Meuleman, W., Welten, M., & Verbeek, F. J. (2006). Construction of correlation networks with explicit time-slices using time-lagged, variable interval standard and partial correlation coefficients. In M. R. Berthold, R. C. Glen, & I. Fischer (Eds.), *Computational Life Sciences II: Second International Symposium, CompLife 2006, Cambridge, UK, September 27-29, 2006. Proceedings* (pp. 236-246). (Lecture Notes in Computer Science; Vol. 4216). Springer Berlin Heidelberg. https://doi.org/10.1007/11875741_23

Peer reviewed version

Link to published version (if available):
10.1007/11875741_23

Link to publication record in Explore Bristol Research
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Springer at http://dx.doi.org/10.1007/11875741_23. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: http://www.bristol.ac.uk/pure/user-guides/explore-bristol-research/ebr-terms/

Construction of correlation networks with explicit time-slices using time-lagged, variable interval standard and partial correlation coefficients

Wouter Meuleman[1,3], Monique C.M. Welten[1,2], Fons J. Verbeek[1]

[1] Leiden Institute of Advanced Computer Science (LIACS), Leiden University, Niels Bohrweg 1, 2333CA Leiden, The Netherlands
[2] Division of Molecular Cell Biology, Institute for Biology, Leiden University, Wassenaarseweg 64, 2333AL Leiden, The Netherlands
{wmeulema,mwelten,fverbeek}@liacs.nl
[3] Current address W. Meuleman: Information and Communication Theory group, Delft University of Technology, P.O. Box 5031, 2600GA Delft, The Netherlands

Abstract. The construction of gene regulatory models from microarray time-series data has received much attention. Here we propose a method that extends standard correlation networks to incorporate explicit time-slices. The method is applied to a time-series dataset of a study on gene expression in the developmental phase of zebrafish. Results show that the method is able to distinguish real relations between genes from the data. These relations are explicitly placed in time, allowing for a better understanding of gene regulation. The method and data normalisation procedure have been implemented using the `R` statistical language and are available from `http://zebrafish.liacs.nl/supplements.html`.

1 Introduction

Microarray data potentially disclose relations between genes. To that end a lot of research effort is spend on revealing networks of genes from microarray time-series experiments. Two established approaches for doing so are by using correlation networks ([1, 2]) and dynamic Bayesian networks ([3–5]).

Correlation networks reveal only global associations between genes (via control mechanisms) over all time-points. No exact indication of time and interaction, i.e., interaction control, is given in these networks. The advantage of correlation networks is however, that they can be built in a deterministic manner, based on relatively simple calculations.

In general, dynamic Bayesian networks give more insight in the data by providing information on when certain genes are related, i.e., over which time-points. The major drawback of dynamic Bayesian networks however, is that there is no polynomial time algorithm known for finding the optimal network structure for a particular dataset, i.e., this problem is NP-hard ([6]). Network structure candidates have to be tried one by one and since the number of possibly suitable

structures grows exponentially with the number of variables (genes) used, this quickly becomes computationally infeasible[1].

In this paper, we propose a method that combines these approaches so that correlation networks with explicit time-slices are obtained. This is achieved by using time-lagged, variable interval, standard and partial correlation coefficients. We build onto the notion of correlation networks. However, in order to avoid ambiguity with biological terms, from this point onwards the term 'correlation models' will be used.

2 Approach

The model building process involves the selection of strong correlations between gene expression profiles of which one profile is lagged in time and the interval over which the correlation is calculated is shortened. More specifically, correlations are calculated for all possible gene-pairs over transitions in time which are supported by the dataset. A t-test is used as a cutoff filter to obtain strong correlations only. These correlations are subsequently used to build an initial model incorporating explicit time-slices.

Indirect relations are removed from the model by calculating first order partial correlations for each strong correlation in the model. For two arbitrary genes g_1 and g_2 this works as follows. For each strong correlation between the profiles of g_1 and g_2 over time-points t_i and t_j, where $i < j$, partial correlations are calculated. This is done by controlling the influence of each of the genes in t_i, excluding g_1, on the correlation between g_1 and g_2.

