Spatial gene expression quantification in changing morphologies

Botman, D.

Publication date
2016

Document Version
Final published version

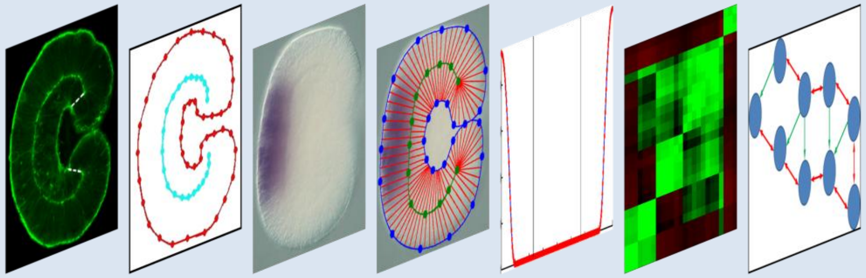
Link to publication

Citation for published version (APA):
Botman, D. (2016). *Spatial gene expression quantification in changing morphologies*. [Thesis, fully internal, Universiteit van Amsterdam].

Spatial gene expression quantification in changing morphologies

Daniël Botman

Spatial gene expression quantification in changing morphologies



Daniël Botman

Spatial gene expression quantification in changing morphologies

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. D.C. van den Boom
ten overstaan van een door het College voor Promoties
ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op dinsdag 14 juni 2016, te 12:00 uur

door

Daniël Botman

geboren te ‘s-Gravenhage

Promotiecommissie:

Promotor:	prof. dr. P.M.A. Sloot
	Universiteit van Amsterdam
Copromotor:	dr. J.A. Kaandorp
	Universiteit van Amsterdam
Overige leden:	dr. E. Röttinger
	Université Nice Sophia Antipolis
	prof. dr. F. Baas
	Universiteit van Amsterdam
	prof. dr. A.H.C. van Kampen
	Universiteit van Amsterdam
	prof. dr. A.G. Hoekstra
	Universiteit van Amsterdam
	prof. dr. M.K. Richardson
	Universiteit Leiden
Faculteit:	Faculteit der Natuurwetenschappen,
	Wiskunde en Informatica

The work described in this thesis was performed at the Computational Science research group of the University of Amsterdam.
The research has received funding from the Honda Research Institute (HRI) and from the European Union Seventh Framework Programme (FP7) under grant agreement Nr. 289434 (project 'BioPreDyn' – New Bioinformatics Methods and Tools for Data-Driven, Predictive Dynamic Modelling in Biotechnological Applications, www.biopredyn.eu).

Contents

1 Introduction

Embryogenesis is the developmental process of generating an organism from a fertilized egg. Early development varies widely among animals, and many different mechanisms for development have been observed in the animal kingdom. Embryogenesis has been studied in model species like humans [1,2], mice [3], chickens [4,5], frogs [6], flies [7], sea urchins [8] and, more recently, sea anemones [9]. Despite the large variation, early development is usually dissected in various general stages: cleavage, gastrulation and organogenesis. First, the nucleus is copied many times in the cleavage stage. The cleavage process in an animal is classified by the extent of membrane formation and yolk distribution and usually results in the formation of a blastula (Figure 1.1). Next, the germ layers are formed during gastrulation, as the interior cells (endoderm) are distinguished from the periphery (ectoderm). Gastrulation is a rearrangement of cells that occurs through a combination of mechanical mechanisms (Figure 1.2), and a wide variation is observed among model animals. Embryogenesis is completed with organogenesis, the formation of functional organs. Cells differentiate into their final tissue types and end up in their correct locations.

1.1 Pattern formation in *Drosophila melanogaster*

A change in cell type occurs in two steps. First, the correct set of genes is activated; this process is called determination. Before germ layers, tissues and organs start to form, they are distinguished from

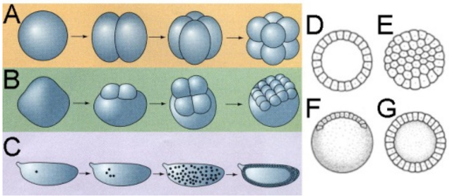


Figure 1.1: Overview of the main cleavage modes and the main types of blastula. A) Holoblastic cleavage results in complete cells; this is observed in a wide range of animals, like mammals, amphibians, snails and worms. B) Discoidal cleavage results in a layer of cells on top of a yolk cell; this occurs in fish, reptiles and birds. C) During superficial cleavage, the nuclei divide within a single cell membrane; this process is unique for insects. D) A coeloblastula is a hollow sphere that forms from a clump of cells. E) A stereoblastula is a densely packed sphere of cells. F) A discoblastula contains the cleaved cells on top of a yolk cell. G) A periblastula consists of a cell layer that formed around a yolk sac. (A-C adapted from [35] with permission of Sinauer Associates; D-G adapted from [36])

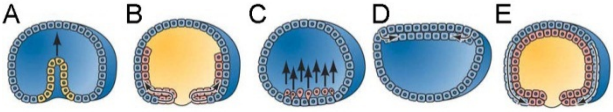


Figure 1.2: Different mechanisms of gastrulation. A) Invagination is the infolding of a cell sheet. B) Involution is the expansion of a cell sheet on the inner surface. C) Ingression is the detachment of individual cells. D) Delamination is the splitting of a single sheet into a double sheet. E) Epiboly is the expansion of a cell sheet on the outer surface. (Adapted from [35] with permission of Sinauer Associates)

the surrounding area by their sets of active genes. Gene activity is measured by the amount of gene products such as RNA and proteins.The production of RNA and proteins mainly depends on the presence of regulating molecules and their access to the DNA. Second, the structure, function or mechanical properties are adapted; this process is called differentiation. Differentiation is the observed change of cells from one type to another, while determination is the prior expression of the necessary set of genes. The specification of the body axis in fruit flies is a good example of determination by a regulatory gene expression cascade. The embryonic development in the fruit fly *Drosophila melanogaster*

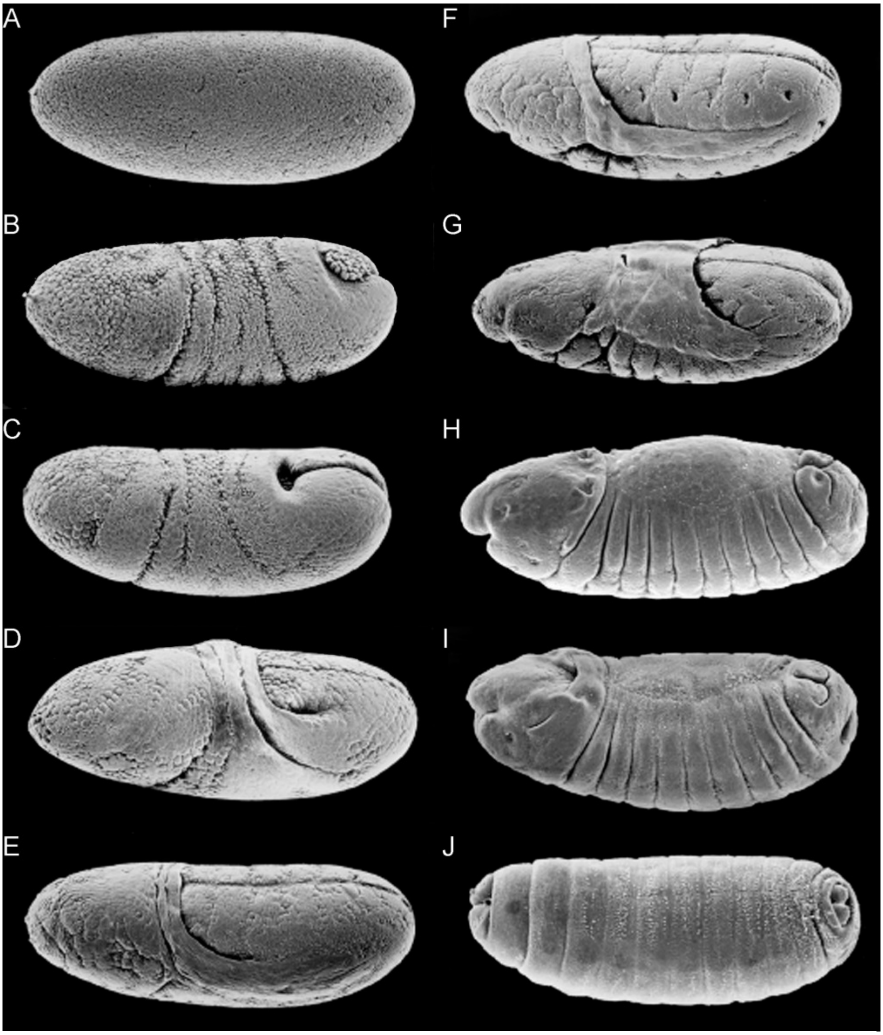


Figure 1.3: Embryogenesis in the fruit fly. The blastula (A) consists of a sheet of cells around a yolk bubble. Gastrulation occurs at the bottom (ventral) side. The ventral cell sheet then expands over the rear (posterior) end to the top (dorsal) side (B-E) in a process called germ band extension. The segment boundaries are formed while this tissue shrinks again (F-I), which is referred to as germ band retraction. Segment formation has finished with the larva (J). The future head is oriented to the left. (Adapted from [37]; © 2008 Canadian Science Publishing or its licensors. Reproduced with permission)

from the blastula until the segmented larva is displayed in Figure 1.3. The position of a cell in the blastula determines the segment

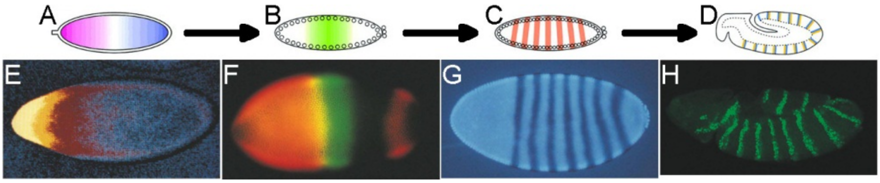


Figure 1.4: Schematic overview and fluorescence micrographs of patterning cascade in the fruit fly. The future head is oriented to the left. Expression patterns with increasing resolution arise after maternal morphogen gradients (A) regulate the gap genes (B). The gap genes, that are expressed in broad domains, initiate the pair rule genes (C), that are characterized by 7 narrow stripes. Combined regulation by the pair rule genes defines the exact boundaries of the 14 segments (D). E) Fluorescence view of the maternal bicoid protein distribution; the protein concentration is largest in the head region. F) The protein domains of gap genes hunchback (orange) and kruppel (green) display a small region of overlapping expression (yellow). G) Striped protein pattern of pair rule gene fushi tarazu (dark signal). H) The protein pattern of engrailed indicates the segment boundaries (green). (Adapted from [35] with permission of Sinauer Associates for A-D, Christiane Nüsslein-Volhard for E, Chris Rushlow for F, Timothy Karr for G, Steve Paddock and Sean Carroll for H)

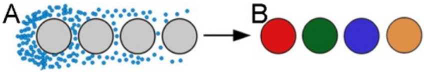


Figure 1.5: A) Cells can detect different concentrations of a morphogen (blue dots). In this example, the morphogen concentration decreases with distance from the morphogen source on the left. B) Based on the local morphogen concentration, different regulatory programs are initiated in the cells. The different colors represent different regulatory programs. (Adapted from [10] with permission of The Company of Biologists)

where this cell arrives, and the segment where a cell is located determines into which part of the fly's body this cell develops. This may seem obvious, but a cell cannot sense its spatial coordinates. A cell can only sense signals from its direct environment, and these signals can trigger reactions that activate genes within the cell's nucleus. In the fruit fly, a spatial axis is specified by a chemical gradient (Figure 1.4A). Various developmental pathways are induced in different cells by different concentrations of this chemical called a morphogen (Figure 1.5) [10]. In this fashion, morphogen gradients specify the embryo's body axes.

The positions and functions of segments in the larva are determined early in the fly's development, before the segments are visible. A smooth maternal morphogen gradient is already established in the *D. melanogaster* egg (Figure 1.4E). The formation of the final segmentation gene expression pattern occurs in three stages. First, the morphogen gradient activates the gap genes after a number of cleavages, resulting in overlapping gene expression domains (Figure 1.4F) [11]. Second, the gap genes activate the pair rule genes, which form patterns of seven sharp stripes at the blastula stage (Figure 1.4G). Third, the pair rule genes cause the segment polarity genes to be expressed to delineate the fourteen segments. These three sets of genes with different regulatory functions in segment formation were initially inferred from mutation experiments [12]. One gene at a time was switched off and the morphology at the larval stage was observed. Mutants with a gap gene switched off were missing a group of adjacent segments. A blocked pair rule gene caused the removal of alternating segments. Without the expression of a segment polarity gene, all segments in the mutant were reduced and mirrored.

The regulatory interactions that govern pattern formation have been studied in largest detail in the fruit fly. Initially, interactions between pairs of genes were derived from protein measurements in mutants [13]. In these experiments, one gene is artificially switched off or switched on in the whole embryo, while the concentration of a protein from another gene is measured. An activating influence is detected if the knockdown of one gene causes a decreased concentration of the other protein, while an increased concentration indicates repression.

This method of deriving regulatory interactions by comparing patterns in mutants to natural patterns clarified the main regulatory mechanisms of segmentation pattern formation in *D. melanogaster*.

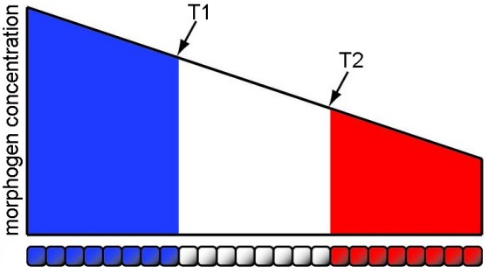


Figure 1.6: Early French Flag model. Thresholds T1 and T2 in the concentration of a morphogen gradient determine which target gene is expressed in each cell (indicated by blue, white, and red). (Adapted from [38] with permission of Johannes Jaeger)

Still, the exact network of regulatory interactions could not be determined. Experiments with mutants can not distinguish a direct interaction from an indirect interaction. In an indirect interaction, a regulatory protein does not interact with the target gene, but with an intermediary regulatory gene instead. Computational models and computer simulations for gene expression help to solve this issue by evaluating extensive gene regulatory networks.

1.2 Gene regulation network modeling

A gene regulation network model represents a collection of genes with the modeled influences among these genes. Many models have been developed for various purposes. The earliest models for gene regulation are no discriptions of gene networks, but simply propose an explanation for static gene product concentrations. For example, the French flag model aims to explain how a spatial distribution of cells responds to a morphogen gradient [14]. The French flag model basically states that a spatial gene expression pattern is formed if cells can measure the concentration of a morphogen gradient and activation of different genes depends on specific concentration thresholds (Figure 1.6). More complicated models are based on mathematical equations for interactions among genes, on development of expression over time or on adjustable parameters from biochemical measurements [15,16]. Gene circuit models are an example of spatial, dynamic (time inclusive), parametric, gene regulation network models.

The gene circuit method consists of four steps that are regarded as a standard procedure in dynamic, data-driven modelling [17]. First, a mathematical framework is defined, that includes unknown parameters to represent the system properties of interest. The gene circuit formalism is a set of coupled differential equations that represent the changing concentrations of the studied genes:

$$\frac{dv_a^i}{dt} = R_a g(u_a) + D_a\left[\left(v_a^{i-1} - v_a^i\right) + \left(v_a^{i+1} - v_a^i\right)\right] - \lambda_a v_a^i$$

$$u_a = \sum_b T_{ab} v_b^i + m_a v_{Bcd}^i + h_a$$

(eq. 1.1)

The parameters in these equations are *R*, *T*, *m*, *h*, *D* and *λ*. The production rate *R* represents the maximum value for the production of protein *a*. The interaction matrix *T* describes how strong protein *b* regulates gene *a*. The maternal interaction *m* describes how strong the Bicoid gradient regulates gene *a*. Any regulation of gene *a* from constant factors is included in *h*. The diffusion coefficient *D* represents the rate of protein transport between neighbouring nuclei. The decay rate *λ* describes how rapid protein *a* is broken down. The remaining symbols are the concentration *v* of protein *a* at position *i* and the sigmoid function *g* that scales the total regulatory input *u* between 0 and 1.

Second is the collection of systematic, quantitative gene expression data. Spatial expression profiles for five gap genes were measured at nine time points. These profiles were quantified by averaging the expression around the long axis of the embryo (Figure 1.7). Third, the model is fitted to the data by executing the model with many different sets of parameters. An algorithmic optimization strategy

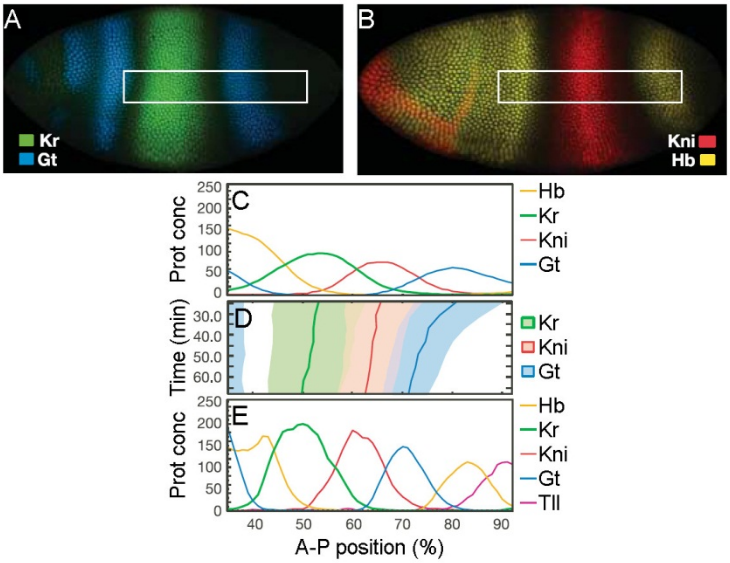


Figure 1.7: Gap gene expression pattern in the fruit fly. A,B) The protein concentrations of the genes kruppel (Kr), giant (Gt), knirps (Kni) and hunchback (Hb) have been visualized with fluorescence microscopy. Spatial protein concentrations have been measured for 9 time points in a region around the primary axis (indicated with the white frames). The profiles are plotted for the first (C) and last (E) time points, along with an overview of progressing expression of three genes (D). (Adapted from [17]; Reprinted by permission from Macmillan Publishers Ltd: Nature 430(6997):368-71, copyright 2004)

attempts to find model parameters that generate simulated expression profiles matching the measured expression profiles. The fourth step is an analysis of the optimized model to evaluate new insights. The rate of change d*v*/d*t* of a gene product at fixed positions was plotted over time, and this graph was decomposed into the contributions of the modelled processes.

1.2.1 Estimation of model parameters

In biological systems, protein levels depend on many regulatory and environmental interactions. These interactions are too many and too complex to be accurately included in a mathematical model. A model can only include a limited number of processes and represent these processes with a limited number of parameters. With these simplifications, finding the set of model parameters that produce the observed protein levels is still a nontrivial task. If a model is optimized by tuning one parameter at a time, a local optimum is found. To arrive at an absolute optimum, a global optimization strategy is required.

Simulated annealing is the oldest method for a global parameter search, and has been applied to find the gene circuit model with the closest fit to a systematic set of spatiotemporal gene expression data [18]. From a random parameter set, simulated annealing starts with adjusting parameter values and solving the mathematical equations, including combinations that produce poor results. As the search progresses, the tolerance on bad solutions is slowly lowered and the accessible parameter sets become restricted while the parameter values are fine-tuned towards an increasingly better solution. This optimization strategy is inspired from the physical process of slowly cooling a heated object, that results in an improved material strength [19].

Another approach to solve a problem with many parameters is evolutionary computing [20]. This strategy is inspired from the theory of evolution. The computational procedure can be described in terms of the agricultural processes of reproduction and selection, which result in crops with favorable traits. From a population of initial parameter sets, those with the best solutions to the mathematical equations are selected as seeds for reproduction. Parameter values from the seeds are exchanged to produce a new generation of parameter sets. From this new generation, the sets that provide the best solutions to the numerical model are selected as new seeds. This procedure is repeated for many generations until the solutions stop improving, which may suggest that a global optimum has been reached. Gene circuit parameters have been estimated for the *D. melanogaster* segment patterning with an evolution strategy, a subgroup of evolutionary algorithms [21].

1.3 Development in *Nematostella vectensis*

Fruit fly studies have revealed many principles of gene regulation. However, *D. melanogaster* exhibits biological and morphological properties that are not representative for other animals. Therefore, another model animal is needed to discover morphological principles and develop procedures that are generally applicable. Flies develop in a fashion that is quite unique across the animal kingdom, and their shape makes them an ideal model system. First, flies exhibit superficial cleavage, which means that the nuclei containing the DNA divide within a yolk sac that is enclosed by a single membrane (Figure 1.1C). A regulatory protein can diffuse toward another nucleus and directly bind to the DNA, while other animals require complex signaling between cells across membranes to achieve regulation in a neighboring nucleus. Second, the outline of fly embryos is uniform among individuals and remains constant during development because the embryo is encapsulated by a hard eggshell (chorion). Image segmentation and gene expression quantification is straightforward in this simple oval shape. Embryo images of other animals require tedious manual preprocessing to extract their complex, changing and less uniform morphology before gene expression is quantified.

The starlet sea anemone *Nematostella vectensis* is a recent model animal in the study of embryogenesis and gene regulation [22]. The name refers to its nematocysts, pressurized pockets that release poisonous threads to immobilize its prey. This animal is easily collected from salt marsh pools and cultured in a petri dish [23] and its small size and transparent body wall allow observation with all kinds of microscopic techniques. *N. vectensis* has a simple body

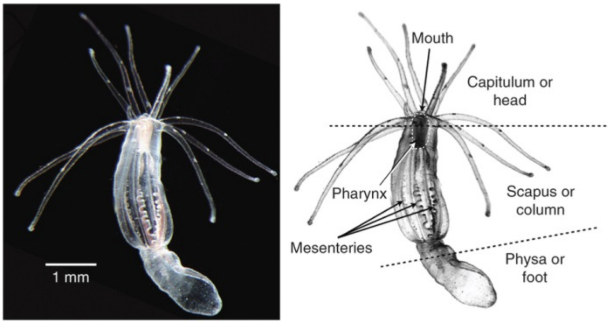


Figure 1.8: Anatomy of the adult *N. vectensis*. The mouth and tentacles are located on top of the body, while the throat (pharynx) and septa (mesenteries) are within a double-layered body wall. The body is divided in a head, a column and a foot. (Taken from [23]; Reprinted by permission from Macmillan Publishers Ltd: Nature Protoc 8(5):916-23, copyright 2013)

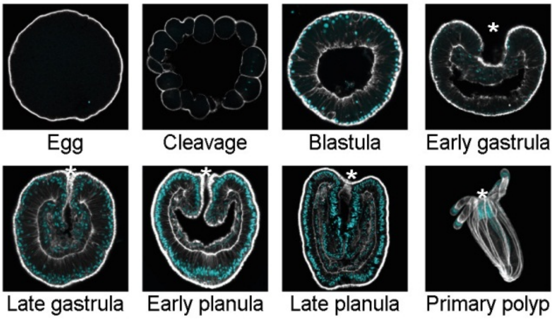


Figure 1.9: Embryonic development of the starlet sea anemone. The asterisk indicates the site of gastrulation, which develops into the mouth. (Adapted from [39] with permission of Eric Röttinger)

plan consisting of tentacles around a mouth on the outside and a throat and eight septa on the inside (Figure 1.8). *N. vectensis* belongs to the phylum Cnidaria, the group of aquatic animals distinguished by their nettle cells called cnidocytes. Cnidarians form two germ layers and display approximate or exact radial (rotational) symmetry, as opposed to bilaterians that form three germ layers and display clear bilateral symmetry.

Despite its simple body plan, *N. vectensis*’ early development and mode of gastrulation (Figure 1.9) are common across the animal kingdom. The cleavage produces cells of equal size and shape,

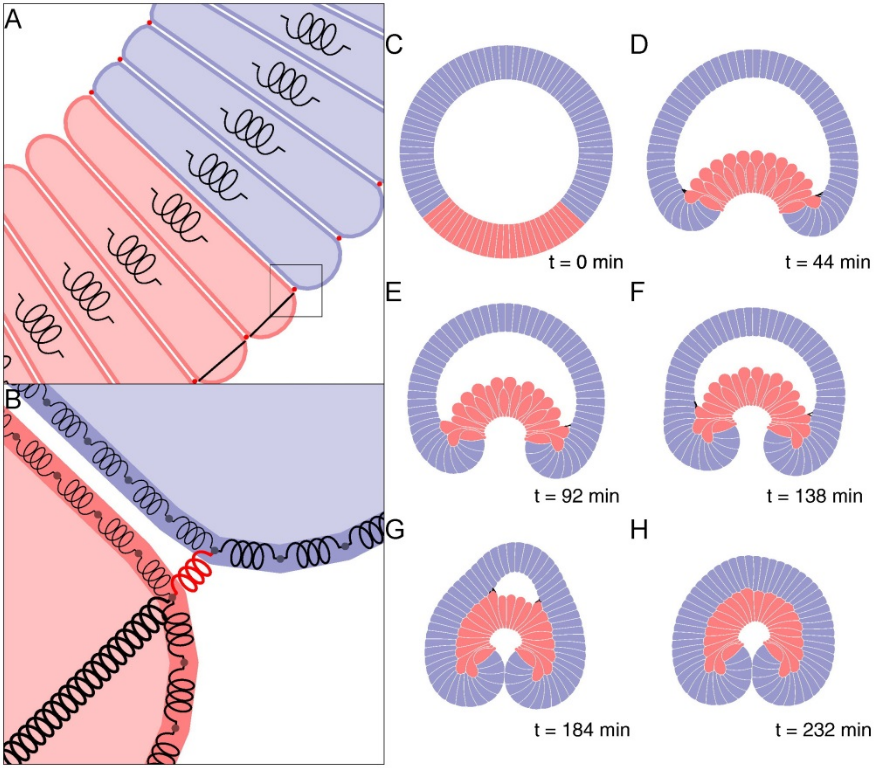


Figure 1.10: Modeled gastrulation of *N. vectensis*. A) A cell is represented as a string of 84 connected edges, filled with a slightly elastic fluid. The cells are connected with tight junctions (red dots) and the cells moving inwards contain an internal connection (black lines). The blue edges exert an adhesive force on edges of surrounding cells, while the red cells have no mutual adhesion. B) The edges behave like elastic springs, while the additional connections are modeled as stiff springs. C) In the initial setup, 87 wedge-shaped cells are positioned in a circle. The two cell types have distinct colors. D,E) The loss of adhesion and the constriction cause the red cells to fold inwards. F,G) Contraction of filopodia causes both layers of cells to zip up. H) Gastrulation has finished when the gap between both layers has closed. Simulation time is indicated at the bottom right. (Adapted from [25]; Reprinted from [25], Copyright 2011, with permission from Elsevier)

resulting in a spherical coeloblastula. During gastrulation, the embryo remains a continuous cell layer with cylindrical symmetry [24]. In the planula (larval) phase the pharynx, the first septa and the first tentacle buds appear. After settling on a surface, the larva transforms into a polyp and matures towards its final adult shape and size.

During gastrulation, the *N. vectensis* embryo changes from a blastula with a single layer of cells to a planula with a double cell layer. This major shape change is induced by the mechanical properties of individual cells. A recent mechanical model achieves invagination as a result of reduced stickiness and one-sided contraction in a patch of cells (Figure 1.10) [25]. Constriction and partial loss of adhesion, combined with the formation of filopodia that pull these cells against the opposite side, generate a complete closure of the blastocoel. This model shows that a change of properties in a set of cells is a sufficient requirement for gastrulation, without a need for preprogrammed trajectories.

Cellular properties are determined by the set of genes expressed in a cell [26]. Various regulatory genes are expressed exclusively in the invaginating cells, so gastrulation is regulated by a gene network [27,28].

Gene regulation in *N. vectensis*' embryonic development could be modeled in a similar fashion as *D. melanogaster*'s gap gene network. This would include constructing a formalism of coupled differential equations, measuring gene product distributions, estimating the model parameters with a suitable algorithm and analyzing the properties of the optimized model. However, obtaining quantified gene expression profiles is more difficult for *N. vectensis* than for *D. melanogaster*. *D. melanogaster*'s gap gene expression is quantified along a straight line, because the gene expression patterns are generated within a single cell membrane along the main axis. Additionally, no shape changes occur during the superficial cleavage stage and all samples are shaped identically due to the chorion encapsulation. This stable and reliable embryo shape allows a single automated image processing algorithm to quantify gene expression from all raw microscopy pictures [29].

Gene expression quantification in *N. vectensis* must take cell division, shape changes and shape variation into account.

1.4 Imaging techniques for visualizing cell movement and gene expression

Accurate models of cell dynamics or gene regulation in developing animals require knowledge of cell movement and gene expression in various stages of development. These processes are visualized with various imaging techniques.

Examples of developmental processes that have been visualized, are changing tissue shape during *N. vectensis* gastrulation [25] and gene activity in *N. vectensis* embryos [30].

1.4.1 Changing tissue properties in *N. vectensis* gastrulation

A cell's shape results from the configuration of its cytoskeleton. The cytoskeleton is a network of filaments and tubules, that connect all parts of a cell. The filaments, that mainly consist of the protein actin, are often found anchored to proteins in the cell membrane. Therefore, a fluorescent marker that binds actin is used to visualize the cell membrane [31]. For high-resolution detection, a confocal microscope scans through the sample to produce a three-dimensional map of points that produce a fluorescent signal. An image is constructed from a cross section of this map [32].

This technique allowed accurate measurements of a gastrulating *N. vectensis* embryo [25]. Volume and surface area increased for the endoderm and decreased for the ectoderm, suggesting that ectoderm cells continued to transform into endoderm cells during gastrulation. Moreover, the high-resolution images (Figure 1.11) served as a guide for the mechanical model of *N. vectensis* gastrulation (Figure 1.10).

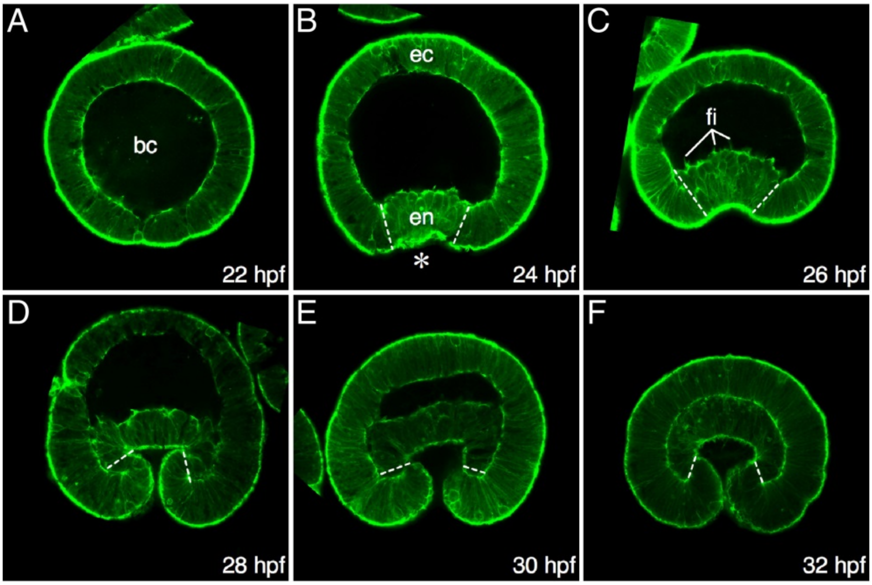


Figure 1.11: Gastrulation in *N. vectensis*. The times of development are indicated at the bottom right in hours after fertilization (hpf). Dashed lines indicate the endoderm-ectoderm boundary, bc = blastocoel, ec = ectoderm, en = endoderm, *= site of invagination, fi = filopodia. (Adapted from [25]; Reprinted from [25], Copyright 2011, with permission from Elsevier)

1.4.2 Differential gene expression during *N. vectensis* development

DNA is the genetic material located in the nucleus of nondividing *N. vectensis* cells. DNA molecules are long, double strands that contain the code for the amino acid sequence of proteins. The production of proteins occurs outside the nucleus in ribosomes, where free amino acids are catenated. The information is transported from the DNA to the ribosomes through short, single-strand copies called messenger RNA. As an RNA transcript is a single strand, it will bind uniquely to a complementary RNA fragment, forming a hybrid. If the complementary RNA strand is attached to a marker that activates a chemical or fluorescent agent,

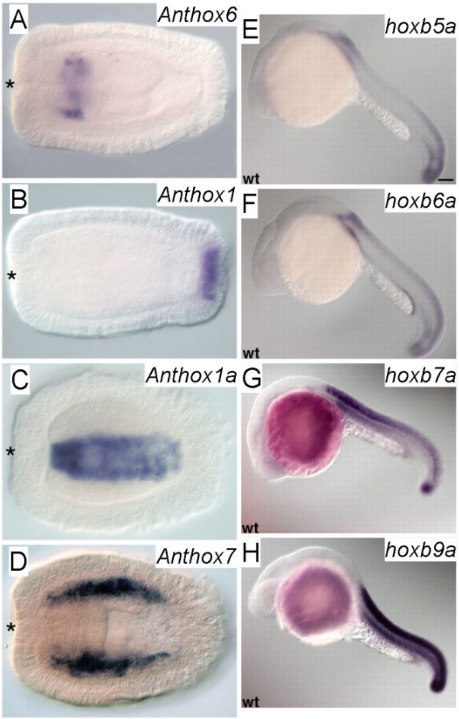


Figure 1.12: Hox gene expression patterns in *N. vectensis* and zebrafish. The probed genes are indicated in the top right of each panel; the prefix "Ant" refers to the taxonomic class Anthozoa that *N. vectensis* belongs to. In *N. vectensis* (A-D), hox genes are differentially expressed along both the main body axis (A,B) and the second body axis (C,D). *N. vectensis* embryos are in the planula stage, oriented with their future mouth (indicated with an asterisk) to the left. In zebrafish (E-H), hox genes are differentially expressed along the main body axis. Zebrafish embryos at 22 hours after fertilization are oriented with their future head to the left. (A-D adapted from [30] with permission of John Finnerty and E-H adapted from [34] with permission of The Company of Biologists)

it is visible under an optical or fluorescence microscope. In this fashion, RNA hybridization is applied to visualize the spatial distribution of RNA transcripts for a single protein [33].

This technique revealed that homeobox (hox) genes in *N. vectensis* are expressed in patterns along both the primary and secondary body axes (Figure 1.12A-D) [30]. Expression of hox genes is only observed along the main body axis in many other model animals, such as the zebrafish (Figure 1.12E-H) [34]. This complex hox pattern in *N. vectensis* is surprising, because only few organs in *N. vectensis* are affected by its second axis. In comparison, almost all zebrafish organs display bilateral symmetry and need to form at a single correct location on the zebrafish second body axis.

1.5 Purpose of this thesis

In this thesis we explore how spatial gene expression images can increase our knowledge and understanding of gene regulation in animals during embryonic development. Computational methods to infer gene regulatory networks from the spatial distribution of gene products have already been described and validated for fruit flies without shape changes during axis patterning. We use the sea anemone *N. vectensis* as a biological model for animals with changing morphologies during embryonic development.

To find out how validated computational tools for analyzing gene expression can be applied to developing animals in general, we used three approaches. First, we want to compare spatial gene expression across developmental stages. For this purpose, we designed a method to quantify spatial gene expression from *in situ* mRNA hybridization pictures and put these expression patterns in standardized profiles. Second, we want to derive gene regulatory networks from spatial gene product distributions. For this purpose, we modeled the regulatory interactions among *N. vectensis* gut formation genes based on quantified spatial gene expression profiles. Third, we want to obtain new insights from spatial gene expression databases. For this purpose, we performed hierarchical clusterings and a statistical analysis of gene expression patterns from an *in situ* hybridization database.

1.6 Thesis outline

A detailed description of our method for processing images of spatial *N. vectensis* gene expression patterns is provided in Chapter 2. Digital embryo geometries are produced from high resolution fluorescence microscopy pictures. A suitable digital geometry is fitted to the outline of an RNA hybridization image. The color intensity, which decreases with dark RNA staining, is measured along the cell layer outline. Based on the symmetry of the gene expression domain, a one-dimensional description of the expression pattern can be sufficient. Otherwise, a sensible transformation is performed and multiple 1D profiles are combined into a 2D profile. Chapter 3 proposes a computational workflow for producing a gene interaction network, and this procedure is applied on the *N. vectensis* gut formation. The construction of digital geometries, the quantification of spatio-temporal gene expression patterns and the computational inference of a gene regulation network are the basic parts of the proposed workflow. The first result is a preliminary regulation network with genes that appear in hugely different animals.

Chapter 4 is an analysis of *N. vectensis* gene expression images from the Kahi Kai (Hawaiian for ‘one ocean’) database. Digital gene expression profiles have been derived for over a hundred RNA hybridization images, and the profiles are clustered for each stage of early *N. vectensis* development. Based on the cluster analysis, three major regions are identified in each stage of development that likely contain corresponding cell lineages. For many genes, a shift of gene expression across these regions is initiated during the short period of gastrulation. Such a large shift in gene expression occurs during the longer larval stage for relatively few genes.

2 Spatial gene expression quantification: a tool for analysis of *in situ* hybridizations in sea anemone *Nematostella vectensis*

This chapter provides a detailed description of the spatial gene expression quantification method. The chapter is based on the following publication: Botman D, Kaandorp JA (2012), Spatial gene expression quantification: a tool for analysis of *in situ* hybridizations in sea anemone *Nematostella vectensis*, *BMC Res Notes* **5**: 555.

Abstract

Spatial gene expression quantification is required for modeling gene regulation in developing organisms. The fruit fly *Drosophila melanogaster* is the model system most widely applied for spatial gene expression analysis due to its unique embryonic properties: the shape does not change significantly during its early cleavage cycles and most genes are differentially expressed along a straight axis. This system of development is quite exceptional in the animal kingdom.

In the sea anemone *Nematostella vectensis* the embryo changes its shape during early development; there are cell divisions and cell movement, like in most other metazoans. *N. vectensis* is an attractive case study for spatial gene expression since its transparent body wall makes it accessible to various imaging techniques.

Our new quantification method produces standardized gene expression profiles from raw or annotated *N. vectensis in situ* hybridizations by measuring the expression intensity along its cell

layer. The procedure is based on digital morphologies derived from high-resolution fluorescence pictures. Additionally, complete descriptions of nonsymmetric expression patterns have been constructed by transforming the gene expression images into a three-dimensional representation.
We created a standard format for gene expression data, which enables quantitative analysis of *in situ* hybridizations from embryos with various shapes in different developmental stages. The obtained expression profiles are suitable as input for optimization of gene regulatory network models, and for correlation analysis of genes from dissimilar *N. vectensis* morphologies. This approach is potentially applicable to many other metazoan model organisms and may also be suitable for processing data from three-dimensional imaging techniques.

2.1 Introduction

Spatial gene expression assays are a substantial tool for verifying predicted regulatory interactions and for predicting properties of missing components in a regulation network [40,41].Their largest potential is in inferring parameters for numerical models of regulatory interaction networks, which is demonstrated for the embryonic development of the fruit fly *Drosophila melanogaster*. To perform accurate simulations, the spatial gene expression patterns are quantified and formatted to consistent profiles.

In systems biology, computational tools have become indispensable for deriving and validating gene regulation networks [15]. In data-driven modeling, parameter estimation can determine which set of rules represents the best network model to match a set of observations. Figure 2.1 shows an overview of the modeling cycle. First, a general mathematical framework is selected. The gene

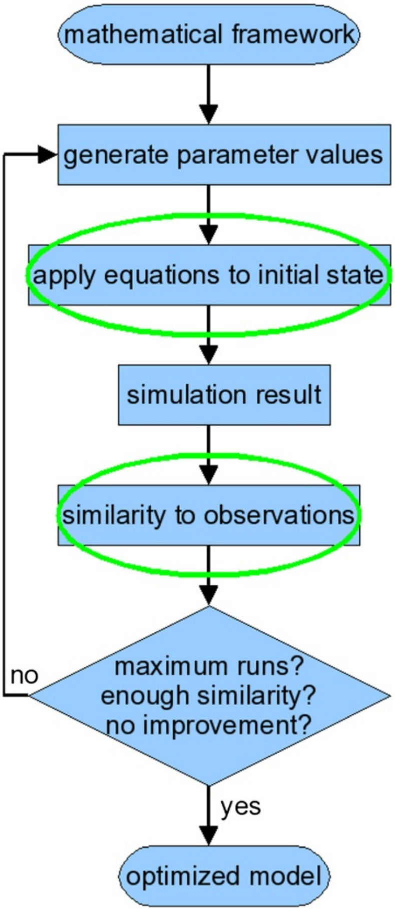


Figure 2.1: Overview of the modeling cycle. The modeling cycle starts with a framework of general mathematical equations. Initial parameter values are randomly generated or manually provided. These values are substituted into the general framework to define a specific set of equations. The equations are applied to the initial state of the system (usually derived from measurements) and produce intermediate and final states. These simulated states are compared to reference data and their similarity is determined. New parameter values are generated and new simulation runs are performed repeatedly, while stopping conditions are tested after each run (such as a maximum number of runs, a target similarity or a lack of improved similarity after multiple runs). As soon as a stopping condition applies, the cycle is terminated and the set of parameter values that results in the closest match with the observations is the optimized model. The steps that require quantitative data are encircled.

circuit model [17,18] (which is derived from the connectionist model [42]) is a convenient formalism that does not require knowledge about interactions or their mechanisms. The actual modeling starts with choosing parameter values for the general equations; the initial parameters are defined by the user or generated by an algorithm. The resulting equations are applied to the concentration profiles at the first timepoint, derived from experimental data. A simulation is basically the repeated application of the parametrized equations to the newly obtained profiles, which produces concentration profiles at various timepoints. The simulated profiles are compared with observed expression profiles and the similarity is calculated with a predefined fitness measure. An optimization algorithm then picks new parameter values and starts a new simulation cycle. The optimization process ends when a stopping criterion is reached, such as a fixed time duration or number of cycles, a satisfactory fitness or a series of cycles without significant fitness improvement. The output is a set of parameters that defines the equations which represent the observed gene expression profiles best. For a meaningful solution, the fitness criterion should be well-defined. Data overfitting occurs if multiple parameter sets provide equally good solutions and this should be avoided [43].

A good computational model for gene regulation accurately replicates the observed gene expression patterns. The modeling cycle serves as a robust procedure for data fitting, without guaranteeing realistic results. Biological mechanisms are included in the process of model building in several ways. First, prior knowledge about biological mechanisms can be incorporated in the design of the mathematical formalism. Second, if a computational model reproduces gene expression without overfitting, this is a strong clue that the formalism approximates biological principles. Third, a thorough parameter analysis can reveal new information about a biological system; for example, correlating parameters may indicate entangled biological processes [44].

