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

Welten, M.C.M.

Summary

In developmental biology, the expression of genes is studied to understand normal development, phenotypes and to construct models to understand disease. In this thesis, we explore and validate biological as well as computerized tools, to address research questions in developmental biology. Based on these techniques, we developed a workflow to generate a large number of 3D spatio-temporal patterns of gene expression.
The analysis of developmental processes has specific problems. For example, many genes are active and many events occur in a short time span, gene expression patterns in differentiating structures do not always correspond with the boundaries of anatomical structures, and expression patterns of different genes need to be compared for colocalization and coexpression. Though several techniques for gene expression analysis are available, most spatial gene expression data are only in 2D. In order to study gene expression and differentiation of structures during development at the same time, both spatial 3D information, and temporal data are essential. Nowadays there is a tendency towards combining molecular biological as well as computational tools and techniques.

In the research group Imaging and Bioninformatics, we have developed a 3D Atlas of zebrafish development. This atlas is intended to serve as an on line reference for researchers. In addition to the anatomical information of the 3D Atlas, a zebrafish gene expression database is under construction (Belmamoune and Verbeek, 2006). This gene expression database can be considered as the molecular counterpart of the 3D atlas. Spatial and temporal expression patterns of genes involved in developmental processes can be mapped on developing anatomical structures in the 3D atlas and compared for co-localization and co-expression of genes; providing a clue about development of complex organs and tissues.
These spatio-temporal patterns of gene expression have to be generated. To that end, we have developed a workflow based on fluorescent *in situ* hybridization (FISH) (ZebraFISH; cf **Chapter 2**), confocal laser scanning microscopy (CLSM) and subsequent three-dimensional modeling with, in our case, TDR-3Dbase software - but comparable results could be obtained with other 3D reconstruction softwar e- resulting in a large amount of 3D spatio-temporal patterns of gene expression obtained in a straightforward and non-destructive manner.

In **Chapter 2,** we describe the methodology of ZebraFISH. This fluorescent *in situ* hybridization method is based on the high-resolution whole mount *in situ* hybridization protocol described by Thisse et al (1993, 2004). The strength of our approach is that it enables 3D imaging with confocal laser scanning microscopy. In our ZebraFISH protocol, Tyramide Signal Amplification (TSA) is used to visualize gene expression patterns. The work in this thesis (Chapter 2 and 3) demonstrates that the ZebraFISH method yields a high signal and relatively low background in a short, defined staining period.
For the work described in this thesis, we applied our workflow to a wide variety of genes in two groups of case studies:

1) *Case study Early zebrafish development*
2) Case study Late zebrafish and cross species development

Case study Early zebrafish development

In **Chapter 3,** the expression of 14-3-3 isoforms was analyzed in zebrafish embryos in a developmental series ranging from 18 hpf up to 48 hpf. In this study, we applied ZebraFISH to investigate complex gene expression patterns with a diffuse appearance. With our workflow, based on ZebraFISH, CLSM and 3D modelling, we were able to detect more gene expression domains and at earlier stages of development than in previous studies. Our methods provided an accurate characterization of genes encoding 14-3-3 γ and τ proteins in subsequent stages of zebrafish development.

In **Chapter 4**, ZebraFISH was applied to analyze gene expression in single white blood cells and precursors thereof. Gene expression patterns of *l-plastin*, a general marker for leukocytes, and *mpx*, a marker for neutrophil granulocytes, were characterized in developmental series of zebrafish. In addition, a pilot study was carried out to analyze the distribution of *l-plastin* and *mpx* expressing cells from 24 hpf up to 72 hpf in relation to haematopoietic events. The schematic 3D modeling presented in this study, allows better inspection of the distribution of the single *l-plastin* and *mpx* expressing cells and their relation to anatomical structures.

Case study Late zebrafish and cross species development

In this case study the focus is on genes involved zebrafish fin and tetrapod limb development.

In **Chapter 5**, ISH and 3D modelling are used to investigate a rudimentary digit in the wing of developing chicken embryos. Gene expression patterns from two markers genes of early cartilage formation, *sox9* and *bmpr1b*, were analyzed in chicken wings and hindlimbs. In the wing, anterior to digit II primordium and close to the ectoderm, we found an expression domain for *sox9* that we identified as a rudimentary digit I domain.
No subsequent *bmpr1b* expression and no cartilage formation were found in this domain, suggesting that the presumptive digit I domain is arrested in development and shows no further differentiation.

In Chapter 6, timing of expression of genes involved in zebrafish pectoral fin and chicken limb development is compared among other vertebrate model systems. Zebrafish and chicken gene expression data were obtained by our own experiments. Data for mouse, axolotl and clawed toad, as well as supplemental data for chicken, were extracted from the literature. A relative time scale was calculated for every model species, from zygote up to hatching, birth or juvenile stages. Gene expression data were processed with the Frequent Episode Mining algorithm (FEDA) to analyze evolutionary relationships between model species, based on heterochrony, i.e. difference in timing, of gene expression. The relative time scale was made part of the computation.
The FEDA algorithm showed to be a powerful and flexible method. Data can be rearranged such that several features can be compared in the analysis; i.e., gene expression, morphological structures, species and timing data.

In **Chapter 7**, a general discussion and conclusions are presented. A perspective of the work described in this thesis is given in relation to recent literature and the most important conclusions from **Chapter 2-6** are summarized