Results are assessed for significance. Those below a certain significance threshold are removed from the model. The result of this is a correlation network with explicit time-slices, where connections between reporters indicate direct relations.

3 Methods

Correlations

In order to establish which genes portray a relation, we can calculate the correlation between their expression profiles. Standard correlation models are based on Pearson's correlation coefficient. However, these are not sufficient here, as we are primarily interested in correlations over time, that is, we would like to find out which expression profiles show a similar trend, with a certain time-lag.

An example of such correlations is given in Figure 1. Profiles 1 and 2 are strongly correlated with a time-lag of 2 units. A method for finding such correlations is used in [4] for grouping genes with possible common pathways. A

[1] Other recent approaches have acknowledged this problem and constructed ways of restricting the search space so that solutions could be generated in reasonable time ([7]).

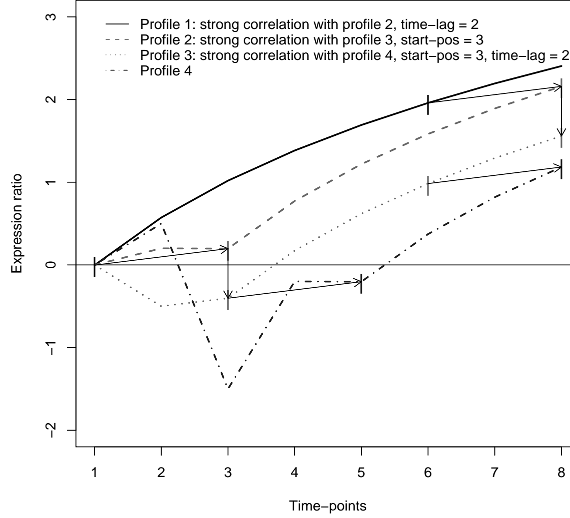

Fig. 1. Profiles with strong time-lagged and/or variable interval correlations. Arrows between profiles indicate the interval over which the correlation was calculated.

drawback of this method is that the correlation calculation is always initiated from the first time-point for one of the profiles. It is therefore not sufficient for finding the strong correlations present between profiles 2 and 3 and between 3 and 4. Profiles 2 and 3 correlate strongly from time-point 3 onwards, but not before that. Another example of this is the correlation between profile 3 and 4, which involves an additional time-lag of 2 units. These correlations must be considered, otherwise valuable information is underutilised. Therefore, we extended the standard formula for Pearson's correlation coefficient to incorporate, in addition to time-lags, a starting position s from which to look for correlations. This extension is formally described as:

$$\rho^*_{s,\ell}(x,y) = \frac{\sum_{i=s}^{|x|-\ell} (x_i - \mu_{x_{s,|x|-\ell}})\,(y_{i+\ell} - \mu_{y_{s+\ell,|x|}})}{\sqrt{\sum_{i=s}^{|x|-\ell} (x_i - \mu_{x_{s,|x|-\ell}})^2 \sum_{i=s}^{|x|-\ell} (y_{i+\ell} - \mu_{y_{s+\ell,|x|}})^2}}\,, \tag{1}$$

where $|x|$ $(= |y|)$ denotes the number of time-points available and $\mu_{x_{s,|x|-\ell}}$ and $\mu_{y_{s+\ell,|x|}}$ are the means of variables x_s to $x_{|x|-\ell}$ and $y_{s+\ell}$ to $y_{|x|}$ respectively.

Significance of correlations

The adjustment of intervals has consequences for the significance of correlations calculated. In order to test the significance of a correlation, a t-test with $n' - 2$

degrees of freedom is used. Here, n' indicates the number of time-points used for calculating the correlation, given by

$$n' = n - (s - 1) - \ell \tag{2}$$

where n is the total number of time-points and ℓ and s the used time-lag and starting position respectively. Because we need at least one degree of freedom, we can define the allowed ranges for s and ℓ as

$$s = 1, \ldots, n - 3 \tag{3}$$
$$\ell = 1, \ldots, n - s - 2\,. \tag{4}$$

We do not look at zero-lag correlations, as we are interested in finding relations over time.

