical projection tomography (Sharpe et al. 2002; http://genex.hgu.mrc.ac.uk/OPT_Microscopy) can be mapped on the anatomical structures of the 3D mouse atlas, thus providing a clue in the genetic pathways underlying developmental processes. For *Xenopus,* images obtained with FISH, ISH and immunohistochemical RNA detection can be viewed and compared with the 3D models built from whole mount confocal images (Gerth et al.2007).

3.1 3D atlas of zebrafish development

In our research group (http://bio-imaging.liacs.nl), we are developing and maintaining a 3D digital atlas of zebrafish development (Verbeek et al, 1999). The 3D atlas is intended for use as an online reference for researchers. In addition it may serve as a template to map gene expression patterns on the developing anatomical structures. Supplementary to the 3D atlas of zebrafish development, currently a zebrafish gene expression database is under construction (Belmamoune and Verbeek, 2006). During early development, changing patterns of gene activity are essential for developmental processes and the final form of anatomical structures (Wolpert et al., 2002). Therefore, the zebrafish gene expression database can be considered as the molecular counterpart of the 3D atlas. It enables mapping spatial and temporal activity of genes on developing anatomical structures, comparison of co- localization and co-expression of genes; providing insight in the genetic pathways underlying the formation of complex anatomical structures.

3.2 Zebrafish gene expression database

The zebrafish gene expression data from the experiments for research described in this thesis are stored in the Gene Expression Management System (GEMS) (Belmamoune and Verbeek 2006). Besides these gene expression data, images using GFP labeling can also be stored in GEMS. This system enables both management and retrieval of 3D spatiotemporal gene expression data. To allow interoperability of the spatiotemporal data in this database with other resources, the annotation system uses two ontologies: the zebrafish developmental anatomy ontology (DAOZ; http://bio-imaging.liacs.nl/liacsontology.html), and the gene ontology (GO; http://geneontology.org). The anatomical terms of the DAOZ are extracted from the standard vocabulary provided by the zebrafish information network (ZFIN; http://zfin.org). It uses approved nomenclature by the zebrafish community and it is intended to serve as an annotation tool for zebrafish scientific images. The GO combines biological processes and corresponding molecular and cellular functions of genes (The Gene Ontology Consortium 2000; Camon et al. 2004) and is based on a dynamic controlled vocabulary. This common, standard vocabulary is updated when information changes or accumulates; making it possible to annotate genes, proteins and biological processes in a wide variety of organisms (The Gene Ontology Consortium 2000). Integration of the zebrafish gene expression database with the GO and other Internet resources facilitates the extraction of information (Belmamoune and Verbeek, 2006).

3.3 Obtaining 3D gene expression data

In order to study developmental processes and the differentiation and patterning of specific organ systems at specific locations, both spatial and temporal information is required. Therefore, 3D data of gene expression patterns as well as anatomical structures provide most information about the state of development. To that end, we developed a straightforward technique for fluorescent *in situ* hybridization (FISH) of whole mount zebrafish embryos: the ZebraFISH protocol. This protocol is the counterpart of the high resolution whole mount ISH protocol described by Thisse et al. (2004). It is developed to enable 3D imaging with the confocal laser-scanning microscope (CLSM). A 3D image

from the CLSM, basically serial optical sections are prepared for processing with TDR-3Dbase software (Verbeek et al. 2000, Verbeek et al. 2002). This allows schematic 3D visualization of the spatial characters of gene expression patterns, as well as analytical approaches in future functional studies. For this thesis the most laborious method of TDR-3Dbase was used (Verbeek et al. 2004) in order to obtain more accurate 3D models. However, we were able to generate a large amount of 3D patterns in a straightforward and non-destructive manner**.**

In order to obtain large amounts of spatial and temporal gene expression data in a straightforward manner, we developed a workflow based on ZebraFISH, CLSM and 3D reconstruction (Fig. 1)

