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

Welten, M.C.M.

Chapter 7

General discussion and conclusions

General discussion and conclusions

In this thesis we explored *in situ* hybridization techniques and computerized tools to study gene expression patterns during embryonic development. We applied these tools to investigate several developmental processes in both zebrafish and chicken embryos and in two groups of case studies.

In order to study differentiation and patterning of organ systems during embryonic development, both spatial, as true 3D data, and temporal information is required. True 3D data of gene expression patterns and anatomical structures provide most information. In this thesis, the focus was on obtaining 3D spatial and temporal gene expression data during embryonic development. Therefore, we have developed a whole-mount fluorescent *in situ* hybridization protocol: ZebraFISH. The ZebraFISH method is particularly suitable for 3D imaging of whole mount zebrafish embryos with confocal laser scanning microscopy (CLSM).

In **Chapter 2** we describe the ZebraFISH protocol and underlying methodology. Our method is based on Tyramide Signal Amplification (TSA) and proved to be a powerful, sensitive method to visualize gene expression patterns. In this detection method the presence of horseradish peroxidase catalyses the deposit of cyanine labelled tyramide at the site of hybridisation. TSA yields a strong expression signal with a relatively low background in a short defined staining time (Clay and Ramakrishnan, 2005; Zaidi et al. 2000). Even weak probes and small gene expression domains can be visualized. Stronger amplification of weak signal can be obtained by using TSA Plus (Perkin Elmer).

In order to obtain a large amount of 3D gene expression data, we developed a workflow based on the use of ZebraFISH, confocal laser scanning microscopy and 3D modelling with three- dimensional reconstruction software TDR-3Dbase (Verbeek et al. 2000).

In this thesis, all FISH experiments were carried out manually. The one-step antibody detection method used in ZebraFISH is a major advantage for application of ZebraFISH in a high throughput setting and for use in a robot. Subsequent imaging with CLSM is a more time-consuming process than imaging with fluorescent or bright-field microscopy. These images however, provide only 2D information. For the research described in this thesis, each CLSM image slice was sampled to 1024 x 1024 pixels, providing high-resolution 3D images. Although this is a time-consuming process, these high-resolution images allow precise 3D reconstructions. Therefore, in comparison with *in situ* hybridisation (ISH) and subsequent serial sectioning, our workflow provides the possibility to produce a large amount of 3D gene expression data in an accurate and non-destructive manner.

The CLSM work described in this thesis was performed with a Leica TCS/ SP DM IRBE confocal laser scanning microscope equipped with an Ar/Kr laser. Dependent on the laser lines available, a wide range of fluorescent labels can be used to discern gene expression, anatomical structures and the outline of the embryo. Also, tissue-specific fluorescent markers can be used to more accurately annotate gene expression patterns on specific tissues and developing organs.

CLSM proved to be suitable for imaging whole mount zebrafish embryos. In order to overcome the limitations in specimen depth scanning of the CLSM, multiphoton microscopy can be applied to obtain images of complete embryos. In like manner, excellent results with Optical Projection Tomography (OPT) have been presented for mouse (Sharpe et al. 2002; http://genex.hgu.mrc.ac.uk/OPT_Microscopy), for human

(Kerwin et al. 2004; http://www.ncl.ac.uk/ihg/EADHB/) and for zebrafish (Bryson-Richardson et al. 2007; http://www.fishnet.org.au). Imaging with CLSM is limited to fluorescently labelled specimens only, whereas OPT allows 3D imaging of chromogenic ISH as well as many other experiments (Sharpe et al. 2002).

The voxel size in OPT is about 10x10x10 μm (http://genex.hgu.mrc.ac.uk/Emage/database). With CLSM, depending on the NA of the lens, a much smaller voxel size can be achieved; this resolution is more suitable for small samples like zebrafish embryos and the visualization of gene expression at the cellular level. For most of the work done in this thesis a 10 x plan apo lens (NA = 0.24, long working distance) was applied. The resolution in the xy-plane was used as a guide for z-axis sampling to achieve isotropic voxels in the image. With given NA, each image slice was sampled in 1024x1024 pixels; next choosing the right z-displacement a voxel size of around 1 μm^3 can be achieved. For subcellular localization of gene expression in zebrafish embryos however, a magnification of 63 x is required.