Partial Correlations

It may be the case that two strongly correlated variables (i.e., genes) have a common controller variable or that there is a moderator variable through which two variables are correlated. In both cases, two variables are strongly correlated but either one does not cause the other.

Partial correlations allow one to investigate the correlation between two variables while controlling (or, excluding the influence of) one or several other variables. That is, the partial correlation between variables x and y, controlling z, is the correlation between the parts of x and y that are uncorrelated with z. This restricts the results to only direct relations between variables, ruling out possible common controller or moderator variables.

The first order partial correlation between variables x and y, controlling z, is formally described by

$$\rho(x, y|z) = \frac{\rho(x, y) - \rho(x, z)\rho(y, z)}{\sqrt{\left(1 - \left(\rho(x, z)\right)^2\right)\left(1 - \left(\rho(y, z)\right)^2\right)}}\,. \tag{5}$$

Note that when testing first order partial correlations for significance, the number of degrees of freedom to be used is $n' - 3$.

Model building

Using Eq. 1, pairwise correlations can be calculated between g selected genes, over ℓ time-lags and with variable start positions s. This results in a correlation matrix, given by

$$M_{s,\ell} = \begin{bmatrix} \rho^*_{s,\ell}(1, 1) & \cdots & \rho^*_{s,\ell}(1, g) \\ \vdots & \ddots & \vdots \\ \rho^*_{s,\ell}(g, 1) & \cdots & \rho^*_{s,\ell}(g, g) \end{bmatrix}\,. \tag{6}$$

For each combination of s and ℓ, supported by the data and allowing for enough degrees of freedom, such a correlation matrix is calculated. Analogous to Eq. 3 and 4, the number of valid correlation matrices m is given by

$$m = \sum_{s=1}^{n-3} \sum_{\ell=1}^{n-s-2} 1 \, . \tag{7}$$

Of each correlation in the correlation matrices, the significance is assessed using a t-test. All correlations with a p-value below a pre-defined value are selected. A p-value of 0.05 is commonly used as a cutoff point to distinguish strong correlations from weak ones. Using these correlations, an initial model can be built. An example of such a model is shown in Figure 2.

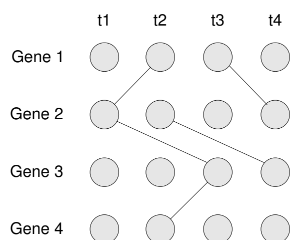

Fig. 2. Example of initial model, with $g = 4$ genes and $n = 4$ time-points

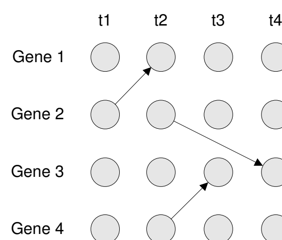

Fig. 3. Example of final model

The initial model consists of a matrix of nodes, made up of g rows (genes) and n columns (time-points). Each column can be seen as a *slice* of a time-series, a common notion in the field of dynamic Bayesian networks and time-series analysis in general. Each slice contains g nodes, depicting the genes at that time-point. Two nodes in different slices are connected if and only if the reporters they represent show a strong enough correlation.

After an initial model has been built, the correlations are verified using partial correlations. For each two reporters connected between time-points t_i and t_j in the initial model, where $i < j$, partial correlations are calculated. This is done in an iterative manner, controlling each of the remaining reporters in time-point t_i one by one.

The significance of each calculated partial correlation is calculated (using a t-test with $n' - 3$ degrees of freedom, as described before) and if one of them turns out not to be significant, this means the controlled reporter is responsible for a substantial part of the strong initial correlation. This could indicate a common controller or indirect relation.

We are particularly interested in direct relations. Therefore, in order to obtain strong direct relations only, if one or more partial correlations do not reach a pre-specified significance threshold, the initial correlation is removed from the model.

Inferring causation from correlation coefficients is discouraged ([8]), even after correction using partial correlations ([9]). However, we can extend the undirected connections in the model to directed ones when it is physically impossible for two-way relationships to occur. Because we are considering correlations between data from different time-points and time is generally directional, we can infer directions. Figure 3 shows an example of a final model, including the directed connections.