In any modeling approach, meaningful results require accurate reference data to initiate the simulation and to evaluate the simulation output. For gene regulation networks, this means that gene expression patterns should be quantified in a consistent manner. A quantification procedure has been created and validated for the fruit fly [29].

Gene expression in *D. melanogaster* is quantified along a straight line during superficial cleavage. In this stage, nuclei are located within an embryo sac that does not significantly change its shape while the nuclei are dividing (Figure 1.1C). In most other animals however, nuclear division is coupled to cell division during cleavage and the early embryo displays rapid cell movement and morphological changes. This is why we developed a method for gene expression quantification that accounts for a complex and changing embryo morphology.

Over the past decade, *Nematostella vectensis* has become an important model organism in the field of evolutionary developmental biology [22]. As a research object, the animal is easy to culture and its small size and transparent body wall make it suitable for all kinds of microscopy. Subsequent gene expression studies and the sequencing of the genome have shown that *N. vectensis*, curiously, shares more genes with humans than either *D. melanogaster* or the roundworm *Caenorhabditis elegans* [45]. Much work on *N. vectensis* has been dedicated to the genetic regulation of development [46].

The early developmental stages of *N. vectensis* are displayed in Figure 2.2 (adapted from [47]). The *N. vectensis* body wall consists of an outer cell layer (ectoderm) and an inner layer (endoderm). We are primarily concerned with the invagination process called gastrulation, when the presumptive endoderm moves inwards and covers the ectoderm. The side of invagination is the location of the

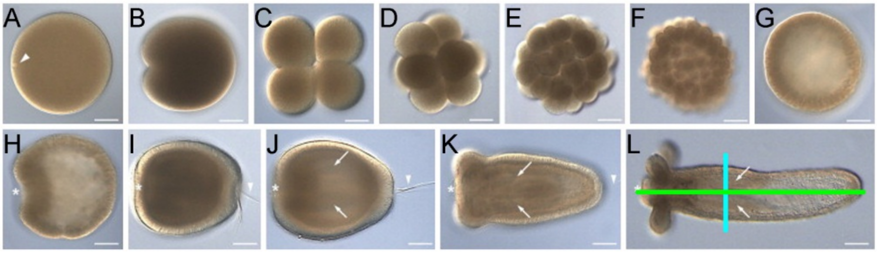


Figure 2.2: Various stages of *N. vectensis* embryonic development. Development stages from egg to polyp are shown, with the oral pole to the left in panels H-L (indicated with an asterisk). A) Unfertilized egg, with the female pronucleus visible near the cell membrane at the oral pole (arrowhead). B) The first cleavage (1-2 hpf) occurs at the oral pole of the fertilized egg. C-F) Cleavage stages, with the 4-cell stage often arising after the first two cleavages finished simultaneously (C). G) Cleavages result in a hollow sphere of cells called a coeloblastula (10-20 hpf). H) Invaginating cells at the oral pole mark the beginning of gastrulation (24 hpf). I) The early larva (72 hpf) displays a doubly layered body wall and swims with the aboral end forward (arrowhead indicates the apical organ). J) The late larva starts to elongate and to form two directive mesenteries (arrows). K) Early polyp, with four tentacle buds around the mouth at the oral pole. L) Juvenile polyp; the primary and secondary axes are indicated with a green and a blue line, respectively. Scale bars are 60 µm in panels A-H and 90 µm in panels I-L. Development times in hpf = hours past fertilization (original image from [47], with development times estimated from [24,63]).

future mouth and is therefore referred to as the oral pole; the opposite side is the aboral pole. In adult polyps, the gastric cavity is subdivided by eight mesenteries. These tissue folds contain retractor muscles, digestive cells, nettle cells and reproduction organs. The first two mesenteries appear in the planula and are referred to as the directive mesenteries. *N. vectensis* displays bilateral symmetry: the line between the oral and aboral ends is the primary body axis, while the line through the directive mesenteries defines the secondary or directive axis. The primary and secondary axis are marked with a green and a blue line, respectively, in Figure 2.2L.

In this paper we discuss a geometric method for extracting quantitative spatio-temporal gene expression data from *in situ* hybridizations in the sea anemone *N. vectensis*. We measure gene

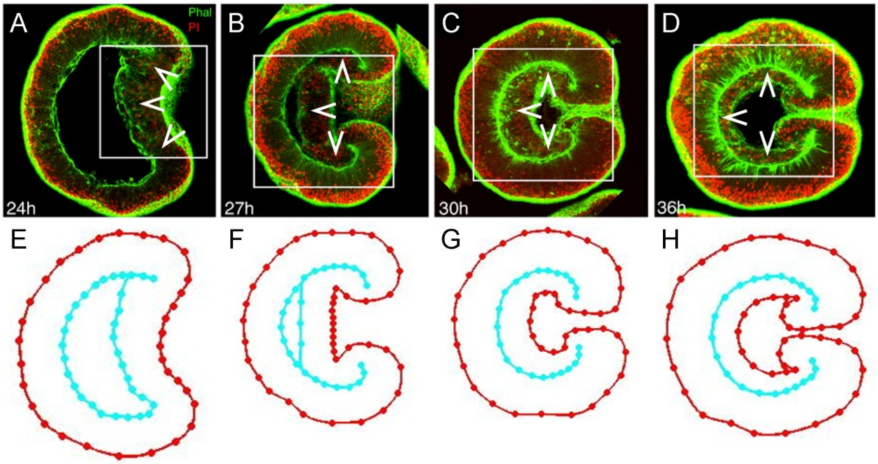


Figure 2.3: Graphical embryo morphologies derived from confocal microscopy images. A-D) *N. vectensis* embryos were stained at various stages of gastrulation with fluorescent markers for filamentous actin (phalloidin in green) and nuclei (propidium iodide in red). Arrowheads indicate the endoderm. The embryos are oriented with the blastopore to the right. Development times after fertilization at 16 °C are indicated in the lower left corner. The images are modified from [24]. A) Cells at the oral pole are invaginating. B, C) The inner cell layer (endoderm) is zipping up with the outer layer (ectoderm). D) The endoderm is flattening against the ectoderm. E-H) Based on these confocal cross sections, average cell layer geometries have been constructed. The cell layer outlines are closed loops; the inner loop overlaps itself where endoderm and ectoderm are zipped up. Only a selection of confocal micrographs and embryo geometries is shown [48].

expression during gastrulation using a gene expression quantification tool developed by de Jong [48]. We show some preliminary results how quantified gene expression profiles have been analyzed with hierarchical clustering.

2.2 Results

2.2.1 Gene expression quantification

Microscopy images accentuate cell layer outlines of *N. vectensis* embryos stained for nuclei and filamentous actin (Figure 2.3A-D, copied from [24]). The latter is bound to the membrane as part of the cell cortex, so it highlights cellular shapes. From these images, average embryo morphologies have been derived for various stages of gastrulation (Figure 2.3E-H). An embryo geometry is derived from a confocal microscopy image (such as Figure 2.3A) by placing nodes on the cell layer boundaries. Node locations from multiple (2 to 5) geometries are averaged to obtain an average embryo geometry (such as Figure 2.3E). This averaging reduces the effect of local irregularities, as the average geometry is supposed representative for embryos of a particular age. From the nodes of the average geometries, a set of points is selected for interpolation of geometries in subsequent stages. These points include main morphological features such as the boundary between the ectoderm and the endoderm, and form an accurate spline. Interpolation of these average morphologies results in a continuous range of embryo morphologies.

In the following example, these graphical shapes are applied to quantify gene expression patterns with the GENEXP program (Appendix A1). Published *N. vectensis* gene expression images are collected in the CnidBase [49] and Kahi Kai [50] databases, and Kahi Kai also contains expression images outside journal publications. Still, many *N. vectensis* expression pictures are found in publications outside these databases.

Example 1: a 1D expression profile from a symmetrical pattern

Figure 2.4A (adapted from [51]) displays an *Nvnos2 in situ* hybridization in a late gastrula. The transcripts were hybridized with digoxigenin-labeled RNA probes. The published image is overlaid with the most similar morphology. The graphical geometry is then adjusted to the outline of this particular embryo by dragging the nodes to their final location (Figure 2.4B). The nodes are connected by a curve (about 10^5 points) calculated with cubic spline

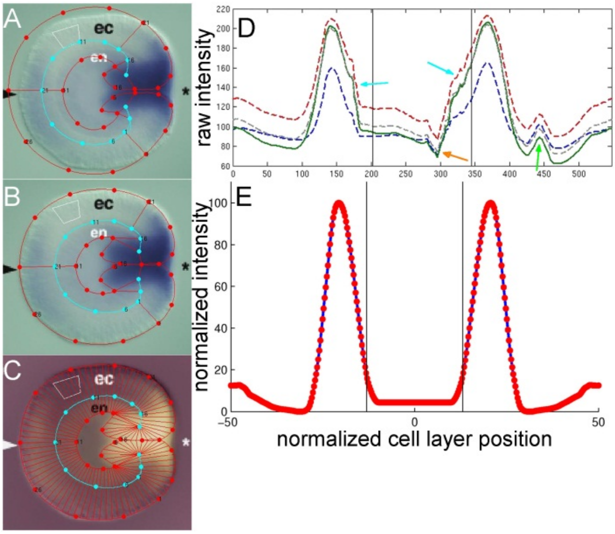


Figure 2.4: Steps in obtaining a quantified gene expression profile from an *N. vectensis in situ* hybridization. A) An unaltered embryo geometry (chosen from a list of interpolated geometries) is projected on an in situ hybridization of *Nvnos2* at the end of gastrulation (original image from [51]). In the original image, the arrowhead highlights gene expression in the aboral ectoderm and the asterisk denotes the future mouth. Endoderm and ectoderm are labeled ‘en’ and ‘ec’, respectively. B) The points in the graphical geometry are manually dragged over the cell layer boundaries of the observed *N. vectensis* embryo. C) After applying a color inversion, the embryo is decomposed into parallel sections along the cell layer. D) For each section in figure 2.4C, the average red, green, blue and greyscale intensities are plotted in an expression profile as a function of its position along the cell layer. The segment on the aboral end of the decomposition corresponds to 0 on the horizontal axis; from here, the decomposition proceeds counterclockwise along the cell layer. After crossing the endoderm-ectoderm boundaries (which correspond to the vertical black lines in the plot), the decomposition again reaches the aboral end, corresponding to the right side of the horizontal axis in the intensity plot. Arrows highlight artefacts (see example 1 in the main text). The main profile is plotted as a solid graph. E) After vertically and horizontally shifting the main profile from panel D, artefacts are removed, noisy regions are smoothed and the profile is symmetrized. On the horizontal axis, the endoderm center lies at normalized cell layer position 0 and the ectoderm center is normalized at −50 and +50, while the maximum intensity is normalized to 100 on the vertical axis.

interpolation [52]. The geometry is automatically decomposed into parallel segments along the cell layer and color intensities are measured for all segments. The decomposition is performed by dividing the outer curve into sections with a user-defined length and calculating the nearest point on the inner spline for each boundary. This is repeated for the inner curve at sections with large gaps. The decomposition is finished by smoothening the boundary points on both curves to obtain segments that are more uniformly spaced. For the expression profile the red, green and blue color intensities of all pixels enclosed within a segment are averaged. These average intensities are plotted against the position on the line through the segments' centers. To make an increasing intensity indicate an increase in concentration, the colors are inverted in Figure 2.4C and the corresponding intensities along the cell layer are displayed in Figure 2.4D.

The graph in Figure 2.4D contains features that do not represent the actual transcript concentration. The "en" and "ec" annotations cause a trough (orange arrow) and a ridge (green arrow), respectively, while an imperfect decomposition has shifted the peaks with regard to the center and introduced some noise (blue arrows). Moreover, the nonuniform background and nonsymmetric lighting cause an asymmetric baseline. The profile is exported to an editor for additional processing to correct for these features. To remove artefacts caused by annotations, the user selects this section of the graph and can choose among linear interpolation, cubic spline interpolation, piecewise cubic Hermite interpolation [52] and replacement with a specified constant. Noise from erroneous decomposition is smoothened with a lowpass filter (with filter coefficients equal to the reciprocal of the span), known as 'moving average' [53]. The graph is lowered with a constant value to subtract the average background and regions without observed expression are put to zero. Both halves are averaged to cancel

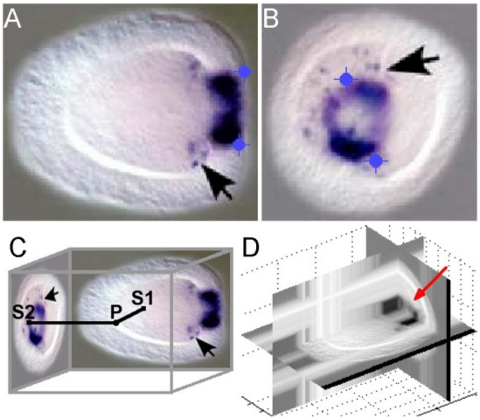


Figure 2.5: A three-dimensional array constructed from perpendicular *N. vectensis* embryo views. A,B) *NvFoxB* expression in the early larva appears focused on two spots around the oral pole (original images from [54]). Reference points are indicated as blue crosshairs on lateral (A) and oral (B) views to shift and scale the oral image's height. In the original images, the black arrows highlight isolated expression in some endodermal cells. C) A coordinate sweep is performed over the three-dimensional array and the corresponding points *S1* and *S2* on the lateral and oral images, respectively, are determined. The point with minimal signal is selected for the new array **P**. D) The greyscale intensity is displayed in three perpendicular planes of the 3D gene expression array. The main features of the expression pattern in panels A and B can be identified in this representation (red arrow).

nonsymmetric influences. The final expression profile is plotted in Figure 2.4E.

2.2.2 Three-dimensional expression pattern reconstruction

For expression patterns that are radially symmetrical around the primary axis, a one-dimensional profile is a complete description. However, most signaling pathways involve genes that are asymmetrically expressed along the secondary axis. A three-dimensional representation is required to fully define the expression pattern of these genes. Depending on the data available for a gene expression pattern, a suitable method is determined for the approximation of this 3D pattern.

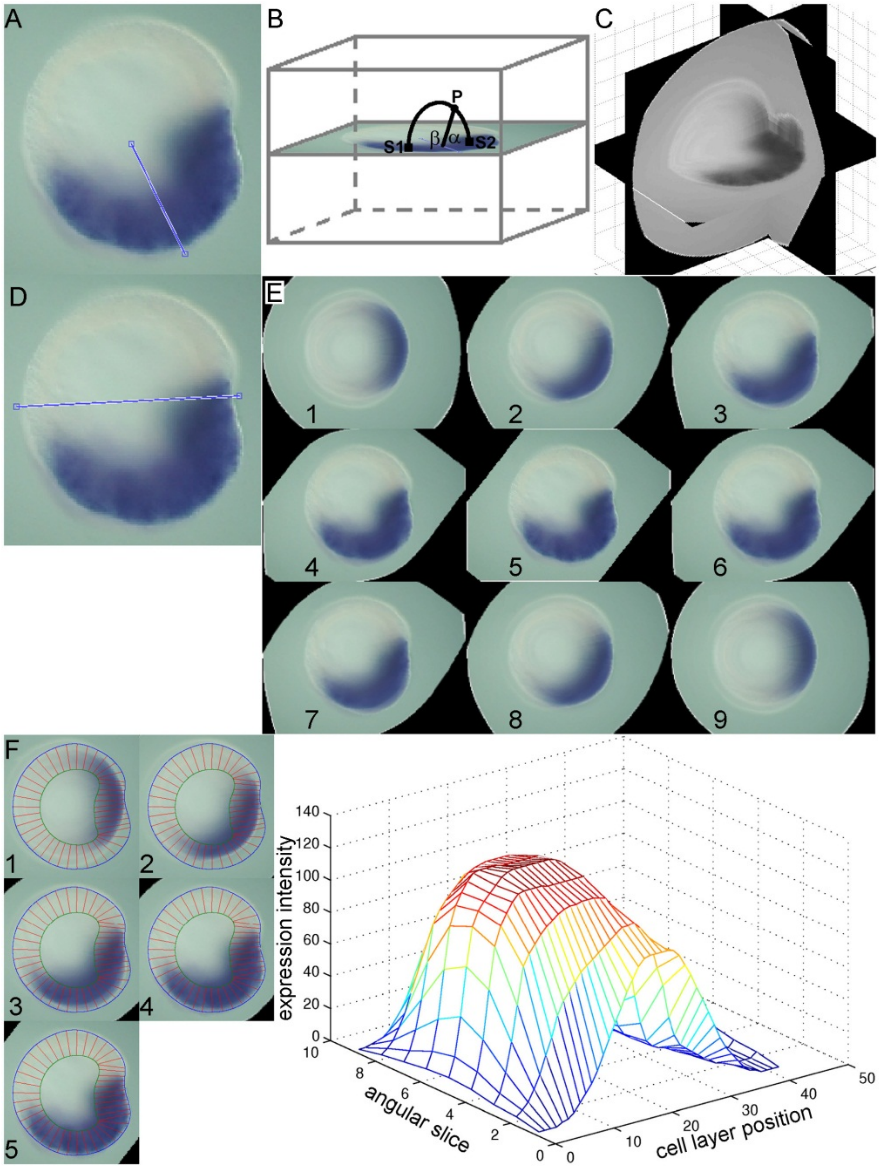


Figure 2.6

If a set of perpendicular expression images shows that an expression pattern is not radially symmetrical, then a 3D representation is constructed by mixing both images into a three-dimensional array (Figure 2.5C) with the TWOVIEWS program (Appendix A3). The 3D array contains the pairwise expression minimum of the points in

Figure 2.6: A 2D quantified expression profile from a single *N. vectensis* gene expression image. A) At the start of gastrulation, the gene *Nvvas2* is expressed roughly symmetrically with respect to the blue axis drawn (image from [51]). B) For all points of the 3D array **P**, the base points *S1* and *S2* of the arc centered on this axis are determined. The intensity is calculated as the weighted average of these base points, as expressed in the algorithm in Pseudocode 2.1. C) The greyscale intensity is displayed in three perpendicular planes of the 3D gene expression array. The horizontal plane is the truncated original. D) The primary axis is indicated on the original image. E) From the 3D expression, slices are calculated at equally incremented angles through the primary body axis. F) The original embryo geometry is overlaid on the first five pattern slices from figure 2.6E (the last four are identical in reverse order). As expected, the embryo geometry does not match the expression outlines. The geometry is not adjusted to the expression, because the expression representation is not meant to approach the embryo's shape. Besides, the expression outline outside the central *x*,*y* plane is too blurred for a precise geometry extraction. (The distortion is more obvious with a dark background, as seen in Figure 2.5D.) G) After processing each 1D expression profile derived from the decomposed angular slices, the complete two-dimensional expression array is plotted as an intensity landscape. The cell layer position follows the segments in the decomposition in counterclockwise direction, starting and finishing at the aboral end. The angular slice designations correspond to the number labels in panel E. Processing of onedimensional profiles is done in a similar fashion as in Example 1 (main text).

each horizontal plane. Otherwise a radial expression pattern symmetry is assumed, and a 3D representation is created by smoothly averaging pairs of points on the image along circular arcs about an approximate axis of symmetry (Figure 2.6B) with the ONEVIEW program (Appendix A2). The pseudocode for this operation is displayed in Pseudocode 2.1.

Example 2: a 3D representation from perpendicular embryo views

Figure 2.5A,B (adapted from [54]) shows the expression pattern of gene *NvFoxB*, which is concentrated on two spots on opposite sides of the oral end. Two sets of reference points are picked to scale and

Pseudocode 2.1: calculating a 3D array from a single expression image with a manually drawn line dividing the expression domain

```
input: line $L$ drawn on the expression image ($dim1*dim3$ pixels)
output: 3D expression array $\mathbf{P}$
for $x$ from 1 to $dim1$ do
  for $y$ from 1 to $dim3$ do
    for $z$ from 1 to $dim3$ do
      [calculate distance $d$ from $\mathbf{P}_{xyz}$ to $L$]
      [calculate coordinates of points $S1$ and $S2$ located
         distance $d$ on both sides of $L$]
      [calculate angle $\alpha$ between $S2$, $L$ and $\mathbf{P}_{xyz}$]
      [calculate angle $\beta$ between $S1$, $L$ and $\mathbf{P}_{xyz}$]
      $\mathbf{P}[x][y][z] = (\alpha * S1 + \beta * S2) / \pi$
return($\mathbf{P}$)
```

align both pictures. The height of the oral image is adjusted to match these points in the lateral image.

Elements for the 3D array that represents the volumetric expression pattern are calculated as the minimum expression from the associated pair of pixels (Figure 2.5C). A greyscale visualization of this array is displayed in Figure 2.5D. Two domains appear in the oral region as expected.

The 3D array is an intermediate step in the quantification process that is completed for the next example.

Example 3: a 2D expression landscape from a single embryo view

Figure 2.6A (based on [51]) shows the expression of *Nvvas2* in the early gastrula stage. The expression domain covers the embryo’s lower half and its future mouth. A line is drawn on the image that divides the expression domain. Each element in the 3D array is the weighted average of the image pixels at both ends of a circular arc around this line (Figure 2.6B,C).

This discrete volumetric expression array is not yet suited to be compared to other patterns, because their shapes do not match or their cell layers are located at different Cartesian positions. To arrive at consistent profiles, slices are cut through the primary axis and decomposed. The primary axis is drawn on the original image and slices of the 3D array through this axis are constructed (Figure 2.6D,E).

These slices are overlaid with a geometry that fits the native image (Figure 2.6F) and through the decomposition procedure described in the one-dimensional example, a set of profiles is produced. These profiles are stacked in a 2D array and displayed as a landscape in Figure 2.6G.

2.2.3 Visualization and clustering

The profiles are easily interpolated and stored in arrays of equal length. Such a collection of standardized arrays can serve as input for conventional comparison and analysis software. For example, Figure 2.7 shows 41 1D expression profiles from 20 genes, ordered with hierarchical clustering, using Pearson correlation and unweighted average linkage. In this fashion, a database is conveniently displayed and correlating patterns can be identified at a glimpse.

From the cluster tree, the patterns are divided in three main groups, and five individuals with little similarity. In green, all four asymmetric expression patterns are included. The purple patterns are restricted to the endoderm. The yellow group is the largest, containing genes that are expressed in the presumptive pharynx and mouth. The remaining genes are expressed in ectoderm away from the mouth.

The clustering displayed in Figure 2.7 might be somewhat artificial as the expression domains clearly overlap. Moreover, some regions are elongating faster during gastrulation than others and even after

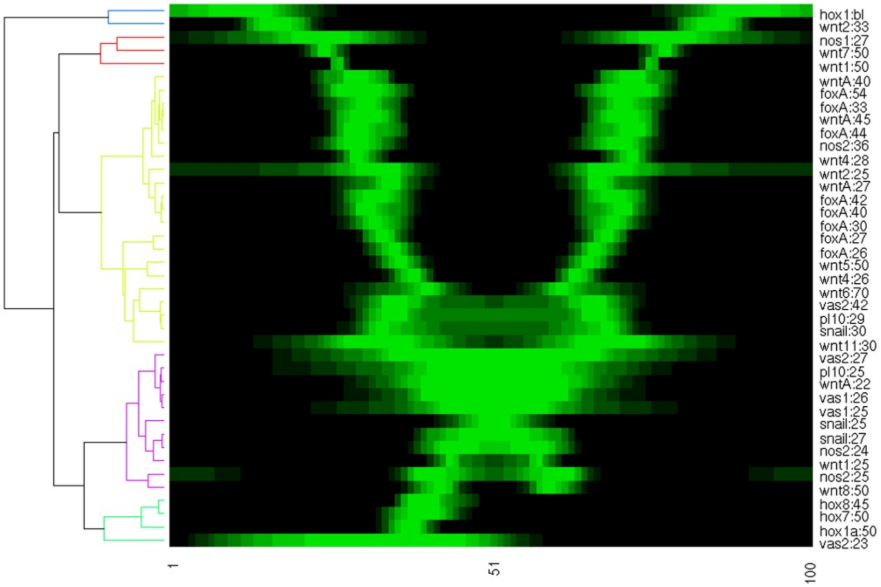


Figure 2.7: Synoptic display and clustering of 1D expression profiles of various genes at different development times. In this hierarchical clustering, the Pearson correlation coefficient was applied for the similarity measure and unweighted average linkage for tree formation. Normalized gene expression profiles are conveniently ordered based on their similarity in spatial intensity distribution. This compact overview contains expression patterns of *hox*, *wnt*, *nanos*, *foxA*, *vasa*, *PL10* and *snail* genes from various sources. Gene name and development time for each profile are displayed on the right in the format {gene}:{hours past fertilization}, except for *hox1* which is in the hollow sphere (blastula, bl) stage and therefore its age cannot be determined from its morphology. The gene expression intensity is scaled from black (no expression) to bright green (maximal expression). The dendrogram on the left shows the correlation between the spatial expression patterns. The dendrogram was cut at similarity 0.6 to obtain five groups (blue, red, yellow, purple and green from top to bottom).

pinning the estimated endoderm-ectoderm boundary, stationary patterns such as *NvFoxA* seem to migrate. Still, the comparison is very helpful in observing correlations and proposing hypotheses. For example, the three main groups may indicate regulatory modules. More specific, the patterns with broad boundaries belong to embryos in a relatively early stage, indicating a regulatory cascade in which fuzzy domain boundaries are sharpened, comparable to the *D. melanogaster* gap genes. An extended and systematic set of profiles would enable an inference of the developmental gene regulation network in *N. vectensis*, based on the modeling techniques and analyses that established many properties of the *D. melanogaster* gap gene network [21,55,56].

2.3 Discussion

Our quantification procedure provides a standardized format for the most diverse spatial visualization techniques. In this paper, hybridized mRNA has been quantified, but the method can be applied to any specific molecular entity. Potential examples include native proteins visualized with antibody staining and overexpressed proteins fused to a fluorescent agent [47,57].

Current limitations arise from the strong assumptions imposed on the images that are used to construct 3D representations. For a rotated pattern, it is formally assumed that only expression in the plane of dissection is visible, while in fact observed expression is not restricted to this plane.

More serious is the requirement that the embryos used for the reverse projection are completely transparent, while the endoderm and aboral ectoderm are hidden on most oral images. Sometimes, only the cumulative signal on the periphery of an expression region is detected, as in Figure 2.5A (the speckled region at the arrow and on the opposite side).

The embryo is also assumed to be viewed from exactly perpendicular angles, but the sample is often rotated imperfectly. Additionally, slight deformations can occur during rotation, causing small domains with granular expression to overlap improperly and thus to be misrepresented. These issues are observed in Figure 2.5 as well.

With the advance of direct three-dimensional imaging the volumetric array construction may become superfluous, and these limitations will be removed. Confocal laser scanning microscopy has already been applied to zebrafish [58], *D. melanogaster* [59] and sea urchin [60] embryos. This method may provide quantitative, spatial expression data for *N. vectensis* as well. Conversely, general methods for mapping these data to the embryo's morphology should be useful for comparison and analysis in these other organisms.

2.4 Future developments

An integrated method has been presented that combines geometry extraction and gene expression quantification. The basic concept is that gene expression is conveniently measured along the cell layer in a morphology that can be viewed as a continuous sheet of cells. This straightforward approach can be applied generally to embryos across the animal kingdom. As confocal laser scans with high three-dimensional spatial resolution are widely applied, application of this method is not limited to symmetrical body shapes.

We have shown how to extract quantified gene expression profiles from *N. vectensis in situ* hybridizations, and how a preliminary comparison and cluster analysis lead to new insights. The next step is to estimate parameters that describe interactions among genes in the *N. vectensis* regulatory network. The powerful methods designed for parameter inference [17,18,21] and network analysis [55,56] in *D. melanogaster* can now be applied to the standardized gene expression profiles of *N. vectensis* and other model species in genetics.

Currently, a database of published images is processed into 1D arrays, annotated with roughly estimated development times based on comparison with high resolution micrographs. This comparison is very subjective, as our designation often differs from the developmental stage originally claimed. Moreover, the embryos change very subtly in the hollow sphere stage (Figure 2.2G) and between invagination and mesentery formation (Figure 2.2I), so timestamps are highly ambiguous. If a gene or a combination of genes is found with continuously changing expression patterns, we can derive labelling protocols to determine the exact developmental time. (Registration techniques like this have already been described and proved useful for *D. melanogaster* [61].)

Spatial gene expression quantification can be combined with modern quantitative polymerase chain reaction (qPCR) techniques [62]. The absolute total amount of transcripts measured with qPCR coupled to the spatial distribution from quantified *in situ* images should enable the calculation of absolute local mRNA concentrations.

3 A computational approach towards a gene regulatory network for the developing *Nematostella vectensis* gut

This chapter describes how the spatial gene expression quantification method has been used to generate a preliminary gene regulatory network. The chapter is based on the following publication:
Botman D, Röttinger E, Martindale MQ, de Jong J, Kaandorp JA (2014), A computational approach towards a gene regulatory network for the developing *Nematostella vectensis* gut, *PLoS One* **9**(7):e103341.

Abstract

The starlet sea anemone *Nematostella vectensis* is a diploblastic cnidarian that expresses a set of conserved genes for gut formation during its early development. During the last decade, the spatial distribution of many of these genes has been visualized with RNA hybridization or protein immunolocalization techniques. However, due to *N. vectensis'* curved and changing morphology, quantification of these spatial data is problematic. A method is developed for twodimensional gene expression quantification, which enables a numerical analysis and dynamic modeling of these spatial patterns.

In this work, first standardized gene expression profiles are generated from publicly available *N. vectensis* embryo images that display mRNA and/or protein distributions. Then, genes expressed during gut formation are clustered based on their expression profiles, and further grouped based on temporal appearance of their

gene products in embryonic development. Representative expression profiles are manually selected from these clusters, and used as input for a simulation-based optimization scheme. This scheme iteratively fits simulated profiles to the selected profiles, leading to an optimized estimation of the model parameters. Finally, a preliminary gene regulatory network is derived from the optimized model parameters.

While the focus of this study is *N. vectensis*, the approach outlined here is suitable for inferring gene regulatory networks in the embryonic development of any animal, thus allowing to comparatively study gene regulation of gut formation *in silico* across various species.

3.1 Introduction

During animal development asymmetric signals set up during the early cleavage stages are utilized to initiate different pathways of cell type specific differentiation. Individual cells undergo a complex sequential and combinatorial pattern of differential activation/repression of gene activity that are causally required for the correct assignment of cell identity [64]. The body plan is thus formed by interactions between genes and proteins. A collection of such interactions defines a gene regulatory network (GRN).

A GRN can be described using mathematical models. The goal of modeling GRNs is to understand the basic properties of these networks. Various mathematical frameworks have been proposed for the description of GRNs [15]. Some models are quantitative, some models include time or spatial compartments, but combined quantitative spatio-temporal models are rare. Dynamic models that simulate quantitative gene expression levels in interacting domains can capture the formation of gene expression patterns during early animal development [65]. These dynamic simulation models are validated by their ability to reproduce spatio-temporal patterns based on experimental measurements.

The general model building process contains three main steps [17]. First, quantitative gene expression data is required, which is extracted from spatio-temporal measurements. Second, a modeling framework is established from a set of mathematical equations. Third, the parameters in the modeling framework are estimated: the optimal parameters produce simulated expression patterns that correspond to the quantitative gene expression data. An overview of the modeling cycle is shown in Figure 2.1.

Modeling GRNs has the advantage that parameters can be investigated without the noise and limited precision of experimental measurements. The influence of the proposed mechanisms can be tested without the interference of many other processes that occur in living systems. Moreover, new hypotheses can be generated from abstract model properties that cannot be obtained from experimental measurements. For example, Manu et al. [66] suggested that anterior and posterior regions in the early fly embryo move towards separate basins of attraction, based on a phase space analysis of their quantitative spatial dynamic model. Biologists may regard the inferred parameters as new hypotheses for conducting further experiments. On a lower level, the quantitative extraction of spatio-temporal gene expression patterns provides a convenient method to systematically organize, analyze and share these data among workers around the world.

When modeling methods are applied to investigate GRNs, some pitfalls should be avoided. The data quality, scope and usefulness of a model should be considered.

The reliability of a numerical model depends on the quality of the data that is supplied as input. For example, RNA *in situ* hybridizations can come from various laboratories, implying that the images could be produced with different light settings, operators and purposes. Moreover, differences may also arise from variation among individual samples.

Conclusions beyond the scope of the model should also be viewed with caution. If a model identifies spatial and temporal correlations between pairs of genes, then these should not be treated as interactions, even though direct or indirect influences would be the most straightforward cause of these correlations. However, these proposed interactions can be directly tested experimentally thus vetting the model's predictions.

Finding an optimal solution in problems with many unknown parameters can be computationally extremely intensive. Besides, the optimal solution is not necessarily the best approximation of the biological system. Analyzing multiple solutions from a repeated stochastic search to determine which parameters are most consistent (and therefore most reliable) is an alternative method. An analysis of many solutions can provide more information than the best solution from a single optimization run.

Currently, the most precisely described spatio-temporal regulation mechanism for early development is the gap gene network in the fruit fly *Drosophila melanogaster* [16,67]. One notable insight is the function of cross-regulatory interactions among gap genes [68]. These interactions are necessary for precise gap gene expression domains to emerge from a larger spread in maternal concentration gradients.

In comparison to most other metazoans, gene regulation in early fly embryos such as *D. melanogaster* is easy to understand, because the regulatory proteins do not require intermediate metabolites to interact with the DNA [69]. These straightforward regulatory interactions are coupled to the early fly morphology: no membranes are present during the first nuclear division cycles, so transcription factors can diffuse between nuclei. In other metazoan embryos, complex signaling pathways operate from the early cleavage stage

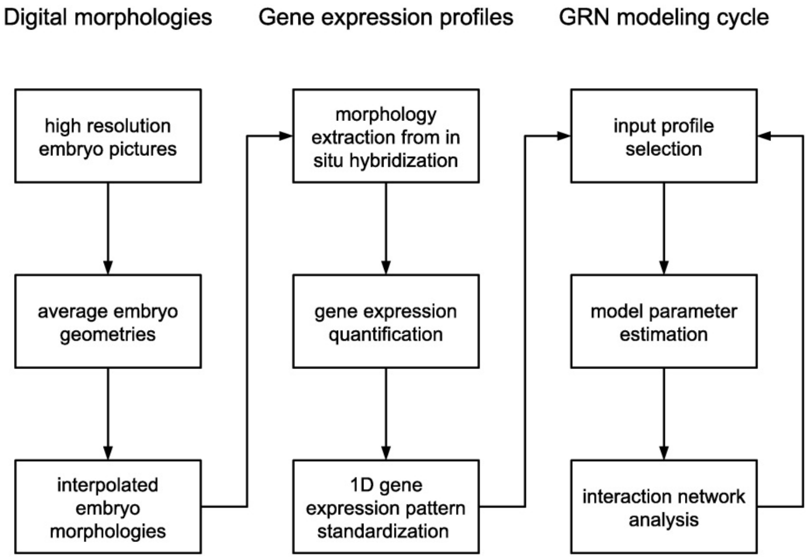

Digital morphologies
Gene expression profiles
GRN modeling cycle
high resolution embryo pictures
average embryo geometries
interpolated embryo morphologies
morphology extraction from in situ hybridization
gene expression quantification
1D gene expression pattern standardization
input profile selection
model parameter estimation
interaction network analysis

Figure 3.1: Overview of the GRN production workflow. The workflow is divided in three main parts, which are required for the study of pattern formation in any system, and nine smaller steps that apply specifically to complex and changing shapes. The main parts are the design of digital morphologies, the preparation of standardized gene expression profiles and the implementation of gene regulation models. For complex and changing morphologies, the particular steps are explicitly mentioned. First, embryo micrographs are prepared with a sufficiently high resolution to observe the tissue outlines. The outlines in every time bin are averaged to obtain representative embryo geometries for all developmental stages of interest. Points for an approximate spline of each geometry are selected and these spline points are interpolated for subsequent geometries to obtain a continuous series of digital embryo morphologies. The second part is the preparation of expression profiles from observed gene expression patterns, starting with the adjustment of a digital morphology to a gene expression image such as an *in situ* hybridization. The spatial gene expression is quantified by measuring the expression signal along the adjusted morphology. The raw signal is edited and interpolated at a fixed number of equidistant points to arrive at a standardized gene expression profile. In the third part, the gene regulatory network is inferred and updated. A set of expression profiles is selected as modeling input. The free parameters of a network model are estimated with an optimization algorithm. The optimized parameters are incorporated in an interaction network that can be analyzed and validated; the modeling cycle is then repeated with new conditions.

cells, and many regulatory interactions are mediated by a chain of inter- and intracellular compounds.

Even after the formation of membranes around the nuclei, the fly embryo outline does not change much due to its encapsulation in the eggshell. This allows a highly automated procedure for image segmentation and expression profile extraction [70]. However, the shape of most other metazoan embryos changes continuously, especially during blastula formation and gastrulation.

An extended workflow is proposed with the purpose of elucidating gene regulation mechanisms in other animals beyond flies. The particular steps in this workflow, summarized in Figure 3.1, already provide means to quantitatively compare external properties like average shapes and expression patterns among different species. The complete procedure may eventually allow the comparison of pattern formation programs.

In the current study, the starlet sea anemone *Nematostella vectensis* is used as a case study to investigate GRNs during embryonic development. As a model organism, *N. vectensis* is very convenient since it is sufficiently small and transparent for use with various microscopy methods, it is easily grown in a petri dish and it can reproduce sexually and asexually in a laboratory environment [22]. Also in terms of development, *N. vectensis* is an interesting model organism, as its mode of gastrulation is common among metazoans and many conserved signaling pathways have been identified, while its body plan is relatively simple (Figure 3.2H) [45].

Many gene expression images have been published for *N. vectensis*, and some papers are listed at the Cnidarian Evolutionary Genomics Database [49]. An increasing number of raw pictures, including unpublished material, is collected in the marine invertebrates database Kahi Kai [71]. While these images show the spatio-temporal progress of gene expression patterns, they do not give insight into how these complex patterns arise.

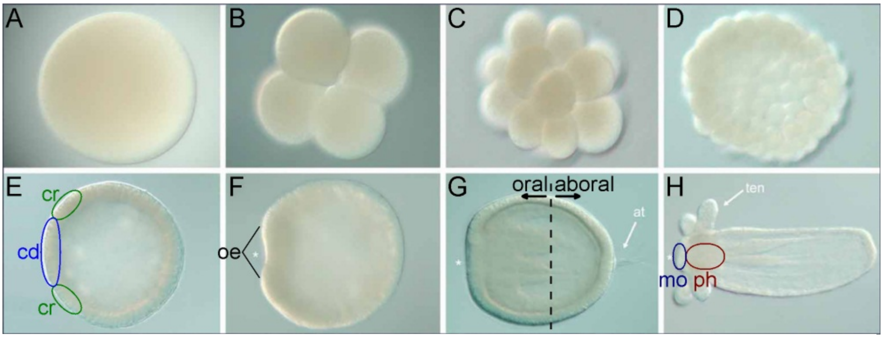


Figure 3.2: Various stages of *N. vectensis* embryonic development. Development stages from egg to polyp are shown, with the oral pole to the left in panels F-H (indicated with an asterisk). A) Fertilized egg (0 h). B) Four-cell stage (3 h), often after two cleavages finish simultaneously. C) 32-cell stage (5 h). D,E) Cleavages result in a hollow sphere called a blastula (10-20 h). F) Invaginating cells at the oral pole mark the beginning of gastrulation (24 h). G) Planula larva (72 h) with a double cell layer and apical tuft (at). Black arrows indicate the oral and aboral directions. H) Juvenile polyp with four early tentacles (ten). (cd = central domain, cr = central ring, mo = mouth, oe = oral end, ph = pharynx. Development times in hours at 16 °C estimated from [24,63].)

Previously, a method was described for the quantitative extraction of gene expression patterns in embryos with a changing morphology [72], such as *N. vectensis*. The availability of such a method is a basic requirement for modeling spatio-temporal gene regulation and forms the first step towards GRNs for morphological development in various animals. Still, an understanding of the dynamical aspects of these GRNs requires a more precise description of the complex signaling between genes and among cells.

Therefore, in this study we first apply the above method on microscopy images of RNA hybridizations and protein antibodies, such as obtained from the Kahi Kai database. Having quantified these images, a high-level mathematical description is sought to understand what properties and interactions are required for gene products to exhibit these spatial distributions during the embryo's progressing development. We start with a minimal set of genes to explain the appearance of characteristic features in the quantified expression patterns. We assume that initially studying a small number of genes will provide a clear view on the major mechanisms, while refining a model by adding more genes should show additional mechanisms responsible for properties like stability and fixed final expression domains.

We focus on *N. vectensis* gut formation. The gut is formed from an embryonic tissue called endoderm, and the delineation of endoderm (internal tissue) from overlying ectoderm (giving rise to the outer epidermis of the animal) is among the first visible developmental events in sea anemone development. To select a set of genes for simulation, we first determine which of the genes that are involved in the process of endoderm formation display similar behavior by clustering genes with similar patterns. We assume that selecting a single member from each cluster of genes is usually sufficient to discover the main mechanism that can then be elaborated on with additional genetic information. When a main mechanism has been elucidated, additional genes can be selected by narrowing the cluster sizes.

Our general approach consists of three basic steps: 1) design of digital morphologies, 2) quantification of spatial expression data and 3) mathematical analysis. For the model organism *N. vectensis*, a range of morphologies are derived from high-resolution confocal microscopy pictures during the first three days of embryonic development (Figure 2.3). These morphologies are then applied as adaptive masks to quantify expression intensity from RNA *in situ* hybridization and protein immunolocalization images of *N. vectensis* development (Figure 3.3). To infer the regulatory gene network, selected expression profiles from four genes at three distinct time points serve as reference data in the gene circuit model. Genetic interactions that show the same sign in many optimization runs are incorporated in a regulation network.