ZebraFISH can also be used in experiments with multiple FISH labelling, in combination with fluorescent immunostaining and Green Fluorescent Protein (GFP) transgenic zebrafish. In multiple FISH experiments only a limited number of fluorescently tagged genes can be visualized in the same embryo. More complex patterns of coexpression and colocalization can be analyzed using dedicated resources such as the zebrafish gene expression database (Belmamoune & Verbeek, 2007).

Other fluorescent techniques comprise transgenic zebrafish lines expressing GFP, and fluorescent immunolabelling. GFP is used as a reporter for gene expression; immunohistochemistry shows expression at the protein level. While FISH and immunolabelling are only possible in fixed samples, GFP can be applied to live samples as well as in combination with FISH, i.e. in fixed embryos (Langford et al., 2006).

In order to further improve the quality of the CLSM images, deconvolution can be applied to images acquired with a higher magnification. This process eliminates the distortions and blurring caused during capture of the image. Deconvolution will result in a sharper and clearer image than could be obtained with the available optics alone (Boutet de Monvel et al. 2003).

Spatio-temporal analysis and 3D modelling of gene expression patterns were applied to study relationships between genes in both spatial and temporal context. It also allowed an accurate characterization of more complex gene expression patterns. This is illustrated through application of methods to two groups of case studies:

1) Early zebrafish development:
   - Characterization of 14-3-3 gene expression during zebrafish development (**Chapter 3**)
   - 3D reconstructions of gene expression patterns in the developing innate immune system (**Chapter 4**)

   In this group of case studies, we analysed complex gene expression patterns as well as gene expression patterns in small domains, i.e. in single cells scattered over the embryo.

2) Late zebrafish and cross species development:
   - Gene expression and digit homology in the chicken wing (**Chapter 5**)
   - Limb/fin heterochrony, analysis of gene expression data with Frequent Episode Mining in Developmental Analysis (FEDA, **Chapter 6**).

   In this group of case studies, the focus was on marker genes, involved in zebrafish pectoral fin and chicken limb development. Gene expression data of these marker genes were compared with data from other tetrapods.

Case study Early zebrafish development
In **Chapter 3**, ZebraFISH and 3D modelling were applied to characterize complex gene expression patterns.

Recently, expression patterns of genes encoding 14-3-3 proteins were studied in zebrafish embryos. The 14-3-3 isoforms exhibited complex gene expression patterns (Besser et al. 2006). These gene expression patterns were not yet studied in 3D. In order to employ zebrafish as a model to study cancer and neurological disorders, the role of 14-3-3 proteins in zebrafish brain development needs to be characterized.

In this chapter, gene expression patterns for genes encoding two 14-3-3 isoforms were analyzed using our workflow (cf. **Chapter 1**). Application of the ZebraFISH protocol yielded more 14-3-3 gene expression domains and at earlier developmental stages than by ISH detection in previous studies. To confirm our ZebraFISH results, samples from ISH detection were overstained for 3-4 days. Serial sections were made from these overstained samples. The serial sections were then processed for 3D reconstruction (Verbeek et al. 2002). Using 3D modelling, we were able to establish an accurate characterization of zebrafish genes encoding 14-3-3 γ and τ in developing anatomical structures.
From earlier work two subtypes for the genes encoding the zebrafish γ and τ isoforms have been described. Data from microarray analysis (Besser et al. (2006) revealed that the two γ subtypes displayed very diverse expression profiles, suggesting different roles for 14-3-3 γ1 and γ2 during development. Microarray expression profiles for the 14-3-3 τ1 and τ2 subtypes were less distinct (Besser et al. 2006) and ZebraFISH as well as ISH yielded almost similar gene expression patterns (Besser et al. 2006, this thesis). Our initial analysis was based on the 14-3-3 γ1 probe used by Besser et al.(2006). In order to specifically analyze gene expression of γ1 and γ2, recently two new probes were synthesized for the 14-3-3 γ1 and γ2 subtypes. Results from ZebraFISH and CLSM imaging with these probes suggest that the expression patterns for the two subtypes of the gene encoding 14-3-3 γ can be distinguished. FISH results for the "new" 14-3-3 γ1 and γ2 subtype display expression in the domains previously described (Besser et al. 2006; this thesis Chapter 3). The expression patterns of the two 14-3-3 γ subtypes are partially overlapping and partially complementary to each other (Welten et al., in preparation). 3D modelling of the CLSM images confirmed these data. Thus, the methods presented stablish the specificity of probes.