In the strict sense, the connections indicate which reporters have strongly correlating expression profiles over a certain interval. A life scientist may interpret these connections as *possible* regulations, or at least as hints towards them. After all, when a reporter shows a certain behaviour from one time-point onwards and another reporter shows similar behaviour from a later time-point onwards, the first could well be regulating the second.

4 Results

The model building process has been tested using the dataset resulting from experiments by Linney et al. ([10]). This dataset is the result of a zebrafish time-series microarray experiment and contains data for 8 equidistant time-points, ranging from 10 to 24 hours post fertilisation (hpf).

Data preparation

The processed data as used in [10] is publicly available from the ArrayExpress[2] microarray data repository (accession number `E-TIGR-17`). However, to have full control over the dataset, we preferred to start from the "raw" readouts instead, provided by Renae Lin Malek ([10]). The normalisation procedure used consisted of spatial, dye and conditional normalisations ([11]). All procedures were implemented in the R programming language.

Variable selection

Using ANOVA and Benjamini & Hochberg false discovery rate analysis ([12]) with a false positive level of 5%, genes significantly differentially expressed over the 8 time-points have been selected for further analysis.

We have focussed on one particular family of genes, so called *hox* genes. These have been recognised ([13, 14]) as important factors in the developmental mechanism of vertebrates and are significant in research involving axial patterning during development. The dataset used contains data of early developmental stages, in this time-frame these genes are expected to show temporal differential expression. Because of this, the 28 *hox* genes present in the normalised dataset have been selected as a test case for further analysis.

[2] ArrayExpress: http://www.ebi.ac.uk/arrayexpress

Model building

Following Eq. 7, for the 28 *hox* genes, $m = 15$ correlation matrices have been calculated. Out of the total amount of 11760 ($28 \times 28 \times 15$) correlations, 677 have been determined to be strong enough to be selected (p-value < 0.05).

Subsequently, partial correlations have been calculated, of which the significance was tested with p-value < 0.05 as well. Out of the initial 677 correlations, only 8 had strong enough partial correlations. This means that they suggest strong direct relations. In Table 1 the 8 selected correlations are listed with the lowest obtained partial correlation. The selected correlations have been used to build the final model as shown in Figure 4.

Table 1. Connections in final *hox* model. The 's' and '$s+\ell$' columns denote the starting points and time-lags of the connections as they are used in Eq. 1. The 'Acc. no.' column contains the accession numbers of reporters and the 'Corr.' and 'Part. corr.' columns contain the values of the standard and partial correlation coefficients respectively. Values have been rounded to 4 decimals for clarity. The last column contains references to literature showing evidence of found connections.

Selected correlations							Lowest partial correlations			
s	**Acc. no.**	**Gene**	$s+\ell$	**Acc. no.**	**Gene**	**Cor.**	**Acc. no.**	**Gene**	**Part. cor.**	**Evidence**
1	U40995	*hoxb1b*	2	AF071261	*hoxc13a*	-0.9173	AF071258	*hoxc11a*	-0.8706	[15, 16]
1	U40995	*hoxb1b*	2	BI705747	*hoxb8a*	-0.9177	AF071258	*hoxc11a*	-0.8538	[15, 17–19]
1	AF071264	*hoxc4a*	3	AF071261	*hoxc13a*	-0.9697	AF071247	*hoxa5a*	-0.9092	[15, 16, 20]
1	AF071245	*hoxa3a*	3	BI705747	*hoxb8a*	0.9895	AF071252	*hoxb4a*	0.9456	[15, 17]
1	AF071241	*hoxa13a*	4	Y14530	*hoxb8a*	0.9880	BI705747	*hoxb8a*	0.9661	
2	AF071251	*hoxb1a*	3	BI705747	*hoxb8a*	-0.9841	AF071264	*hoxc4a*	-0.9121	[15, 17]
2	AF071264	*hoxc4a*	4	X17267	*hoxb6a*	-0.9972	AF071251	*hoxb1a*	-0.9850	[17]
3	AF071251	*hoxb1a*	4	AF071258	*hoxc11a*	-0.9908	AF071252	*hoxb4a*	-0.9669	[15, 17]

Result verification

Hox genes are conserved throughout vertebrate evolution and possess a property commonly referred to as spatial and temporal co-linearity ([21–23]). This property is reflected in the names of the different *hox* genes. Lower numbered genes, e.g., *hoxb1a*, are typically expressed earlier and more anterior (closer to the head) than genes with a higher index, e.g., *hoxc11a*.