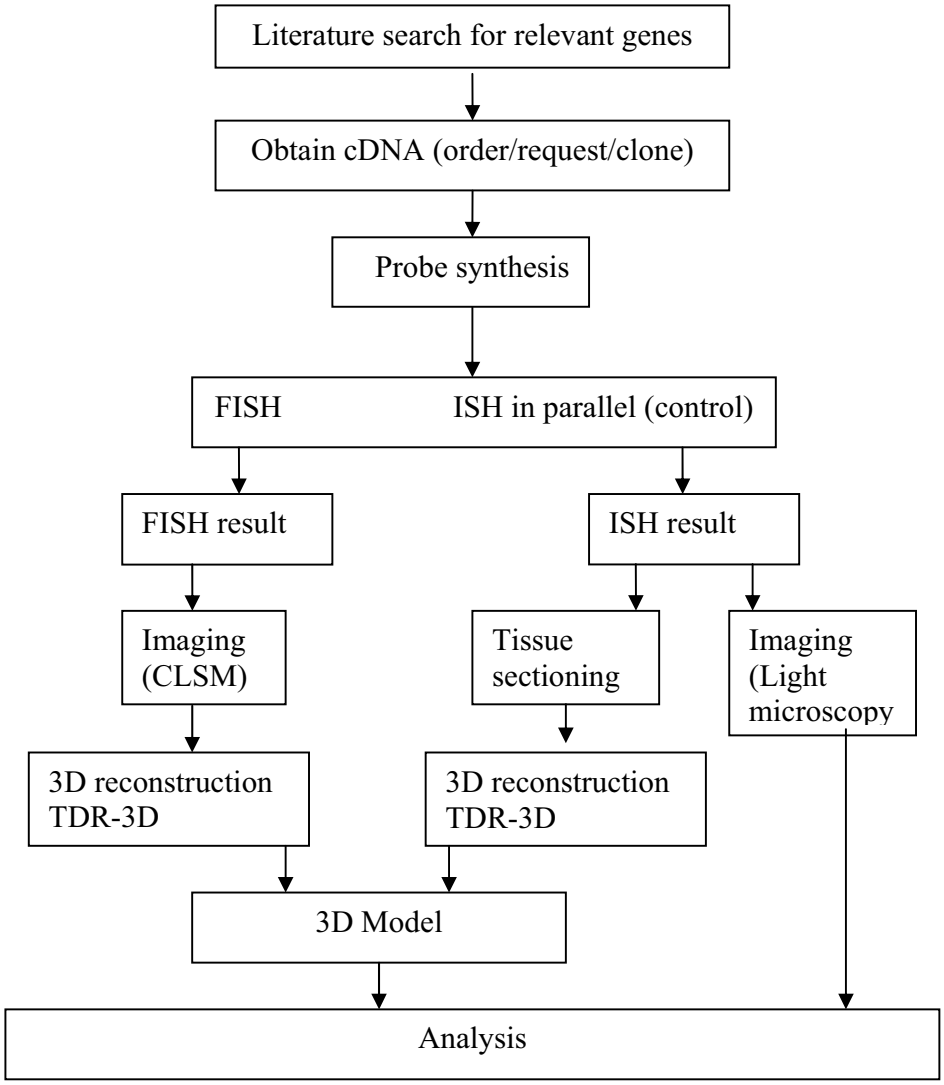

Fig.1: workflow used for obtaining *in situ* gene expression data and eventually, analysis of the data.

Spatial gene expression data are obtained

- by combining ZebraFISH (cf C**hapter 2**) and 3D reconstruction with TDR-3Dbase.

Temporal in situ gene expression data are obtained

- by application of zebraFISH to produce developmental series of zebrafish embryos,
- by analysis of spatiotemporal expression data with Frequent Episode Mining in Developmental Analysis (FEDA, cf. **Chapter 6**)
- by analysis of microarrays. Besides in situ gene expression analyses, we also generated and analyzed temporal (i.e. micro array) gene expression data (Corredor et al. 2006; Meuleman et al. 2006). The microarray data provide a quantitative analysis for a large variety and a large number of genes expressed during time series of development, but are poor in spatial information in the zebrafish embryos studied (Linney et al. 2004).