Although the 14-3-3 proteins are highly conserved among amniotes, a large variation in isoform number can be observed in higher organisms (Wiker and Yaffe, 2004). Alterations in expression or function of the 14-3-3 isoforms in relation to human disorders indicate specific functions for the different isoforms (Wiker and Yaffe, 2004).

For zebrafish, the function of the distinct 14-3-3 isoforms and the cell type in which the isoforms are expressed, are still under study. FEDA (cf. Chapter 6) can be applied to analyze timing data of gene expression of the 14-3-3 isoforms in anatomical structures, within one species and across species. Moreover, application of the FEDA algorithm to location and function of the distinct 14-3-3 isoforms in various model organisms (e.g. fruitfly, zebrafish, mouse, rat) and human may help predicting location and function of the isoforms in zebrafish. However, some basic features of the teleost fish brain are different from those in higher vertebrates. For example, in teleost fish neurogenesis occurs in a wide range of brain structures and continues during adulthood (Miramura and Nakayasu, 2001; Zupanc et al. 2005) whereas in higher vertebrates, adult neurogenesis occurs only in restricted areas (Zupanc et al. 2005). This aspect might involve different roles for 14-3-3 isoforms in zebrafish and higher vertebrates such as mammals.

In **Chapter 4**, we applied ZebraFISH and 3D modelling to study gene expression patterns in single white blood cells (and precursors thereof).
Zebrafish macrophages and granulocytes are essential cell types in the innate immune system. These cells behave strikingly similar to their counterparts in mammals. Marker genes expressed by zebrafish macrophages and granulocytes are found to be homologous to those of mammals. This is one of the reasons that the zebrafish can be employed as a model to study the response of both the innate and the adaptive immnune system to inflammation. Experimental findings from zebrafish can be used for understanding these systems in human.
In this case study we focused on *l-plastin*, a general marker of leukocytes, and *mpx*, a specific marker of neutrophil granulocytes. An exploratory study was carried out to analyze cell distribution and migration behaviour of *mpx* and *l-plastin* expressing cells in relation to vascularization and haematopoietic events.

The distribution patterns of *mpx* and *l-plastin* expressing cells during zebrafish development matched the vascularization and haematopoietic events described in literature (Herbomel et al.1999 and 2001, Isogai et al. 2001, Lieschke et al. 2001, Murayama et al.2006). With the schematic 3D modelling gene expression can be inspected in single cells and give a clue about the distribution patterns of these cells in relation to anatomical structures.
Application of ZebraFISH and schematic 3D modelling allow a quantitative approach in future studies. For example, cell numbers and distribution patterns can be analyzed in infection experiments, or in mutants in which the vascular system is affected. Consequently, further understanding of the innate immune system in zebrafish can be achieved.

3D gene expression data and models from several other marker genes were made, but not used in the analysis presented in Chapter 4. These markers comprise *draculin*, an early marker of macrophages and erytrhroblasts (Herbomel et al.1999); *fms*, a marker of macrophages, early microglia and neural crest derivatives (Herbomel et al. 2001) and *lysozyme C*, previously suggested as macrophage marker (Berman et al. 2005; Liu and Wen 2001) though recent multiple FISH experiments revealed expression of *lysozyme C* in neutrophil granulocytes (Meijer et al. 2007). In order to characterize these overlapping cell populations, multiple FISH experiments as well as FISH combined with fluorescent

immunolabeling or with known GFP lines can be useful in analysis. Subsequent 3D modelling allows a better inspection of these overlapping cell populations and their expression domains in other tissues.

Gene expression data from the developing innate immune system can also be compared over different animal models, using the FEDA algorithm (cf. Chapter 6). This computational approach can be helpful to elucidate characters such as locations of embryonic and adult haematopoietis or differences in function of early macrophages (Shepard and Zon 2000), across zebrafish, amphibians and amniote model systems.