Thus, temporal co-linearity of *hox* genes already implicitly provides a method for verification of the model given in Figure 4. The exact origin of the order of expression of *hox* genes is still subject of research; moderators, regulation through other (*hox*) genes, might be involved ([24]).

Seven (7) out of the eight (8) relations present in the model do indeed directly correspond to the co-linearity pattern; the gene in the earlier time-point is lower-numbered than the gene in the later time-point. One relation, i.e. from *hoxa13a* to *hoxb8a* (cf. Figure 4) is an exception to the co-linearity rule and we have to reason as to whether this is a false positive result or another explanation clarifies this outcome.

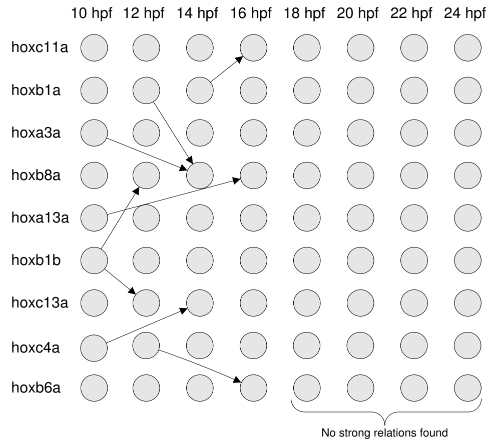

Fig. 4. Correlation model of *hox* genes. Genes are ordered in such a way to provide maximum graphical clarity. No strong relations were found between genes in time-points 18 to 24 hpf.

The forward connections as computed by our method are confirmed by published experimental approaches ([15–20], cf. Table 1). Very pronounced examples from the recent literature are the connections between *hoxb1b* and *hoxb8* and between *hoxa3a* and *hoxb8a*.

5 Discussion

We have elaborated on a methodology for developing correlation models to find good candidates for gene regulatory networks from microarray time-series data. In [25], a method is presented for inferring correlation networks using both standard and partial correlations. The principal difference between their method and the method presented in this paper is that we do incorporate explicit time-slices.

The principle of operation of our method was illustrated with a time-series dataset of zebrafish development and the results on this dataset provide insight in the usefulness of the method. We focussed on the connections between *hox* genes. The use of partial correlations helped to eliminate indirect relations from this subset of genes and thus focus on the more prominent ones over time. One should realise that indirect relations still might play a role, i.e., other genes/factors could influence the gene expression. Moreover, the connections found should not be interpreted as conclusive direct causations or regulations between genes. In our correlation model a connection indicates that gene expressions show a similar trend in behaviour over a certain time-span and with a certain time-lag.

Clearly, the more samples available the more distinct the results will be. Microarray data have some shortcomings that we should realise in the light of understanding the results. First, the samples used in Linney et al. consisted of crushed embryos. According to the co-linearity rule, in addition to a temporal

order, *hox* genes also have a spatial order of expression. Using whole embryos complicates the possibility of detecting such arrangements which has an effect on the relations found by our method. Second, the accuracy with which relations are detected strongly depends on the sampling over time, i.e., time resolution. Given the time resolution of the dataset that we explored, consisting of samples taken with 2 hour intervals, a number of relations between genes might not have been detected. In considering our results one should also take into account that gene relation events do not necessarily adhere to given sampling intervals; the segmentation phase of zebrafish development is particularly profound in terms of morphological changes and therefore changes in gene expression will also arise between the time-points ([26]) and not appear distinct in the model building process.