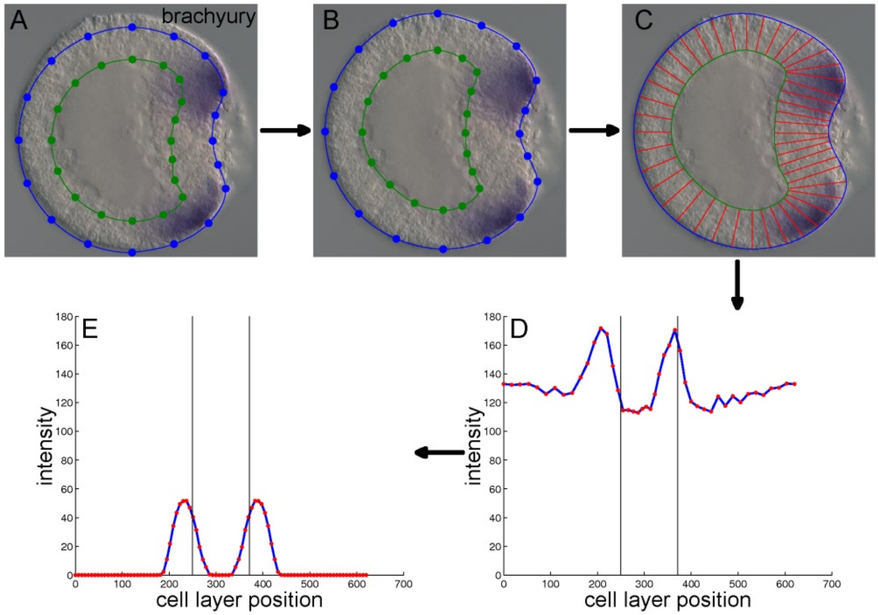


Figure 3.3: Gene expression quantification procedure. A) An in situ hybridization of the gene *brachyury* (blue reaction product) of an *N. vectensis* embryo at the correct developmental stage oriented with the site of gastrulation to the right (modified from www.kahikai.org). The overlaid prototype morphology has not been adapted. B) The morphology has been adjusted to the observed cell layer outlines. C) The cell layer is decomposed into segments of approximately the user-defined width. D) For each segment the average color intensity is calculated (the red intensity profile has been selected in this example). Expression intensity is measured along the cell layer and plotted against the cell layer position. E) Gene expression profile after editing the raw intensity profile. The vertical black lines correspond to the position of the oral pole.

The primary objective of this paper is setting up a flexible and complete workflow to obtain putative regulatory information from gene expression images at multiple time points. Current limitations are stated with propositions for improvements. In this preliminary study, we arrive at a rough network structure of regulatory interactions in the gut formation of *N. vectensis* during early development. This regulation network will be improved based on new expression data that have become available recently.

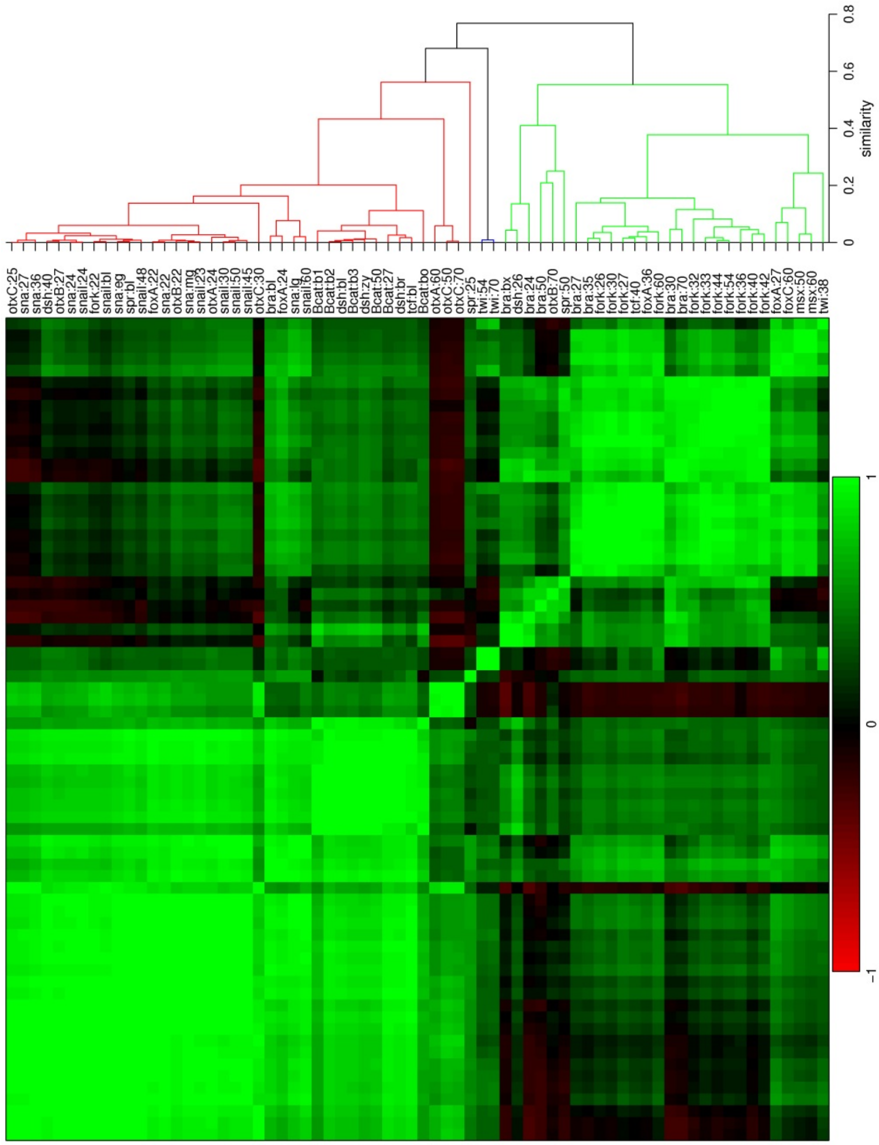


Figure 3.4: Correlation matrix and dendrogram of expression profiles for gut development genes. The 70 profiles listed in Appendix B are clustered using Pearson correlation and unweighted average linkage. The color scale goes from red (negative correlation) to black (no correlation) to green (positive correlation). The dendrogram is cut off at linkage distance 0.6 to obtain the three clusters colored in green, red and blue. (Here, the linkage distance is one minus the correlation coefficient.)

3.2 Results

A set of conserved genes that are expressed in cnidarians and echinoids is provided in [73] (Figure 4 herein) and these genes are ordered in functional modules that are associated with tissue differentiation. A single functional module is an interesting starting point for studying gene regulation, because a functional module may act as a regulatory module as well. The genes that are associated with gut development are useful for two-dimensional quantification, because their expression patterns are cylindrically symmetric. Based on the list of conserved genes for gut development, a total of 70 gene expression micrographs for 13 genes from *N. vectensis* have been retrieved from various sources. No gene expression images have been found for *six1/2* and *blimp,* while *gata* displays a grainy pattern during gastrulation [27], which is unsuitable for quantification.

The 70 gene expression images and the derived spatial expression profiles for the cluster analysis are listed in Appendix B. The correlation matrix and dendrogram are displayed in Figure 3.4. Applying a similarity cut-off of 0.6 (one minus the correlation coefficient), the genes are clustered in three groups, for which the profiles are plotted in Figure 3.5. The largest branch, colored in red, contains 40 profiles characterized by their expression in the endoderm. The green branch contains 28 profiles characterized by expression in the (presumptive) oral region. The two remaining profiles in the blue branch correlate with both clusters, as these consist of sharp peaks at the edges of both regions.

Expression domains for a large set of genes in the *N. vectensis* blastula have recently been analyzed with *in situ* hybridization [28]. Expression was found in a central domain at the oral pole, in rings at various distances from the oral center and in an aboral region. The genes in our cluster analysis correspond to those that are

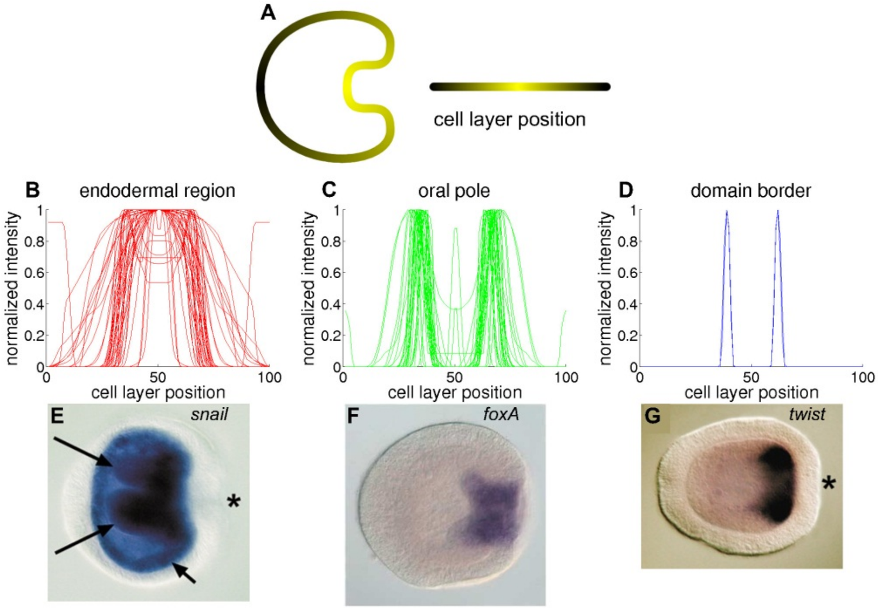


Figure 3.5: Spatial gene expression profiles divided in three hierarchical clusters. A) The embryo cell layer is mapped to the horizontal axis. B-D) The spatial expression profiles plotted in each graph show common features within the clusters from the dendrogram in Figure 3.4. B) Genes in the red cluster are mainly expressed in the endodermal region (segments 40 to 60). C) Genes in the green cluster show expression at the oral pole (segments 30-40 and 60-70). D) The remaining gene is expressed in a narrow domain that roughly corresponds to the border of both regions (right). These clusters are applied for the selection of genes that are used in the simulations. E-G) *In situ* RNA hybridizations from each cluster in the planula stage: *snail* (E), *foxA* (F) and *twist* (G). The annotations appear in the original publications [27,81]; the meaning of these annotations is irrelevant for the quantification procedure.

expressed either in the central domain or in the central ring (these regions are indicated in Figure 3.2E). Even though our analysis contains relatively few measurements from the blastula stage, the strong correlation within the two main clusters agrees with this classification. Note that some genes appear in both clusters; this could be caused by dispersion of the staining agent in older images, while some genes are repressed in the central domain after initially being expressed in this area. The separation into these two clusters

Table 3.1: Expression pattern properties for conserved gut development genes in *N. vectensis*. In the cleavage, gastrula and planula stages, the presence and location of the main expression domain is indicated for each gene. The “endoderm” and “oral pole” designations are based on the hierarchical clusters (see text).

		genes												
		group 1			**group 2**			**group 3**				**group 4**		
stage	**expression**	Bcat	dsh	tcf	otxA	otxC	sna	bra	foxA	otxB	spr	foxC	msx	twi
cleavage	present	X	X	X										
gastrula	present	X	X	X	X	X	X	X	X	X	X			
planula	endoderm	X	X		X	X	X							
	oral pole			X				X	X	X	X	X	X	

Table 3.2: Gene expression profiles selected for simulations. The interacting genes are included in the interaction matrix, while the maternal gene is merely a regulator with a constant profile. The four genes are selected from the groups indicated in Table 3.1 (see text). The expression profiles at 0 hours serve as initialization for the simulations. The profiles at 25 hours and 50 hours are used as fitting targets to rank the simulation models. The table entries refer to the profiles in Appendix B and are plotted in Figure 3.6. “zeros” means no observed expression and represents a list of zeros.

Interacting genes	**Maternal gene**	**Timepoints**		
		0 hours	**25 hours**	**50 hours**
foxA		zeros	fork:26	fork:44
snail		zeros	snail:24	snail:45
twist		zeros	zeros	twi:54
	β-catenin	Bcat:b2	Bcat:b2	Bcat:b2

also suggests that the central domain develops into endodermal tissue, while the central ring becomes the future mouth and pharynx (Figure 3.2H). Cell fate experiments are however required to confirm this observation.

The expression profiles within each cluster are strongly correlated (Figure 3.4) and show substantial overlap (Figure 3.5). Half of the genes have expression profiles in multiple clusters. In the blastula and early gastrula stages, the separate domains have partially overlapping lines of sight.

Moreover, oral views reveal that these expression domains can have irregular shapes [28], causing a variable domain boundary among individuals.

In Table 3.1, the primary spatial expression feature for each gene is indicated at the cleavage, gastrula and planula stages. Because many genes appear in two clusters in the gastrula stage, the expression cluster for each gene is only indicated for the planula stage. Based on the stage they first appear and on their spatial expression in the planula, four groups of genes are identified. First, β-*catenin*, *dishevelled* and *tcf* are expressed already at the early cleavage stage. Second, *otxA*, *otxC* and *snail* are present in the gastrula and expressed in the planula endoderm. Third, *brachyury*,

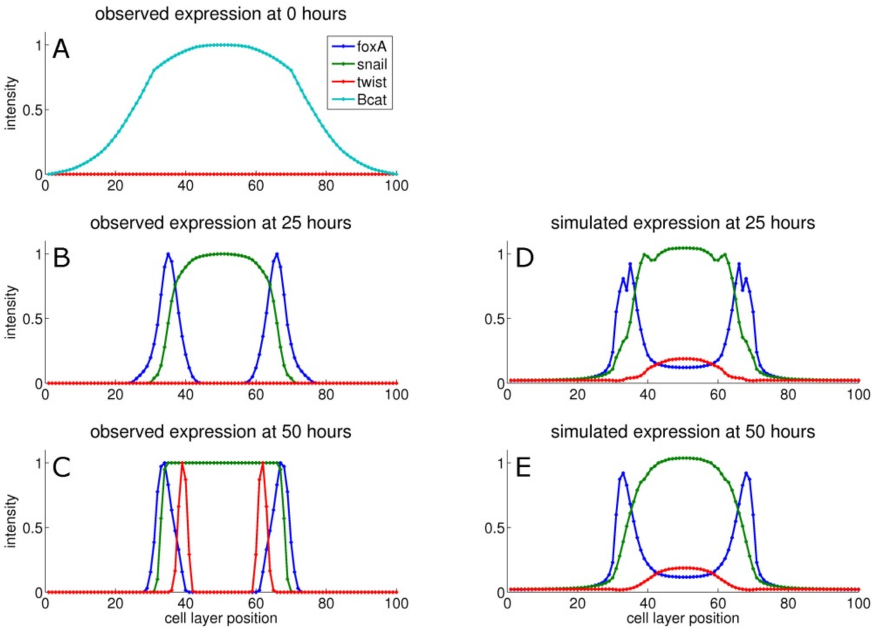


Figure 3.6: Simulated profiles compared to observations. A) All simulations start with the interacting genes *foxA*, *snail* and *twist* unexpressed and a gradient of β-*catenin*. B) At 25 hours, *foxA* and *snail* are expressed in domains near the center. C) At 50 hours, sharply bound *twist* expression occurs within the *snail* domain. The bounds of *foxA* and *snail* expression are sharper as well. D, E) The best simulation model approximates the positions of the *foxA* and *snail* domains, while the late *twist* peaks are not reproduced. The β-*catenin* gradient is kept unchanged in both the reference and the simulations, therefore this profile is not displayed in the middle and bottom plots.

foxA, *otxB* and *sprouty* are expressed at the oral pole in the planula. Fourth, *foxC*, *msx* and *twist* are not yet expressed at the gastrula stage. To arrive at a set of profiles to be used for simulation, we selected from each cluster the single gene with the largest number of expression profiles, namely β-*catenin*, *snail*, *foxA* and *twist*. For β-*catenin*, a profile is selected as a maternal gradient, because its expression precedes the gene expression in the other groups. For the purposes of the model, this means that β-*catenin* is initiated with a nonzero profile that remains constant. The other genes are initiated with an expression level of zero.

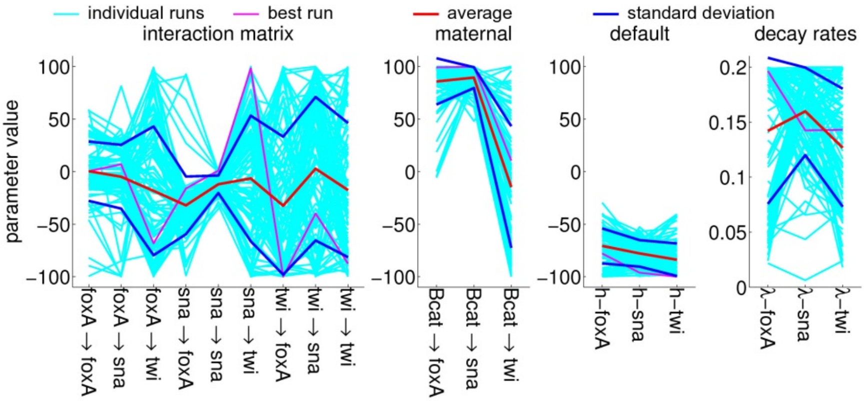


Figure 3.7: Parameter sets from 100 optimization runs. From left to right, the elements in interaction matrix *T*, the maternal influences *m*, the default influences *h* and the decay coefficients λ. The best fit is displayed in Figure 3.6.

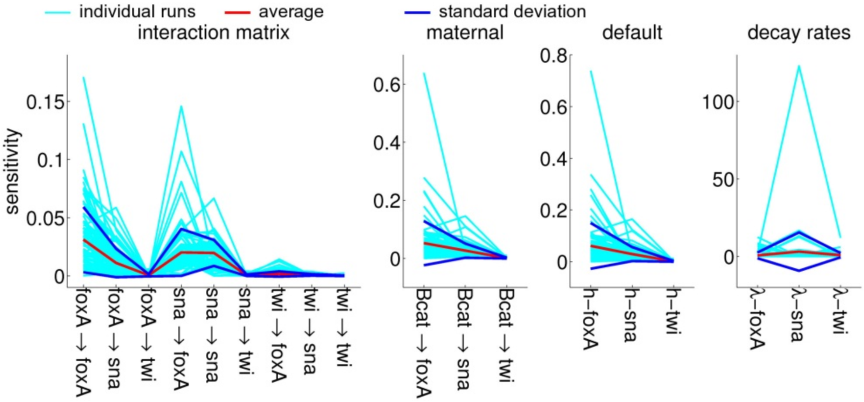


Figure 3.8: Parameter sensitivities for the parameter sets in Figure 3.7. The sensitivity is calculated as the highest average derivative of all concentration profiles to the parameter value (see Methods). From left to right, sensitivities of the elements in interaction matrix *T*, of the maternal influences *m*, of the default influences *h* and of the decay coefficients λ.

The input profiles for parameter estimation are listed in Table 3.2, with all non-constant profiles initialized at zero and a constant maternal profile. These profiles are displayed in Figure 3.6 and compared to the simulated profiles from the model with the highest similarity. The parameter sets from every run (100 runs in total) are collected in Figure 3.7 and the parameter sensitivities are plotted in

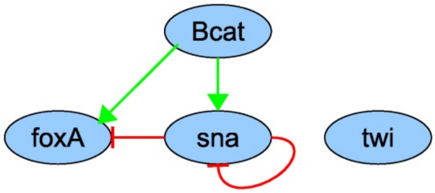


Figure 3.9: Proposed gene regulation network for gut development in *Nematostella vectensis*. The connections in this network are based on the estimated parameters in a simplified gene circuit model from 100 parameter estimations from 0 to 50 hours after fertilization. Interactions with a consistent signal in 90% of the estimated sets are incorporated into the network. The genes in this network represent groups of spatially correlated genes, as indicated in Table 3.1.

Figure 3.8. If an interaction parameter is positive in at least 90% of the estimated parameter sets, a corresponding activation is indicated in the regulation network (Figure 3.9). Likewise, an inhibition is added for an interaction parameter that is negative in 90% of the sets or more.

Figure 3.6 shows that the simulated *twist* expression pattern deviates most from the observed expression pattern: both simulated *foxA* and simulated *snail* patterns display peaks at the observed locations, while a shallow simulated *twist* band appears at the incorrect location. *twist* expression is observed only at the last time point and even then the area under the peaks is smallest for *twist*, so correctly simulated *twist* peaks would contribute the least to the overall similarity. Moreover, the *twist* peaks are located within the central region of both the β-*catenin* and *snail* domains, so no agent is present to induce a separation in the *twist* domain. Compare this to the simulated *foxA* peaks that are induced by activation from β-*catenin* and repression from *snail* (purple line in Figure 3.7). Because the simulated *twist* pattern shows the worst fit with the observed pattern and maintains the lowest expression levels over the whole length, it is expected that the parameters that involve the *twist* gene are the least sensitive. The graphs in Figure 3.8 show that this is indeed the case.

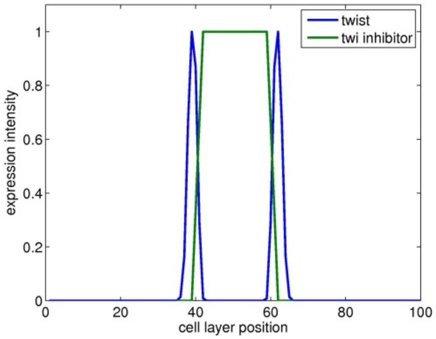


Figure 3.10: Approximated pattern for *twist* inhibitor. A gene that is expressed in the aboral endoderm is necessary to suppress the *twist* gene and limit *twist* expression to the oral endoderm.

In the simulation, *twist* is upregulated early (at 25 hours), and at the wrong location (in the aboral endoderm). A gene that is expressed in the aboral endoderm is needed to limit a *twist* peak to the oral endoderm (Figure 3.10). This role might be fulfilled by *otxA*, *otxB* or *otxC*, but another gene that is not necessarily conserved could serve this purpose as well.

Our results are not conclusive, so a comparison with a gene network from another organism would not yield new knowledge. Still, the model may allow initial comparisons with observations in sea urchins.

For example, the regulation of endomesoderm formation in the sea urchin is intensely studied [74]. The extensive network shares the genes β-*catenin*, *brachyury*, *foxA*, *otx* and *tcf* with our limited study. The reported interactions in the sea urchin system are listed in Table 3.3, along with a comparison to the inferred edges in the sea anemone regulation network. From this comparison, it seems that the regulatory function of *otx* in sea urchins is more similar to *otxB* than to *otxA* or *otxC* in sea anemones. The correspondence of most relations in sea urchin and sea anemone is remarkable, although no strong conclusions can be drawn.

Table 3.3: Experimental influences in the sea urchin [74] compared to inferred interactions in the sea anemone.

Sea urchin	**Sea anemone**	**Agreement?**
bra activates *foxA*	*brachyury* and *foxA* are clustered	yes
bra activates *otxB*	*brachyury* and *otxB* are clustered	yes
foxA represses *foxA*	*foxA* cluster lacks interaction with itself	no
otx activates *bra*	*brachyury* and *otxB* are clustered	ambiguous
	otxA/C cluster likely represses *brachyury* cluster	
otx activates *foxA*	*foxA* and *otxB* are clustered	ambiguous
	otxA/C cluster represses *foxA* cluster	
otx activates *otx*	*otxB* cluster lacks interaction with itself	no
	otxA/C cluster represses itself	
tcf/β-catenin activates *bra*	β-*catenin* cluster activates *brachyury* cluster	yes
tcf/β-catenin activates *foxA*	β-*catenin* cluster activates *foxA* cluster	yes

3.3 Methods

3.3.1 Digital morphologies from high-resolution micrographs

N. vectensis embryos at 16 °C were stained with propidium iodide and phalloidin to visualize nuclei and cell boundaries, respectively, and imaged with a confocal microscope [24] (Figure 2.3A-D). These images were used to generate nodes placed along the cell layer boundaries to indicate their shapes over developmental time points. Multiple samples (2 to 5) were recorded and node locations are averaged to generate representative geometries for all time points (Figure 2.3E-H). This averaging reduces the influence of local irregularities in individual embryos. Strategic points are picked from these average geometries for interpolation of subsequent geometries, to obtain a continuous range of embryo morphologies.

3.3.2 Standardized profiles from gene expression images

Published and raw gene expression images are imported into GenExp, a Matlab interface to quantify gene expression patterns. The expression profile is extracted as described in [72]. A prototype morphology is overlaid with the image (Figure 3.3A). The selected morphology is adapted to the observed embryo’s cell layer by dragging the points of the digital morphology over the cell layer boundaries (Figure 3.3B). The cell layer is decomposed into segments with edges between the inner and outer cell layer boundaries (Figure 3.3C). The average color intensities of the pixels within each segment are plotted as a function of the segment’s position on the cell layer (Figure 3.3D). This plot is edited to compensate for artifacts from the environment, annotations and decomposition (Figure 3.3E). The edited plot is interpolated at a hundred equidistant points and the intensity is scaled to unity to arrive at a standardized expression profile suitable for numerical analysis.

3.3.3 Numerical analysis

3.3.3.1 Spatial correlation and gene selection

All standardized expression profiles are clustered with average linkage and Pearson correlation distance; these measures are straightforward and applied most frequently in co-expression analysis [75]. In a table, the genes are ordered based on the characteristics of their expression profiles at the cleavage, gastrula and planula stages. From each group of genes with similar expression characteristics, a gene is selected for simulation. Expression profiles of the selected genes at roughly 0, 25 and 50 hours after fertilization serve as input for the parameter estimation.

3.3.3.2 Parameter estimation

The gene circuit model [17] (derived from the connectionist model [42]) is a mathematical framework that can simulate gene regulation in flies with no prior knowledge about interaction mechanisms. It is based on the assumption that the gene products influence the production rate of proteins, while diffusion and decay are protein-specific. The general differential equations for the protein concentrations in a one-dimensional chain of nuclei are

$$\frac{\partial P_i^a}{\partial t} = R_i s\left(\sum_j T_{ij} P_j^a + m_i P_{mat}^a + h_i\right) + D_i[(P_i^{a-1} - P_i^a) + (P_i^{a+1} - P_i^a)] - \lambda_i P_i^a \tag{eq. 3.1}$$

with product concentrations *P* of gene *i* in nucleus *a*, interaction matrix *T*, maternal influence *m* of maternal gene *mat*, constant influence *h*, sigmoid function *s*, production rate *R*, diffusion coefficient *D* and decay rate λ. For each gene *i*, the value of parameters T_{ij}, m_i, h_i, R_i, D_i and λ_i need to be determined. A reliable solution does not overfit the data, is sensitive to variation in parameter values, has no strongly coupled parameters and is numerically stable. Various analyses can test an optimized parameter set for these properties [43,55,56].

The modeling cycle is summarized in Figure 2.1. Parameter values are obtained by simulating a system with many sets of trial parameters from an initial state and comparing the simulated concentration profiles to the observed reference data.

We simplified the gene circuit formalism for the simulation model:

$$\frac{\partial P_i^a}{\partial t} = s\left(\sum_j T_{ij} P_j^a + m_i P_{mat}^a + h_i\right) - \lambda_i P_i^a$$

(eq. 3.2)

with parameters *T, m, h* and λ, and sigmoid function *s*(f) = ½ + arctan(f)/π. The main motivation for removing the diffusion and replacing the production rate with a constant is our focus on the interaction parameters *T* and *m*. Moreover, our one-dimensional expression profiles are simplifications of two-dimensional curved surfaces. The relative amount of cells that is mapped to each point depends on the embryo's spacial shape, and this shape is changing during the embryo's development. Therefore, the diffusion function would be time dependent and too complicated for our simple model. Neglecting the diffusion is justified based on the insensitivity of the diffusion coefficients in *Drosophila* gene circuits [55]. The production rate would correlate strongly with the decay rate, and this would result in a superfluous expansion of the search space. The production rates in *Drosophila* gene circuits are insensitive as well.

With this simulation model, a hundred optimization runs are performed, minimizing the root-mean-squared value of differences with the reference profiles at 25 and 50 hours. We applied the enhanced scatter search (eSS) algorithm [76] with at least 10,000 function evaluations for each run and the local search options deactivated. The eSS algorithm performed well for high-dimensional benchmark problems in comparison to other methods [77].

3.3.3.3 Regulation network inference

The best parameter set found in each optimization run is collected for a statistical analysis. The values from these parameter sets are displayed in a scatter plot and standard deviations around the means are indicated.

The parameter sensitivities from the optimized sets are calculated with the iterative approximation based on directional derivatives [78]. A parameter’s sensitivity is the change in the system with a changing parameter value. The derivative of a protein concentration to the parameter value is a measure for the parameter sensitivity. The algorithm calculates the derivatives of all concentrations with respect to the parameter value for every point along the cell layer. The derivatives are averaged along the cell layer for each concentration and the highest average derivative is defined as the system’s sensitivity towards this parameter. The sensitivities are plotted for all parameters in the optimized sets, along with the mean sensitivity values and standard deviations (Figure 3.8).

Those interaction parameters that have an equal sign for 90% of all values are incorporated into the proposed regulation network. These parameters are expected to be most significant, so they should exhibit high sensitivities.

3.4 Discussion

3.4.1 Gene expression quantification issues

Embryonic tissue is expanding during development, but this expansion is not homogeneous. Static points on the embryo geometries are mapped to fixed positions to minimize the apparent shift of expression patterns due to different growth rates in the embryo body. The fixed points are located roughly at the oral end after gastrulation has commenced (this region is indicated in Figure 3.2F), because many genes display a stable expression domain around this point and this location is readily established. Without a correction for inhomogeneous tissue expansion, expression at the oral ectoderm would be displaced toward the aboral ectoderm in the one-dimensional cell layer during gastrulation. The uncorrected patterns would exhibit less correlation over time and model parameters would be inferred to accommodate the imaginary shift, while the expression remains at the same position in three-dimensional space.

All quantified expression intensity is normalized to unity, because the raw intensity of *in situ* RNA hybridizations depends on the duration of hybridization, which is different for separate measurements. As a consequence of the normalization, the absolute expression levels between genes and developmental stages cannot be compared. All analysis is based on the differential gene expression within the individual embryos, which means that the strength of regulatory interactions between genes cannot be determined. The sign of the interaction parameters can only be inferred if the regulatory interactions do not influence the maximum expression. A justification for this strong assumption is that the simulated genes are selected for their expression in different domains. To enable more accurate simulations, the quantified gene expression patterns can be combined with information from quantitative PCR measurements [62].

If spatial information is available for proteins, then this is used rather than mRNA distributions, because proteins are the compounds with actual regulatory function. For *dishevelled*, both protein antibody stainings, fluorescent protein constructs and *in situ* RNA hybridizations in *N. vectensis* were available, and therefore only the protein pictures have been analyzed. Moreover, *dishevelled* transcripts are uniformly expressed throughout embryonic development [47], so *dishevelled in situs* do not provide information on differential regulation. For β-*catenin*, only protein distributions are available.

From the *in situ* measurements consulted in this study, the initial time of expression is hard to determine for each gene. The measurements are not part of a systematic time series, so an approximate time of development is derived from the embryo morphology. For the majority of genes that make their first appearance during the blastula stage, this time can only be classified as roughly as 10 to 20 hours after fertilization. Besides this lack of precise timing, the *in situ* hybridization technique is quite insensitive to low expression levels. These limitations are reduced with the availability of systematic measurements in the blastula stage and highly sensitive qPCR data. These quantitative measurements can be applied to define the total amount of mRNA in an embryo. This amount is then approximated as the total expression intensity in quantified gene expression patterns from *in situ* hybridizations.

3.4.2 Proposed improvements for geometry extraction

Currently, the geometry extraction procedure is very labor intensive and time consuming. For fly embryos, algorithmic image segmentation speeds up this task substantially [70]. For *N. vectensis* embryos, an extended image segmentation method would be required to take irregular shapes and low-contrast internal structures into account. Such an extended image segmentation method could significantly reduce the manual effort to identify the cell layer boundaries. Algorithm-guided image segmentation would also reduce subjective human judgment in estimating the boundaries, especially for rough or blurred edges. Raw RNA hybridization images of *N. vectensis* and other marine invertebrates are available in large numbers and high quality at Kahi Kai [71]. The embryo images in this database can serve as a benchmark for general image processing methods.

3.4.3 Selection of representative genes

One gene was selected for the simulations from each gene cluster in Table 3.1 based on the number of available expression profiles for each gene. While the availability of a large number of profiles reduces the uncertainty in reconstructing the spatio-temporal expression patterns for that particular gene, it does not guarantee that this gene is representative of all other genes in the same cluster. To address this issue, it needs to be established, for each selected gene, whether this gene is an outlier in its own cluster.

3.4.4 Effects of simplifications in the simulation model

The simulation model is formally fitted to 50 unique spatial points for 2 time points per gene (Figure 3.6B,C). However, all differential dynamics is observed between segments 25 and 45, while each gene reaches its final expression pattern from zero in roughly one time frame due to the intensity normalization. In this way, the useful information contained in the reference profiles is reduced to about 20 points for each of the 3 genes. The total information content is thus 60 points.

The original gene circuit formalism (Equation 1) includes n(n+5) optimization parameters for n fitted genes, which amounts to 24 optimization variables for our 3+1-gene simulation model. To limit the search space, i.e. the set of models that could potentially be evaluated during the optimization, the production rates and diffusion coefficients are excluded from the variation parameters. These modifications to the model reduce the number of variables to n(n+3), or 18 for the studied system. A lower number of variation parameters reduces the effects of overfitting, but also limits the solution space, i.e. the set of models that could be considered a good fit to the data.

The production rate in equation 3.1 determines the range of the production term. This means that a higher production rate allows a larger increase in gene product, and a higher production will generally result in higher product concentrations. In our simulations the reference input is normalized, so solutions with varying concentration maxima for the different genes do not contribute to a lower score. Therefore the variation of the production rates would not result in new regions of accessible solutions.

The diffusion coefficient determines the exchange rate of gene products between adjacent cells. A higher diffusion coefficient will smoothen the concentration profiles by decreasing large concentration differences between neighboring cells. In our best simulation the effect of removing diffusion is clearly visible for the spiked *FoxA* pattern at 25 hours (Figure 3.6D). Increasing the smoothness of the simulated patterns would result in a better fit.

A small, constant value for the diffusion coefficient that is the same for all genes can be a good alternative to completely neglecting diffusion. This will generate smooth profiles without expanding the search space, while inconsistencies with the real three-dimensional embryo morphology remain limited.

3.4.5 Comparison to experiments

Tcf is an effector of the canonical Wnt pathway that forms a complex with β-catenin for its regulatory action. The effects of Tcf on the expression patterns of other genes in the blastula stage has been studied with knockdown experiments [28]. The genes *brachyury* and *foxA* are downregulated by NvTcf knockdown, while no significant effect is observed in the expression patterns of *snailA, snailB, sprouty, otxA, otxB* and *otxC*. Based on these knockdown experiments, this means that β-catenin/Tcf likely activates *brachyury* and *foxA*, but does not interact with the other genes. Our inferred GRN correctly includes activation of *foxA* and incorrectly predicts activation of *snail* by β-catenin.

The influence of Tcf on *twist* expression has not been addressed, because the knockdown study was limited to targets that are expressed in the blastula stage, and *twist* expression has not been observed before the late gastrula. The prediction from our regulation network that β-catenin/Tcf does not influence *twist* can be tested with a functional study.

3.4.6 Comparison to another quantification approach

Our approach is similar to that of Crombach et al. [79] for *D. melanogaster*. The main features of their inferred gene networks are more reliable than ours, even though they include more genes and more parameters per gene in their optimizations. This is caused by different features in the source images, besides the obvious differences in image processing.

Their expression images are systematic time series, while our *in situ* hybridizations are obtained from several sources. Slight differences in staining procedures and microscopy settings can result in images with dissimilar properties.

Furthermore, the amount of images per gene in our sources is highly variable. This causes the clustering to be biased towards genes with many profiles. This bias could be diminished by averaging the patterns of each gene at identical time points before clustering, but for most images the exact development time is not available.

Both sources of uncertainty are diminished with the increased amount of hybridization images. A new series of hybridization images has recently become available from the Martindale lab in the Kahi Kai database. A repeated study including recent contributions would allow a more balanced clustering, a less biased gene selection and more accurate approximations of selected time points.

Because the new measurements are systematic, it can even be sufficient to discard the nonsystematic sources.

3.4.7 Robust results generated by the model

Some reliable results are obtained from the GRN model, despite its shortcomings. The cluster analysis of gene expression patterns confirms recent findings that genes in endomesoderm formation are mainly expressed in two regions. Moreover, many parameter sets in the gene circuit formalism are capable of simulating the major gene expression features, so the correct interactions that appear in the inferred GRN are probably necessary regulatory interactions to describe the main mechanism of *N. vectensis* gut patterning.

The main patterning mechanism includes maternal activation of genes in the oral pole (represented by *foxA*) and in the presumptive endoderm (represented by *snail*), and repression of the oral pole genes by the presumptive endoderm genes. β-catenin/Tcf is not necessary for *snail* expression, so *snail* requires another maternal activator. The interaction between *snail* and *foxA* has not been explored yet; a knockdown of *snail* is expected to upregulate *foxA* in the endodermal region.

3.4.8 Suggested improvements to the model

The simulated genes have been selected based on a clustering of expression patters. However, this cluster analysis is unbalanced, with almost half of the profiles belonging to *snail* and *foxA*. A more homogeneous spread of observations over the genes selected for the correlation analysis should provide clusters that are populated more evenly. More systematic measurements would also provide more intermediate time points, allowing a more precise emergence of simulated profiles.

The proposed mechanism does not rely on the presence of *twist*; removing this gene from the simulation restricts the search space and may increase the sensitivity of the remaining parameters. Another strategy would be to increase the importance of a correctly simulated *twist* pattern to the overall fitness value. An increased weight for the simulated *twist* pattern can reduce the negative bias caused by the limited *twist* expression. Increasing the *twist* weight can be achieved by setting the *twist* similarity as a separate objective in a multi-objective optimization approach.

Replacing *twist* with a more suitable gene for simulation input will probably yield smoother simulated profiles and a better understanding of the core regulatory interactions than adding more genes to the current simulated system. Adding more genes will increase the number of parameters, and the simulated profiles can contain artifacts from irrelevant parameters.

3.4.9 Gene expression quantification in other animals

With the spatio-temporal RNA data available for *N. vectensis,* a general procedure for improving GRNs can be designed. First, *in situs* from a gene expression database are quantified and a correlation analysis is performed. From this analysis, genes are selected that represent major correlation clusters. For these genes, spatial distributions at fixed time points are constructed and prepared as input for network inference. Many optimization runs are performed and the resulting models are analyzed for their targeted properties, such as statistical relevance and parameter sensitivities. Based on this information, regulatory interactions among the simulated genes are proposed. These interactions, or their absence, are compared to experiments in literature or evaluated by additional experiments. This validation is then a new starting point to adjust the modeling framework or to change the set of simulated genes. A repeated series of parameter optimizations and model analyses should result in an improved gene interaction network.

The genes involved in the formation of various tissue types and organ systems have a high similarity across various organisms. Comparison of developmental regulation networks in different organisms can determine whether the regulatory interactions among these common genes are similar as well. These results may also test the hypothesis that an organism’s outward complexity correlates with its number of regulatory interactions. In this light, it is interesting to note that previous observations have indicated that apparent complexity is independent from the organism’s gene count [80].

Qualitative spatial expression maps have been drawn up for a wide variety of organisms in the blastula stage. If digital morphologies were constructed for these organisms, their spatial gene expression distribution could be quantified as well.

4 Statistical analysis of a spatio-temporal gene expression database for sea anemone *Nematostella vectensis*

This chapter describes how the spatial gene expression quantification method has been used for a cluster analysis of gene expression patterns across a database. The chapter is based on the following publication: Botman D, Jansson F, Röttinger E, Martindale MQ, de Jong J, Kaandorp JA (2015), Analysis of a spatial gene expression database for sea anemone *Nematostella vectensis* during early development, *BMC Syst Biol* **9**(1):63.

Abstract

The spatial distribution of many genes has been visualized during the embryonic development in the starlet sea anemone *Nematostella vectensis* in the last decade. *In situ* hybridization images are available in the Kahi Kai gene expression database, and a method has been developed to quantify spatial gene expression patterns of *N. vectensis*. In this paper, gene expression quantification is performed on a wide range of gene expression patterns from this database and descriptions of observed expression domains are stored in a separate database for further analysis.

Spatial gene expression from suitable *in situ* hybridization images has been quantified with the GenExp program. A correlation analysis has been performed on the resulting numerical gene expression profiles for each stage. Based on the correlated clusters of spatial gene expression and detailed descriptions of gene

expression domains, various mechanisms for developmental gene expression are proposed.

In the blastula and gastrula stages of development in *N. vectensis*, its continuous sheet of cells is partitioned into correlating gene expression domains. During progressing development, these regions likely correspond to different fates. A statistical analysis shows that genes generally remain expressed during the planula stages in those major regions that they occupy at the end of gastrulation.

Observed shifts in gene expression domain boundaries suggest that elongation in the planula stage mainly occurs in the vegetal ring under the influence of the gene *Rx*. The secondary body axis in *N. vectensis* is proposed to be determined at the mid blastula transition.

Early gene expression domains in *N. vectensis* appear to maintain a positional order along the primary body axis. Early determination in *N. vectensis* occurs in two stages: expression in broad circles and rings in the blastula is consolidated during gastrulation, and more complex expression patterns appear in the planula within these broad regions. Quantification and comparison of gene expression patterns across a database can generate hypotheses about collective cell movements before these movements are measured directly.

4.1 Introduction

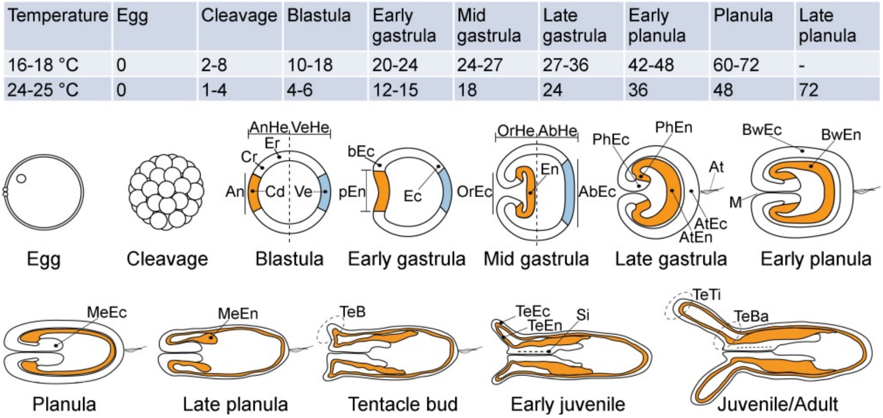
Spatial gene expression assays can be used as a tool for verifying predicted regulatory interactions between genes and for predicting properties of missing components in a gene regulation network [40,41]. The largest potential of spatial gene product distribution datasets, is in verifying numerical models of regulatory interaction networks, which has been demonstrated for the embryonic development of fruit fly *Drosophila melanogaster* [17]. Also the formation of digits in early mouse limbs has been replicated with mechanistic models with the help of gene expression patterns [82].

To perform accurate simulations of such processes, the spatial gene expression patterns need to be quantified and formatted to standardized profiles. A procedure for gene expression quantification has been described [72] and applied [83] for organisms with changing morphologies during embryonic development. Similarities among gene expression profiles can provide information about co-expression relationships [75]. Similarity metrics are a common tool for classifying time series expression data to identify correlating dynamics among genes. These similarity measures can identify correlating spatial expression among genes from quantified expression patterns as well. To use quantified gene expression patterns in dynamic simulations, reliable time points for gene expression patterns are required.