4. Case studies

In this thesis, ISH and ZebraFISH have been applied on a wide variety of genes. We analyzed gene expression patterns of more than 35 genes, in 4 different functional systems, during 4-10 developmental stages (cf. Chapter 3 – 6). In Table 1, a summary of the functional systems, developmental stages and genes is given.

Consequently, we have obtained a large amount of data from both standard marker genes and less specific genes. We were able to study relationships between genes in a spatial and temporal context. 3D spatiotemporal analysis can be used to characterize expression patterns of certain genes very precisely. We have applied ZebraFISH to two case studies to support this:

1) gene expression during early zebrafish development

2) gene expression during late zebrafish and cross species development, with a focus on limb and pectoral fin formation.

4.1 Case study early zebrafish development

In this case study, two processes in early zebrafish development were investigated:
Case study 1a) Gene expression analysis of 14-3-3 proteins in brain development.
In this study we aimed to accurately map gene expression patterns that have a diffuse appearance over a specific part of the body, to developing anatomical structures in zebrafish embryos during the early stages of development.
The 14-3-3 protein family is a highly conserved family of small dimeric proteins, found in eukaryotes. 14-3-3 proteins are involved in numerous cellular processes (cf. **Chapter 3**). Recent studies have indicated that the 14-3-3 proteins play a role in human disorders such as cancer and neurological disorders (Dougherty and Morrison, 2004), Expression of 14-3-3 during zebrafish development was previously analysed by Besser et al. (2006). In this study, the ISH results for the 14-3-3 genes exhibited diffuse expression patterns that were difficult to annotate on anatomical structures. Moreover, gene expression patterns in zebrafish were not yet localized in detail, i.e. by means of tissue sectioning, and gene expression patterns appeared difficult to characterize from 2D images. To study the zebrafish 14-3-3 family members in relation to human development, neurological disorders and cancer, it is important to accurately characterize the gene expression patterns of 14-3-3 isoforms in the zebrafish brain.

Table 1. Summary of functional systems, organ systems, genes and developmental stages used for this thesis.

Zebrafish Functional system	**Locomotion**		**Innate Immune system**	**Central nervous system**	**Develop-mental**
Organ	Fin	Skeleton & Musculature	Blood, haematopoietic tissues	Brain, spinal cord	Body axis
Process	initiation development patterning identity	development	development distribution	development	development patterning
Developmental Stages Kimmel et al. 1995	24-120 hpf	36-120 hpf	18-96 hpf	10-48 hpf	10-24 hpf
Number of stages	7	6	7	10	8
Genes	*fgf8* *tbx5* *shh* *msx-b* *hoxa9a* *hoxc4a* *hoxd9a* *hoxd11a*	*sox9a* *sox9b* *bmpr1b* *runx2a* *runx2b* *chm 1* *myoD*	*l-plastin* *mpx* *draculin* *fms* *lysC*	*hoxb1a* *hoxb3a* *krox20* *otx2* *pax2* *six3.1* *14-3-3 ε* *14-3-3ι* *14-3-3ζ* *14-3-3γ* *14-3-3τ*	*hoxb6a* *hoxb8a* *hoxb13a* *hoxc12a*
Chicken Organ		wing hindlimb			
Developmental Stages Hamburger & Hamilton, 1951		24-34			
Number of stages		10			
Process		development			
Genes		*sox9* *bmpr-1b* *wnt-14*			

Case study 1b) Gene expression analysis of markers of the innate immune system.
In this case study we annotate gene expression in single white blood cells -or precursors thereof- to larger anatomical structures. We analyze the temporal distribution of marker gene expression in these cells over the embryo during early stages of development. In this study the focus is on *l-plastin*, a general marker of leukocytes, and *mpx*, a specific marker of neutrophil granulocytes. These marker genes are expressed in single cells and show a very distinct expression pattern.