Models resulting from using our method provide an overview of some of the more profound connections between genes. These connections are explicitly situated in time-slices, providing valuable information about their behaviour. The resulting models can be used for the generation of new research hypotheses, as well as for a starting point for heuristic approaches to the construction of further models, such as dynamic Bayesian networks (DBN). Such networks are able to model more complex relations between genes, which may yield more precise and complete models. On the other hand, using DBNs will find much more relations making it more difficult to interpret the results, especially by realising that noise is an intricate component of microarray data. Using microarrays in the onset of new experiments, a simple and straightforward outcome will help the researcher in a better way and this is exactly what we set out to accomplish with our method.

Spatio-temporal analysis approaches such as Whole Mount ISH ([27, 28]) could complement the information gained using this method. Spatial, rather than temporal, information about the expression of genes yields information on co-location in addition to that of co-expression. The unresolved, rather differing, connection between *hoxa13a* and *hoxb8a* could be explained by understanding that these genes co-localize in distinct domains ([29, 30]), which is something that can be shown using such co-locational analysis methods. Further research is therefore directed towards the combination of machine learning approaches on both spatial and time resolved gene expression patterns.

Our results were confirmed in the literature from analysis of wild types in zebrafish ([15, 17]) and additional data from explant as well as knock-out experiments in other model systems (e.g., [20, 19]). We have claimed that the dataset had some limitations; in addition we argue having data, i.e., microarrays, for more time points would enrich our findings so as to get more confirmation on the relations found and disclose other more transient relations. At the time of our experimentation these data were not available, however, the repositories that are currently being populated with microarrays can provide additional data necessary for such analysis.

We have shown that this method allows hypothesis generation and extraction of seed networks for computationally more intensive approaches. It makes use

of multiple correlation matrices which can be calculated independently. As a consequence, the method allows for easy parallelization.

6 Conclusion

We have presented a method for building correlation networks with explicit time-slices which uses a novel way to find relations over time. The essential ingredients are time-lagged, variable interval, correlation coefficients.

The method was tested for genes of the zebrafish *hox*-family and has produced good results which are biologically meaningful and in correspondence with the literature.

Acknowledgements

The authors wish to acknowledge Renae Lin Malek for providing the dataset used to test our methodology. Many thanks to Veronica Vinciotti for providing invaluable support during normalisation of the used data and to Walter Kosters and Miranda Mandjes-van Uitert for helpful comments. This work has been partially supported by the ZF-Models consortium and the University Fund of Leiden University.

References

1. Wu, X., Ye, Y., Subramanian, K.: Interactive analysis of gene interactions using graphical Gaussian model. ACM SIGKDD Workshop on Data Mining in Bioinformatics **3** (2003) 63–69
2. Schfer, J., Strimmer, K.: An empirical Bayes approach to inferring large-scale gene association networks. Bioinformatics **21**(6) (2005) 754–764
3. Friedman, N.: Inferring cellular networks using probabilistic graphical models. Science **303** (2004) 799–805
4. Kellam, P., Liu, X., Martin, N., Orengo, C., Swift, S., Tucker, A.: A framework for modelling virus gene expression data. Intelligent Data Analysis **6**(3) (2002) 267–279
5. Kim, S.Y., Imoto, S., Miyano, S.: Inferring gene networks from time series microarray data using dynamic Bayesian networks. Brief Bioinform **4**(3) (2003) 228–235
6. Chickering, D., Geiger, D., Heckerman, D.: Learning Bayesian networks is NP-hard. Technical Report MSR-TR-94-17, Microsoft Research (1994)
7. Zou, M., Conzen, S.D.: A new dynamic Bayesian network (DBN) approach for identifying gene regulatory networks from time course microarray data. Bioinformatics **21**(1) (2005) 71–79
8. Hill, A.: The environment and disease: association or causation? In: Proc R Soc Med. Volume 58. (1965) 295–300
9. Phillips, C., Goodman, K.: The missed lessons of Sir Austin Bradford Hill. Epidemiologic Perspectives & Innovations **1**(3) (2004)
10. Linney, E., Dobbs-McAuliffe, B., Sajadi, H., Malek, R.L.: Microarray gene expression profiling during the segmentation phase of zebrafish development. Comp Biochem Physiol C Toxicol Pharmacol **138**(3) (2004) 351–362