For example, in *D. melanogaster*, the spatial gene expression of *even skipped* (*eve*) has been measured precisely for many time points. The *eve* pattern is employed as a time reference: *eve* is assayed together with the queried gene to establish the development time for the sample [61]. For many other animal models, a time reference gene is not (yet) available and other embryo properties are applied to estimate the time of development. In these cases the subsequent stages of development can be qualitatively identified from the changing embryo morphology. These changes in morphology are caused by division and migration of cells, processes that are absent during the early cleavage cycles of flies.

In the comparative gene expression database Kahi Kai, *in situ* RNA hybridization assays are collected for many marine invertebrates [71] and are classified according to the embryo morphologies. This database thereby allows for an analysis of spatial expression features for all gene entries.

In this study, many genes from the Kahi Kai database are compared at various stages of development, based on their expression in different embryonic regions. First, the majority of *in situ*



Temperature	Egg	Cleavage	Blastula	Early gastrula	Mid gastrula	Late gastrula	Early planula	Planula	Late planula
16-18 °C	0	2-8	10-18	20-24	24-27	27-36	42-48	60-72	-
24-25 °C	0	1-4	4-6	12-15	18	24	36	48	72

Figure 4.1: Progressing embryo morphology during *N. vectensis* development. The table estimates the time of development at two different temperatures for the stages until the late planula. Table entries indicate the hours after fertilization derived from [24,47,63]. The annotations in the schematic morphologies are guidelines for researchers to describe expression domains in their hybridization images. AnHe = animal hemisphere, VeHe = vegetal hemisphere, An = animal pole, Cd = central domain, Cr = central ring, Er = external ring, Ve = vegetal pole, pEn = presumptive endoderm, bEc = blastoporal ectoderm, Ec = ectoderm, OrHe = oral hemisphere, AbHe = aboral hemisphere, OrEc = oral ectoderm, En = endoderm, AbEc = aboral ectoderm, PhEc = pharyngeal ectoderm, PhEn = pharyngeal endoderm, AtEn = apical tuft endoderm, AtEc = apical tuft ectoderm, At = apical tuft, M = mouth, BwEc = body wall ectoderm, BwEn = body wall endoderm, MeEc = mesentery ectoderm, MeEn = mesentery endoderm, TeB = tentacle bud, TeEc = tentacle ectoderm, TeEn = tentacle endoderm, Si = siphonoglyph, TeTi = tentacle tip, TeBa = tentacle base (The original nomenclature in [71] has been adapted to the more detailed denotations for the blastula stage in [28].).

hybridization images is quantified and the quantified gene expression patterns are collected in a list of digital expression profiles. Stage-specific correlation analyses are performed on these spatial gene expression profiles to discover the embryo’s major division in expression domains.

Second, a subset of genes from the database is listed with a detailed description of the spatial expression in the stages for which data is available. This list provides an overview of the developmental stages with spatial hybridization images for each gene and allows a detailed description of expression properties beyond the general classifications. Progression of spatial expression is compared for subsequent available stages, and the main periods of gene expression dynamics are identified.

A large set of gene expression patterns in the starlet sea anemone *Nematostella vectensis* is analyzed. The determination of the secondary axis in *N. vectensis* is one aspect of gene expression that requires spatial localization. The database contains various genes that are expressed along this axis.
The change in *N. vectensis* morphology during development is schematically displayed in Figure 4.1. The nucleus in the egg is located at the future oral pole, which means that the primary (oral-aboral) axis is already determined before fertilization [47,73]. The determination of the secondary axis, which is defined by the location of the syphonoglyph, is unclear. The first structures that appear along this axis are the primary mesenteries, but differential gene expression is already observed during gastrulation [84,85]. Based on the early symmetry break in various gene expression patterns and on early *N. vectensis* morphogenesis, a mechanism is proposed for secondary axis determination. Spatial gene expression patterns in early stages of development are necessary to study the determination and formation of the secondary axis.

We analyzed spatial gene expression patterns in various stages of development in *N. vectensis*. Changes in these patterns are due to gene expression dynamics within stationary cells or due to migrating cells that retain their gene expression state. Fate mapping experiments can conclusively determine migratory behavior, but these data are not available for *N. vectensis*. Our solution to deal with this lack of fate map data is the assumption that the expression state in migrating cells likely remains unchanged for many genes. In

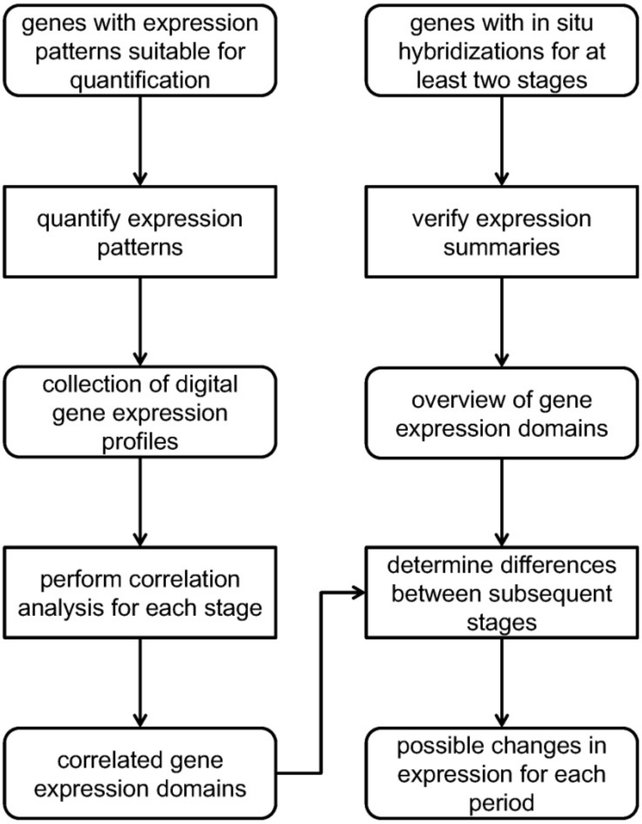

genes with expression patterns suitable for quantification
quantify expression patterns
collection of digital gene expression profiles
perform correlation analysis for each stage
correlated gene expression domains
genes with in situ hybridizations for at least two stages
verify expression summaries
overview of gene expression domains
determine differences between subsequent stages
possible changes in expression for each period

Figure 4.2: Workflow overview. The information stored in the Kahi Kai gene expression database has been processed into convenient formats for two partly overlapping sets of genes. These processing methods and the methods used for additional analyses are described in the text. While this workflow may seem to converge to a single final result, all intermediate results can be explored for multiple purposes.

this fashion, we estimate major cell movements based on the spatial gene expression data, which are available.

In conclusion we demonstrate the application of correlation analysis to quantified spatial gene expression patterns in order to identify co-expressed spatial domains.

A possible application of these correlation matrices is the selection of gene clusters for regulatory experiments, functional studies [86]

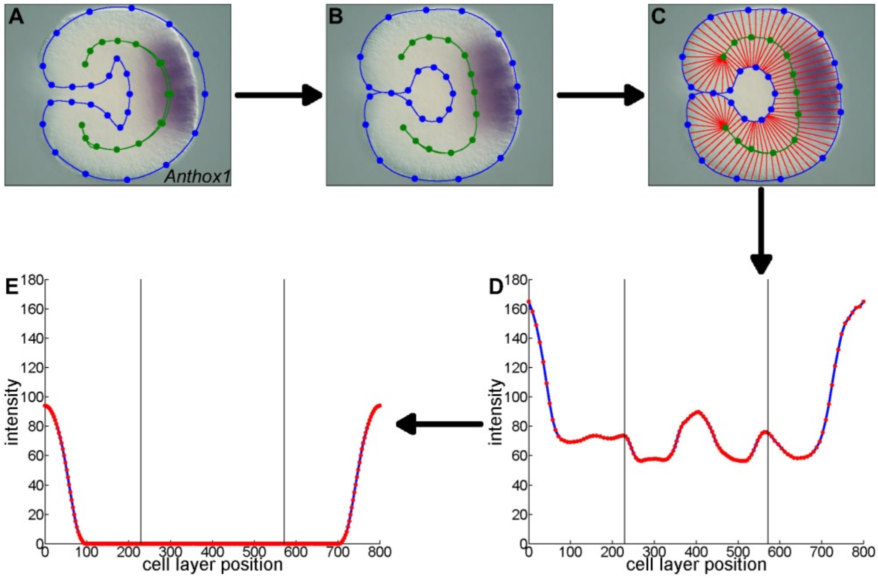


Figure 4.3: Gene expression quantification pipeline. A digital morphology is overlaid with the gene expression image (A) and the points are manually dragged over the embryo’s cell layer (B). After the cell layer is decomposed into segments (C) and the intensity is plotted as a function of cell layer position (D), the profile is manually edited to correct for artifacts (E).

and computational gene regulation network models [83], because co-expressed genes often share regulators or biological functions.

4.2 Methods

The order of application for the described methods is displayed in the diagram in Figure 4.2. Note that the intermediate results provide new information on their own and can be subjected to additional analyses.

4.2.1 Cluster analysis of quantified *in situ* hybridizations

The genes listed in the Kahi Kai database for *N. vectensis* are screened for useful expression patterns. For one-dimensional quantification, suitable genes are genes that display cylindrical expression in broad domains up to the late planula stage. Other genes, such as those that are expressed on the syphonoglyph side only or in individual cells, require a two-dimensional or three-dimensional quantification method for a complete description. *In situ* hybridization images are imported into GenExp, a Matlab interface designed to extract and quantify gene expression patterns [72]. A continuous series of digital morphologies has been derived from a confocal microscopy study on *N. vectensis* gastrulation. A digital morphology is selected and overlaid with the hybridization image (Figure 4.3A). To get a correct alignment, the points of the digital morphology are dragged over the observed cell layer boundaries (Figure 4.3B). The cell layer is divided into segments with edges between the inner and outer cell layer boundaries (Figure 4.3C). The color intensities of the pixels within each segment are averaged and plotted as a function of the segment's position on the cell layer (Figure 4.3D). This plot is edited to compensate for artifacts from the environment, annotations and imperfections in the segmentation (Figure 4.3E). The edited plot is interpolated at a hundred equidistant points and the intensity is scaled to unity to arrive at a standardized expression profile. The standardized profiles are ordered in seven groups from blastula to late planula, based on the development label of the source images. The profiles within each group are clustered with unweighted average linkage and Pearson correlation distance. The groups are divided into two main clusters, or into three clusters if the second split reduces the largest branch. The profiles within these main clusters are displayed in combined plots. Profiles in any apparent subclusters are plotted together as well.

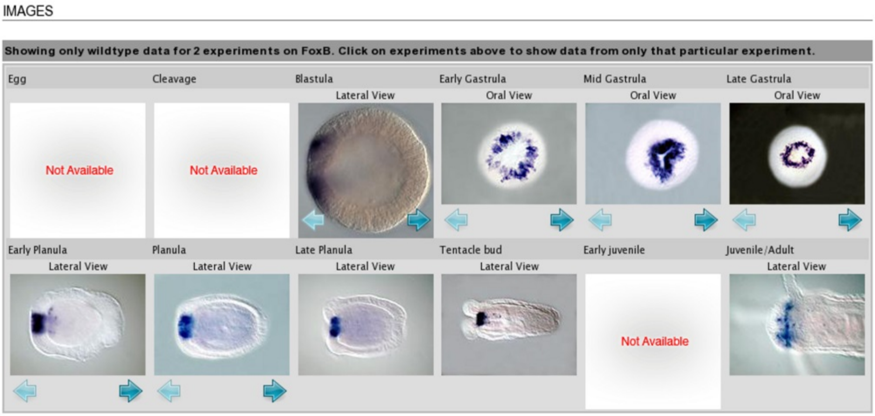
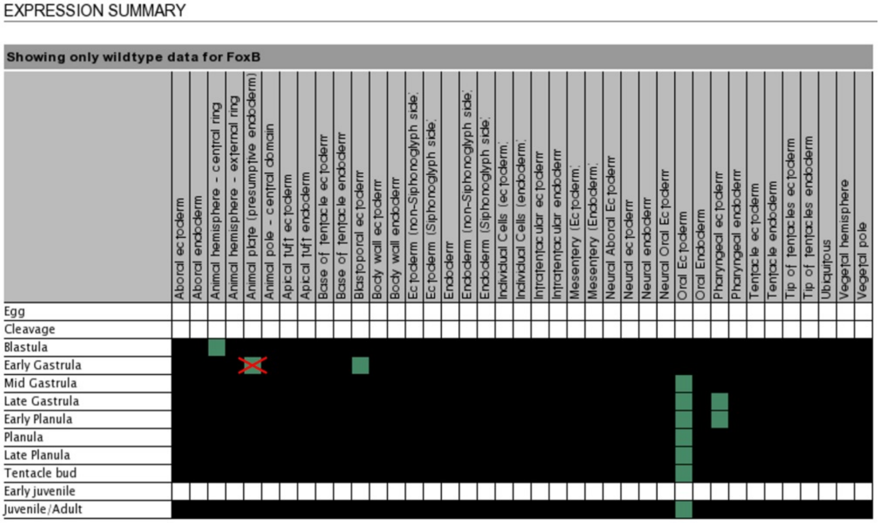


Figure 4.4: Hybridization images for *FoxB* in the Kahi Kai database. If multiple images are available for a developmental stage, these are accessed with the blue arrow buttons.



EXPRESSION SUMMARY

Showing only wildtype data for FoxB

	Aboral ectoderm	Aboral endoderm	Animal hemisphere - central ring	Animal hemisphere - external ring	Animal plate (presumptive endoderm)	Animal pole - central domain	Apical tuft ectoderm	Apical tuft endoderm	Base of tentacle ectoderm	Base of tentacle endoderm	Blastoporal ectoderm	Body wall ectoderm	Body wall endoderm	Ectoderm (non-Siphonoglyph side)	Ectoderm (Siphonoglyph side)	Endoderm	Endoderm (non-Siphonoglyph side)	Endoderm (Siphonoglyph side)	Individual Cells (ectoderm)	Individual Cells (endoderm)	Intratentacular ectoderm	Intratentacular endoderm	Mesentery (Ectoderm)	Mesentery (Endoderm)	Neural Aboral Ectoderm	Neural ectoderm	Neural endoderm	Neural Oral Ectoderm	Oral Ectoderm	Oral Endoderm	Pharyngeal ectoderm	Pharyngeal endoderm	Tentacle ectoderm	Tentacle endoderm	Tip of tentacles ectoderm	Tip of tentacles endoderm	Ubiquitous	Vegetal hemisphere	Vegetal pole
Egg																																							
Cleavage																																							
Blastula	■	■	🟩	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Early Gastrula	■	■	■	■	☒	■	■	■	■	■	🟩	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Mid Gastrula	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	🟩	■	■	■	■	■	■	■	■	■	■
Late Gastrula	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	🟩	■	🟩	■	■	■	■	■	■	■	■
Early Planula	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	🟩	■	🟩	■	■	■	■	■	■	■	■
Planula	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	🟩	■	■	■	■	■	■	■	■	■	■
Late Planula	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	🟩	■	■	■	■	■	■	■	■	■	■
Tentacle bud	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	🟩	■	■	■	■	■	■	■	■	■	■
Early juvenile																																							
Juvenile/Adult	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	🟩	■	■	■	■	■	■	■	■	■	■

Figure 4.5: Expression summary for *FoxB* in the Kahi Kai database. The indicated expression domains are derived from the available *in situ* RNA hybridizations (some hybridization images are displayed in Figure 4.4). Expression in the presumptive endoderm at the early gastrula is incorrect.

4.2.2 Overview and analysis of expression summaries

A list of all genes in the Kahi Kai database with *in situs* available from the blastula to the late planula stage was retrieved with the built-in search tool. From this list, those genes were selected with images available in at least two different stages of development. The selected genes are listed in a database table, with descriptions of the expression patterns in available stages. The descriptions are derived from the Kahi Kai expression summary matrix, while correcting possible inconsistencies with the *in situs* (illustrated for the gene *FoxB* in Figures 4.4 and 4.5 as an example). Expression in the endoderm wall and ectoderm wall is specified in more detail. If a gene expression pattern exhibits noncylindrical symmetry, this is briefly indicated.

All possible pairwise combinations of developmental stages are listed in a spreadsheet, and the expression pattern descriptions from the database table are inserted for subsequent available stages. All combinations with identical expression domains in both the first and the second stage are added up. The extent to which the expression pattern has changed is indicated in a separate column. The instances in which a pattern has remained within the same major region (minor change), shifted across major regions (major change) or vanished are catalogued for genes that start in a single major domain. For each major domain, all possible stages in which a pattern can display a minor change, display a major change or vanish are counted. The relative occurrences of these events in each stage are derived from their counts. The instances that a pattern has displayed minor or major changes with respect to the major expression regions are registered for the complete set of available genes as well, along with their possible first appearance and disappearance.

4.3 Results

4.3.1 Cluster analysis of quantified *in situ* hybridizations

For all seven stages of development from blastula to late planula, *in situs* of suitable genes have been quantified. The expression profiles in Matlab format are provided as a supplement (see Appendix C). The quantified patterns are ordered in dendrograms and correlation matrices. The number of analyzed patterns from each stage are: 112 from the blastula (Figure 4.6); 52 from the early gastrula (Figure 4.8); 18 from the mid gastrula (Figure 4.10); 25 from the late gastrula (Figure 4.12); 15 from the early planula (Figure 4.14); 17 from the planula (Figure 4.16) and 13 from the late planula (Figure 4.18). From the correlation matrices, major and minor blocks are identified. The profiles in these blocks are combined in separate plots for the blastula (Figure 4.7), early gastrula (Figure 4.9), mid gastrula (Figure 4.11), late gastrula (Figure 4.13), early planula (Figure 4.15), planula (Figure 4.17) and late planula (Figure 4.19) stages. Correlating expression domains from the blastula to the late gastrula stages are summarized in Figure 4.20.

For the blastula stage, gene expression is present in the central domain in the 101 profiles in the blue cluster, while expression is excluded from the central domain in the 11 profiles in the red cluster. From the correlation matrix in Figure 4.6, two smaller blocks with strong correlation and an isolated sample are identified within the blue cluster. The profiles within each subcluster are combined in three small plots next to the major blue cluster in Figure 4.7. The first subcluster contains profiles with gene expression limited to the central domain, while gene expression in the second subcluster is extended to the central ring. The isolated profile displays gene expression in a narrow spot within the central

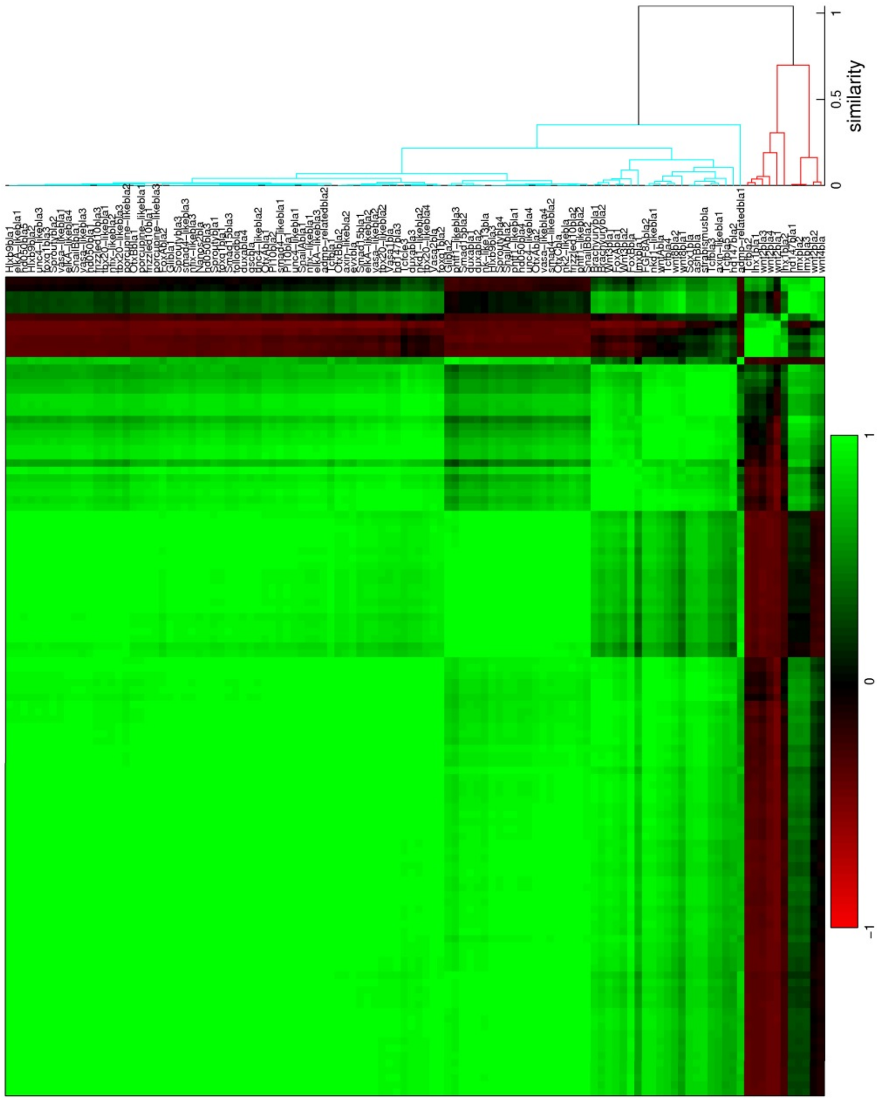


Figure 4.6: Hierarchical clustering with Pearson correlation and unweighted average linkage of gene expression profiles in the blastula stage. Tcfcle3 is labeled as cleavage, but with a large blastocoel the sample is suitable for quantification. The dendrogram is cut off at a similarity (1 minus the correlation coefficient) of 0.7.

domain. The red cluster in the correlation matrix contains two smaller blocks as well. These subclusters, the small plots next to the major red cluster in Figure 4.7, seem to separate the central ring and the external ring. An earlier study of whole mount *in situ*

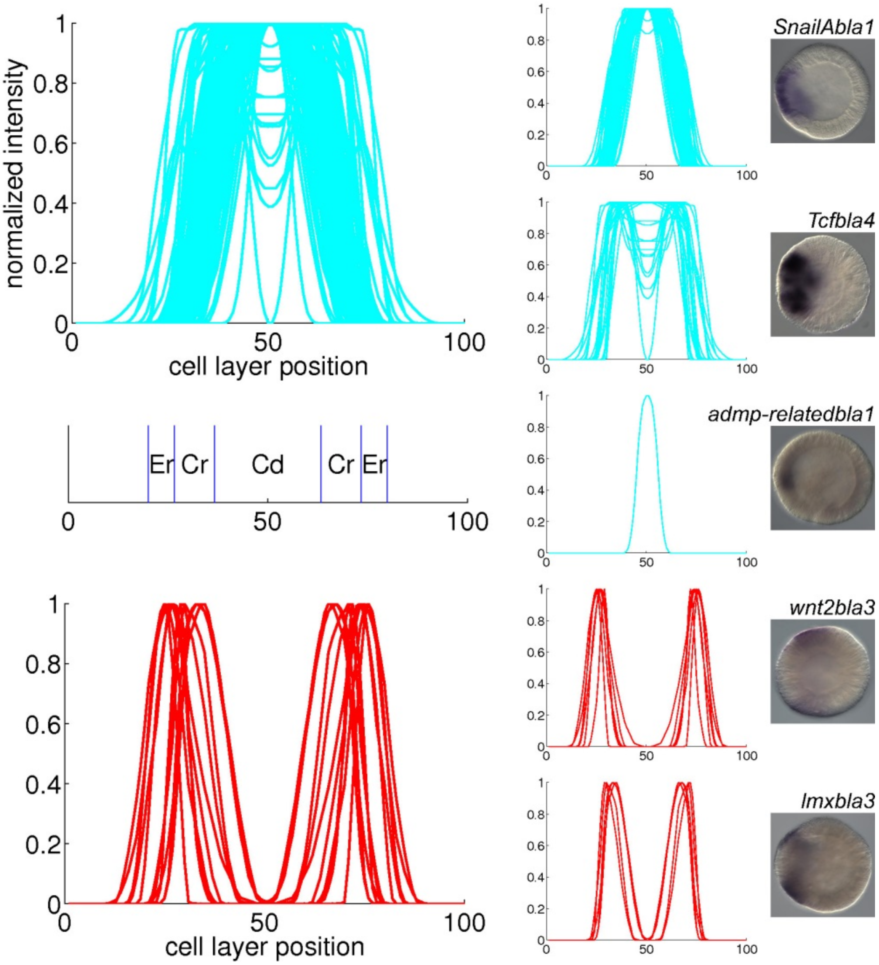


Figure 4.7: Combined plots of quantified gene expression patterns in the blastula stage. The main clusters from Figure 4.6 are plotted in large diagrams. The small diagrams on the right are subclusters within the large plot. For one profile in each subcluster, the original *in situ* hybridization is displayed. The expression domains that arise from the clustering are: central domain (Cd), central ring (Cr) and external ring (Er).

hybridizations in the *N. vectensis* blastula distinguished four co-expression domains in the animal hemisphere [28]. The present cluster analysis confirms the existence of these four domains (central domain, central domain + central ring, central ring and external ring).

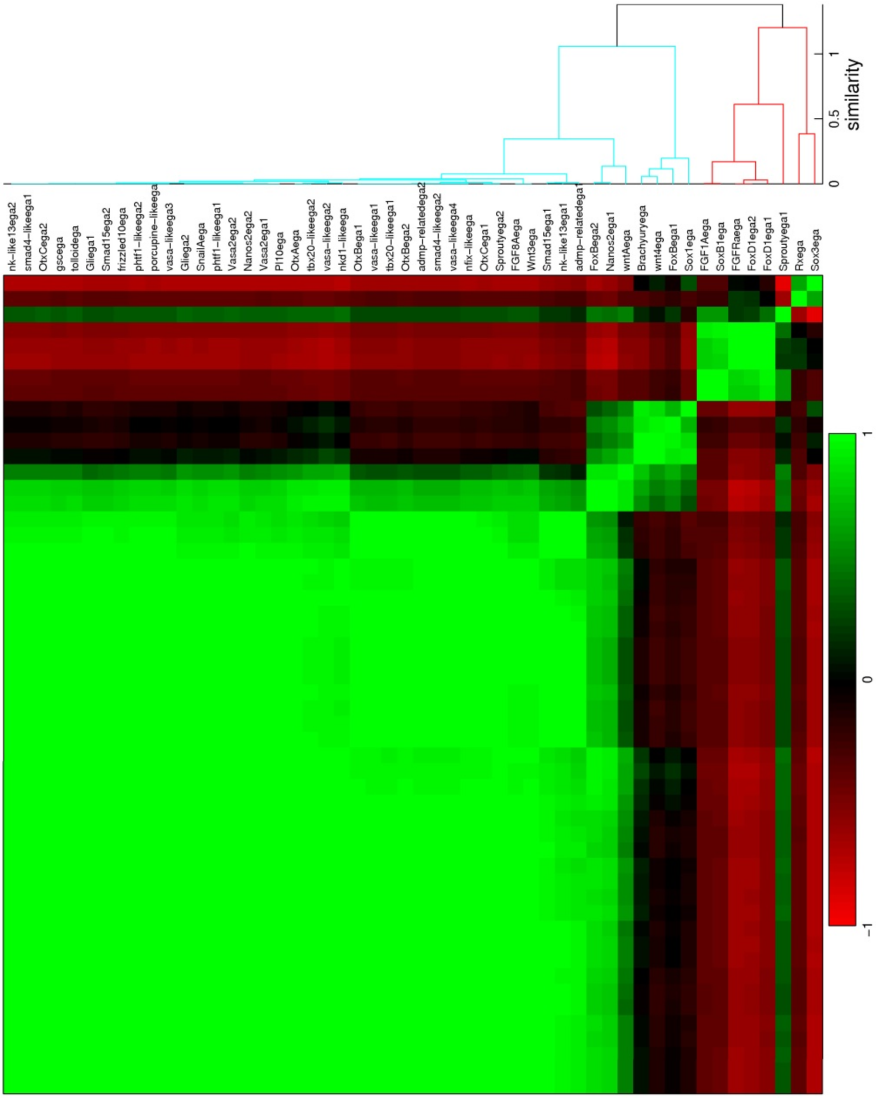


Figure 4.8: Hierarchical clustering with Pearson correlation and unweighted average linkage of gene expression profiles in the early gastrula stage. The dendrogram is cut off at a similarity of 1.3.

For the early gastrula stage, gene expression is absent in the aboral ectoderm in the 44 profiles in the blue cluster, while this region is included in the expression patterns of the 8 profiles in the red cluster. The blue cluster of the correlation matrix (Figure 4.8) contains three smaller blocks; two outer blocks are clearly

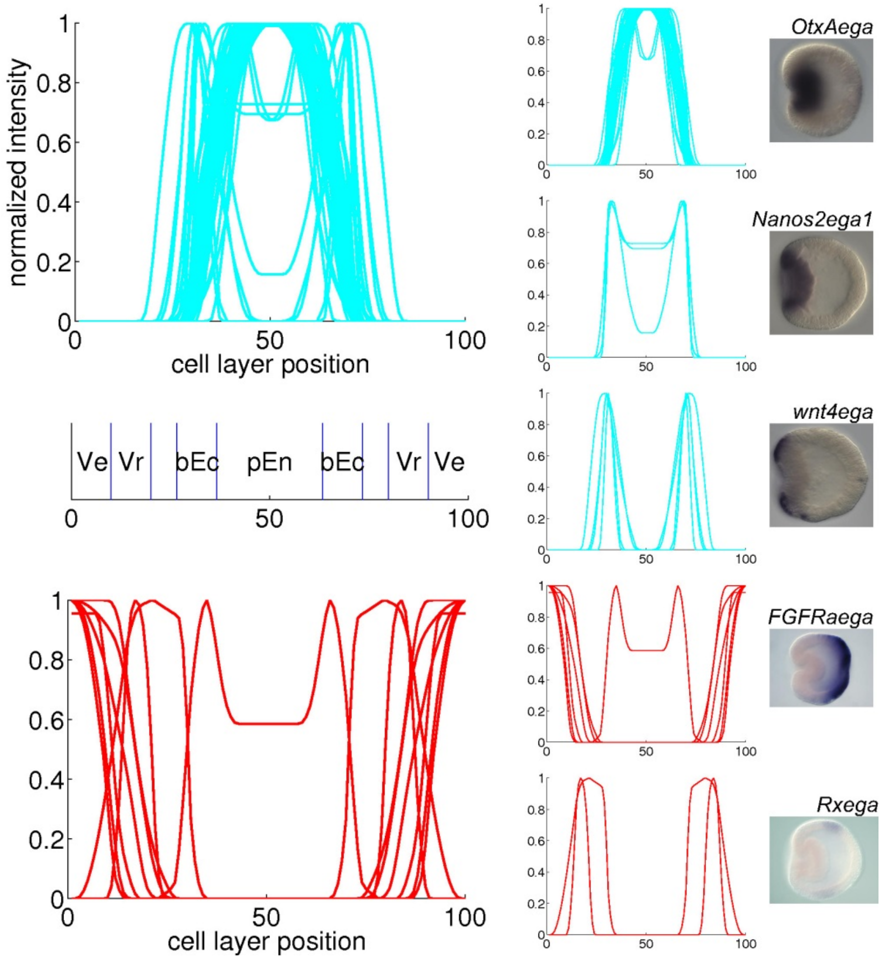


Figure 4.9: Combined plots of quantified gene expression patterns in the early gastrula stage. The main clusters from Figure 4.8 are plotted in large diagrams. The small diagrams on the right are subclusters within the large plot. For one profile in each subcluster, the original *in situ* hybridization is displayed. The expression domains that arise from the clustering are: presumptive endoderm (pEn), blastoporal ectoderm (bEc), vegetal ring (Vr) and vegetal pole (Ve).

separated, while the middle block is positively correlated to both other blocks. The profiles within each subcluster are combined in three small plots next to the major blue cluster in Figure 4.9. Expression in the first and third subclusters is limited to the presumptive endoderm and blastoporal ectoderm, respectively.

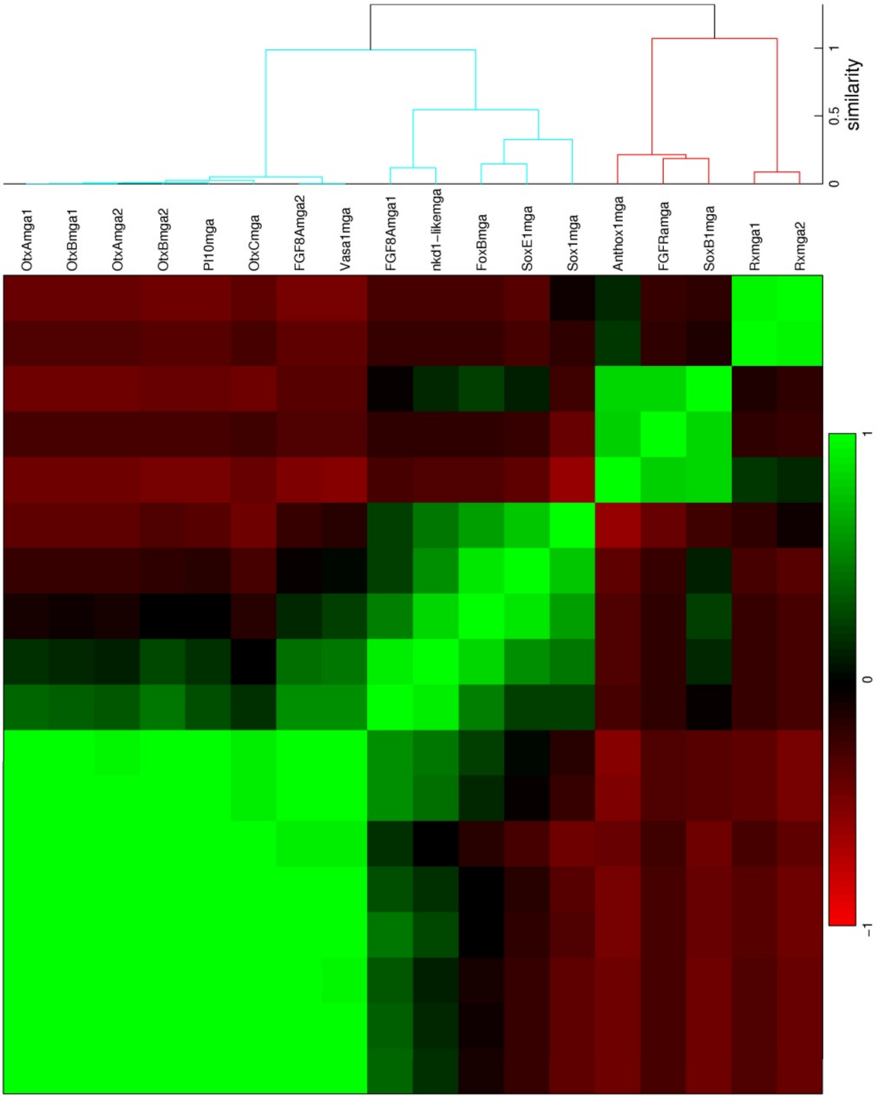


Figure 4.10: Hierarchical clustering with Pearson correlation and unweighted average linkage of gene expression profiles in the mid gastrula stage. The dendrogram is cut off at a similarity of 1.1.

Expression in the second subcluster covers both regions. The red cluster of the correlation matrix contains two smaller blocks; these subclusters are plotted in Figure 4.9 next to the major red cluster. The first subcluster exhibits expression at the vegetal pole, while

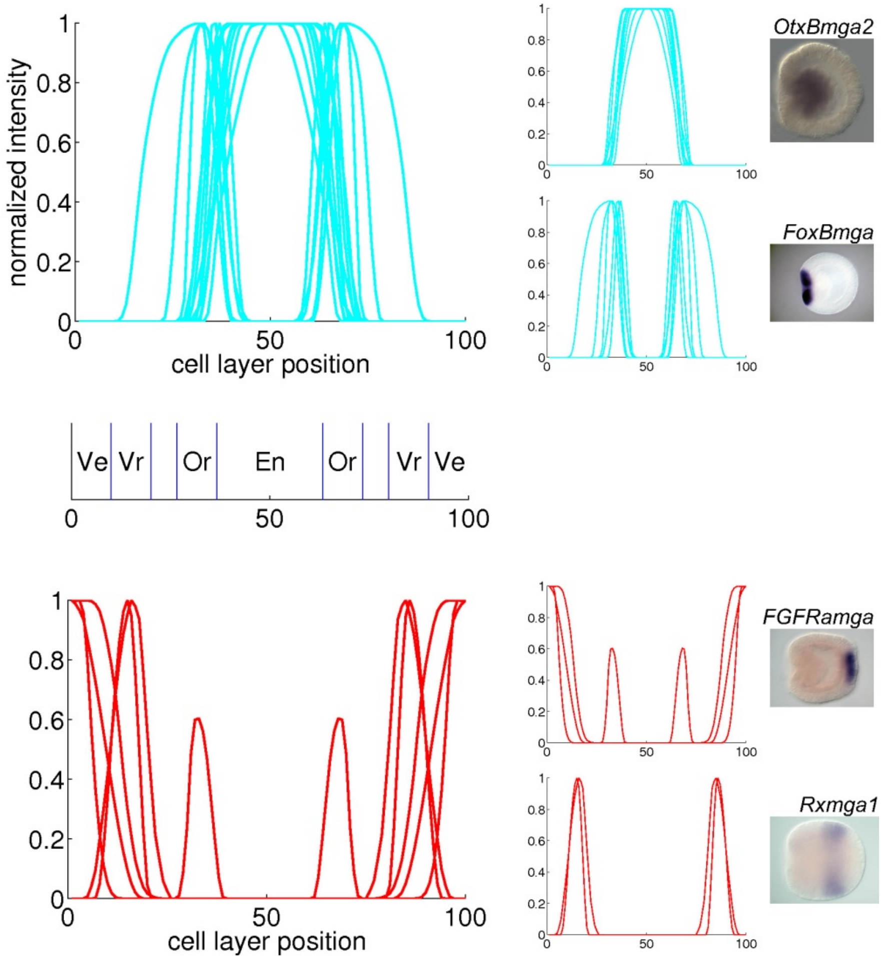


Figure 4.11: Combined plots of quantified gene expression patterns in the mid gastrula stage. The main clusters from Figure 4.10 are plotted in large diagrams. The small diagrams on the right are subclusters within the large plot. For one profile in each subcluster, the original *in situ* hybridization is displayed. The expression domains that arise from the clustering are: endoderm (En), oral pole (Or), vegetal ring (Vr) and vegetal pole (Ve).

genes in the second subcluster are expressed in a ring around the vegetal pole.

For the mid gastrula stage, the blue cluster consists of 13 profiles with expression in the blastopore, while gene expression appears in

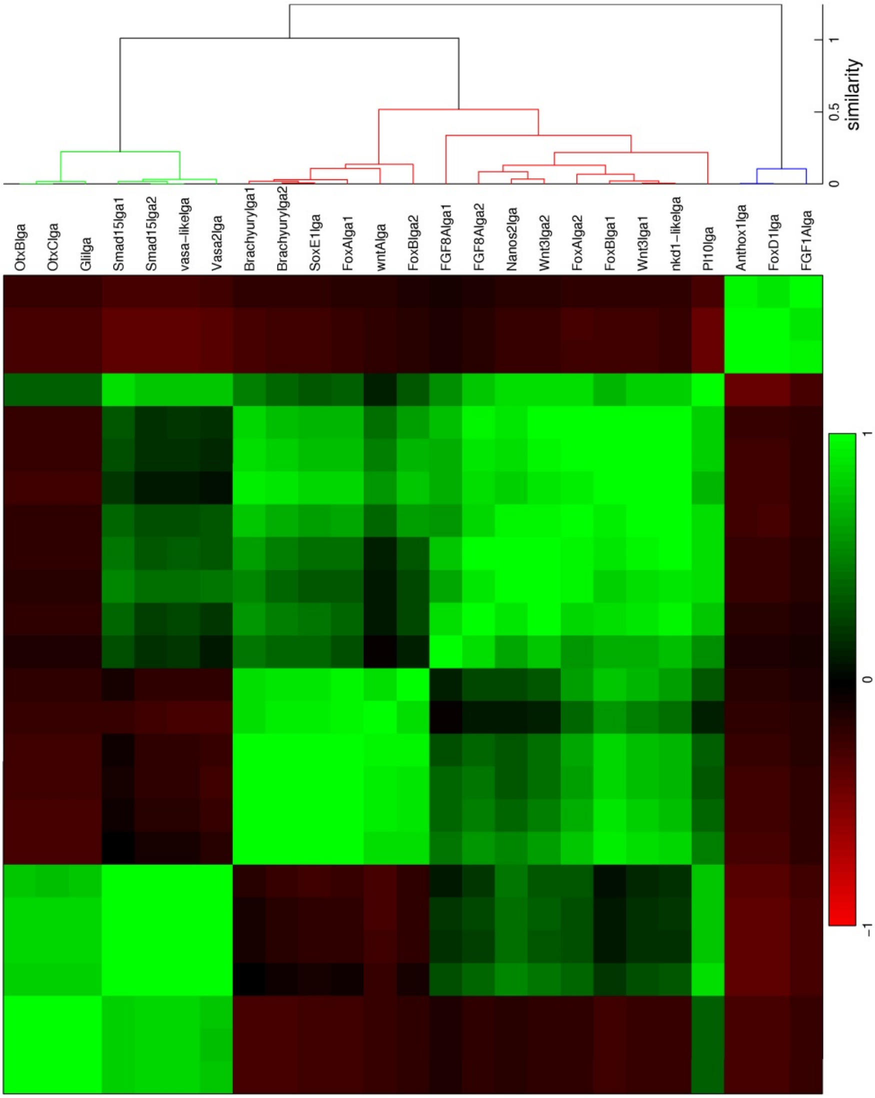


Figure 4.12: Hierarchical clustering with Pearson correlation and unweighted average linkage of gene expression profiles in the late gastrula stage. The dendrogram is cut off at a similarity of 0.6.

the aboral ectoderm in the 5 profiles in the red cluster. The blue cluster of the correlation matrix in Figure 4.10 contains a block with strongly correlated profiles, while the remaining profiles are correlated somewhat weaker. The first subcluster represents endodermal expression, while expression in the second subcluster is

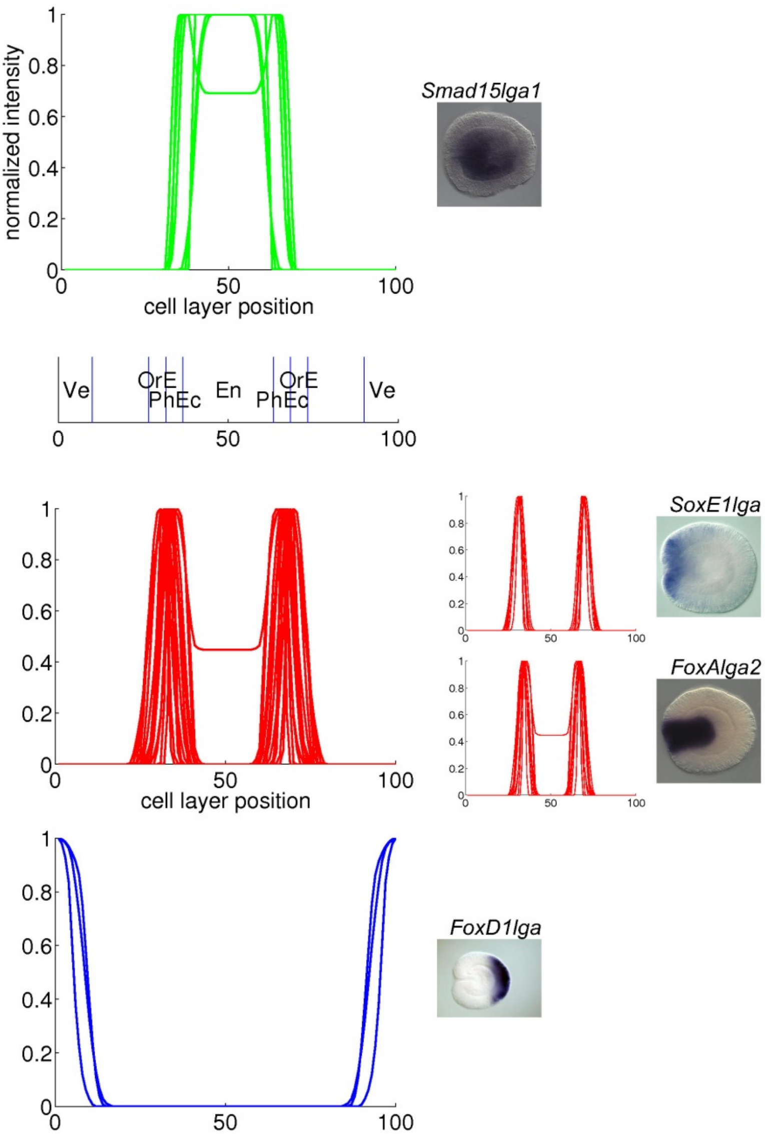


Figure 4.13: Combined plots of quantified gene expression patterns in the late gastrula stage. The main clusters from Figure 4.12 are plotted in large diagrams. The small diagrams on the right are subclusters within the large plot. For one profile in each (sub)cluster, the original *in situ* hybridization is displayed. The expression domains that arise from the clustering are: endoderm (En), pharyngeal ectoderm (PhEc), oral end (OrE) and vegetal pole (Ve).