4.2 Case study late zebrafish and cross species development
In this case study, the focus is on marker genes involved in zebrafish pectoral fin as well as in chicken limb development. Marker genes involved in zebrafish fin and chicken limb development are analyzed and compared with limb development in other tetrapods. The chicken is another well-studied model system, and a large amount of gene information is available to study chicken limb development in a developmental as well as in an evolutionary context.
The teleost pectoral fin and the tetrapod limb are well studied in evolutionary - developmental biology. Data from the fossil record suggest that the transition from fin to limb occurred approximately 410 million years ago (Shubin et al, 2006). However, many similarities can be observed in both the teleost fin and the tetrapod limb. Highly conserved organizing structures are found in limbs as well as in paired fins, and the same genetic pathways involved in patterning and outgrowth are found in teleost fins and in tetrapod limbs (Hinchliffe, 2002; Tickle, 2002).
Though skeletal structures in tetrapod limbs are more complex than in teleost fish such as zebrafish (Coates and Cohn, 1998; Grandel and Schulte-Merker, 1998), the molecular marker genes present in early cartilage formation are conserved in both tetrapods and teleosts.
The timing of limb and fin development, however, displays a lot of variation (Richardson, 1995). All considered, it is interesting to compare gene expression data involved in fin and limb development in vertebrates. Such studies are an example to facilitate extrapolation between model systems as well as to human.
In this case study on limb development, we analyze

1) Gene expression patterns of three early cartilage marker genes in chicken embryo wings and hindlimbs. In this study ISH in combination with tissue sectioning was used to reveal a rudimentary structure (cf. **Chapter 5**) and to compare timing of gene expression within one species.
2) Gene expression of marker genes involved in pectoral fin development and cartilage formation in zebrafish embryos from 30-120 hpf., using ISH and zebraFISH.
3) Gene expression data from known orthologues in other tetrapod model species. Here, gene expression patterns for clawed toad, axolotl, mouse, as well as supplemental data for chicken are retrieved from literature.

Eventually, the relative timing of gene expression during fin and limb development is compared between all five model species, using the frequent episode mining algorithm (FEDA).

In this case study, we have specifically focused on difference in timing of gene expression, in one species i.e. chicken, as well as cross – species, i.e. in zebrafish and four tetrapod model species; using ISH, FISH, as well as data acquisition.

5. Outline of the thesis

In **Chapter 2** the methodology of the ZebraFISH is presented and discussed. The FISH method is based on the high resolution whole- mount ISH protocol developed by Thisse et al. (1993, 2004). This protocol is used for high throughput genomic screens; excellent results have been shown (www.ZFIN.org). To test and optimize our ZebraFISH protocol, we used a panel of five standard marker genes, with a wide range of gene expression patterns (cf. **Chapter 2**)

Following the application with standard markers in **Chapter 2**, we have worked out two case studies with more specific markers.
In **Chapter 3**, we present a spatiotemporal characterization of the 14-3-3 γ and τ isoforms with the focus on the developing zebrafish brain, using ZebraFISH, TDR-3Dbase software and the 3D atlas of zebrafish development as a reference. In addition, we demonstrate that the techniques presented facilitate a precise localization of complex gene expression patterns.

In **Chapter 4**, cell-based gene expression patterns of *mpx* and *l-plastin* are analysed in their spatial relation to anatomical structures such as developing blood vessels and heart. Distribution of these cells over the embryo is visualized in 3 dimensions and at subsequent time points during embryonic development, using TDR-3Dbase software.

In **Chapter 5**, *in situ* hybridization is applied to investigate a rudimentary structure in another model system, the chicken. Traditional histological methods might detect a rudimentary cartilage structure too late, when it already starts to disappear. Marker genes used in **Chapter 5** are expressed before condensation of mesenchyme in future skeletal elements (Akiyama et al, 2002; Chimal- Monroy et al, 2003), or in early differentiating chondrocytes (Karsenty and Wagner, 2002; Pizette and Niswander, 2002).
In **Chapter 6,** *in situ* gene expression data from marker genes involved in limb and pectoral fin development are compared between five model species. A new method to analyze limb and fin gene expression data with a focus on heterochrony in gene expression, is presented: Frequent Episode Mining in Developmental Analysis (FEDA).
A general discussion is presented in **Chapter 7**, followed by the conclusions from the work described in this thesis. Finally, in **Chapter 8** a summary and a summary in Dutch are given.