Case study Late zebrafish and cross species development

In this case study we focused on marker genes involved in zebrafish pectoral fin and tetrapod limb development. Though data from the fossil record indicate that the fin / limb transition occurred about 410 million years ago (Shubin et al, 2006), many similarities can be found during both teleost fin and tetrapod limb development (Hinchliffe, 2002; Tickle, 2002). First, we made a side step to another model organism: the chicken **(Chapter 5)**. The chicken is extensively studied in a developmental as well as in an evolutionary context and a large number of molecular data has become available. Gene expression data from zebrafish pectoral fin and chicken limb development obtained by our own experiments were then compared with data from known orthologues in other tetrapod model systems **(Chapter 6)**. In **Chapter 5,** *in situ* hybridization and 3D reconstruction were used to investigate a rudimentary structure in the chicken wing in relation to bird evolution. In the wing, the transcription factor *sox9* exhibits a transient gene expression pattern anterior to the digit II primordium, which we identified as a rudimentary digit I domain. The receptor molecule *bmpr1b* displayed no hybridisation signal in the same region at the same stage, suggesting that the presumptive digit I domain is arrested during development and shows no further differentiation into cartilage

In our study, ISH and 3D reconstruction were applied to whole mount chicken wings and hindlimbs. Recent studies reveal excellent FISH results for whole mount chicken embryos (Denkers et al. 2004). Repeating our experiments with FISH and subsequent imaging with CLSM will result in an accurately aligned stack of optical sections, in a relatively short time. Since the TSA detection proved to be more sensitive, we may be able to detect the onset and offset of *sox9* and *bmpr-1b* gene expression patterns in the digit primordia and cartilage elements more precisely. During later stages of chicken development, CLSM may not be capable to scan through tissues like cartilage. To overcome this problem, thick sections of the samples can be scanned with CLSM (Sharpe, 2003).

To compare differences in timing of gene expression in wing and hindlimb, a computational method was applied: FEDA (**Chapter 6**). Though a wide variety of genes are involved in digit formation, analysis with FEDA revealed a delay in *sox9* expression in the presumptive digit I domain in the chicken wing, compared with that of the hindlimb. The pattern shift diagram in **Chapter 6** also reveals that FEDA can visualize heterochrony at different levels within one analysis. Investigation of more gene expression data from in chicken wing and hindlimb development with FEDA might provide a clue about genetic networks involved in digit formation and identity.

In the skink *Chalcides chalcides*, a similar conflict between embryonic and adult digit identity exists as in birds (Wagner, 2005). Anatomical evidence suggests that the digits of *Chalcides chalcides* fore and hindlimb are digit I, II and III. Embryological evidence however, suggests that these digits develop from the digit II, III and IV primordia. The reduction of digits in *Chalcides chalcides* as well as in theropod dinosaurs might be the result of an adaptive modification during evolution. Digit reduction is also observed in *Hemiergis*, another skink genus. In this lizard, loss of digits is correlated with shifts in timing, i.e. heterochrony - in sonic hedgehog expression (Shapiro et al. 2003, Smith et al. 2003). Since FEDA showed to be a very flexible method in data organisation, it can be applied to data from at different levels, such as anatomical structures, gene expression data and developmental events. Analysis of data from skinks, birds and other tetrapods might reveal a link between digit reduction and heterochrony in gene expression, chondrogenesis or other processes.
To solve the problem of homology of avian wing digits with those of other tetrapods, several models are proposed (reviewed by Wagner, 2005). The Frame Shift Hypothesis (Wagner and Gauthier, 1999) proposes that embryonic digit domains II-IV have undergone a homeotic transformation and adopted the identity of digit I-III. The Pyramid Reduction Hypothesis (Kundrat et al. 2002) solves the problem by proposing a bilateral reduction of digits, and that the remaining digits (II-IV) have been remodelled during evolution, into the I-III phenotype. In **Chapter 5** we discuss an alternative model, the polydactyly model. This model provides continuity between digit position and identity across the theropod and bird phylogeny. However, recent findings point towards a possible frame shift in bird digit identity in the theropod –bird lineage. This might be caused by a change of *hox* gene expression in the wing (Vargas and Fallon 2004; Wagner, 2005), though a wide range of genes are involved in the specification of digit identity (Dahn and Fallon 2000, Litingtung et al. 2002; Shapiro et al. 2003; Tiecke et al. 2007). Further studies of the molecular mechanisms underlying digit development and identity, as well as new fossils from bird ancestors will give more insight in the evolution of birds.