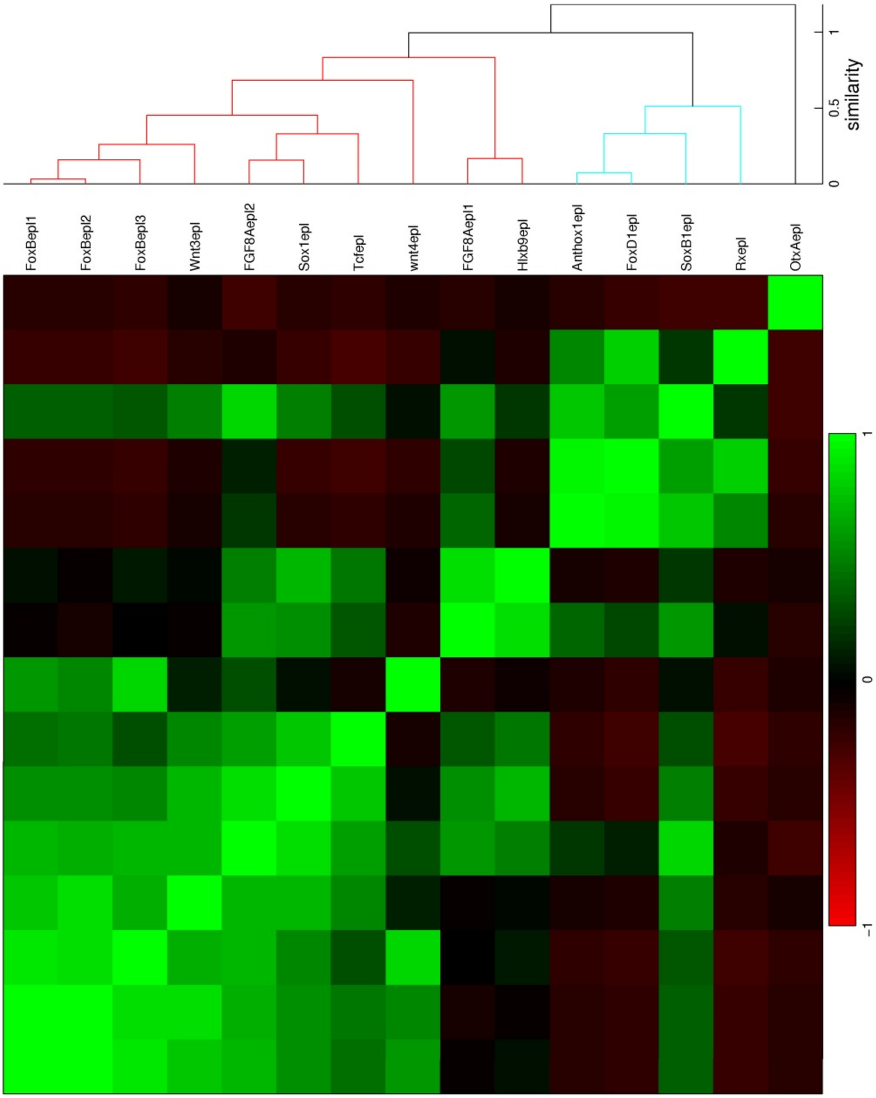


Figure 4.14: Hierarchical clustering with Pearson correlation and unweighted average linkage of gene expression profiles in the early planula stage. The dendrogram is cut off at a similarity of 0.9.

in the oral pole at various ranges from the center. The profiles of both subclusters are displayed in Figure 4.11 next to the major blue cluster. The red cluster of the correlation matrix contains two blocks with strongly correlated profiles. The profiles in these blocks, combined in two small plots next to the major red cluster in Figure

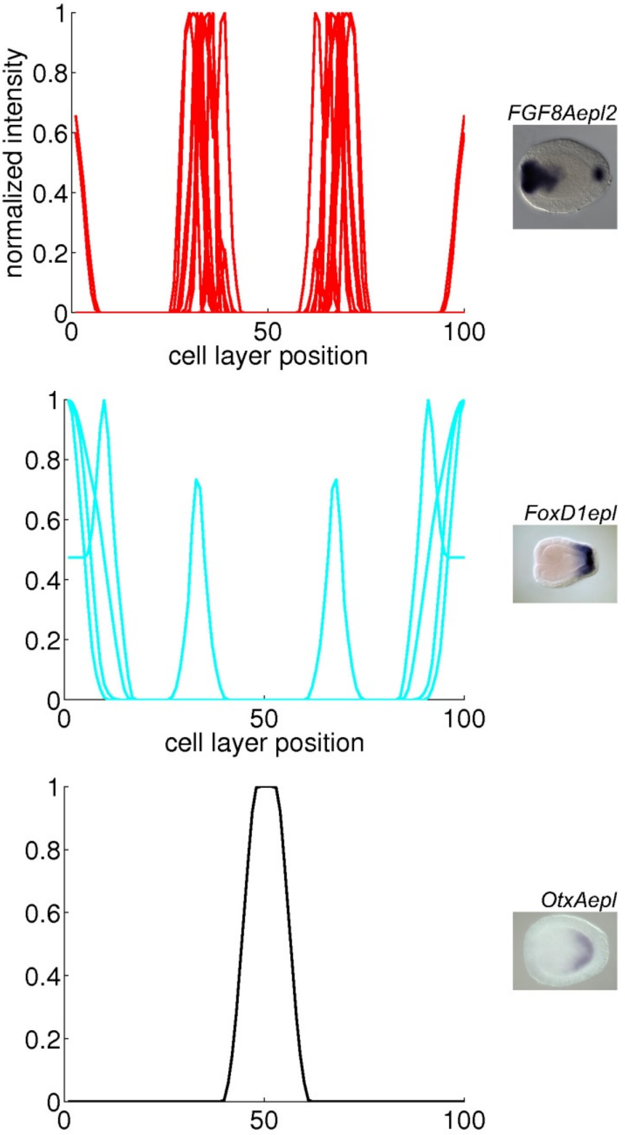


Figure 4.15: Combined plots of quantified gene expression patterns in the early planula stage. The plots represent the clusters in Figure 4.14. For one profile in each cluster, the original *in situ* hybridization is displayed.

4.11, represent gene expression at the vegetal pole and in a ring around this pole, respectively.

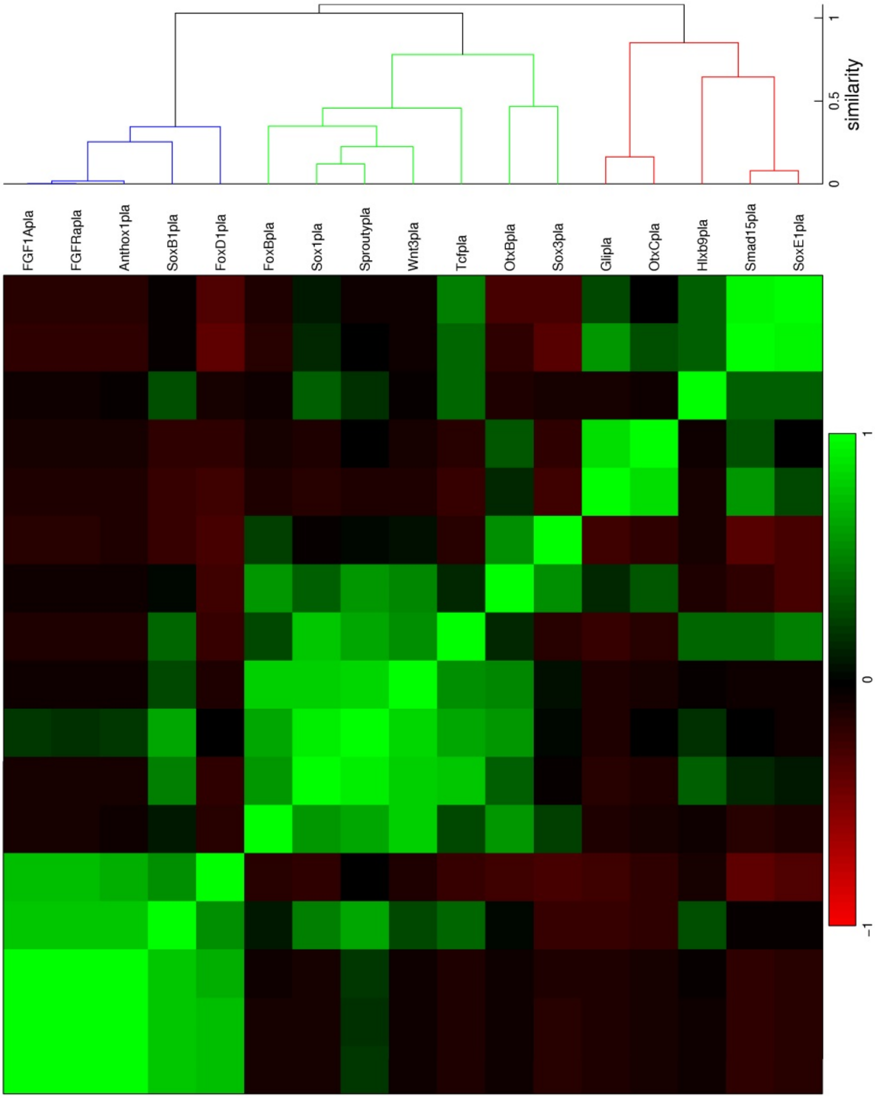


Figure 4.16: Hierarchical clustering with Pearson correlation and unweighted average linkage of gene expression profiles in the planula stage. The dendrogram is cut off at a similarity of 0.9.

For the late gastrula stage, the green cluster represents endodermal expression, the red and blue clusters represent ectodermal expression in the oral and aboral pole, respectively. Two smaller blocks are visible within the red cluster; the profiles in these subclusters are displayed in two small plots in Figure 4.13. Gene

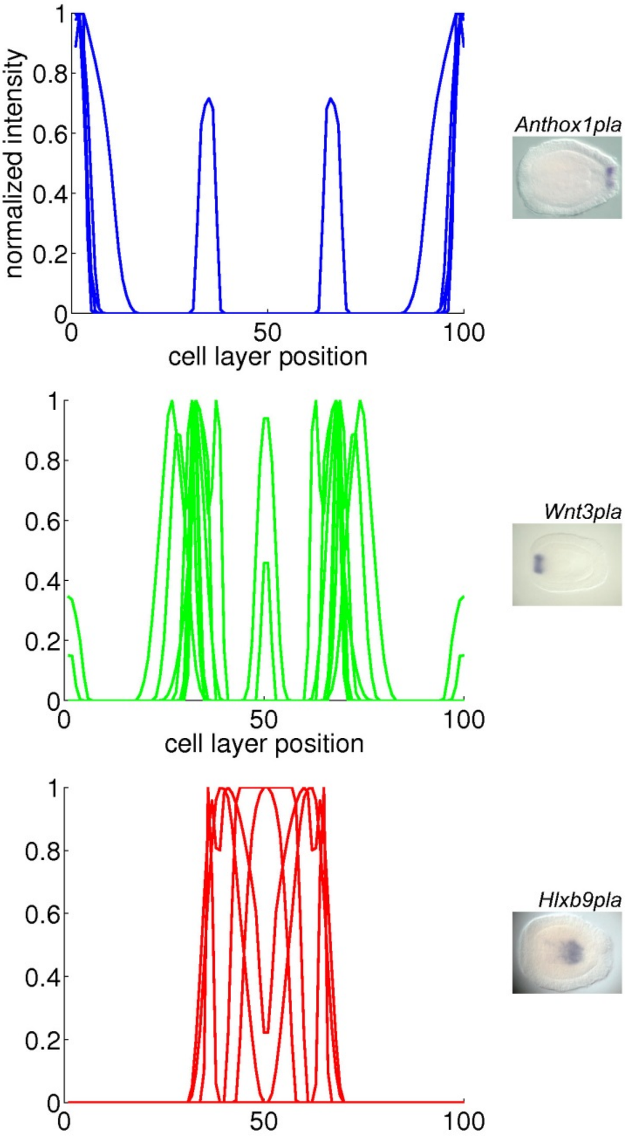


Figure 4.17: Combined plots of quantified gene expression patterns in the planula stage. The plots represent the clusters in Figure 4.16. For one profile in each cluster, the original *in situ* hybridization is displayed.

expression in the first subcluster is limited to the oral end, while expression in the second subcluster is extended inwards to the pharynx.

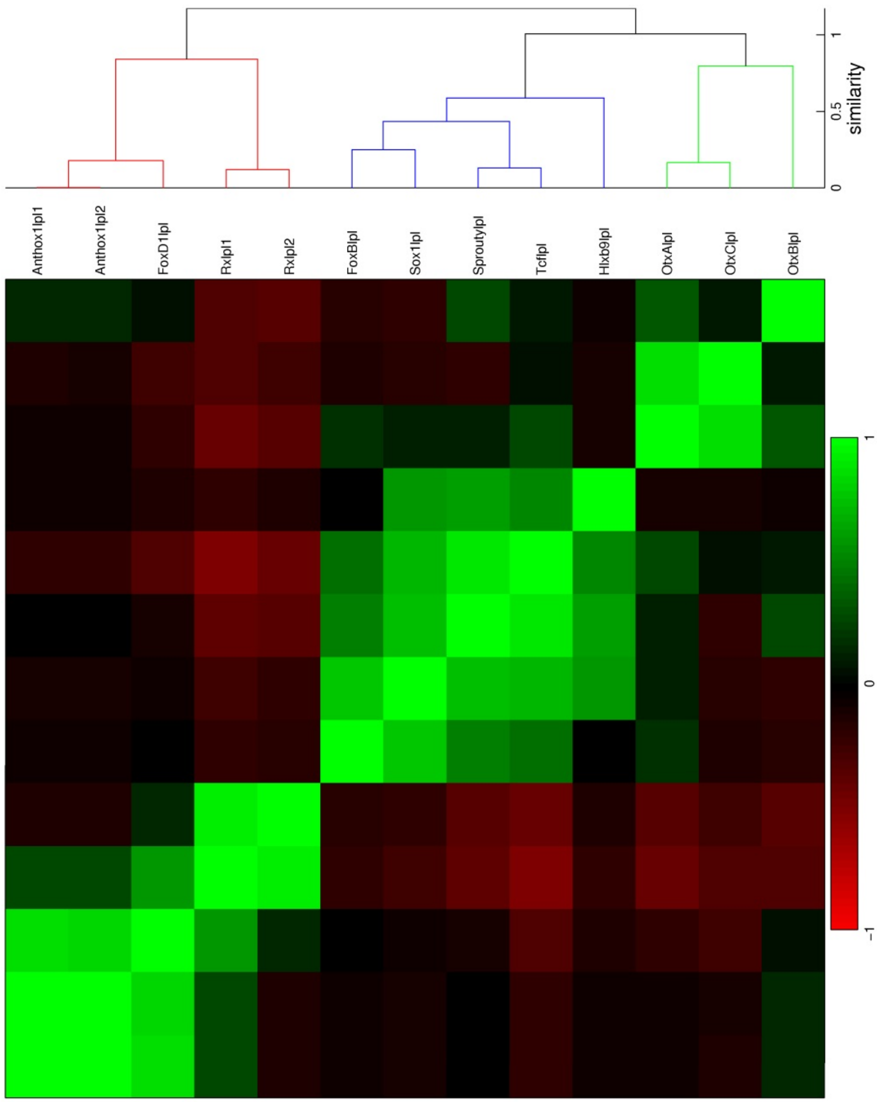


Figure 4.18: Hierarchical clustering with Pearson correlation and unweighted average linkage of gene expression profiles in the late planula stage. The dendrogram is cut off at a similarity of 0.9.

For the early planula stage, the three clusters represent the oral pole, the aboral ectoderm and the aboral endoderm. The profiles in these clusters are plotted in red, blue and black in Figure 4.15, respectively.

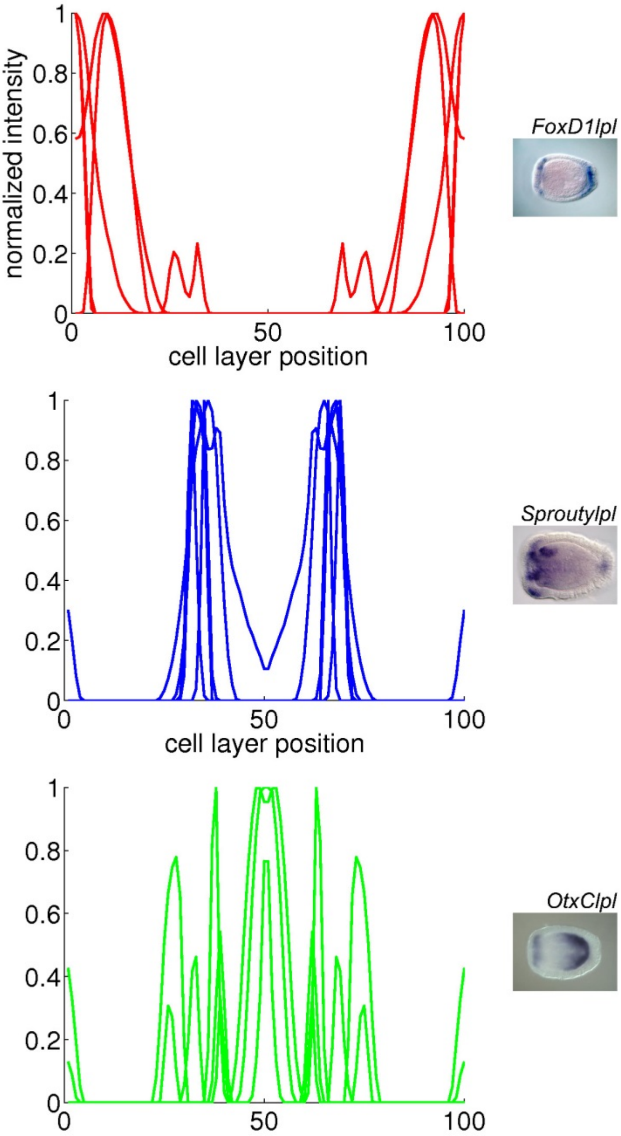


Figure 4.19: Combined plots of quantified gene expression patterns in the late planula stage. The plots represent the clusters in Figure 4.18. For one profile in each cluster, the original *in situ* hybridization is displayed.

For the planula stage, the characteristic expression domains in each of the three clusters are the aboral pole ectoderm, the oral pole ectoderm and the endoderm, respectively. The profiles in these

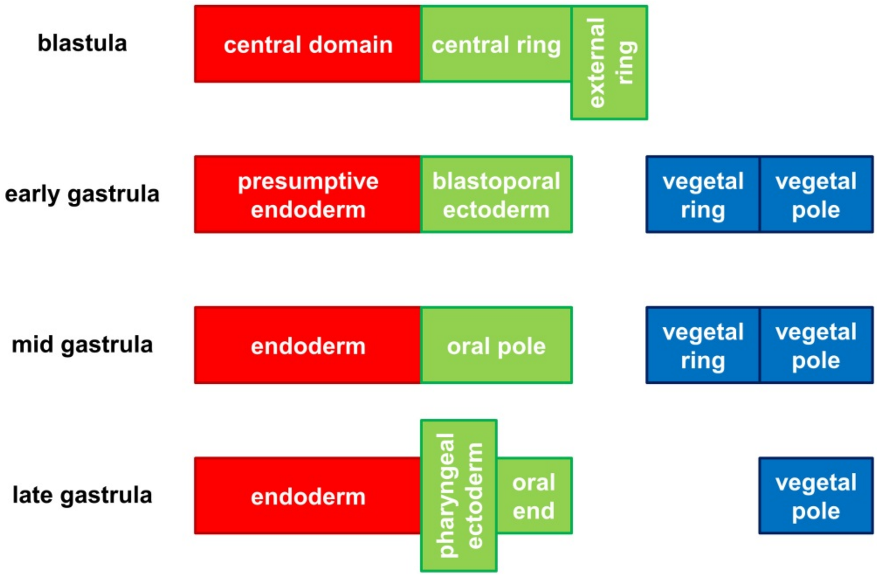


Figure 4.20: Overview of gene expression regions at various stages. The clusters and subclusters of correlating standardized expression profiles have been divided in three major regions: central domain/endoderm (red), central ring/external ring/oral ectoderm (green) and vegetal hemisphere/aboral ectoderm (blue).

clusters are collected in the blue, green and red plots in Figure 4.17, respectively.

For the late planula stage, the first two clusters show expression in the aboral and oral ectoderm, respectively, while the profiles in the third cluster show expression in multiple locations. The clusters are displayed in the red, blue and green plots in Figure 4.19, respectively.

4.3.2 Overview and analysis of expression summaries

From the Kahi Kai gene expression database, 73 genes have hybridization images available for at least two stages from blastula to late planula. These genes are listed with descriptions of their expression in Appendix D. Counts of pairwise expression domains

Table 4.1: Gene expression behavior with initial expression limited to the central domain/endoderm. Entries from the pairwise gene expression spreadsheet (Appendix E) that start with expression only in the central domain/endoderm are included in this table. For all combinations of stages, the modes of progression of gene expression (minor change:major change:vanished) are counted. The highlighted entries include the period from the early gastrula to the mid gastrula.

	initial stage	blastula	early gastrula	mid gastrula	late gastrula	early planula	planula
final stage							
early gastrula		(21:1:1)					
mid gastrula		(1:0:0)	(4:2:0)				
late gastrula		(0:0:0)	(4:1:0)	(3:0:0)			
early planula		(0:1:0)	(0:0:0)	(1:0:0)	(0:0:0)		
planula		(0:0:0)	(0:0:0)	(0:0:0)	(4:1:0)	(7:0:0)	
late planula		(0:0:0)	(0:0:0)	(0:0:0)	(0:0:0)	(0:1:0)	(9:0:0)

Table 4.2: Sums of possible periods for gene expression behavior in the central domain/endoderm. For each period, the possible changes in gene expression (minor change:major change:vanished) are added in the second column and expressed as percentages in the third column. As an example, the period from the early gastrula to the mid gastrula (grey background) is included in the combinations of stages highlighted in Table 4.1.

two-stage period	**possible counts**	**percentages**
blastula - early gastrula	(22:2:1)	(88:8:4)
early gastrula - mid gastrula	(9:4:0)	(69:31:0)
mid gastrula - late gastrula	(8:2:0)	(80:20:0)
late gastrula - early planula	(5:2:0)	(71:29:0)
early planula - planula	(11:2:0)	(85:15:0)
planula - late planula	(9:1:0)	(90:10:0)

are listed in Appendix E. The central domain/endoderm, central ring/external ring/oral ectoderm and vegetal hemisphere/aboral ectoderm are selected as major expression regions.

Table 4.3: Gene expression behavior with initial expression limited to the central ring/external ring/oral ectoderm. Entries from the pairwise gene expression spreadsheet (Appendix E) that start with expression only in the central ring/external ring/oral ectoderm are included in this table. For all combinations of stages, the modes of progression of gene expression (minor change:major change:vanished) are counted.

	initial stage	blastula	early gastrula	mid gastrula	late gastrula	early planula	planula
final stage							
early gastrula		(8:0:0)					
mid gastrula		(0:0:0)	(5:0:0)				
late gastrula		(1:0:0)	(2:2:0)	(8:0:0)			
early planula		(0:1:0)	(1:0:0)	(1:0:0)	(4:1:0)		
planula		(0:0:0)	(0:0:0)	(0:0:0)	(2:1:0)	(4:0:1)	
late planula		(0:0:0)	(0:0:0)	(0:0:0)	(0:0:0)	(1:0:0)	(3:0:0)

Table 4.4: Sums of possible periods for gene expression behavior in the central ring/external ring/oral ectoderm. For each period, the possible changes in gene expression (minor change:major change:vanished) are added in the second column and expressed as percentages in the third column.

two-stage period	**possible counts**	**percentages**
blastula - early gastrula	(9:1:0)	(90:10:0)
early gastrula - mid gastrula	(9:3:0)	(75:25:0)
mid gastrula - late gastrula	(13:3:0)	(81:19:0)
late gastrula - early planula	(8:3:0)	(73:27:0)
early planula - planula	(7:1:1)	(78:11:11)
planula - late planula	(4:0:0)	(100:0:0)

The changes or lack of change in expression patterns starting in a single major expression region are shown in Table 4.1 (central domain/endoderm), Table 4.3 (central ring/external ring/oral ectoderm) and Table 4.5 (vegetal hemisphere/aboral ectoderm).

Table 4.5: Gene expression behavior with initial expression limited to the vegetal hemisphere/aboral ectoderm. For pairs of subsequent stages, the modes of progression of gene expression (minor change:major change:vanished) are counted.

	initial stage	blastula	early gastrula	mid gastrula	late gastrula	early planula	planula
final stage							
early gastrula		(0:0:0)					
mid gastrula		(0:0:0)	(2:1:0)				
late gastrula		(0:0:0)	(2:0:0)	(1:0:0)			
early planula		(0:0:0)	(0:0:0)	(1:0:0)	(3:0:0)		
planula		(0:0:0)	(0:1:0)	(1:0:0)	(1:0:0)	(2:0:0)	
late planula		(0:0:0)	(0:0:0)	(0:0:0)	(0:0:0)	(1:0:0)	(1:1:0)

Table 4.6: Sums of possible periods for gene expression behavior in the vegetal hemisphere/aboral ectoderm. For each period, the possible changes in gene expression (minor change:major change:vanished) are added in the second column and expressed as percentages in the third column.

two-stage period	**possible counts**	**percentages**
blastula - early gastrula	(0:0:0)	(0:0:0)
early gastrula - mid gastrula	(4:2:0)	(67:33:0)
mid gastrula - late gastrula	(5:1:0)	(83:17:0)
late gastrula - early planula	(6:1:0)	(86:14:0)
early planula - planula	(5:1:0)	(83:17:0)
planula - late planula	(2:1:0)	(67:33:0)

Sums of blocks from these matrices are displayed in Table 4.2, Table 4.4 and Table 4.6, respectively.

A total of 25 genes for which *in situs* are available in the blastula stage are exclusively expressed in the central domain. From these genes, 22 are expressed in the endoderm in the next stage with available *in situs*. Expression of 2 genes has changed beyond the

Table 4.7: Gene expression behavior for all regions combined. All entries from the pairwise gene expression spreadsheet (Appendix E) are included in this table. For all combinations of stages, the modes of progression of gene expression (minor change:major change:appeared:vanished:none) are counted.

	initial stage	blastula	early gastrula	mid gastrula	late gastrula	early planula	planula
final stage							
early gastrula		(29:3:1:1:1)					
mid gastrula		(1:0:2:0:0)	(12:3:0:0:1)				
late gastrula		(1:0:0:0:0)	(8:3:2:0:0)	(12:0:1:0:0)			
early planula		(0:2:0:0:0)	(1:0:1:0:0)	(4:1:4:0:0)	(8:2:1:0:0)		
planula		(0:0:0:0:0)	(0:3:3:0:0)	(2:0:3:0:0)	(7:2:2:0:0)	(17:0:0:1:0)	
late planula		(0:0:0:0:0)	(0:0:0:0:0)	(0:2:1:0:0)	(0:0:0:0:0)	(2:1:0:0:0)	(16:2:0:0:0)

endoderm in the next available stage and 1 gene is no longer expressed at all. This means that 88% of the genes expressed only in the central domain in the blastula is subsequently expressed only in the endoderm. From all combinations of subsequent available stages that include the early gastrula-mid gastrula period

Table 4.8: Sums of possible periods for gene expression behavior in all regions. For each period, the possible changes in gene expression (minor change:major change:appeared:vanished:none) are added in the second column and expressed as percentages in the third column.

two-stage period	**possible counts**	**percentages**
blastula - early gastrula	(31:5:3:1:1)	(76:12:7:2:2)
early gastrula - mid gastrula	(23:11:8:0:1)	(53:26:19:0:2)
mid gastrula - late gastrula	(28:11:15:0:0)	(52:20:28:0:0)
late gastrula - early planula	(22:12:15:0:0)	(45:24:31:0:0)
early planula - planula	(28:8:9:1:0)	(61:17:20:2:0)
planula - late planula	(18:5:1:0:0)	(75:21:4:0:0)

(highlighted in Table 4.2), 13 genes are initially only expressed in the central domain/endoderm. From these genes, 4 expression patterns have changed beyond the endoderm, which is 31%. Likewise, out of the 10 genes initially limited to the central domain/endoderm, the major change of 2 (20%) could possibly occur in the mid gastrula-late gastrula period. Out of the 7 genes expressed between the late gastrula and early planula, 2 (29%) could have changed beyond the endoderm during this interval.

Out of the 10 genes expressed only in the central ring or external ring in the blastula, 9 (90%) are subsequently expressed only in the oral ectoderm.

No blastula *in situs* are available for genes expressed in the vegetal hemisphere. Out of the 6 genes expressed only in the early gastrula vegetal hemisphere, 4 (67%) are subsequently expressed only in the aboral ectoderm.

The changes or lack of change in expression in all regions are shown in Table 4.7. Sums of blocks from this matrix are displayed in Table 4.8. Out of all 41 genes with images stored in the blastula stage, 31 (76%) display expression in the next available stage in the same major region(s). The added percentages for major changes and first appearance of gene expression are highest in the early gastrula to early planula stages (n = 214, p = 0.03, two-tail Fisher's exact test).

The latter trend was also observed for the subsets of genes initially expressed exclusively in the endoderm and in the oral ectoderm, respectively. This could not be statistically confirmed though (p = 0.31 and p = 0.42, respectively), due to lower numbers (n = 53 and n = 51, respectively). The data points for genes expressed in the aboral ectoderm are too few (n = 28) to observe any trend.

4.4 Discussion

A gene with expression profiles from multiple samples at one stage can belong to more than one cluster. This may be due to the noise in gene expression among individuals. Another possible explanation is that the expression patterns are different across individual *in situs* at the beginning and at the end of a developmental stage. This change could be caused by cells migrating from one region to another, or by dynamic regulatory interactions. In the current study, this issue is handled by performing cluster analyses to all expression profiles for all genes within a broad time window instead of analyzing blurred averages of each gene. A solution to this uncertainty would be an increased time resolution for the expression profiles, resulting in more precise regions for narrower time windows. In *D. melanogaster* embryos, the definition of narrow time classes allowed the observation of significant domain shifts [17]. In the sea urchin, precise timing resulted in a sequence of spatial regulatory states [87].

Central domain expression is generally persistent in the endoderm, while expression in the central and external rings is often limited to the oral ectoderm during and after gastrulation. The first entry in Table 4.1 shows that the early gastrula expression pattern is known

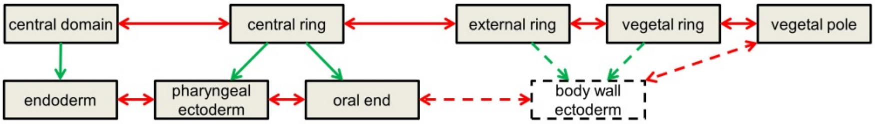


Figure 4.21: Proposed interactions between gene clusters in various expression regions. Early expression clusters activate or develop into later expression clusters in corresponding regions (green arrows). Neighboring expression clusters inhibit each other (red arrows). The blastula domains developing into the body wall ectoderm region are derived from a single gene expression pattern (see text). For this reason, the proposed interactions with the body wall ectoderm are indicated with dashed arrows.

for 23 genes that are expressed exclusively in the central domain in the blastula stage. Out of these 23 genes, 21 are expressed in the presumptive endoderm in the early gastrula stage. Similarly, all 8 genes expressed exclusively in the central or external ring in the blastula with known expression in the early gastrula, are expressed in the blastoporal ectoderm in the latter stage as summarized in the first entry in Table 4.3. Moreover, expression is observed exclusively in the same major domain in the next available stage for at least 69% of all instances of expression limited to either of these two major domains, as indicated in the last column of Tables 4.2 and 4.4. These observations are a strong indication that the central domain differentiates into the later endoderm and that the central and external rings become oral ectoderm. The vegetal hemisphere likely becomes aboral ectoderm, but this is based on a small number of expression patterns (Table 4.6). These differentiation regions are in agreement with the locations of a dye injected into an *N. vectensis* egg and recorded during embryonic development [47]. Injection of lineage tracers into individual cells of developing *Nematostella* embryos results in contiguous clones of labeled cells without long distant migration of individual cells (Martindale, unpublished observations). Based on the statistical stability of the main gene expression regions, a regulation mechanism is proposed where genes in early regions activate genes in corresponding later regions, while genes in adjacent regions repress one another (Figure 4.21).

If changes in expression patterns occur consistently in many genes, this might indicate a collective cellular motion in the embryo. During gastrulation, expression in the region between the oral and vegetal poles is observed for the quantified profiles of *Anthox1*, *FGFRa*, *FoxD.1*, *Sox1*, *Sox3* and *Rx*. Due to intense *Sox1* expression in the oral pole, the *Sox1* profiles are outside the vegetal ring clusters. *Anthox1*, *FGFRa* and *FoxD.1* are likewise classified as members of the vegetal pole cluster because their strongest expression occurs in the corresponding region. For these six genes, a single *in situ* in the blastula stage has been stored: this *Sox1* image shows staining throughout the animal hemisphere. This hybridization experiment is a weak indication that genes in the body wall ectoderm are first expressed in the external ring or vegetal ring of the blastula. This is basically the null hypothesis; in the absence of any known collective cell movements in the aboral ectoderm during gastrulation, expression in this region has likely remained stationary from the blastula. The body wall has been included in the proposed regulatory interactions among clustered gene expression regions (Figure 4.21). The vegetal domains of *Anthox1*, *FGFRa* and *FoxD.1* become restricted to the aboral end in the planula stage, while *Sox1* and *Sox3* move towards the oral pole in the planula. Meanwhile, the *Rx* domain has significantly expanded in the late planula. These changes in gene expression domains may indicate that ectoderm elongation in the planula stage is most pronounced in the initially narrow *Rx* expression domain. Even though this hypothesis is based on few observations, it could be tested with fate mapping experiments. In the frog *Xenopus laevis*, the gene *rax* promotes cell proliferation in developing retinal tissue [88]; *Rx* may similarly induce tissue growth in *N. vectensis*.

Many hierarchical clusters include profiles with strong pairwise correlations to profiles in other clusters. This is caused by the partial overlap of the expression domains that characterize the clusters and by expression of some genes in both regions. The average cutoff value for the main clusters is 0.9; this value is quite high and indicates fuzzy boundaries between many expression clusters. These fuzzy boundaries between expression clusters are likely due to dynamic changes in gene expression. Out of the 25 genes expressed only in the central domain, 3 (12%) will exhibit major expression changes, or terminate their expression (Table 4.2, first row). Out of the 10 genes expressed in the central ring or external ring, 1 (10%) is not expressed in the oral ectoderm (Table 4.4, first row). This explains why some genes are expressed in multiple clusters at the same stage, and why some genes appear in clusters for different expression domains between developmental stages.

According to the hypothesis that the position of cells in the planula is largely determined by their position in the blastula, the change of gene expression patterns across major regions must be the result of regulatory action. The added percentages for possible major change and first appearance of gene expression in the early gastrula to early planula stages are 45%, 48% and 55% (Table 4.8). In comparison, these percentages are 37% and 25% in the early planula to late planula stages. These observations suggest that more dynamic changes in expression occur during gastrulation than in the period that the planula transforms into a polyp. For the study of pattern formation in *Nematostella vectensis*, recording gene expression during the relatively short period of gastrulation should therefore be more informative than monitoring the planula stages. In general, during gastrulation it is decided in which major region(s) the genes are expressed. Detailed expression patterns arise during the planula stages, generally within the bounds of the major regions determined for each gene at the end of gastrulation. The appearance of differential details explains the decrease in correlation among gene expression patterns after gastrulation (Figures 4.14-4.19). This loss of correlation could also be caused by the lower number of quantified patterns in the planula stages; with additional gene expression quantifications in these stages of development it could be tested whether the expression domains diverge or whether they form a new set of clusters.

One aspect of gene expression that requires spatial localization, is the determination of the secondary axis in *N. vectensis*. Individual cells appear indistinguishable during the early cleavage cycles until the early blastula starts oscillating at the animal pole [63]. This oscillation stops at the mid blastula transition, when the synchrony of cell divisions is lost and the blastula remains spherical. The spherical blastula symmetry is permanently broken at the onset of gastrulation [89]. During gastrulation, Bmp ligands and antagonists are asymmetrically expressed along the secondary axis [85]. Based on these observations, determination of the secondary axis may coincide with the mid blastula transition. The asynchrony in cell divisions then produces a stochastic perturbation in a morphogen concentration which would define the secondary axis.

The gene expression database contains various genes that are differentially expressed along the secondary axis, although most are not identified in the expression summaries. From the incomplete list of genes included in the expression pattern overview in this study (Appendix D), various genes involved in secondary axis formation can already be identified. The genes *Anthox7, Anthox8b, Bmp2/4, chordin*, *Gbx*, *Hex*, *Msx*, *Msx2, NvHD060, Smad1/5* and *Vent1* exhibit noncylindrical expression patterns. Quantification of these patterns requires a two-dimensional or three-dimensional template and eventually a three-dimensional detection method. Especially *Vent1* asymmetric expression in the early gastrula stands out. In various animals, *vent* genes are involved in a Bmp signaling feedback loop [90]. Curiously, the genes *Bmp2/4*, *chordin* and *Smad1/5* in this signaling pathway are still symmetrically expressed in the *N. vectensis* early gastrula (or at least their asymmetry is less pronounced).

The *N. vectensis in situ* hybridization collection contains expression data for less than half of all developmental stages for most genes stored. Despite the sparsity of this dataset, an analysis results in meaningful insights. Sparse data is common for databases that contain labor intensive measurements, and standardization allows global analysis approaches to be applied to incomplete biological databases. Developmental gene expression databases contain large sets of genes with many regulatory interactions. Our clustering-by-region approach is convenient to select genes from a large set for computational regulatory network modeling.

4.5 Conclusions

Our cluster analysis indicates that early gene expression domains in *N. vectensis* are spatially separated in a stable sequence along the primary body axis. An additional statistical analysis indicates that precise gene expression domains in *N. vectensis* are generally formed by two processes. Genes that are expressed in the blastula appear in broad expression regions. During gastrulation and planula development, the expression domains are refined within the boundaries of these broad regions.

It should be stressed that no additional experiments have been performed for our study. Spatial expression data in a public gene expression database have been quantified and analysed, and these analyses resulted in new hypotheses on cellular migration in *N.*

vectensis. Spatial gene expression patterns have been collected in extensive databases for other model animals as well, and a similar computational approach can generate hypotheses about cellular migration before fate maps are available for these animals.

5 Discussion

The main purpose of this thesis is to derive gene regulatory networks for animals with changing morphologies during embryonic development. With the starlet sea anemone *Nematostella vectensis* as a biological model, three strategies towards this goal are described in the main chapters.

First, a method for spatial gene expression quantification has been designed to compare expression patterns across developmental stages and among multiple genes. One-dimensional gene expression profiles have been included in a preliminary hierarchical clustering. This cluster analysis is a proof of concept that the standardized format for spatial gene expression can be used for numerical analyses. A one-dimensional profile is only a complete description for gene expression patterns with radial symmetry about the primary body axis of *N. vectensis*. Therefore, two transformations have been described to allow a complete two-dimensional description for expression patterns with lower symmetry. The construction of a two-dimensional profile is not very precise and makes some strong assumptions on the properties of *in situ* hybridization images. With the emergence of three-dimensional imaging techniques for gene expression, these limitations might be removed, as the transformations become superfluous.

Second, regulatory interactions have been inferred for a small set of gut formation genes in *N. vectensis* to demonstrate a computational workflow for the production of gene regulatory networks. Quantified gene expression patterns have been included as references for parameter estimation. The optimization is a proof of concept that the standardized gene expression profiles are suitable input for numerical modeling. The correlation matrix confirmed the existence of two previously discoved gene expression domains in the blastula stage. The strong clustering reveals that these domains are maintained in later stages of *N. vectensis* development. The *in situ* hybridizations exhibit some limitations for numerical simulations. For samples in the blastula and planula stages, the development time can not be derived from the embryo morphology. Although the staining technique does not provide the amount of gene expression, the relative boundaries of gene expression domains are accurately captured and correct domain boundaries are sufficient for computing gene regulatory networks [79]. The sets of estimated parameters confirm the main patterning mechanism for the sea anemone gut. Selecting genes that are representative for the clustered expression patterns is likely the most effective improvement to the gene regulatory model.