11. Wit, E., McClure, J.: Statistics for microarrays: design, analysis and inference. John Wiley & Sons Ltd, Chichester, West Sussex, U.K. (2004)
12. Pavlidis, P.: Using ANOVA for gene selection from microarray studies of the nervous system. Methods **31**(4) (2003) 282–289
13. Duboule, D.: The vertebrate limb: a model system to study the Hox/HOM gene network during development and evolution. Bioessays **14**(6) (1992) 375–384
14. Krumlauf, R.: Hox genes in vertebrate development. Cell **78**(2) (1994) 191–201
15. Prince, V.E., Moens, C.B., Kimmel, C.B., Ho, R.K.: Zebrafish hox genes: expression in the hindbrain region of wild-type and mutants of the segmentation gene, valentino. Development **125**(3) (1998) 393–406
16. Thummel, R., Li, L., Tanase, C., Sarras, M.P., Godwin, A.R.: Differences in expression pattern and function between zebrafish hoxc13 orthologs: recruitment of Hoxc13b into an early embryonic role. Dev Biol **274**(2) (2004) 318–333
17. Prince, V.E., Joly, L., Ekker, M., Ho, R.K.: Zebrafish hox genes: genomic organization and modified colinear expression patterns in the trunk. Development **125**(3) (1998) 407–420
18. Bel-Vialar, S., Itasaki, N., Krumlauf, R.: Initiating Hox gene expression: in the early chick neural tube differential sensitivity to FGF and RA signaling subdivides the HoxB genes in two distinct groups. Development **129**(22) (2002) 5103–5115
19. Forlani, S., Lawson, K.A., Deschamps, J.: Acquisition of Hox codes during gastrulation and axial elongation in the mouse embryo. Development **130**(16) (2003) 3807–3819
20. Gaunt, S.J., Strachan, L.: Temporal colinearity in expression of anterior Hox genes in developing chick embryos. Dev Dyn **207**(3) (1996) 270–280
21. Duboule, D., Doll, P.: The structural and functional organization of the murine HOX gene family resembles that of Drosophila homeotic genes. EMBO J **8**(5) (1989) 1497–1505
22. Graham, A., Papalopulu, N., Krumlauf, R.: The murine and Drosophila homeobox gene complexes have common features of organization and expression. Cell **57**(3) (1989) 367–378
23. Duboule, D.: Vertebrate hox gene regulation: clustering and/or colinearity? Curr Opin Genet Dev **8**(5) (1998) 514–518
24. Kmita, M., Duboule, D.: Organizing axes in time and space; 25 years of colinear tinkering. Science **301**(5631) (2003) 331–333
25. de la Fuente, A., Bing, N., Hoeschele, I., Mendes, P.: Discovery of meaningful associations in genomic data using partial correlation coefficients. Bioinformatics **20**(18) (2004) 3565–3574
26. Kimmel, C.B., Ballard, W.W., Kimmel, S.R., Ullmann, B., Schilling, T.F.: Stages of embryonic development of the zebrafish. Dev Dyn **203**(3) (1995) 253–310
27. Verbeek, F.J., Lawson, K.A., Bard, J.B.: Developmental bioinformatics: linking genetic data to virtual embryos. Int J Dev Biol **43**(7) (1999) 761–771
28. Welten, M., De Haan, S., Bertens, L., Noordermeer, J., Lamers, G., Spaink, H., Meijer, A., Verbeek, F.: ZebraFISH: Fluorescent *in situ* hybridization protocol and 3D images of gene expression patterns. Zebrafish (Accepted) (2006)
29. Knosp, W.M., Scott, V., Bchinger, H.P., Stadler, H.S.: HOXA13 regulates the expression of bone morphogenetic proteins 2 and 7 to control distal limb morphogenesis. Development **131**(18) (2004) 4581–4592
30. Sakaguchi, S., Nakatani, Y., Takamatsu, N., Hori, H., Kawakami, A., Inohaya, K., Kudo, A.: Medaka unextended-fin mutants suggest a role for Hoxb8a in cell migration and osteoblast differentiation during appendage formation. Dev Biol **293**(2) (2006) 426–438