Besides the analysis of gene expression patterns in single species, in situ gene expression data were compared across species, with zebrafish as key model.
In **Chapter 6**, gene expression patterns were analyzed using Frequent Episode Mining in Developmental Analysis (FEDA). The FEDA algorithm was applied in two case studies:

1) Gene expression data from zebrafish fin and chicken limb development were compared with data from other tetrapod model systems to analyse heterochrony in gene expression during fin and limb development.
2) Heterochrony in gene expression was analyzed within one species.

We demonstrated that heterochrony in gene expression in limb and fin development could be elucidated very well using the FEDA algorithm, in a phylogenetic context as well as within one species. The use of relative time scales is essential for analysis with FEDA.

We demonstrated that FEDA is flexible in the arrangement of data and capable of handling large datasets such as 14 species, 25 developmental events and 17 gene expression data. Sparseness of gene expression data from a certain species might influence the accuracy of phylogenetic trees. FEDA is capable of finding a solution, even

if data are sparse. The computations with FEDA produce a numerical value to express distance between species, the Jaccard distance. This provides a true objective measure to express differences between species or other entities.
Comparison of ZebraFISH and ISH data from zebrafish as well as extraction of ISH data from literature might influence the accuracy for construction of phylogenetic trees. For zebrafish pectoral fin development, data were generated using ZebraFISH and ISH in parallel. The probes used for detection of the fin markers were strong and specific; no discrepancies between FISH and ISH data were found for the fin markers and the projections of the FISH were consistent with ISH data from literature. The data that were extracted from literature were generated in different research groups where different ISH protocols are used. Though ISH protocols are optimised for every model species, these gene expression data may show slight differences in sensitivity in relation to each other. However, we did not observe differences in timing of gene expression when we compared images from one species that were generated in different research groups. Timing of gene expression and development of morphological characters is compared using a relative time axis; no absolute data were used.
FEDA showed to be a very flexible method in the both arrangement and type of the data. It is not only useful to study heterochrony, but also to compare location and functions of different subtypes of genes within species or across species.
In the near future, in situ gene expression data will be combined with microarray as well as morphological data. Also, the computational framework of FEDA will be employed in spatiotemporal analysis of gene expression data from structured repositories (Belmamoune and Verbeek, 2006).

Conclusions and future work
The work described in this thesis is concerned with the analysis of gene expression and tools to establish spatiotemporal analysis of gene expression.

- We demonstrated that the fluorescent mRNA *in situ* hybridisation (ISH) technique is an efficient and powerful tool to study developmental processes at the level of gene expression. Fluorescent *in situ* hybridisation (FISH) by means of Tyramide Signal Amplification improves the sensitivity of the method and goes hand in hand with 3D CLSM imaging.
- 3D modelling using TDR-3Dbase enabled us to establish an accurate characterization of gene expression in relation to anatomical structures. The methods presented, produced relevant biological data. For the zebrafish 14-3-3 γ proteins, we were able to distinguish and accurately characterize gene expression domains for two isoforms.
  In functional studies, 3D modelling allows analytical approaches.
- The FEDA algorithm has shown to be a very flexible and powerful method to analyse gene expression, timing, species, morphological characters and combinations of all these data. Moreover, FEDA is very flexible in the arrangement of data.

Optimisation of imaging techniques will contribute to further improvement of the analysis of whole mount specimens such as zebrafish embryos. Given the size of the embryo and the limitations of our CLSM it was not possible to produce 3D images of

complete embryos. In future studies, multiphoton microscopy can be used to accomplish a higher depth penetration so that complete embryos can be scanned (Centonze and White, 1998) and a more complete analysis of cell numbers over time can be obtained. For multiphoton microscopy other fluorescent labels with longer wavelength are required, usually in the infrared spectrum (Centonze and White, 1998). This needs to be tuned with our ZebraFISH method.

Application of FEDA to spatial gene expression data in combination with microarray data and morphological characters will provide a clue about the genetic networks present developmental processes in zebrafish as well as in other model systems. In the near future, FEDA can be applied to a wide variety of characters such as subtypes, location and function of genes, and comparison of these with other model systems.