Third, analyses have been performed on spatial gene expression patterns from an *in situ* hybridization database to obtain new knowledge from this large collection of experimental data. Hierarchical clusterings for seven stages of *N. vectensis* embryonic development revealed three major gene expression domains. The clustered expression domains in the blastula stage correspond to the gene expression domains in a previous experimental study. The major expression domains likely correspond to spatial gene regulatory modules. Based on this assumption, a network has been proposed with regulatory interactions among the various regions in the developing sea anemone embryo. A statistical analysis has been performed on all genes in the database with expression patterns measured for at least two stages. The result suggests that more gene expression patterns shift across major domain boundaries during gastrulation than during planula formation.

6 Conclusion

Chapter 2 describes a method for spatial gene expression quantification in the developing sea anemone *Nematostella vectensis*. This method succeeds to produce standardized one-dimensional profiles based on *in situ* hybridization images of radially symmetrical gene expression patterns. From an accurate three-dimensional gene expression pattern, the quantification method can produce an accurate two-dimensional stack of expression profiles. Major limitations prevent the construction of a precise three-dimensional gene expression pattern from RNA hybridization images. However, new experimental techniques can measure gene expression in three dimensions. This means that in the future expression patterns during the embryonic development of *N. vactensis* could be quantified for all genes.

Chapter 3 demonstrates a workflow for the production of gene regulatory networks during the embryonic development of *N. vectensis*. The hierarchical clustering successfully showed strong correlation among gene expression patterns in two spatial domains that were previously discovered by experiments. Quantified gene expression patterns were suitable as input for a parameter estimation based on numerical simulations. The accuracy of the simulated gene expression patterns was limited by weaknesses of the *in situ* hybridization data. Still, the resulting gene regulatory network exhibited the main mechanism for gut patterning. Alternative sets of selected genes were proposed to improve the reliability of the simulations.

Chapter 4 demonstrates how numerical analyses can provide new insights from an *in situ* hybridization database. Three major gene expression domains were identified by hierarchical clusterings for seven stages of embryonic development. A statistical analysis showed that more genes change their expression pattern across these major domains during gastrulation than during the larval stage. A domain-based gene regulatory network has been proposed from the minor domains identified in the cluster analysis.

The various computational methods described in this thesis can be used to produce gene regulatory networks for the developing sea anemone. The numerical methods already exist and spatial gene expression quantification is necessary to provide numerical gene expression profiles. The procedure that we followed to arrive at this quantification method can be generally applied to design similar methods for other model animals. High-resolution microscopy techniques are used to visualize embryo morphologies that are suitable to derive morphological templates. Moreover, techniques to measure spatial RNA and protein concentrations are applied to many model animals in developmental biology. Combined with techniques that provide quantitative gene expression levels, systematic spatial measurements can improve the accuracy of gene regulatory network simulations. In turn, more precise and sophisticated gene regulation models will result in a better understanding of regulatory interactions.

Appendix A: Matlab software

The code that has been produced for Chapter 2 of this thesis, is part of the corresponding publication entitled “Spatial gene expression quantification: a tool for analysis of *in situ* hybridizations in sea anemone *Nematostella vectensis*”. The additional file to this paper contains various Matlab programs, subfunctions, library files and manuals.

A1: GENEXP program “genexp.m”
This is the major program for decomposing the cell layer of gene expression images and for profile editing. The program runs on Matlab (2008), but unfortunately it fails on Matlab (2010) and later. The most likely cause of the failure is the adaptation of feedback functions in the later versions. Redesigning the user interface in GUIDE (the Matlab environment for developing graphical user interfaces) should solve this issue.

A2: ONEVIEW program “oneview.m”
This is the program for generating a 2D gene expression array from a single decomposed image.

A3: TWOVIEWS program “twoviews.m”
This is the program for generating a 2D array from a decomposed image and a corresponding perpendicular view.

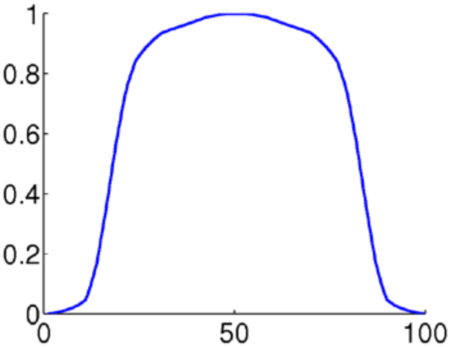
A4: Supporting files
The Matlab user interface “genexp.fig” is necessary to run GENEXP. The directory “embryodata” contains geometry definitions for developing embryos. The directory “functions”

contains Matlab functions used in the main programs A1, A2 and A3. The Matlab script “stanpro.m” creates a standardized 1D numerical array from a GENEXP profile figure. The text files “readme_genexp.m”, “readme_oneview.m” and “readme_twoviews.m” are the manuals for the main programs A1, A2 and A3, respectively.

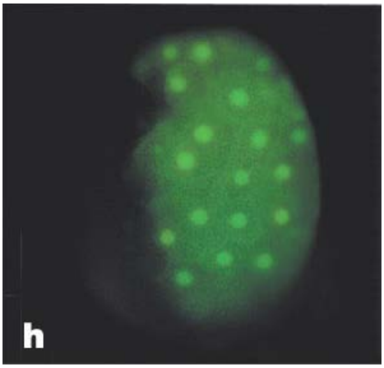
Appendix B: Quantified gene expression profiles

For Chapter 3 of this thesis, microscopy images have been collected that visualize spatial expression patterns of gene products in *Nematostella vectensis* embryos. These gene expression images and their quantified profiles are listed below. This dataset is included as supporting information in the corresponding publication entitled “A computational approach towards a gene regulatory network for the developing *Nematostella vectensis* gut”.

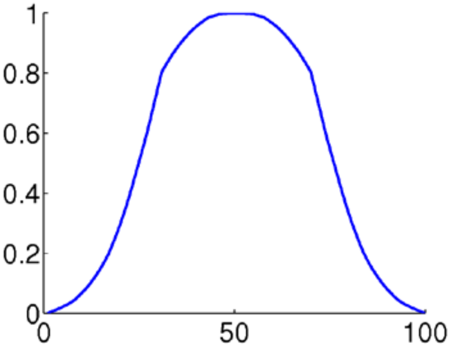
For each entry, the spatial gene expression quantification is displayed on the left. The image on the right is the original gene product visualization. Any annotations in the embryo images appear in the original publications; the meaning of these annotations is irrelevant for quantification.



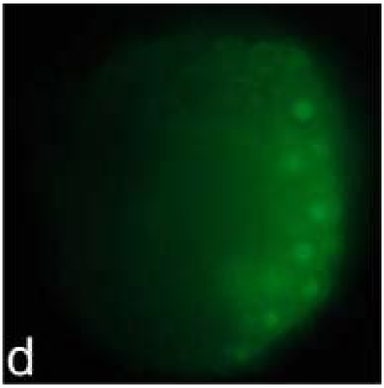
Bcat:b1



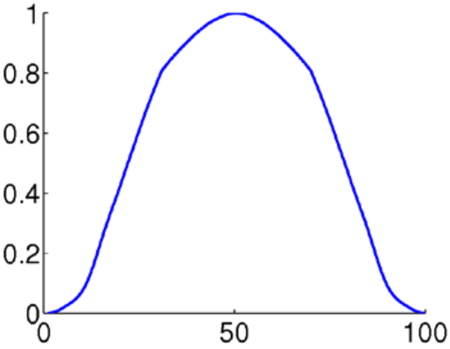
β-catenin in blastula
oral right [57]



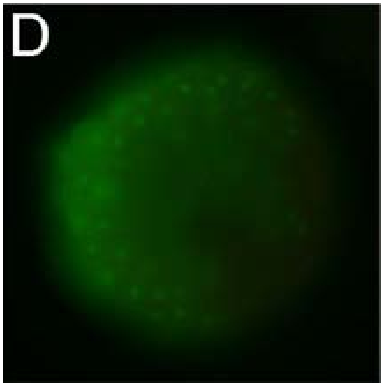
Bcat:b2



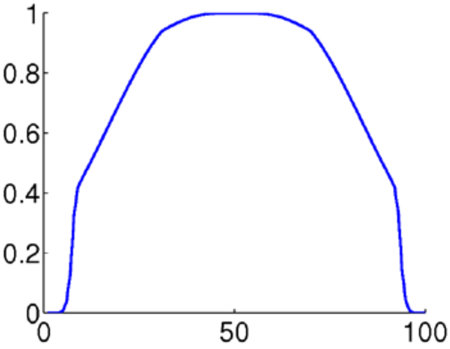
β-catenin in blastula
oral right [91]



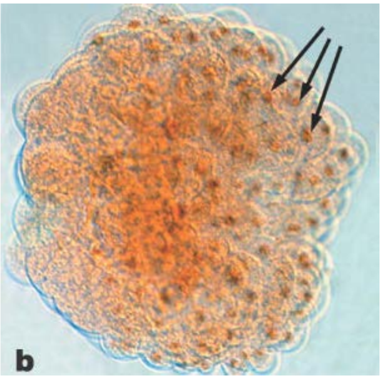
Bcat:b3



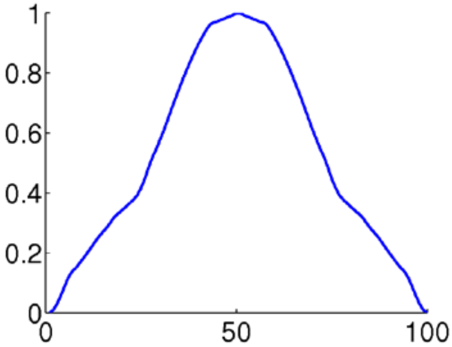
β-catenin in blastula
oral left [47]



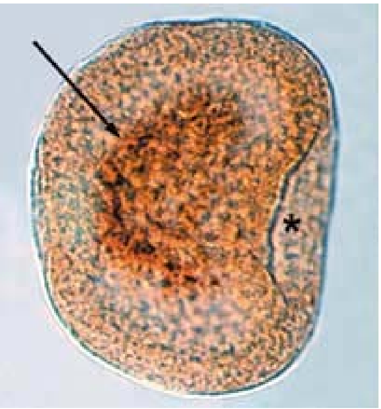
Bcat:bo



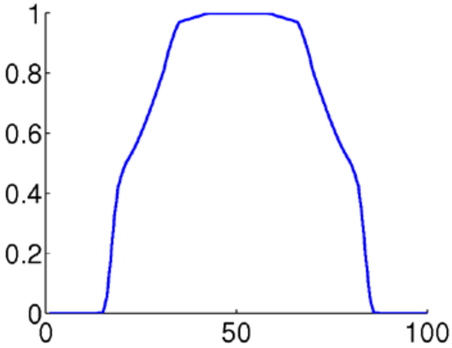
β-catenin in blastula
oral right [57]



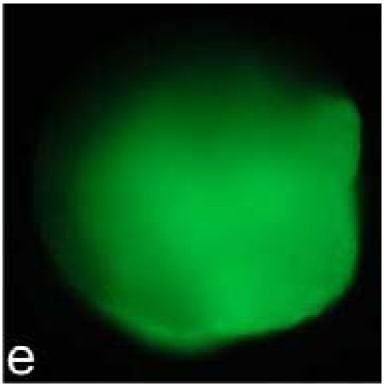
Bcat:27



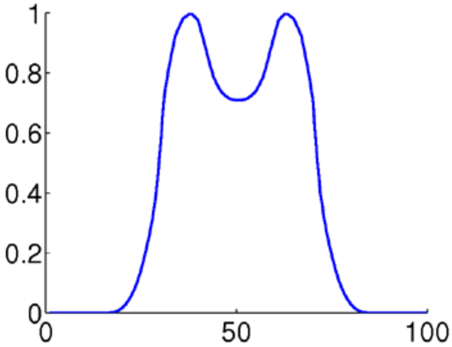
β-catenin at 27 hpf
oral right [57]



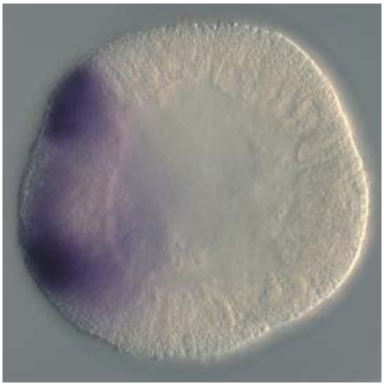
Bcat:50



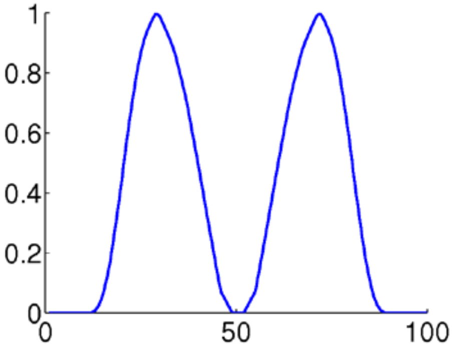
β-catenin at 50 hpf
oral right [91]

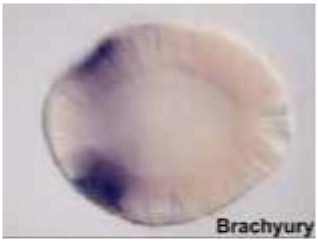


bra:bl



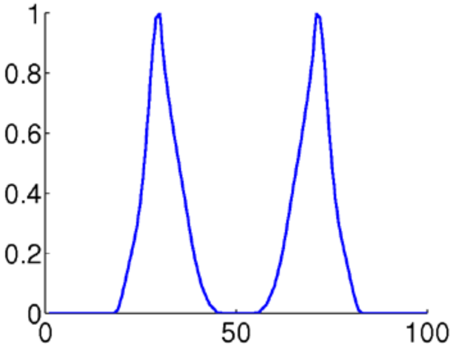
brachyury in blastula
oral left [50,71]

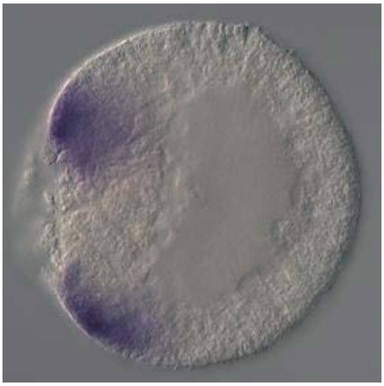




bra:bx

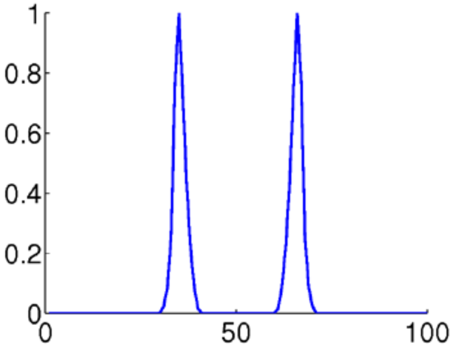
brachyury in blastula
oral left (Martindale Lab, unpublished)

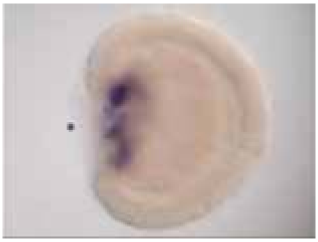




bra:24

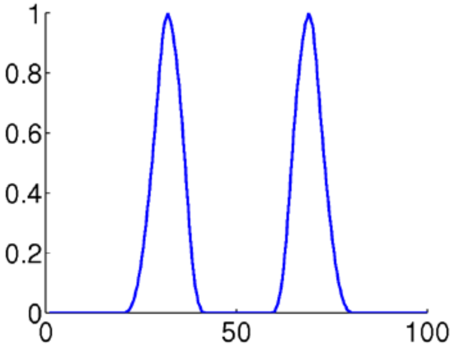
brachyury at 24 hours
oral left [50,71]



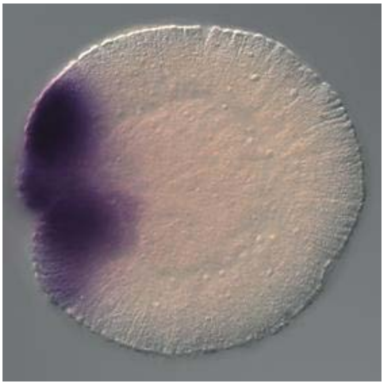


bra:27

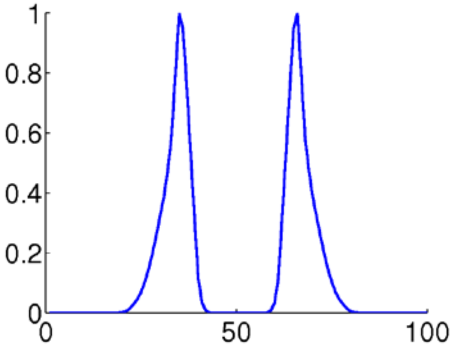
brachyury at 27 hours
oral left (Martindale Lab, unpublished)



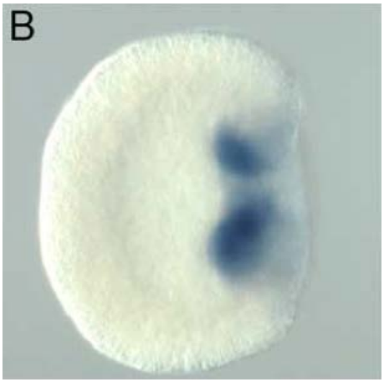
bra:30



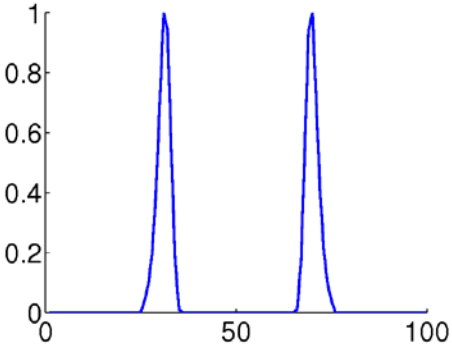
brachyury at 30 hours
oral left [50,71]



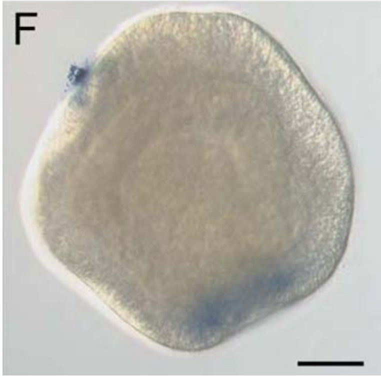
bra:35



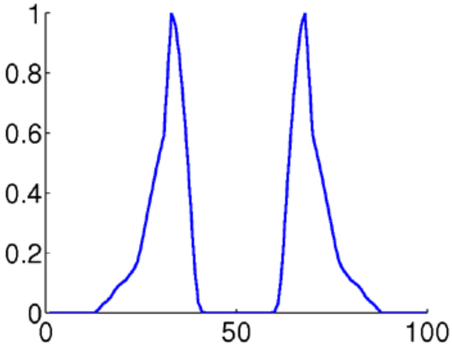
brachyury at 35 hours
oral right [81]



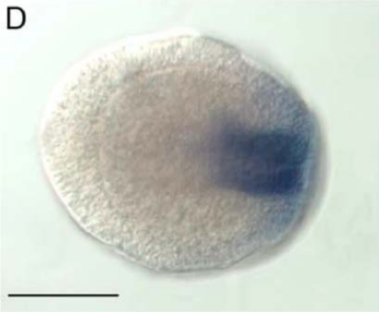
bra:50



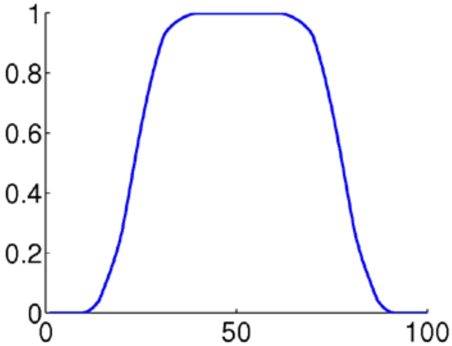
brachyury at 50 hours
oral lower right [92]



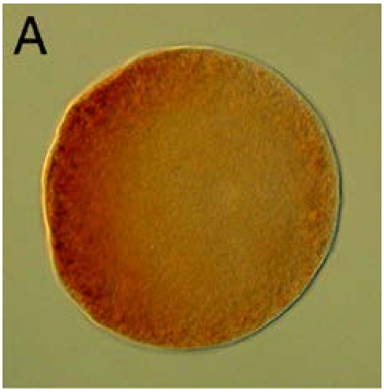
bra:70



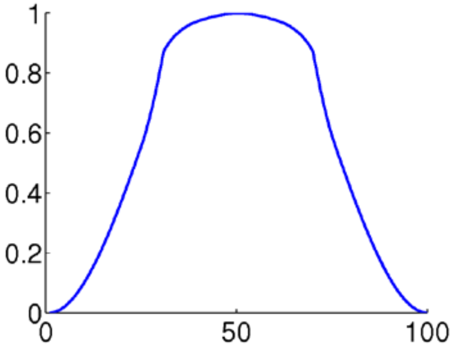
brachyury at 70 hours
oral right [81]



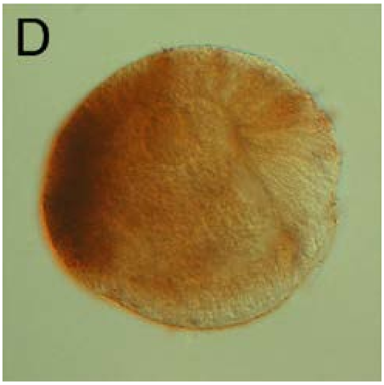
dsh:zy



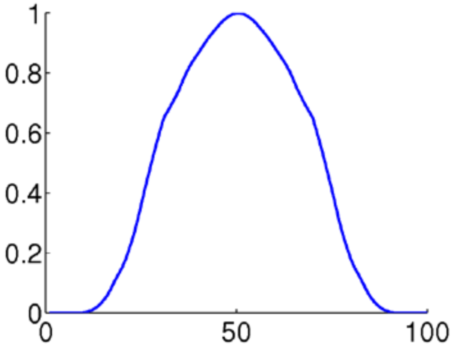
dishevelled in zygote
oral left [47]



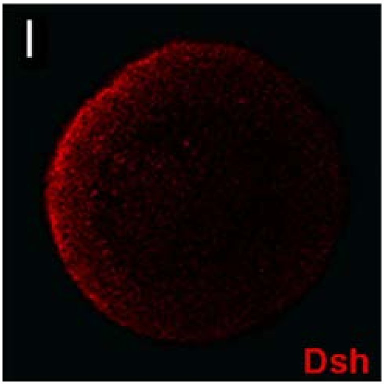
dsh:bl



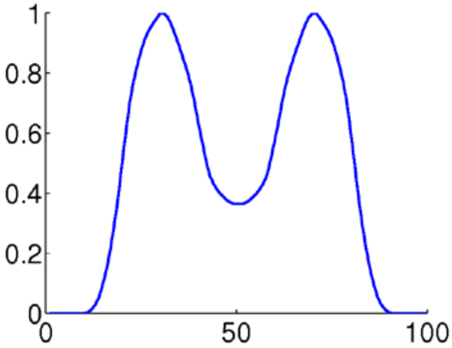
dishevelled in blastula
oral left [47]



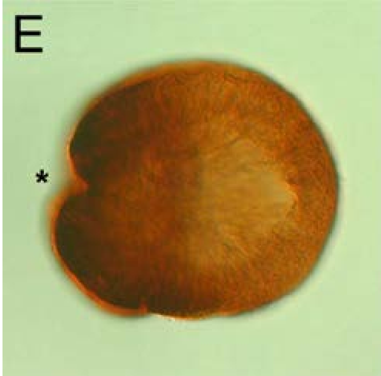
dsh:br



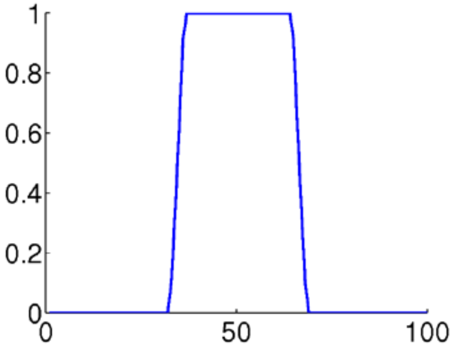
dishevelled in blastula
oral left [47]



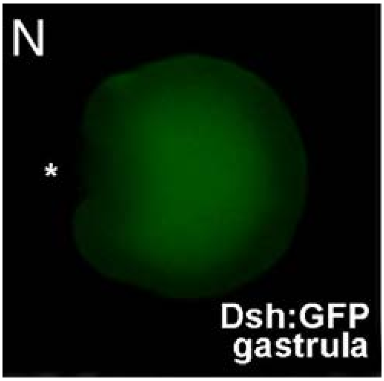
dsh:26



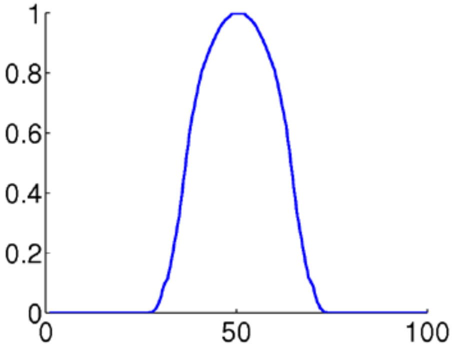
dishevelled at 26 hours
oral left [47]



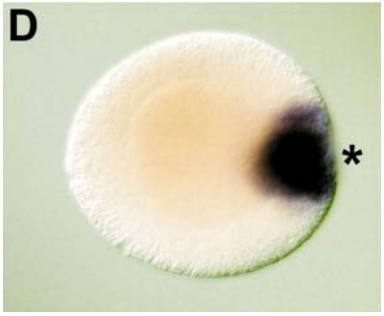
dsh:40



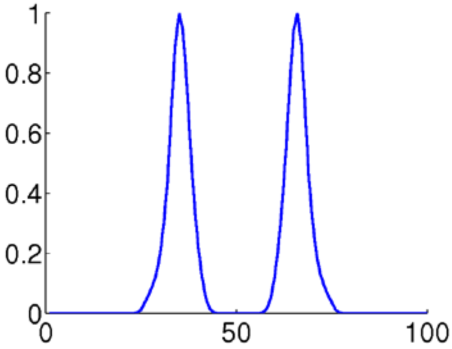
dishevelled at 40 hours
oral left [47]



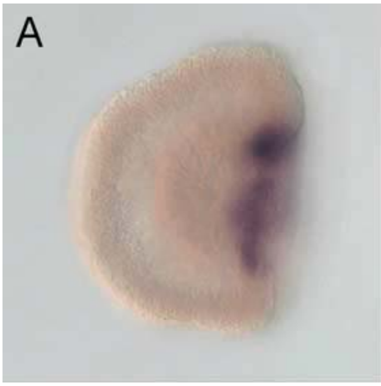
fork:22



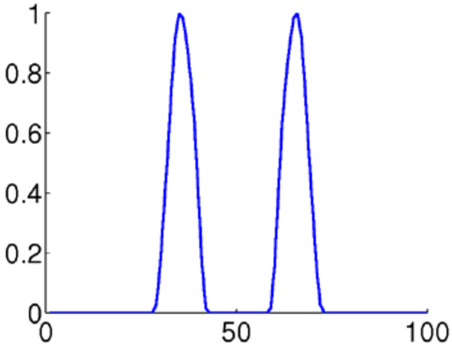
foxA at 22 hours
oral right [27]



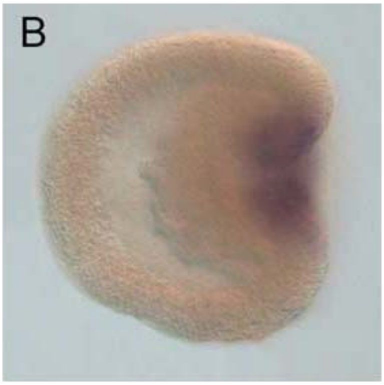
fork:26



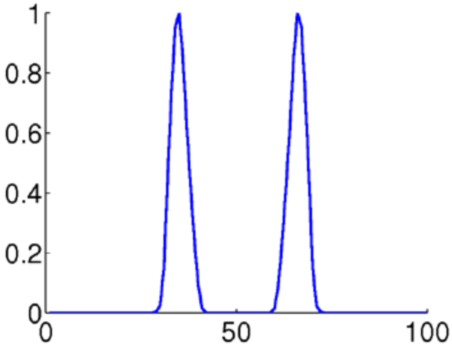
foxA at 26 hours
oral right [81]



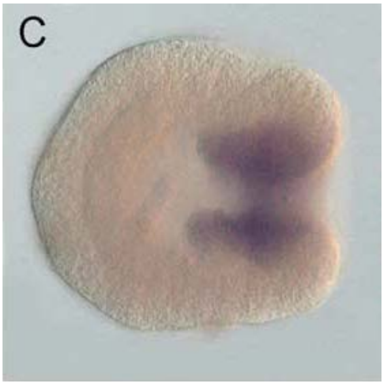
fork:27



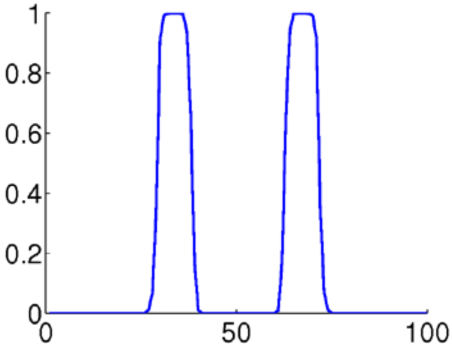
foxA at 27 hours
oral right [81]



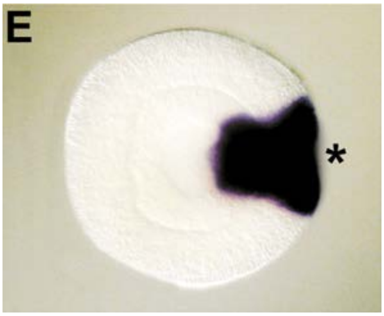
fork:30



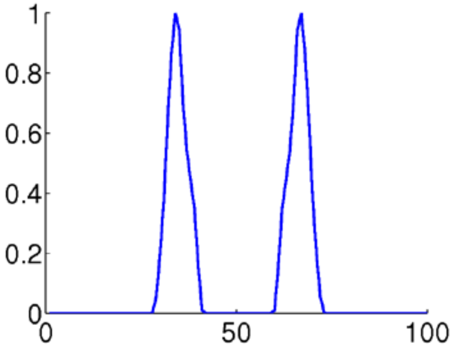
foxA at 30 hours
oral right [81]



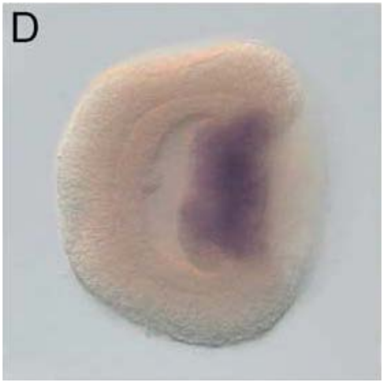
fork:32



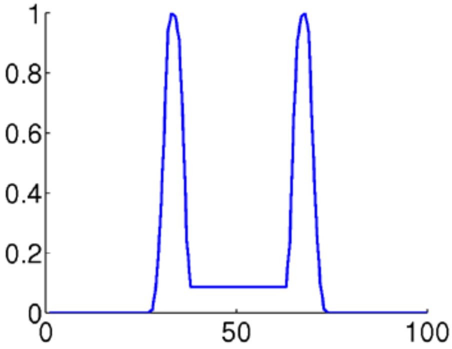
foxA at 32 hours
oral right [27]



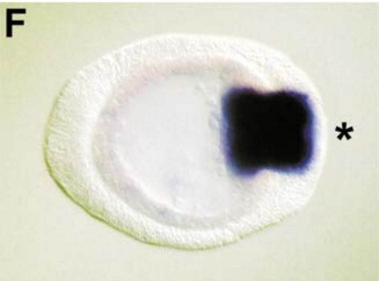
fork:33



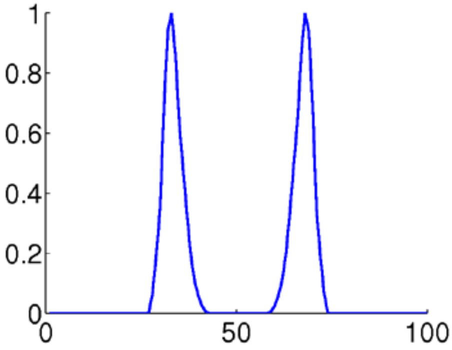
foxA at 33 hours
oral right [81]



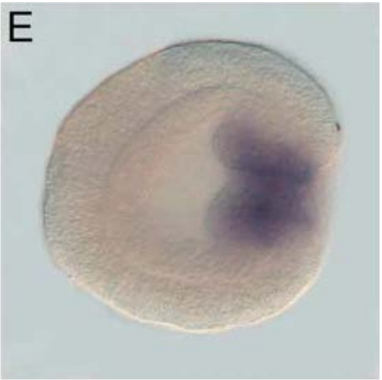
fork:36



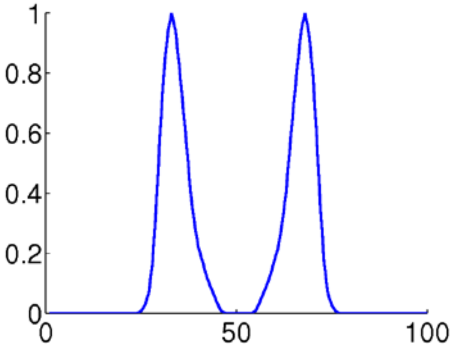
foxA at 36 hours
oral right [27]



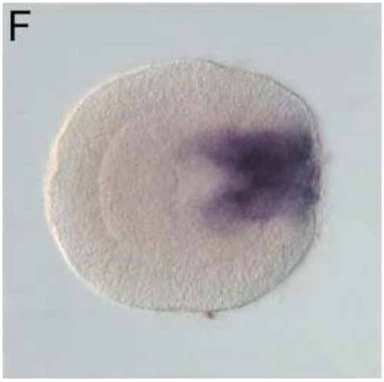
fork:40



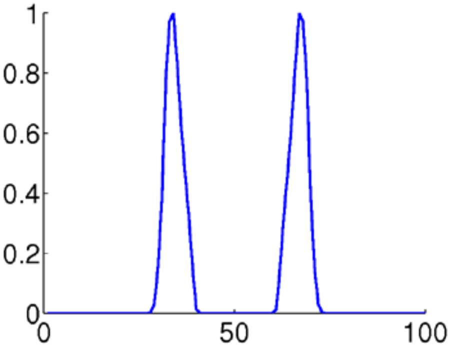
foxA at 40 hours
oral right [81]



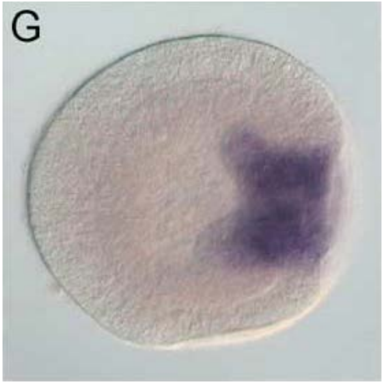
fork:42



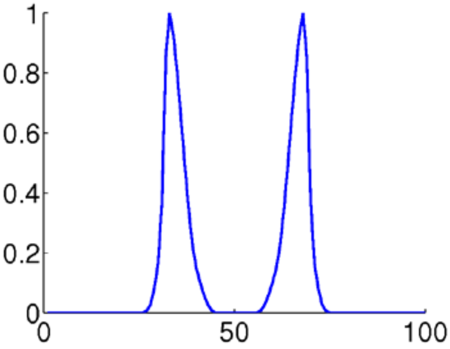
foxA at 42 hours
oral right [81]



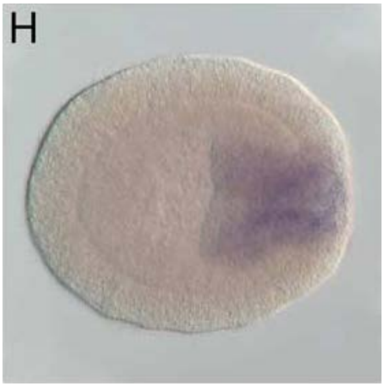
fork:44



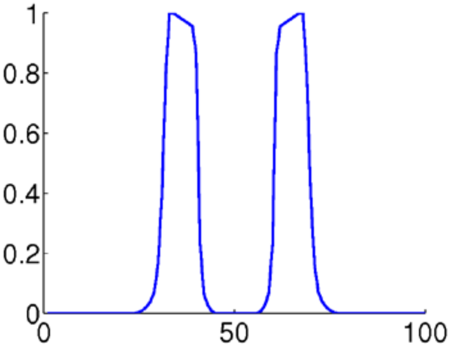
foxA at 44 hours
oral right [81]



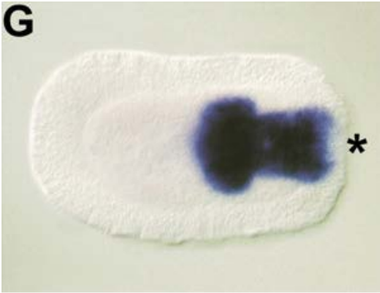
fork:54



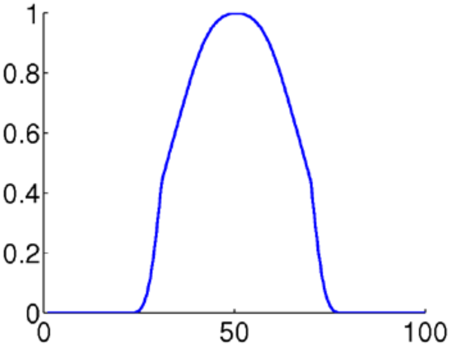
foxA at 54 hours
oral right [81]



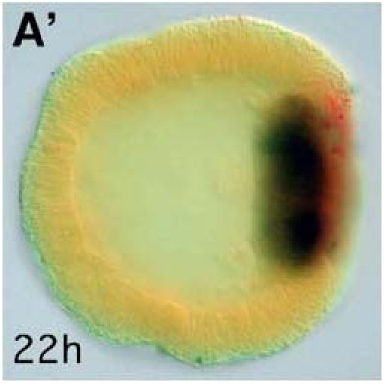
fork:60



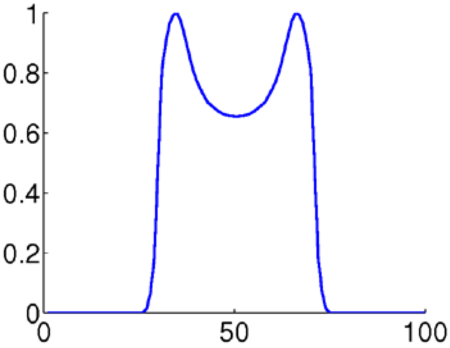
foxA at 60 hours
oral right [27]



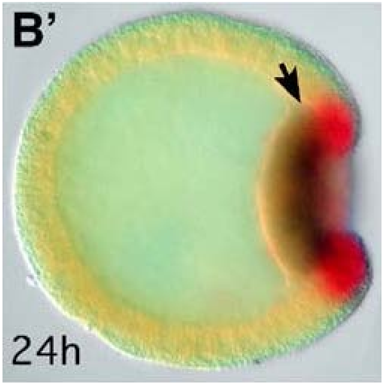
foxA:22



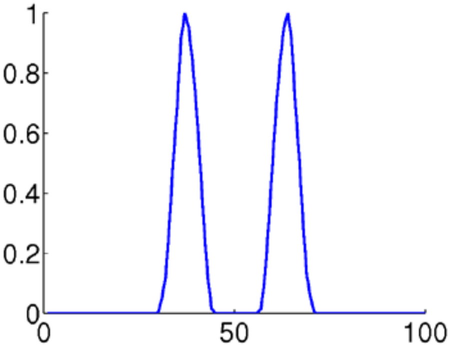
foxA at 22 hours
red staining, oral right [24]



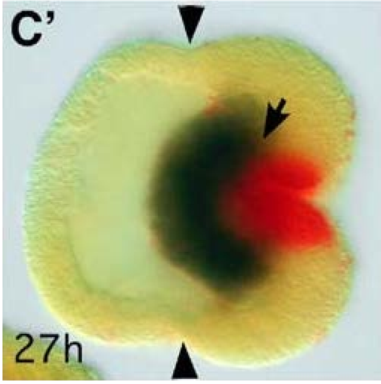
foxA:24



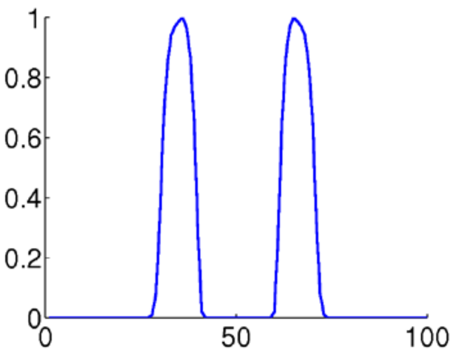
foxA at 24 hours
red staining, oral right [24]



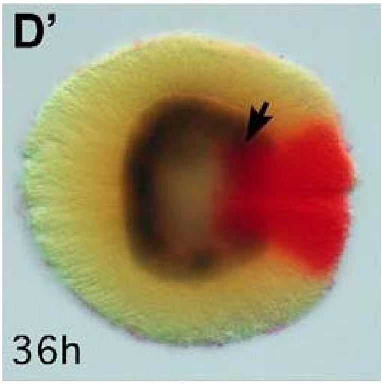
foxA:27



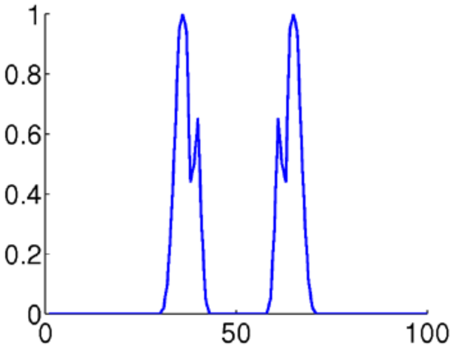
foxA at 27 hours
red staining, oral right [24]



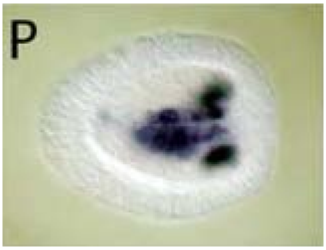
foxA:36



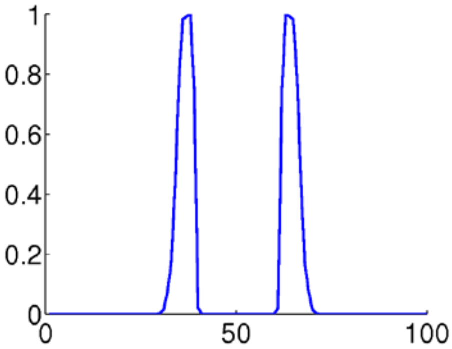
foxA at 36 hours
red staining, oral right [24]



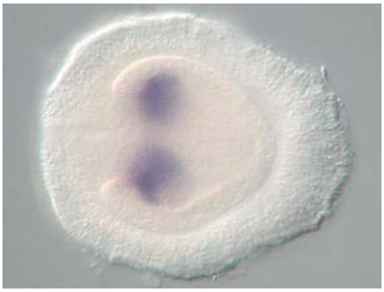
foxC:60



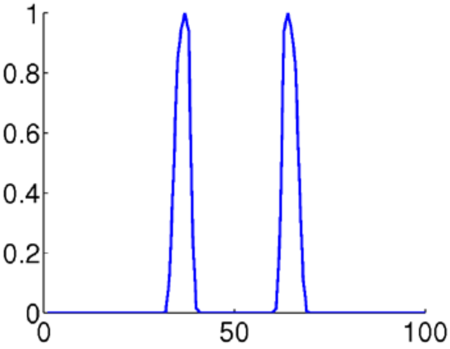
foxC at 60 hours
oral right [54]



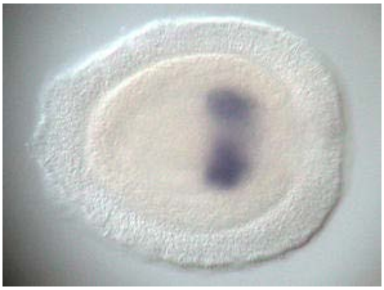
msx:50



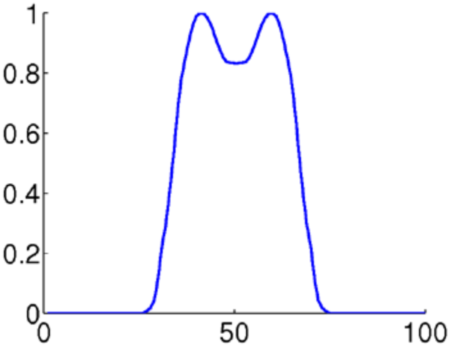
msx at 50 hours
oral left [50,71]



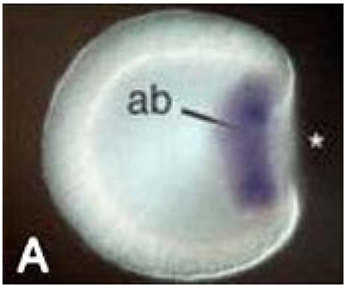
msx:60



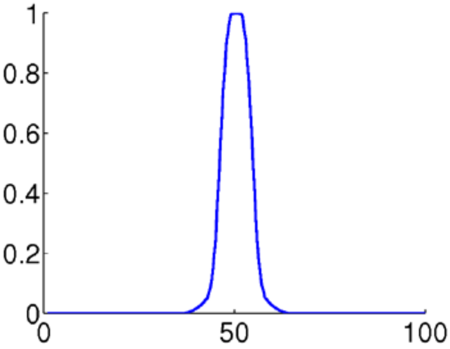
msx at 50 hours
oral right [50,71]



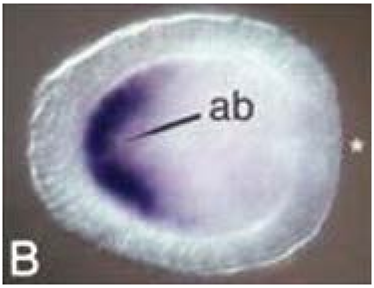
otxA:24



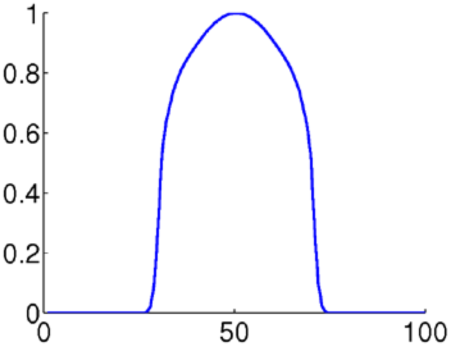
otxA at 24 hours
oral right [93]



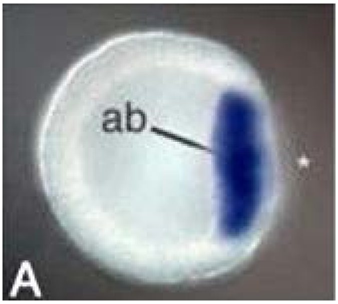
otxA:60



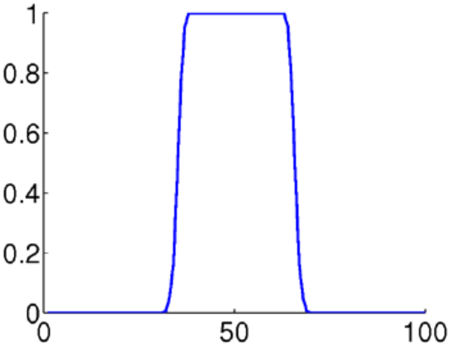
otxA at 60 hours
oral right [93]



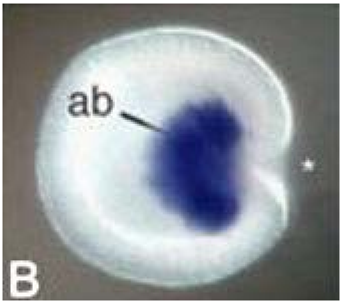
otxB:22



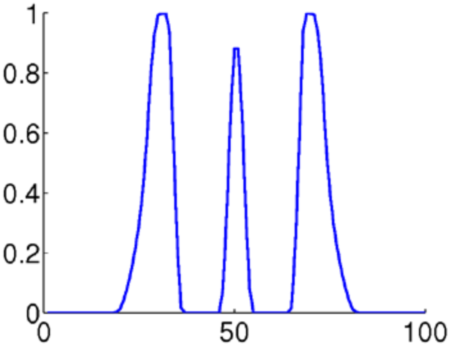
otxB at 22 hours
oral right [93]



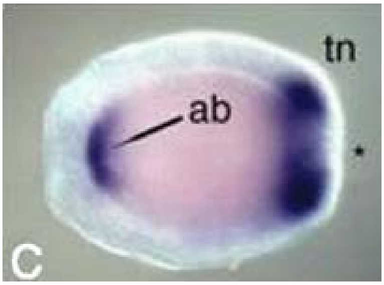
otxB:27



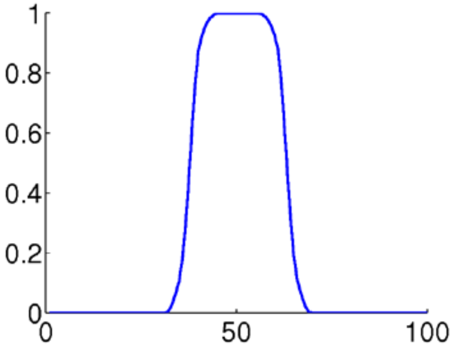
otxB at 27 hours
oral right [93]



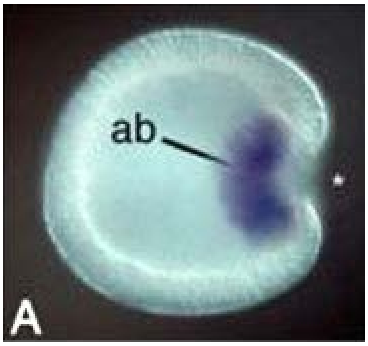
otxB:70



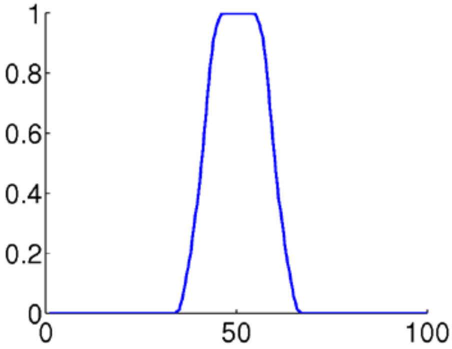
otxB at 70 hours
oral right [93]



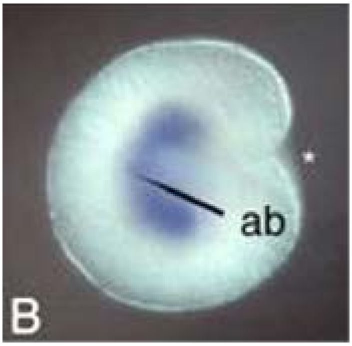
otxC:25



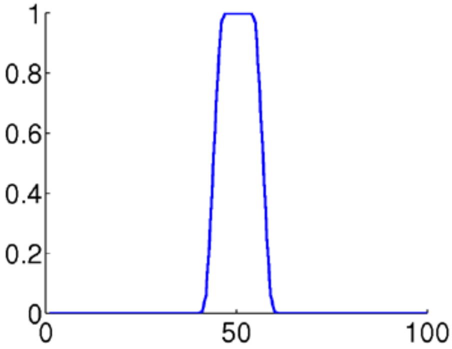
otxC at 25 hours
oral right [93]



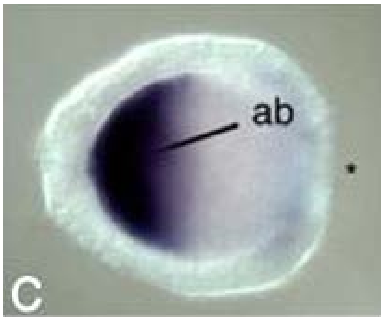
otxC:30



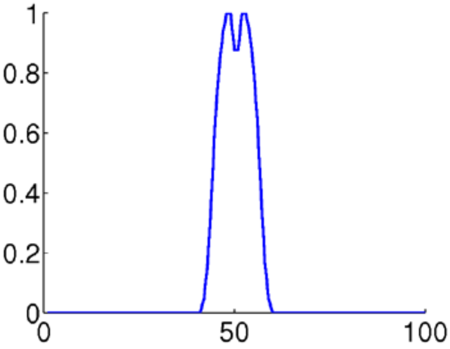
otxC at 30 hours
oral right [93]



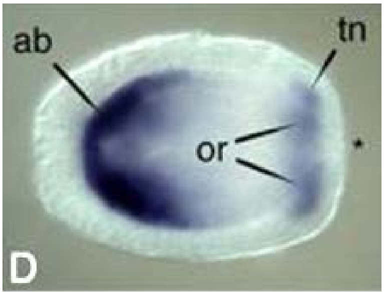
otxC:50



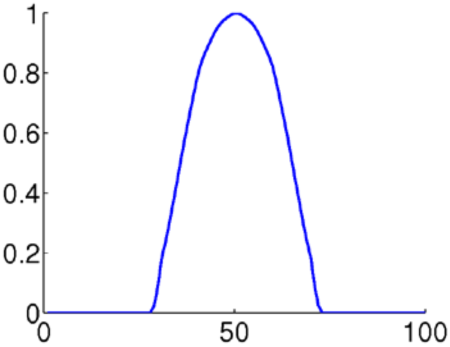
otxC at 50 hours
oral right [93]



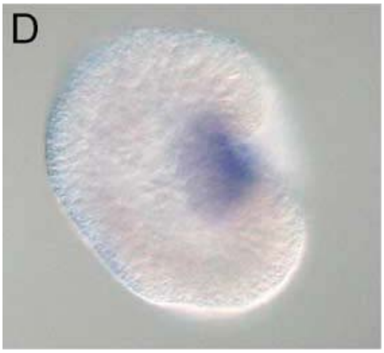
otxC:70



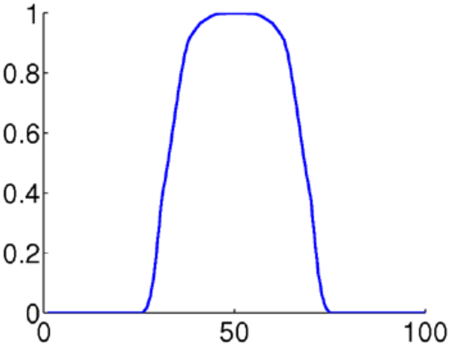
otxC at 70 hours
oral right [93]



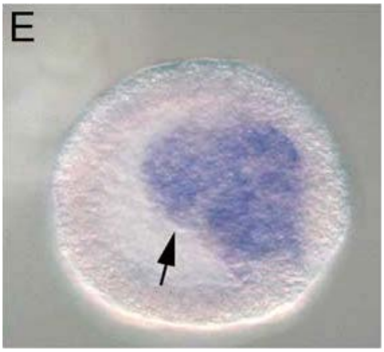
sna:eg



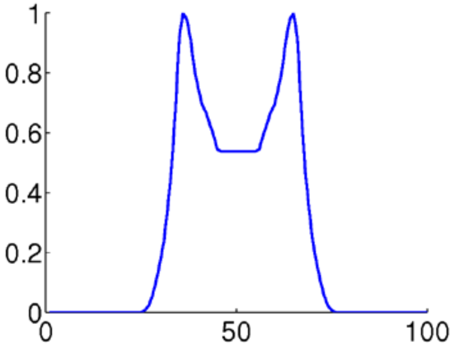
snail in early gastrula
oral right [81]



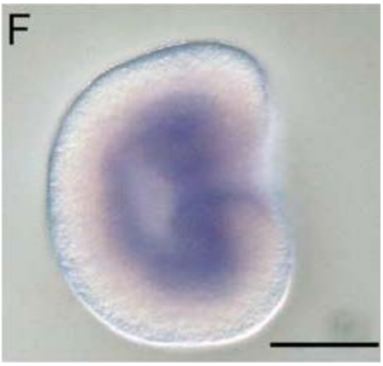
sna:mg



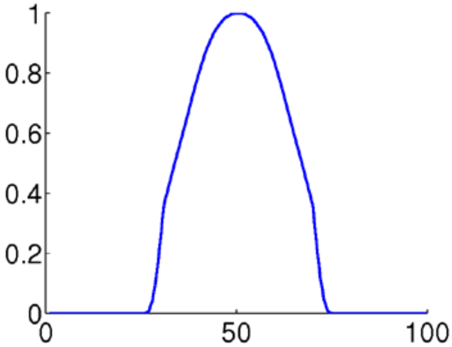
snail in mid-gastrula
oral right [81]



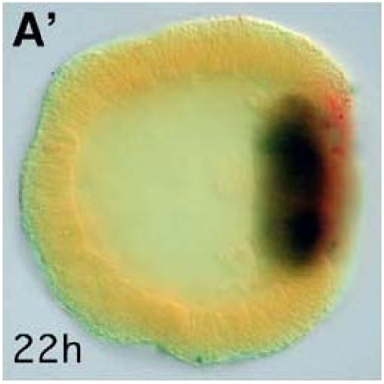
sna:lg



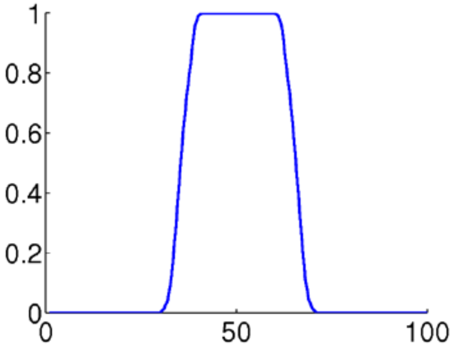
snail in late gastrula
oral right [81]



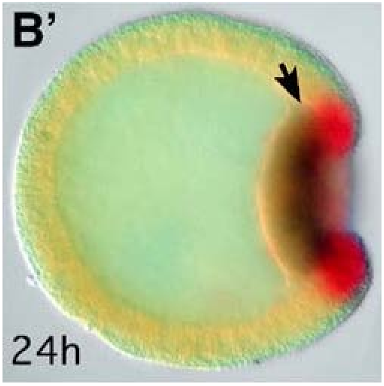
sna:22



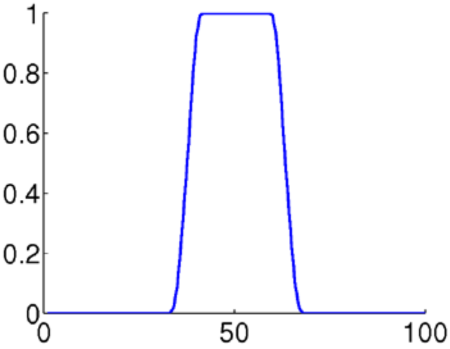
snail at 22 hours
black staining, oral right [24]



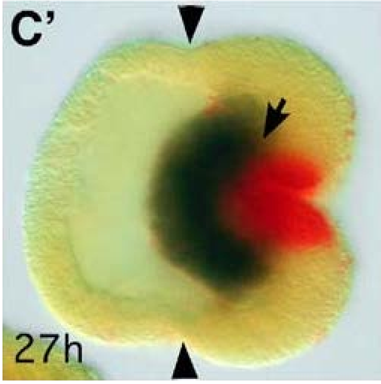
sna:24



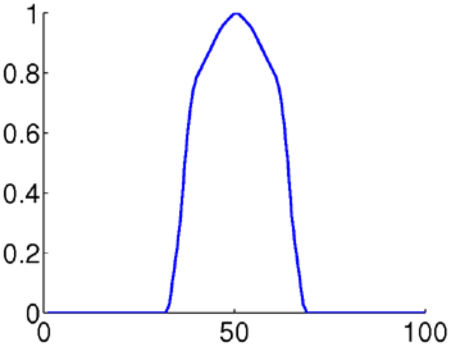
snail at 24 hours
black staining, oral right [24]



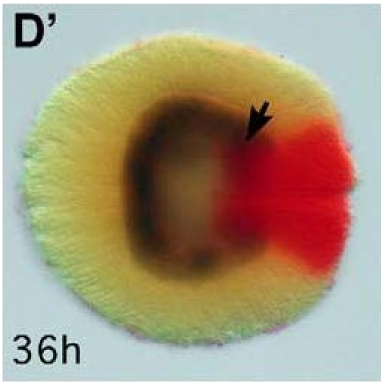
sna:27



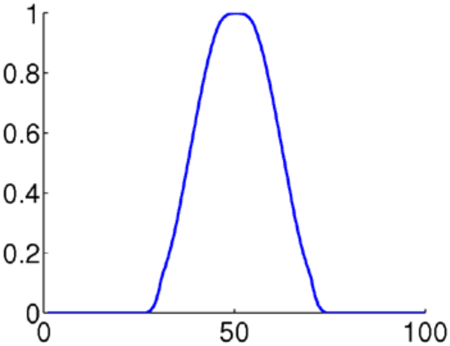
snail at 27 hours
black staining, oral right [24]



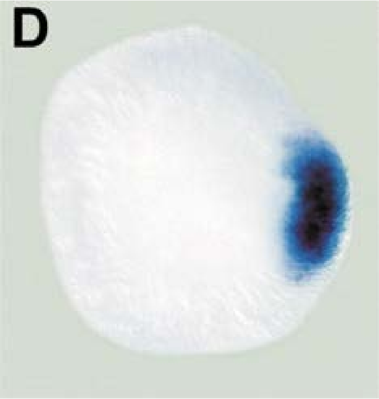
sna:36



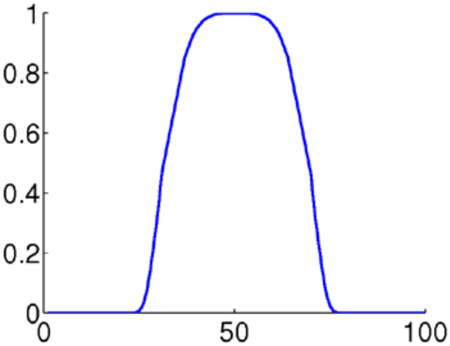
snail at 36 hours
black staining, oral right [24]



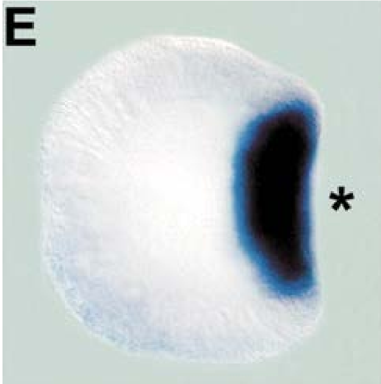
snail:bl



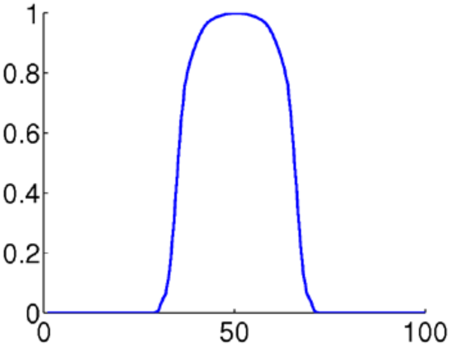
snail in blastula
oral right [27]



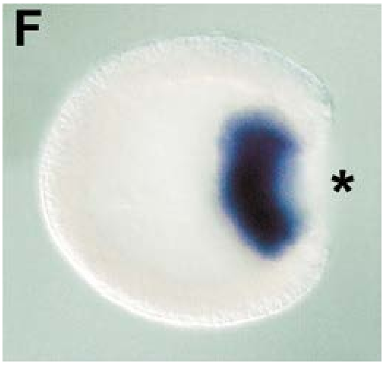
snail:23



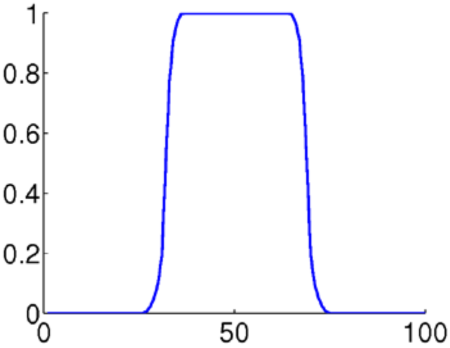
snail at 23 hours
oral right [27]



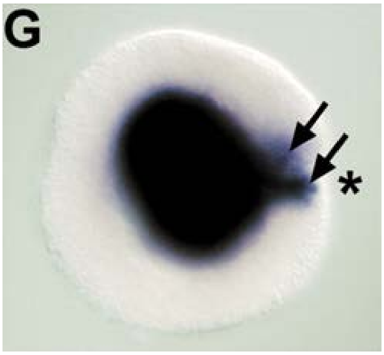
snail:24



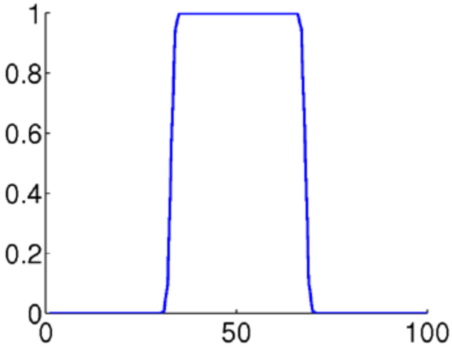
snail at 24 hours
oral right [27]



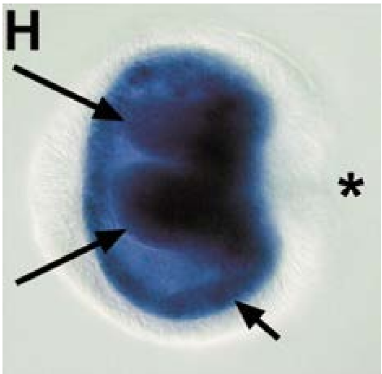
snail:30



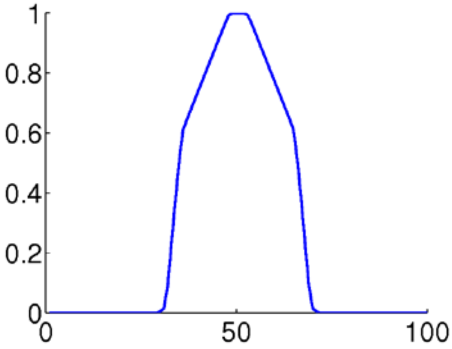
snail at 30 hours
oral right [27]



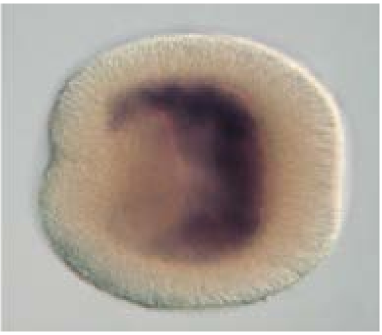
snail:45



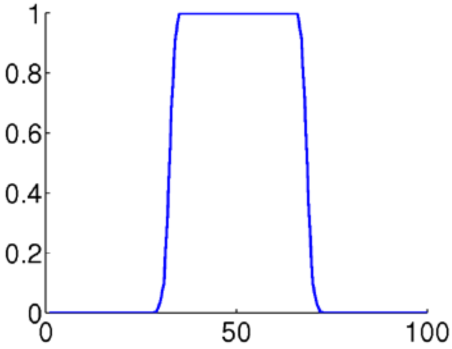
snail at 45 hours
oral right [27]



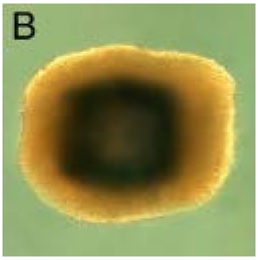
snail:48



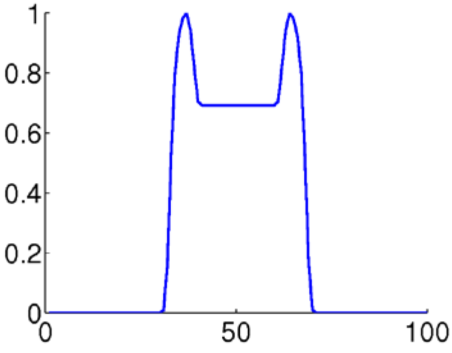
snail at 48 hours
oral left [47]



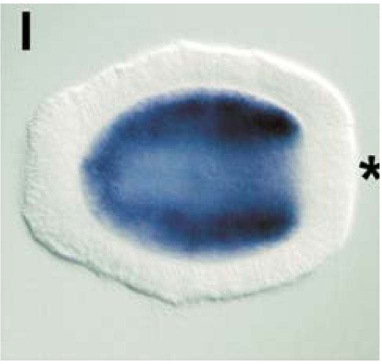
snail:50



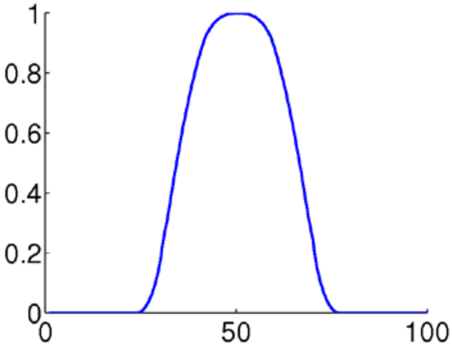
snail at 50 hours
oral left [47]



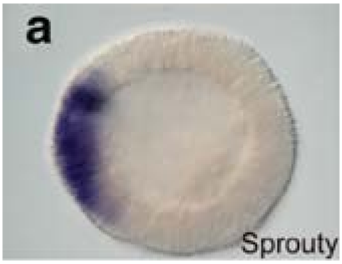
snail:60



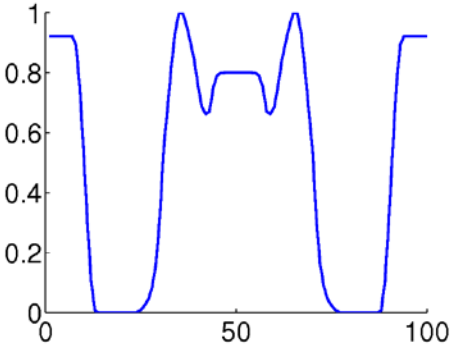
snail at 60 hours
oral right [27]



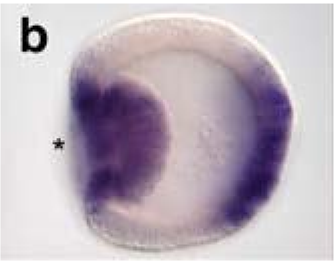
spr:bl



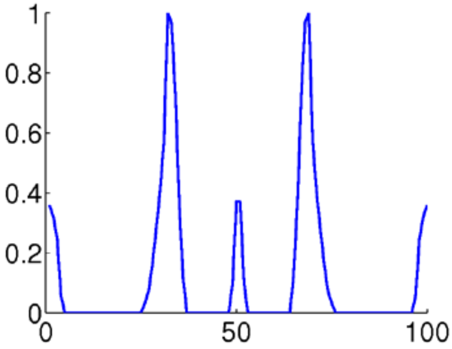
sprouty in blastula
oral left [94]



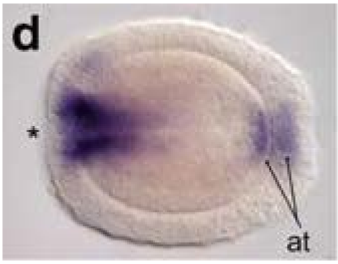
spr:25



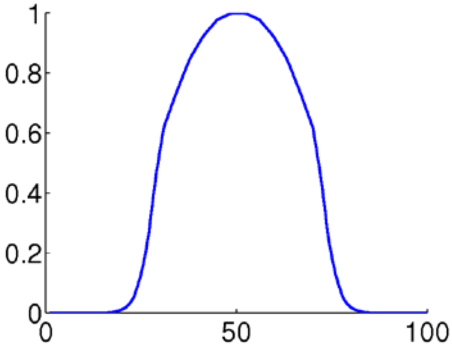
sprouty at 25 hours
oral left [94]



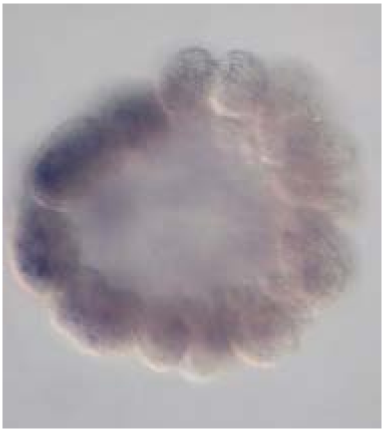
spr:50



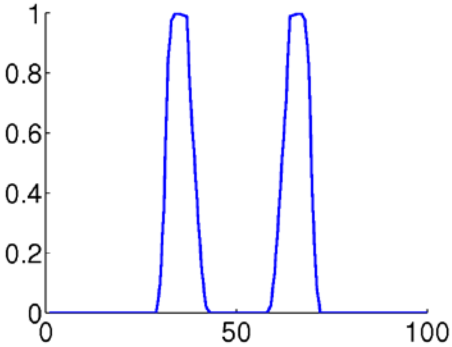
sprouty at 50 hours
oral left [94]



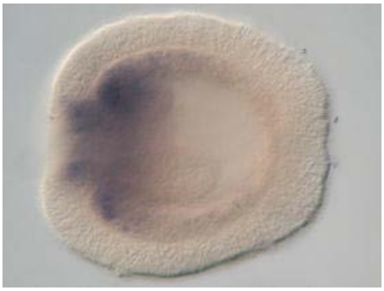
tcf:bl



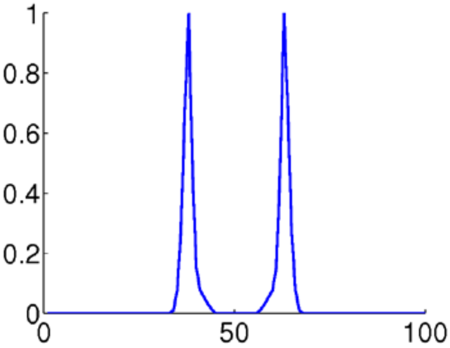
tcf in blastula
oral left [50,71]



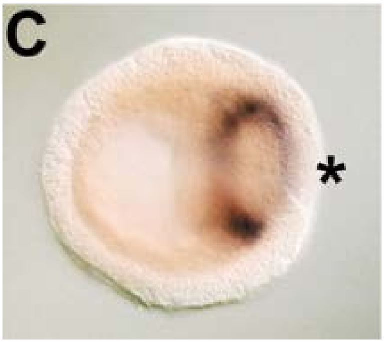
tcf:40



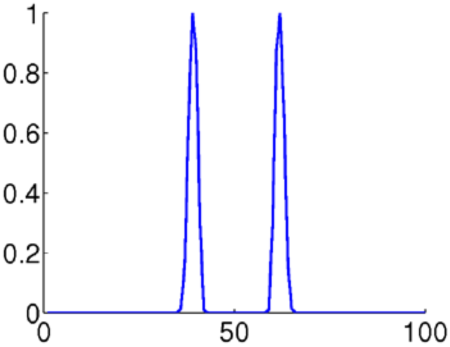
tcf at 40 hours
oral left [50,71]



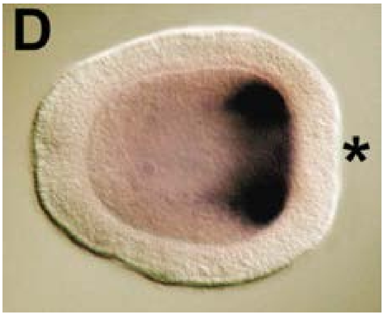
twi:38



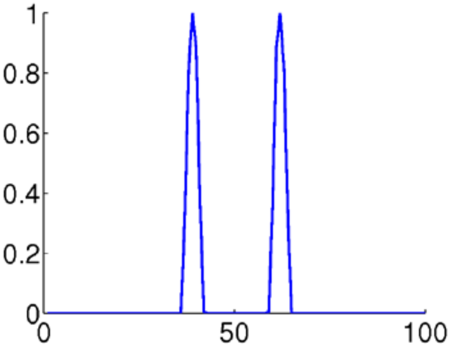
twist at 38 hours
oral right [27]



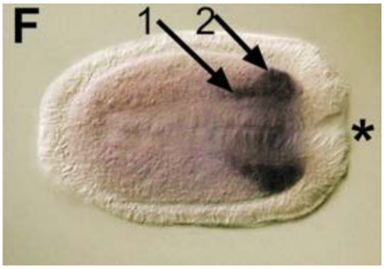
twi:54



twist at 54 hours
oral right [27]



twi:70



twist at 70 hours
oral right [27]

Appendix C: Standardized expression profiles of *in situ* hybridizations in the Kahi Kai database

From the blastula to the planula stages of development, the *Nematostella vectensis* embryo is a continuous sheet of cells with cylindrical symmetry. Strictly, the cylindrical symmetry is broken as soon as the primary mesenteries start to form. However, the growing mesenteries are not immediately visible in the *in situs* and the stage designations are sometimes ambiguous. Therefore, all *N. vectensis* gene expression images from the blastula to the late planula stage in the Kahi Kai database have been evaluated for their suitability of two-dimensional quantification.
Eventually, 252 *in situs* have been processed into standardized gene expression profiles for the analysis in Chapter 4. These profiles are included as numerical Matlab arrays in the corresponding publication entitled “Statistical analysis of a spatio-temporal gene expression database”.

The profiles are provided both as separate numerical arrays in labeled .mat files and as a two-column array of cells in the file “allarrs.mat”. For the separate arrays, the labels are the filenames; for the array of cells, the labels are the character arrays in the first column. The labels include the gene name, the developmental stage and, if applicable, the sequence number. For example, the file “admp-relatedbla1.mat” contains the variable “profile”, which is a 1x100 numerical array from the first image during the blastula of admp-related. This numerical array is also located in the second column of the 252x2 cell array called “expressiondata” in the file “allarrs.mat”, behind character array “admp-relatedbla1” in the first column. (Developmental stage abbreviations: cle = cleavage, bla =

blastula, ega = early gastrula, mga = mid gastrula, lga = late gastrula, epl = early planula, pla = planula, lpl = late planula.)

Appendix D: Spatio-temporal gene expression overview

In the Kahi Kai gene expression database, twelve developmental stages are defined for *Nematostella vectensis*. Seven stages have been used for the analysis in Chapter 4 and the table below is an overview of available expression domains in this database for 73 genes. This table is included as a Microsoft Access database sheet in the corresponding publication entitled “Statistical analysis of a spatio-temporal gene expression database”.

Legend: ? = entry without image, ab = aboral, at = apical tuft, bi = biradial, bl = blastoporal, bt = base of tentacle, bw = body wall, cd = central domain, cr = central ring, ec = ectoderm, en = endoderm, er = external ring, fu = full, ic = individual cells, me = mesentery, no = no expression observed, oc = possibly octoradial, or = oral, os = one side, ph = pharyngeal, pr = presumptive, ri = ring, ss = syphonoglyph side, te = tentacle, vh = vegetal hemisphere, vp = vegetal pole. Expression in the body wall is specified in more detail.

gene	blastula	early gastrula	mid gastrula	late gastrula	early planula	planula	late planula
Admp-related	cd	pr en	N/A	N/A	N/A	N/A	N/A
AnthoRFamide	N/A	N/A	ec (ic)	N/A	N/A	N/A	ec (or + te)
Anthox1	N/A	N/A	vp	vp	ec (at)	ec (at)	ec (at)
Anthox6	N/A	N/A	no	N/A	N/A	en (ph)	en (ph)
Anthox7	N/A	N/A	N/A	no	N/A	en (bw, os)	en (bw, os)
Anthox8b	N/A	N/A	N/A	no	en (bw, os)	en (bw, os)	N/A
bicaudalC-like1	cr	bl ec	N/A	ec (or + ic)	N/A	N/A	N/A
Bmp2/4	cr	bl ec	N/A	en, ss + ec, ss	en, ss	en, ss	N/A
Brachyury	cr	bl ec	N/A	ec (or)	N/A	N/A	N/A
chordin	er	bl ec	bl ec, os	N/A	N/A	N/A	N/A
Churchill	N/A	N/A	N/A	no	N/A	en (bw, or)	en (ph)
dkk1/2/4	no	no	N/A	ec (ab)	ec (ab + at)	N/A	N/A

gene	blastula	early gastrula	mid gastrula	late gastrula	early planula	planula	late planula
dopa beta-monooxygenase	N/A	no	N/A	N/A	N/A	ec (or)	N/A
FGF1A	N/A	vp	N/A	vp	N/A	ec (at)	N/A
FGF8A	cd	pr en	bl ec	ec (ph)	ec (ph + or + at)	N/A	N/A
FGF8B	N/A	N/A	no	N/A	N/A	N/A	en (at)
FGFRa	N/A	vh	vp	N/A	N/A	ec (at)	N/A
FoxA	cr	N/A	N/A	ec (ph + or)	ec (ph + or)?	N/A	N/A
FoxB	cr	bl ec	ec (or)	ec (ph + or)	ec (ph + or)	ec (or)	ec (or)
FoxD.1	N/A	vh	N/A	vh	ec (at)	ec (at)	ec (bt + at)
frizzled 10	cd	pr en	N/A	N/A	N/A	N/A	N/A
Gbx	N/A	N/A	no	N/A	en (bw, bi)	en (bw, bi)	en (bw, bi)
Gli	cd	pr en	N/A	en (bw, fu)	N/A	en (bw, fu)	N/A
gsc	cd	pr en	N/A	N/A	N/A	N/A	N/A

gene	blastula	early gastrula	mid gastrula	late gastrula	early planula	planula	late planula
Hedgehog1	no	N/A	ec (ph)	ec (ph)	N/A	ec (me + ph)	N/A
Hes2	N/A	N/A	ec (ic)	N/A	ec (or)	N/A	ec (or)
Hex	N/A	N/A	ec (ph + or, os)	ec (ph)	N/A	ec (ph)	N/A
Hint1	N/A	N/A	N/A	N/A	ec (ic)	ec (ic)	N/A
Hint3	N/A	N/A	N/A	N/A	N/A	ec (bw, ic)	ec (bw, ic)
Hlxb9	cd	no	no	N/A	en (ph)	en (ph)	en (me + ph)
MoxA	N/A	N/A	no	N/A	en (ph)	en (ph)	N/A
Msx	N/A	N/A	no	N/A	en (ph, bi) + ec (ph, bi)	en (ph, oc) + ec (ph, oc)	N/A
Msx2	N/A	no	N/A	N/A	N/A	en (ph, os)	en (ph, os)
Nanos 2	cd	pr en	N/A	ec (ph)	N/A	N/A	N/A
nfix-like	cd	pr en	N/A	N/A	N/A	N/A	N/A
nkd1-like	cd + cr	pr en	ec (ph)	ec (ph)	N/A	N/A	N/A

gene	**blastula**	**early gastrula**	**mid gastrula**	**late gastrula**	**early planula**	**planula**	**late planula**
nk-like 13	cd	pr en	N/A	N/A	N/A	N/A	N/A
NvHD021	N/A	N/A	N/A	N/A	ec (or)	ec (or)	ec (or)
NvHD060	N/A	no	N/A	N/A	en (bw, os)	en (bw, os)	en (bw, os)
Otp	N/A	N/A	no	N/A	N/A	ec (or)	N/A
OtxA	cd	pr en	pr en	N/A	en (at)	N/A	en (at+or+ph) + ec (or+at)
OtxB	cd	pr en	pr en	en (bw, fu)	N/A	en (at + te) + ec (te)	en (at + te) + ec (te)
OtxC	cd	pr en	pr en	en (bw, fu)	N/A	en (at)	en (at + or)
Patched	N/A	N/A	no	en (bw, or + ph)	N/A	en (bw, fu + ph)	N/A
phtf1-like	cd	pr en	N/A	N/A	N/A	N/A	N/A
Pl10	cd	pr en	pr en	en	N/A	N/A	N/A
porcupine-like	cd	pr en	N/A	N/A	N/A	N/A	N/A
Rough	N/A	no	N/A	ec (bw, ic)	ec (bw, ic)	ec (bw, ic)	en (bw, ic) + ec (bw, ic)

gene	blastula	early gastrula	mid gastrula	late gastrula	early planula	planula	late planula
Rx	N/A	ec (bw, ri)	ec (bw, ri)	N/A	ec (bw, ri)	N/A	ec (bw, ic)
Rx1	no	N/A	ec (bw, ic)	N/A	N/A	ec (bw, ic)	N/A
Smad1/5	cd	pr en	N/A	en, ss	N/A	en (bw, fu + me)	N/A
smad4-like	cd	pr en	N/A	N/A	N/A	N/A	N/A
Snail A	cd	pr en	N/A	N/A	N/A	N/A	N/A
Sox1	cd + cr	ec (or)	ec (or)	N/A	ec (ph)	ec (ph)	ec (ph)
Sox2	N/A	ec (ic)	N/A	N/A	N/A	ec (te)	N/A
Sox3	N/A	ec (bw, ri + or)	N/A	N/A	N/A	ec (bt)	N/A
SoxB1	N/A	vp	ec (ph + at)	N/A	ec (ph + at)	ec (ph + at)	N/A
SoxB2	N/A	ec (ic)	ec (ic)	N/A	N/A	N/A	en (ic) + ec (ic)
SoxE.1	N/A	N/A	ec (or)	ec (or)	N/A	en (bw, fu)	N/A
SoxF.1	N/A	no	N/A	N/A	N/A	en (bw, fu + ph)	N/A

gene	blastula	early gastrula	mid gastrula	late gastrula	early planula	planula	late planula
Sprouty	cd	pr en + ec (at)	N/A	N/A	N/A	en (at) + ec (ph + or + at)	en (at) + ec (or + at)
tbx20-like	cd	pr en	N/A	N/A	N/A	N/A	N/A
Tcf	cr	N/A	N/A	N/A	en (me + ph)	en (me + ph)	en (me + ph)
tolloid	cd	pr en	N/A	N/A	N/A	N/A	N/A
unc4-like	cd	N/A	N/A	N/A	ec (or)	N/A	N/A
Vasa 1	cd	N/A	pr en	N/A	N/A	N/A	N/A
Vasa 2	cd	pr en	N/A	en	N/A	N/A	N/A
vasa-like	cd	pr en	N/A	en	N/A	N/A	N/A
Vent1	no	ec (or, os)	ec (or, os)	ec (ph + or, os)	ec (or, os)	no	N/A
Wnt16	N/A	N/A	no	N/A	N/A	ec (ph + or)	N/A
Wnt3	cr	bl ec	bl ec	ec (ph + or)	ec (or)	ec (or)	N/A
wnt4	er	bl ec	N/A	N/A	ec (or)	N/A	N/A

gene	**blastula**	**early gastrula**	**mid gastrula**	**late gastrula**	**early planula**	**planula**	**late planula**
wntA	cr	bl ec	N/A	ec (or)	N/A	N/A	N/A

Appendix E: Subsequent gene expression domains from recorded stages of development

For the statistical analysis in Chapter 4, the occurrences of subsequent expression patterns have been counted. The database contains gaps in expression data for almost every gene, so the progression during development can not be compared directly. Still, the pattern of a missing developmental stage in between two available stages is expected to look like the earlier or the later stage. In order to find significant trends in spatial gene expression progression, all genes with *in situs* stored for at least two stages from the blastula to the late planula are taken into account.
The following table is included as a Microsoft Excel spreadsheet in the corresponding publication entitled “Statistical analysis of a spatio-temporal gene expression database”. Pairs of subsequently available expression domains (third and fourth columns) are ordered by their initial and final stages of development (first and second columns, respectively). The percentages (sixth column) are the relative frequencies (from the gene counts in the fifth column) for each pair of expression domains. A change from cylindrical expression pattern symmetry to biradial or bilateral symmetry is indicated in the seventh column with “symmetry break”. Changes of expression over major regions are indicated in the eighth column (appeared = expression is observed only in the second stage, major = expression moved to another major region, minor = expression in the second stage is limited to the same major region(s) as in the first stage, none = no expression is observed in both stages, vanished = expression is observed only in the first stage). Most expression domain designations belong to a single major region (color coded in Figure 4.20), so the change between two stages is usually derived

from the third and fourth columns. If the major region is unclear from designations such as “individual cells”, the original images in the Kahi Kai database have been consulted.

Legend: ? = entry without image, ab = aboral, at = apical tuft, bi = biradial, bl = blastoporal, bt = base of tentacle, bw = body wall, cd = central domain, cr = central ring, ec = ectoderm, en = endoderm, er = external ring, fu = full, ic = individual cells, me = mesentery, no = no expression observed, oc = possibly octoradial, or = oral, os = one side, ph = pharyngeal, pr = presumptive, ri = ring, ss = syphonoglyph side, te = tentacle, vh = vegetal hemisphere, vp = vegetal pole.

interval		**expression**		**count**		
initial stage	*final stage*	*initial domain*	*final domain*			
blastula	early gastrula	cd	pr en	21		minor
		cr	bl ec	6		minor
		er	bl ec	2		minor
		no	no	1		none
		cd	no	1		vanished
		cd + cr	pr en	1		major
		cd + cr	ec (or)	1		major
		cd	pr en + ec (at)	1		major
		no	ec (or, os)	1	symmetry break	appeared
blastula	mid gastrula	no	ec (ph)	1		appeared
		cd	pr en	1		minor
		no	ec (bw, ic)	1		appeared
blastula	late gastrula	cr	ec (ph + or)	1		minor
blastula	early planula	cr	en (me + ph)	1		major
		cd	ec (or)	1		major
early gastrula	mid gastrula	bl ec	bl ec, os	1	symmetry break	minor
		pr en	bl ec	1		major
		vh	vp	1		minor
		bl ec	ec (or)	1		minor
		no	no	1		none
		pr en	ec (ph)	1		major
		pr en	pr en	4		minor
		ec (bw, ri)	ec (bw, ri)	1		minor
		ec (or)	ec (or)	1		minor
		vp	ec (ph + at)	1		major
		ec (ic)	ec (ic)	1		minor
		ec (or, os)	ec (or, os)	1		minor

		bl ec	bl ec	1		minor
early gastrula	late gastrula	bl ec	ec (or + ic)	1		major
		bl ec	en, ss + ec, ss	1	symmetry break	major
		bl ec	ec (or)	2		minor
		no	ec (ab)	1		appeared
		vp	vp	1		minor
		vh	vh	1		minor
		pr en	en (bw, fu)	1		minor
		pr en	ec (ph)	1		major
		no	ec (bw, ic)	1		appeared
		pr en	en, ss	1	symmetry break	minor
		pr en	en	2		minor
early gastrula	early planula	no	en (bw, os)	1	symmetry break	appeared
		bl ec	ec (or)	1		minor
early gastrula	planula	no	ec (or)	1		appeared
		no	en (ph, os)	1	symmetry break	appeared
		ec (ic)	ec (te)	1		major
		ec (bw, ri + or)	ec (bt)	1		major
		no	en (bw, fu + ph)	1		appeared
		pr en + ec (at)	en (at) +ec (ph + or + at)			
mid gastrula	late gastrula	vp	vp	1		minor
		bl ec	ec (ph)	1		minor
		ec (or)	ec (ph + or)	1		minor
		ec (ph)	ec (ph)	2		minor
		ec (ph + or, os)	ec (ph)	1		minor
		pr en	en (bw, fu)	2		minor

		no	en (bw, or + ph)	1		appeared
		pr en	en	1		minor
		ec (or)	ec (or)	1		minor
		ec (or, os)	ec (ph + or, os)	1		minor
		bl ec	ec (ph + or)	1		minor
mid gastrula	early planula	no	en (bw, bi)	1	symmetry break	appeared
		ec (ic)	ec (or)	1		major
		no	en (ph)	2		appeared
		no	en (ph, bi) + ec (ph, bi)	1	symmetry break	appeared
		pr en	en (at)	1		minor
		ec (bw, ri)	ec (bw, ri)	1		minor
		ec (or)	ec (ph)	1		minor
		ec (ph + at)	ec (ph + at)	1		minor
mid gastrula	planula	no	en (ph)	1		appeared
		vp	ec (at)	1		minor
		no	ec (or)	1		appeared
		ec (bw, ic)	ec (bw, ic)	1		minor
		no	ec (ph + or)	1		appeared
mid gastrula	late planula	ec (ic)	ec (or + te)	1		major
		no	en (at)	1		appeared
		ec (ic)	en (ic) + ec (ic)	1		major
late gastrula	early planula	vp	ec (at)	1		minor
		no	en (bw, os)	1	symmetry break	appeared
		en, ss + ec, ss	en, ss			major
		ec (ab)	ec (ab + at)	1		minor
		ec (ph)	ec (ph + or + at)	1		major
		ec (ph + or)	ec (ph + or)	2		minor
		vh	ec (at)	1		minor

		ec (bw, ic)	ec (bw, ic)	1		minor
		ec (ph + or, os)	ec (or, os)	1		minor
		ec (ph + or)	ec (or)	1		minor
late gastrula	planula	no	en (bw, os)	1	symmetry break	appeared
		no	en (bw, or)	1		appeared
		vp	ec (at)	1		minor
		en (bw, fu)	en (bw, fu)	1		minor
		ec (ph)	ec (me + ph)	1		minor
		ec (ph)	ec (ph)	1		minor
		en (bw, fu)	en (at + te) + ec (te)	1		major
		en (bw, fu)	en (at)	1		minor
		en (bw, or + ph)	en (bw, fu + ph)	1		minor
		en, ss	en (bw, fu + me)	1		minor
		ec (or)	en (bw, fu)	1		major
early planula	planula	ec (at)	ec (at)	2		minor
		en (bw, os)	en (bw, os)	2		minor
		en, ss	en, ss	1		minor
		ec (ph + or)	ec (or)	1		minor
		en (bw, bi)	en (bw, bi)	1		minor
		ec (ic)	ec (ic)	1		minor
		en (ph)	en (ph)	2		minor
		en (ph, bi) + ec (ph, bi)	en (ph, oc) + ec (ph, oc)	1		minor
		ec (or)	ec (or)	2		minor
		ec (bw, ic)	ec (bw, ic)	1		minor
		ec (ph)	ec (ph)	1		minor
		ec (ph + at)	ec (ph + at)	1		minor
		en (me + ph)	en (me + ph)	1		minor
		ec (or, os)	no	1		vanished

early planula	late planula	ec (or)	ec (or)	1		minor
		en (at)	en (at + or + ph) + ec (or + at)	1		major
		ec (bw, ri)	ec (bw, ic)	1		minor
planula	late planula	ec (at)	ec (at)	1		minor
		en (ph)	en (ph)	1		minor
		en (bw, os)	en (bw, os)	2		minor
		en (bw, or)	en (ph)	1		minor
		ec (or)	ec (or)	2		minor
		ec (at)	ec (bt + at)	1		major
		en (bw, bi)	en (bw, bi)	1		minor
		ec (bw, ic)	ec (bw, ic)	1		minor
		en (ph)	en (me + ph)	1		minor
		en (ph, os)	en (ph, os)	1		minor
		en (at + te) + ec (te)	en (at + te) + ec (te)	1		minor
		en (at)	en (at + or)	1		minor
		ec (bw, ic)	en (bw, ic) + ec (bw, ic)	1		major
		ec (ph)	ec (ph)	1		minor
		en (at) + ec (ph + or + at)	en (at) + ec (or + at)	1		minor
		en (me + ph)	en (me + ph)	1		minor

Bibliography

References

[1] Mall FP (1914), On stages in the development of human embryos from 2 to 25mm long, *Anat Anz* **46**(3-4):78-84.

[2] O'Rahilly R (1979), Early human development and the chief sources of information on staged human embryos, *Eur J Obstet Gynecol Reprod Biol* **9**(4):273-80.

[3] Theiler K (1989), *The House Mouse: Development and Normal Stages from Fertilization to 4 Weeks of Age*. 2nd ed. New York, Springer-Verlag.

[4] Hamburger V, Hamilton HL (1951), A series of normal stages in the development of the chick embryo, *J Morphol* **88**(1):49-92.

[5] Hamburger V, Hamilton HL (1992), A series of normal stages in the development of the chick embryo, *Dev Dyn* **195**(4):231-72.

[6] Nieuwkoop PD, Faber J (1967), *Normal Table of Xenopus laevis (Daudin)*. Amsterdam, North Holland Publishing Company.

[7] Campos-Ortega JA, Hartenstein V (1985), *The embryonic development of Drosophila melanogaster*. Berlin, Springer-Verlag.

[8] Gustafson T, Wolpert L (1967), Cellular movement and contact in sea urchin morphogenesis, *Biol Rev Camb Philos Soc* **42**(3):442-98.

[9] Hand C, Uhlinger KR (1992), The culture, sexual and asexual reproduction, and growth of the sea anemone *Nematostella vectensis*, *Biol Bull* **182**(2):169-76.

[10] Ashe HL, Briscoe J (2006), The interpretation of morphogen gradients, *Development* **133**(3):385-94.

[11] Jaeger J (2011), The gap gene network, *Cell Mol Life Sci* **68**(2):243-74.

[12] Nüsslein-Volhard C, Wieschaus E (1980), Mutations affecting segment number and polarity in *Drosophila, Nature* **287**(5785):795-801.

[13] Gaul U, Jäckle H (1989), Analysis of maternal effect mutant combinations elucidates regulation and function of the overlap of *hunchback* and *Krüppel* gene expression in the *Drosophila* blastoderm embryo, *Development* **107**(3):651-62.

[14] Wolpert L (1968), The French Flag problem: a contribution to the discussion on pattern development and regulation. In: Waddington CH, editor. *Towards a Theoretical Biology*, Vol. 1. Edinburgh, Edinburgh University Press. pp. 125-33.

[15] de Jong H (2002), Modeling and simulation of genetic regulatory systems: a literature review, *J Comput Biol* **9**(1):67-103.

[16] Jaeger J (2009), Modelling the *Drosophila* embryo, *Mol Biosyst* **5**(12):1549-68.

[17] Jaeger J, Surkova S, Blagov M, Janssens H, Kosman D, Kozlov KN, Manu, Myasnikova E, Vanario-Alonso CE, Samsonova M, Sharp DH, Reinitz J (2004), Dynamic control of positional information in the early *Drosophila* embryo, *Nature* **430**(6997):368-71.

[18] Reinitz J, Sharp DH (1995), Mechanism of *eve* stripe formation, *Mech Dev* **49**(1-2):133-58.

[19] Kirkpatrick S, Gelatt CD Jr, Vecchi MP (1983), Optimization by simulated annealing, *Science* **220**(4598):671-80.

[20] Eiben AE, Smith JE (2003), *Introduction to evolutionary computing*. Berlin, Springer.

[21] Fomekong-Nanfack Y, Kaandorp JA, Blom J (2007), Efficient parameter estimation for spatio-temporal models of pattern formation: case study of *Drosophila melanogaster, Bioinformatics* **23**(24):3356-63.

[22] Darling JA, Reitzel AR, Burton PM, Mazza ME, Ryan JF, Sullivan JC, Finnerty JR (2005), Rising starlet: the starlet sea anemone, *Nematostella vectensis, Bioessays* **27**(2):211-21.

[23] Stefanik DJ, Friedman LE, Finnerty JR (2013), Collecting, rearing, spawning and inducing regeneration of the starlet sea anemone, *Nematostella vectensis, Nat Protoc* **8**(5):916-23.

[24] Magie CR, Daly M, Martindale MQ (2007), Gastrulation in the cnidarian *Nematostella vectensis* occurs via invagination not ingression, *Dev Biol* **305**(2):483-97.

[25] Tamulonis C, Postma M, Marlow HQ, Magie CR, de Jong J, Kaandorp J (2011), A cell-based model of *Nematostella vectensis* gastrulation including bottle cell formation, invagination and zippering, *Dev Biol* **351**(1):217-28.

[26] Kumburegama S, Wijesena N, Xu R, Wikramanayake AH (2011), Strabismus-mediated primary archenteron invagination is uncoupled from Wnt/β-catenin-dependent endoderm cell fate specification in *Nematostella vectensis* (Anthozoa, Cnidaria): Implications for the evolution of gastrulation, *Evodevo* **2**(1):2.

[27] Martindale MQ, Pang K, Finnerty JR (2004), Investigating the origins of triploblasty: 'mesodermal' gene expression in a diploblastic animal, the sea anemone *Nematostella vectensis* (phylum, Cnidaria; class, Anthozoa), *Development* **131**(10):2463-74.

[28] Röttinger E, Dahlin P, Martindale MQ (2012), A framework for the establishment of a cnidarian gene regulatory network for "endomesoderm" specification: the inputs of β-catenin/TCF signaling, *PLoS Genet* **8**(12):e1003164.

[29] Janssens H, Kosman D, Vanario-Alonso CE, Jaeger J, Samsonova M, Reinitz J (2005), A high-throughput method for quantifying gene expression data from early *Drosophila* embryos, *Dev Genes Evol* **215**(7):374-81.

[30] Ryan JF, Mazza ME, Pang K, Matus DQ, Baxevanis AD, Martindale MQ, Finnerty JR (2007), Pre-bilaterian origins of the Hox cluster and the Hox code: evidence from the sea anemone, *Nematostella vectensis*, *PLoS One* **2**(1):e153.

[31] Lazarides E, Weber K (1974), Actin antibody: the specific visualization of actin filaments in non-muscle cells, *Proc Natl Acad Sci U S A* **71**(6):2268-72.

[32] Warn RM, Robert-Nicoud M (1990), F-actin organization during the cellularization of the *Drosophila* embryo as revealed with a confocal laser scanning microscope, *J Cell Sci* **96**(Pt 1):35-42.

[33] Wolenski FS, Layden MJ, Martindale MQ, Gilmore TD, Finnerty JR (2013), Characterizing the spatiotemporal expression of RNAs and proteins in the starlet sea anemone, *Nematostella vectensis*, *Nat Protoc* **8**(5):900-15.

[34] Shimizu T, Bae YK, Hibi M (2006), Cdx-Hox code controls competence for responding to Fgfs and retinoic acid in zebrafish neural tissue, *Development* **133**(23):4709-19.

[35] Gilbert SF (2000), *Developmental Biology*. 6th ed. Sunderland (MA), Sinauer Associates.

[36] Etienne L (2013), Early Development chapter 5. Quizlet Flashcards, quizlet.com/20488579/early-development-chapter-5-flash-cards.

[37] Johndrow JE, Magie CR, Parkhurst SM (2004), Rho GTPase function in flies: insights from a developmental and organismal perspective, *Biochem Cell Biol* **82**(6):643-57.

[38] Jaeger J, Martinez-Arias A (2009), Getting the measure of positional information, *PLoS Biol* **7**(3):e1000081.

[39] Kahi Kai comparative database for embryonic development of marine invertebrates, www.kahikai.org/index.php ?content=embryology_nvectensis

[40] Li E, Davidson EH (2009), Building developmental gene regulatory networks, *Birth Defects Res C Embryo Today* **87**(2):123-30.

[41] Chan TM, Longabaugh W, Bolouri H, Chen HL, Tseng WF, Chao CH, Jang TH, Lin YI, Hung SC, Wang HD, Yuh CH (2009), Developmental gene regulatory networks in the zebrafish embryo, *Biochim Biophys Acta* **1789**(4):279-98.

[42] Mjolsness E, Sharp DH, Reinitz J (1991), A connectionist model of development, *J Theor Biol* **152**(4):429-53.

[43] Ashyraliyev M, Fomekong-Nanfack Y, Kaandorp JA, Blom JG (2009), Systems biology: parameter estimation for biochemical models, *FEBS J* **276**(4):886-902.

[44] Ashyraliyev M, Jaeger J, Blom JG (2008), Parameter estimation and determinability analysis applied to *Drosophila* gap gene circuits, *BMC Syst Biol* **2**:83.

[45] Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, Terry A, Shapiro H, Lindquist E, Kapitonov VV, Jurka J, Genikhovich G, Grigoriev IV, Lucas SM, Steele RE, Finnerty JR, Technau U, Martindale MQ, Rokhsar DS (2007), Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization, *Science* **317**(5834):86-94.

[46] Byrum CA, Martindale MQ (2004), Gastrulation in the Cnidaria and the Ctenophora. In: Stern CA, editor. *Gastrulation: From Cells to Embryo*. New York, Cold Spring Harbor Laboratory Press. p. 33-50.

[47] Lee PN, Kumburegama S, Marlow HQ, Martindale MQ, Wikramanayake AH (2007), Asymmetric developmental potential along the animal-vegetal axis in the antozoan cnidarian, *Nematostella vectensis*, is mediated by Dishevelled, *Dev Biol* **310**(1):169-86.

[48] de Jong J (2009), *Quantitative analysis of gene expression in Nematostella vectensis*. MSc thesis. Section Computational Science, University of Amsterdam.

[49] Ryan JF, Finnerty JR (2003), CnidBase: The Cnidarian Evolutionary Genomics Database, *Nucleic Acids Res* **31**(1):159-63.

[50] Kahi Kai comparative gene expression database for marine invertebrates, www.kahikai.org/index.php?content=genes.

[51] Extavour CG, Pang K, Matus DQ, Martindale MQ (2005), *vasa* and *nanos* expression patterns in a sea anemone and the evolution of bilaterian germ cell specification mechanisms, *Evol Dev* **7**(3):201-15.

[52] MathWorks documentation, www.mathworks.com/help/matlab/ref/interp1.html.

[53] MathWorks documentation, www.mathworks.com/help/curvefit/smooth.html.

[54] Magie CR, Pang K, Martindale MQ (2005), Genomic inventory and expression of *Sox* and *Fox* genes in the cnidarian *Nematostella vectensis*, *Dev Genes Evol* **215**(12):618-30.

[55] Fomekong-Nanfack Y, Postma M, Kaandorp JA (2009), Inferring *Drosophila* gap gene regulatory network: a parameter sensitivity and perturbation analysis, *BMC Syst Biol* **3**:94.

[56] Fomekong-Nanfack Y, Postma M, Kaandorp JA (2009), Inferring *Drosophila* gap gene regulatory network: pattern analysis of simulated gene expression profiles and stability analysis, *BMC Res Notes* **2**:256.

[57] Wikramanayake AH, Hong M, Lee PN, Pang K, Byrum CA, Bince JM, Xu R, Martindale MQ (2003), An ancient role for nuclear beta-catenin in the evolution of axial polarity and germ layer segregation, *Nature* **426**(6965):446-50.

[58] Welten MCM, de Haan SB, van den Boogert N, Noordermeer JN, Lamers GEM, Spaink HP, Meijer AH, Verbeek FJ (2006), ZebraFISH: fluorescent *in situ* hybridization protocol and three-dimensional imaging of gene expression patterns, *Zebrafish* **3**(4):465-76.

[59] Luengo Hendriks CL, Keränen SV, Fowlkes CC, Simirenko L, Weber GH, DePace AH, Henriquez C, Kaszuba DW, Hamann B, Eisen MB, Malik J, Sudar D, Biggin MD, Knowles DW (2006), Three-dimensional morphology and gene expression in the *Drosophila* blastoderm at cellular resolution I: data acquisition pipeline, *Genome Biol* **7**(12):R123.

[60] Flynn CJ, Sharma T, Ruffins SW, Guerra SL, Crowley JC, Ettensohn CA (2011), High-resolution, three-dimensional mapping of gene expression using GeneExpressMap (GEM), *Dev Biol* **357**(2):532-40.

[61] Myasnikova E, Samsonova A, Kozlov K, Samsonona M, Reinitz J (2001), Registration of the expression patterns of *Drosophila* segmentation genes by two independent methods, *Bioinformatics* **17**(1):3-12.

[62] Csako G (2006), Present and future of rapid and/or high-throughput methods for nucleic acid testing, *Clin Chim Acta* **363**(1-2):6-31.

[63] Fritzenwanker JH, Genikhovich G, Kraus Y, Technau U (2007), Early development and axis specification in the sea anemone *Nematostella vectensis, Dev Biol* **310**(2):264-79.

[64] Wolpert L (1969), Positional information and the spatial pattern of cellular differentiation, *J Theor Biol* **25**(1):1-47.

[65] Reeves GT, Muratov CB, Schüpbach T, Shvartsman SY (2006), Quantitative models of developmental pattern formation, *Dev Cell* **11**(3):289-300.

[66] Manu, Surkova S, Spirov AV, Gursky VV, Janssens H, Kim AR, Radulescu O, Vanario-Alonso CE, Sharp DH, Samsonova M, Reinitz J (2009), Canalization of gene expression and domain shifts in the *Drosophila* blastoderm by dynamical attractors, *PLoS Comput Biol* **5**(3):e1000303.

[67] Jaeger J, Manu, Reinitz J (2012), *Drosophila* blastoderm patterning, *Curr Opin Genet Dev* **22**(6):533-41.

[68] Manu, Surkova S, Spirov AV, Gursky VV, Janssens H, Kim AR, Radulescu O, Vanario-Alonso CE, Sharp DH, Samsonova M, Reinitz J (2009), Canalization of gene expression in the *Drosophila* blastoderm by gap gene cross regulation, *PLoS Biol* **7**(3):e1000049.

[69] Li XY, MacArthur S, Bourgon R, Nix D, Pollard DA, Iyer VN, Hechmer A, Simirenko L, Stapleton M, Luengo Hendriks CL, Chu HC, Ogawa N, Inwood W, Sementchenko V, Beaton A, Weiszmann R, Celniker SE, Knowles DW, Gingeras T, Speed TP, Eisen MB, Biggin MD (2008), Transcription factors bind thousands of active and inactive regions in the *Drosophila* blastoderm, *PLoS Biol* **6**(2):e27.

[70] Crombach A, Cicin-Sain D, Wotton KR, Jaeger J (2012), Medium-throughput processing of whole mount *in situ* hybridisation experiments into gene expression domains, *PLoS One* **7**(9):e46658.

[71] Ormestad M, Martindale MQ, Röttinger E (2011), A comparative gene expression database for invertebrates, *Evodevo* **2**:17.

[72] Botman D, Kaandorp JA (2012), Spatial gene expression quantification: a tool for analysis of *in situ* hybridizations in sea anemone *Nematostella vectensis*, *BMC Res Notes* **5**:555.

[73] Martindale MQ, Hejnol A (2009), A developmental perspective: changes in the position of the blastopore during bilaterian evolution, *Dev Cell* **17**(2):162-74.

[74] Peter IS, Davidson EH (2011), A gene regulatory network controlling the embryonic specification of endoderm, *Nature* **474**(7353):635-9.

[75] Song L, Langfelder P, Horvath S (2012), Comparison of co-expression measures: mutual information, correlation, and model based indices, *BMC Bioinformatics* **13**:328.

[76] Egea JA, Balsa-Canto E, Garcia MSG, Banga JR (2009), Dynamic optimization of nonlinear processes with an enhanced scatter search method, *Ind Eng Chem Res* **48**(9):4388-401.

[77] Egea JA, Martí R, Banga JR (2010), An evolutionary method for complex-process optimization, *Comput Oper Res* **37**(2):315-24.

[78] Maly T, Petzold LR (1996), Numerical methods and software for sensitivity analysis of differential-algebraic systems, *Appl Numer Math* **20**(1):57-79.

[79] Crombach A, Wotton KR, Cicin-Sain D, Ashyraliyev M, Jaeger J (2012), Efficient reverse-engineering of a developmental gene regulatory network, *PLoS Comput Biol* **8**(7):e1002589.

[80] Pennisi E (2005), Why do humans have so few genes? *Science* **309**(5731):80.

[81] Fritzenwanker JH, Saina M, Technau U (2004), Analysis of *forkhead* and *snail* expression reveals epithelial-mesenchymal transitions during embryonic and larval development of *Nematostella vectensis*, *Dev Biol* **275**(2):389-402.

[82] Sheth R, Marcon L, Bastida MF, Junco M, Quintana L, Dahn R, Kmita M, Sharpe J, Ros MA (2012), *Hox* genes regulate digit patterning by controlling the wavelength of a Turing-type mechanism, *Science* **338**(6113):1476-80.

[83] Botman D, Röttinger E, Martindale MQ, de Jong J, Kaandorp JA (2014), A computational approach towards a gene regulatory network for the developing *Nematostella vectensis* gut, *PLoS One* **9**(7):e103341.

[84] Finnerty JR, Pang K, Burton P, Paulson D, Martindale MQ (2004), Origins of bilateral symmetry: *Hox* and *dpp* expression in a sea anemone, *Science* **304**(5675):1335-7.

[85] Rentzsch F, Anton R, Saina M, Hammerschmidt M, Holstein TW, Technau U (2006), Asymmetric expression of the BMP antagonists *chordin* and *gremlin* in the sea anemone *Nematostella vectensis*: implications for the evolution of axial patterning, *Dev Biol* **296**(2):375-87.

[86] Dittmar WJ, McIver L, Michalak P, Garner HR, Valdez G (2014), EvoCor: a platform for predicting functionally related genes using phylogenetic and expression profiles, *Nucleic Acids Res* **42**(W1):W72-5.

[87] Li E, Cui M, Peter IS, Davidson EH (2014), Encoding regulatory state boundaries in the pregastrular oral ectoderm of the sea urchin embryo, *Proc Natl Acad Sci U S A* **111**(10):E906-13.

[88] Terada K, Kitayama A, Kanamoto T, Ueno N, Furukawa T (2006), Nucleosome regulator Xhmgb3 is required for cell proliferation of the eye and brain as a downstream target of *Xenopus rax/Rx1*, *Dev Biol* **291**(2):398-412.

[89] Manuel M (2009), Early evolution of symmetry and polarity in metazoan body plans, *C R Biol* **332**(2-3):184-209.

[90] Kozmikova I, Candiani S, Fabian P, Gurska D, Kozmik Z (2013), Essential role of Bmp signaling and its positive feedback loop in the early cell fate evolution of chordates, *Dev Biol* **382**(2):538-54.

[91] Lee PN, Pang K, Matus DQ, Martindale MQ (2006), A WNT of things to come: evolution of Wnt signaling and polarity in cnidarians, *Semin Cell Dev Biol* **17**(2):157-67.

[92] Scholz CB, Technau U (2003), The ancestral role of *Brachyury*: expression of *NemBra1* in the basal cnidarian *Nematostella vectensis* (Anthozoa), *Dev Genes Evol* **212**(12):563-70.

[93] Mazza ME, Pang K, Martindale MQ, Finnerty JR (2007), Genomic organization, gene structure, and developmental expression of three clustered *otx* genes in the sea anemone *Nematostella vectensis*, *J Exp Zool B Mol Dev Evol* **308**(4):494-506.

[94] Matus DQ, Thomsen GH, Martindale MQ (2007), FGF signaling in gastrulation and neural development in *Nematostella vectensis*, an anthozoan cnidarian, *Dev Biol* **217**(2):137-48.

Image sources

The figures in Chapter 1 have been adapted from the sources listed below. The figures in the other chapters appeared in their corresponding publications. The sources of the microscopy pictures in Appendix B are indicated below the images and included in the references.

Figure 1.1A-C

Gilbert SF (2000), *Developmental Biology*. 6th ed. Sunderland (MA), Sinauer Associates.

Figure 1.1D-G

Etienne L (2013), Early Development chapter 5. Quizlet Flashcards, quizlet.com/20488579/early-development-chapter-5-flash-cards.

Figure 1.2

Gilbert SF (2000), *Developmental Biology*. 6th ed. Sunderland (MA), Sinauer Associates.

Figure 1.3

Johndrow JE, Magie CR, Parkhurst SM (2004), Rho GTPase function in flies: insights from a developmental and organismal perspective, *Biochem Cell Biol* **82**(6):643-57.

Figure 1.4

Gilbert SF (2000), *Developmental Biology*. 6th ed. Sunderland (MA), Sinauer Associates.

Figure 1.5

Ashe HL, Briscoe J (2006), The interpretation of morphogen gradients, *Development* **133**(3):385-94

Figure 1.6

Jaeger J, Martinez-Arias A (2009), Getting the measure of positional information, *PLoS Biol* **7**(3):e1000081.

Figure 1.7

Jaeger J, Surkova S, Blagov M, Janssens H, Kosman D, Kozlov KN, Manu, Myasnikova E, Vanario-Alonso CE, Samsonova M, Sharp DH, Reinitz J (2004), Dynamic control of positional information in the early *Drosophila* embryo, *Nature* **430**(6997):368-71.

Figure 1.8

Stefanik DJ, Friedman LE, Finnerty JR (2013), Collecting, rearing, spawning and inducing regeneration of the starlet sea anemone, *Nematostella vectensis*, *Nat Protoc* **8**(5):916-23.

Figure 1.9

Kahi Kai comparative database for embryonic development of marine invertebrates, www.kahikai.org/index.php?content=embryology_nvectensis

Figure 1.10

Tamulonis C, Postma M, Marlow HQ, Magie CR, de Jong J, Kaandorp J (2011), A cell-based model of *Nematostella vectensis* gastrulation including bottle cell formation, invagination and zippering, *Dev Biol* **351**(1):217-28.

Figure 1.11

Tamulonis C, Postma M, Marlow HQ, Magie CR, de Jong J, Kaandorp J (2011), A cell-based model of *Nematostella vectensis* gastrulation including bottle cell formation, invagination and zippering, *Dev Biol* **351**(1):217-28.

Figure 1.12A-D

Ryan JF, Mazza ME, Pang K, Matus DQ, Baxevanis AD, Martindale MQ, Finnerty JR (2007), Pre-bilaterian origins of the Hox cluster and the Hox code: evidence from the sea anemone, *Nematostella vectensis*, *PLoS One* **2**(1):e153.

Figure 1.12E-H

Shimizu T, Bae YK, Hibi M (2006), Cdx-Hox code controls competence for responding to Fgfs and retinoic acid in zebrafish neural tissue, *Development* **133**(23):4709-19.

Summary

The purpose of this thesis, entitled “Spatial gene expression quantification in changing morphologies”, is threefold. First, a spatial gene expression quantification method for the starlet sea anemone *Nematostella vectensis* is described. Second, new insights are provided by applying this method on collected gene expression microscopy images and analyzing the resulting numerical profiles. Third, the feasibility of quantifying spatial gene expression for other animals is discussed.

Our method for quantifying *N. vectensis* gene expression pictures consists of three basic steps. The first step is a time series of graphical embryo outlines derived from microscopy pictures of *N. vectensis* embryos. The changing shape is visualized in detail by fluorescent markers attached to the cell membranes. These embryo micrographs are converted into digital templates. In the second step, the cell layer is extracted from images that display the spatial distribution of gene products. If the gene expression pattern is cylindrically symmetric about the main body axis, then a graphical outline is adjusted to the observed cell layer outline. Otherwise a geometric transformation is performed before the cell layer is indicated. After the third step, a one- or two-dimensional numerical profile is obtained. The software program divides the cell layer into segments and plots the average color values of the pixels in each segment. Regions where the selected color profile visibly deviates from the actual expression pattern are manually indicated and corrected; the result is a 1D gene expression profile. For transformed images, the color profiles are measured for multiple cross sections and collected into a 2D profile.

Additional knowledge has been obtained from studying a set of genes involved in gut formation and from another study of all genes in a database for *N. vectensis* gene expression. Spatial gene expression images of gut formation genes have been collected from publications in scientific journals. The 1D numerical profiles extracted from these expression images have been systematically compared to classify the genes based on similarity of their spatial expression patterns. Four genes with dissimilar patterns have been selected for computational simulations. The parameter estimations of a mathematical gene interaction model resulted in a preliminary gene regulation network for *N. vectensis* gut formation.

One-dimensional numerical profiles have been extracted from all suitable *N. vectensis* images in the Kahi Kai gene expression database. These profiles have been compared separately for each distinct stage of development. The genes were found to be clustered in groups that display expression in different spatial domains. A statistical analysis showed that the domain clusters are persistent during early development. This result suggests that very little mixing occurs between cells and that a localized tissue region in a later stage can be mapped to a region in an earlier stage.

To quantify gene expression in other animals, high-resolution images are required to generate accurate shape templates. Moreover, spatial gene expression images are needed to produce numerical expression profiles for comparison among various genes and for input data in computational simulations. Microscopy techniques to obtain both types of images are already applied to many model animals. Therefore, gene expression quantification is expected to be used with other animals besides *Nematostella* and to increase our general understanding of regulatory interactions.

Samenvatting

De doelstelling van dit proefschrift, getiteld “Kwantificering van ruimtelijke genexpressie in veranderende morfologieën”, is drieledig. Allereerst is een methode beschreven om ruimtelijke genexpressie te kwantificeren in de zeeanemoon *Nematostella vectensis*. Ten tweede zijn nieuwe inzichten voortgekomen door deze methode toe te passen op een verzameling microscoopplaatjes met genexpressie en de resulterende numerieke profielen te analyseren. Als derde is de haalbaarheid besproken om ruimtelijke genexpressie te kwantificeren in andere dieren.

Onze methode om genexpressieplaatjes van *N. vectensis* te kwantificeren bestaat uit drie basisstappen. De eerste stap is een tijdreeks van grafische embryo-omlijningen die zijn afgeleid van microscoopbeelden van *N. vectensis*-embryo’s. De veranderende vorm is nauwkeurig zichtbaar gemaakt door fluorescente labels die zich hechten aan de celmembranen. Deze embryoafbeeldingen worden omgezet naar digitale sjablonen. In de tweede stap wordt de cellaag afgeleid uit plaatjes die de ruimtelijke verspreiding van genproducten weergeven. Als het genexpressiepatroon cylindersymmetrisch is rond de lengteas, dan wordt een grafische omlijning aangepast aan de omtrek van de cellaag op de foto. Anders wordt de cellaag gemarkeerd na een geometrische transformatie. De derde stap resulteert in een één- of tweedimensionaal numeriek profiel. Het softwareprogramma verdeelt de cellaag in segmenten en plot de gemiddelde kleurwaarden van de pixels in elk segment. Plekken waar het kleurprofiel zichtbaar afwijkt van het werkelijke expressiepatroon worden handmatig aangeduid en gecorrigeerd; het resultaat is een 1D genexpressieprofiel. Bij getransformeerde plaatjes worden de

kleurprofielen gemeten voor meerdere doorsnedes en verzameld in een 2D profiel.

Er is nieuwe kennis vergaard met het bestuderen van een set genen die betrokken zijn bij de darmvorming en met een studie van alle genen in een genexpressiedatabase voor *N. vectensis*. Plaatjes met ruimtelijke genexpressie van darmvormingsgenen zijn verzameld uit wetenschappelijke publicaties. Uit deze expressieplaatjes zijn 1D numerieke profielen afgeleid, die systematisch vergeleken zijn om de genen te rangschikken naar overeenkomsten in hun ruimtelijke expressiepatronen. Vier genen met de meest verschillende patronen zijn gekozen voor simulatieberekeningen. De parameterschattingen van een wiskundig model voor geninteracties hebben een voorlopig genregulatienetwerk voortgebracht voor de darmvorming in *N. vectensis*.

Uit alle bruikbare *N. vectensis* plaatjes in de genexpressiedatabase Kahi Kai zijn ééndimensionale numerieke profielen opgesteld. Deze profielen zijn afzonderlijk vergeleken voor ieder ontwikkelingsstadium. De genen bleken clusters te vormen met ruimtelijke expressie in verschillende domeinen. Volgens een statistische analyse blijven de clusters gehandhaafd gedurende de vroege ontwikkeling. Deze uitkomst doet vermoeden dat cellen zich nauwelijks mengen en dat een plaatselijk groepje cellen herleid kan worden naar een specifiek groepje cellen in een vroeger stadium.

Om genexpressie in andere dieren te kwantificeren, moeten sjablonen gemaakt worden uit plaatjes met hoge resolutie. Daarnaast zijn plaatjes met ruimtelijke genexpressie nodig als bron voor numerieke expressieprofielen om genen met elkaar te vergelijken en om te dienen als referentie in computersimulaties. Beide soorten plaatjes zijn mogelijk met de microscopische technieken die nu al toegepast worden op een heleboel modeldieren. Daarom zal de kwantificering van genexpressie waarschijnlijk ook gebruikt worden bij andere dieren dan *N. vectensis* en hiermee verbetert ons algemene begrip van de interacties waardoor genen aangestuurd worden.

Publications

Journal papers

Botman D, Kaandorp JA (2012), Spatial gene expression quantification: a tool for analysis of *in situ* hybridizations in sea anemone *Nematostella vectensis*, *BMC Res Notes* **5**:555.
DB performed the gene expression quantifications, drafted the manuscript and designed the 3D expression representation. JAK designed and coordinated the study and helped to draft the manuscript.

Botman D, Röttinger E, Martindale MQ, de Jong J, Kaandorp JA (2014), A computational approach towards a gene regulatory network for the developing *Nematostella vectensis* gut, *PLoS One* **9**(7):e103341.
DB designed the study, performed the gene expression quantifications, carried out the parameter optimizations and drafted the manuscript. ER designed and carried out many of the *in situ* hybridization experiments. MQM provided the original confocal microscopy pictures. JdJ developed the prototype GenExp software and performed prior gene expression quantifications. JAK coordinated the study and helped to draft the manuscript.

Botman D, Jansson F, Röttinger E, Martindale MQ, de Jong J, Kaandorp JA (2015), Analysis of a spatial gene expression database for sea anemone *Nematostella vectensis* during early development, *BMC Syst Biol* **9**(1): 63.

DB designed the study, performed the gene expression quantifications, produced the overview of gene expression domains, carried out the correlation and statistical analysis and drafted the manuscript. FJ participated in the coordination of the study and helped to draft the manuscript. ER and MQM designed the Kahi Kai database and helped to draft the manuscript. ER performed many in situ hybridization experiments for the Kahi Kai database. MQM coordinated many experiments for the Kahi Kai database. JdJ provided the prototype GenExp software, suggested to perform the hypothesis tests and helped to draft the manuscript. JAK coordinated the study and helped to draft the manuscript.

Book chapter

Kaandorp JA, Botman D, Tamulonis C, Dries RM (2012), Multi-scale modeling of gene regulation of morphogenesis. In: Cooper SB, Dawar A, Lowe B, editors, *How the world computes, Turing Centenary Conference and 8th Conference on Computability in Europe*, in series *Lecture Notes in Computer Science*, nr 7318. Cambridge, Springer. pp. 355-362.

Conference proceedings

Botman D, Kaandorp JA, Sloot PMA (2011) An integrated approach to infer the gene network in early development of the cnidarian *Nematostella venctensis*. In: Bosch TCG, Holstein TW, editors. *Searching for Eve: Basal metazoans and the evolution of multicellular complexity*, (abstract) International Workshop: Searching for Eve. Tutzing, Germany, 2011. pp. 101.

Botman D, Kaandorp JA (2011) An integrated approach to infer the gene network in early development of the cnidarian *Nematostella vectensis*. *ISNB 2011: The 8th International Symposium on Networks in Bioinformatics*. Amsterdam, 2011.

Botman D, Kaandorp JA (2010) An integrated approach to infer and model the gene network in early development of the cnidarian *Nematostella vectensis*. In: Kaandorp JA, Meesters E, Osinga R, Verreth JAJ, Wijgerde THM, editors. *Euro ISRS symposium 2010: Reefs in a changing environment, book of abstracts*, (abstract). Wageningen, The Netherlands, 2010. pp. 62.

Acknowledgements

The writing of this thesis could not be accomplished without the support of many people around me. The people involved with the Section for Computational Science at the UvA provided me with a pleasant and inspiring environment to perform my research. My gratitute goes to the staff, teachers, postdocs and my fellow PhD candidates, as well as the technicians, the secretaries and the internship students I worked with. I am confident that this group will continue to produce excellent scientific research under the new name of Computational Science Lab.

Next to this, I also received moral support from close and distant relatives and people from various churches in Amsterdam and Venlo. Especially the couple in Amstelveen, who rented out a room to me, cared a lot about my daily wellbeing. During the six years I have lived here, this family went through various hardships (including the husband's funeral) and managed my capricious behaviour on top of everything.

I take this opportunity to mention the individuals who I met in person only a few times, while their efforts contributed substantially to my achievements. First are Marten Postma, Johann de Jong and Yves Fomekong, who left the SCS department before I arrived. Their labor provided the foundation for my research project and projects of other group members.

Other contributors are my co-authors from outside the Netherlands, who gave me various insights during our e-mail discussions. I profited much from the experience of Eric Röttinger and Mark Martindale, and they undoubtedly increased the quality of our publications.